

Advanced modelling and analytics for effective change and anomaly detection in hyperspectral images.

LI, Y.

2024

The author of this thesis retains the right to be identified as such on any occasion in which content from this thesis is referenced or re-used. The licence under which this thesis is distributed applies to the text and any original images only – re-use of any third-party content must still be cleared with the original copyright holder.



Advanced Modelling and Analytics for Effective Change and Anomaly Detection in Hyperspectral Images

Yinhe Li

A thesis submitted in partial fulfilment of the
requirement of Robert Gordon University for the
degree of

Doctor of Philosophy

November 2024

Acknowledgements

The past four years of my Ph. D. have been the most unforgettable period in my life. During this academic journey, I have not only accumulated invaluable knowledge and skills but have also discovered the profound potential of personal growth in facing various challenges towards success. I would like to extend my heartfelt gratitude and deepest respect to everyone who has supported and guided me throughout this journey.

First, I would like to express my deepest gratitude to my supervisor: Professor Jinchang Ren for his kind academic guidance and very strong support in my personal life. His patience and encouragement have been essential in helping me navigate the challenges of my research. His rigorous scientific approach and profound academic insights have profoundly influenced me, and his meticulous attention to every detail in my research was invaluable. I am deeply grateful for his support, whether in guiding my research, refining my paper or providing support in my daily life. I would also like to take this opportunity to thank Prof. Vladimir Stankovic and Dr. Pamela Johnston, my VIVA external and internal examiners. Thanks for their helpful comments and suggestions.

Besides, thanks National Subsea Centre (NSC) and Robert Gordon University for offering financial support of my PhD. This research would have been impossible without the funding from the University and my supervisor. I also wish to thank all colleagues working in the NSC, especially Dr. Yijun Yan, though he left and is now a Lecturer in the University Dundee, Dr. Ping Ma, Dr. Md Junayed Hasan, Mr Hamidreza Farhadi Tolie and others. I am thankful for their helpful assistance in my personal life, helping me through these four challenging years.

Finally, I would like to thank my parents for their moral support and constant encouragement.

Abstract

The main objective of this research is to design and implement novel models and analytics techniques for hyperspectral change detection and anomaly detection. With the widespread applications of the hyperspectral imagery (HSI) in fields such as remote sensing, environmental monitoring, and agriculture, the need for accurate and efficient change detection and anomaly detection has become increasingly critical. However, existing methods often face huge challenges related to the complexity of processing high-dimensional HSI data, especially the severe sensitivity to noise that cause low detection accuracy, and high computational costs.

To address these issues, this thesis first provides a comprehensive literature review of the current state of research in hyperspectral change detection and anomaly detection, systematically organising the representative algorithms and analysing their trends and advancements in the past, especially in the recent three years. Building on this foundation, the thesis proposes a novel accumulated band-wise binary distancing (ABBD) model for unsupervised parameter-free HCD, which requires no parameter setting and can maintain high detection accuracy across different scenarios, thereby simplifying the operational complexity in practical applications. Additionally, this study introduces a novel 2D self-attention module, leading to the development of two lightweight deep learning networks focused on extracting local spatial-spectral features for more accurate change detection. The first network, namely CBANet, integrates a cross-band feature extraction module with the 2D self-attention, achieving higher detection accuracy and fewer hyperparameters compared to other advanced deep learning-based methods. The second lightweight network, SSA-LHCD, combines the singular spectrum analysis (SSA) as a preprocessing step with a 2-D self-attention module, further improving the detection accuracy while reducing the number of the hyperparameters of the model. Experimental results demonstrate that these two proposed techniques outperform a few state-of-the-art methods on several commonly used hyperspectral change detection datasets, highlighting their superiority in practical applications. Moreover, this thesis introduces a novel deep learning-based model called GASSM, marking the first exploration of combining the state-space-model (SSM) based Mamba model with the global attention for hyperspectral change detection. GASSM effectively overcoming the limitations of traditional convolutional neural networks in terms of the limited receptive field and the high computational complexity associated with transformer-based methods, offering new directions for future

research. Additionally, this study proposes a background reconstruction-based hyperspectral anomaly detection method, which has been shown to exhibit robustness and high detection accuracy across six different scenario datasets.

Overall, this study significantly advances the field of hyperspectral change detection and anomaly detection by proposing and validating several novel models and analytics methods, laying a solid foundation for further research and applications in this area.

Publications

Journal papers:

- [1] **Y. Li**, J. Ren, Y. Yan, Q. Liu, P. Ma, A. Petrovski, H. Sun. “CBANet: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-11, 2023.
- [2] **Y. Li**, J. Ren, Y. Yan, P. Ma, M. Assaad, Z. Gao. “ABBD: Accumulated Band-wise Binary Distancing for Unsupervised Parameter-Free Hyperspectral Change Detection”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp.9880-9893, 2024.
- [3] **Y. Li**, J. Ren, Y. Yan, G. Sun, P. Ma. “SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection”, *Remote Sensing*, vol. 16(13), pp. 2353, 2024.

Conference Papers:

- [4] **Y. Li**, J. Ren, Y. Yan, Q. Liu, A. Petrovski, J. McCall. “Unsupervised change detection in hyperspectral images using principal components space data clustering,” *6th Int. Conf. on Machine Vision and Information Technology (CMVIT)*, Haikou, China, Feb. 2022.
- [5] **Y. Li**, Y. Yan, J. Ren, Q. Liu, H. Sun. “MLM-LSTM: Multi-layer Memory Learning Framework Based on LSTM for Hyperspectral Change Detection”, *Int. Conf. on Brain Inspired Cognitive Systems*, pp. 51-61, Kuala Lumpur, Malaysia, Aug. 2023.
- [6] **Y. Li**, J. Ren, Z. Gao, G. Sun. “Sparse autoencoder-based hyperspectral anomaly detection with the singular spectrum analysis-based spectral denoising”, *IEEE Int. Geoscience and Remote Sensing Symposium*, Athens, Greece, Jul. 2024.

Other Publications:

- **Y. Li**, J. Ren, H. Fu, G. Sun. “GASSM: Global Attention and State Space Model-Based End-to-End Hyperspectral Change Detection”, *Journal of the Franklin Institute* (Accepted), 2024.
- P. Ma, J. Ren, Z. Gao, **Y. Li**, R. Chen. “Hyperspectral imagery quality assessment and band reconstruction using the Prophet model”, *CAAI Transactions on Intelligence Technology*, 1-15, 2024. DOI: 10.1049/cit2.12373
- X. Li, S. Weiss, Y. Yan, **Y. Li**, J. Ren, J. Soraghan, M. Gong. “Siamese residual neural network for musical shape evaluation in piano performance assessment”, *31st European Signal Processing Conference (EUSIPCO)*, pp. 1-5, Helsinki, Finland, Sep. 2023.

- Y. Yan, J Ren, B. Harrison, O. Lewis, **Y. Li**, P. Ma. “Non-Destructive Peat Analysis using Hyperspectral Imaging and Machine Learning”, *arXiv preprint arXiv:2405.02191*, 2024
- H Fu, G Sun, **Y Li**, J Ren, A Zhang, C Jing, P Ghamisi. “HDMba: Hyperspectral Remote Sensing Imagery Dehazing with State Space Model”, *arXiv preprint arXiv:2406.05700*, 2024

Contents

Acknowledgements.....	I
Abstract.....	II
Publications	IV
List of Figures.....	X
List of Tables.....	XIII
Acronyms.....	XV
Chapter 1 Introduction	1
1.1 Hyperspectral Remote Sensing.....	1
1.2 Aims and Objectives	4
1.2.1 Significance of Hyperspectral Change Detection	4
1.2.2 Significance of Hyperspectral Anomaly Detection	5
1.2.3 Challenges and Motivations.....	6
1.3 Thesis Contributions	7
1.4 Thesis Structure.....	9
Chapter 2 Related Work in Hyperspectral Change Detection	12
2.1 Hyperspectral change detection algorithms	15
2.1.1 Traditional Unsupervised Algorithms	15
2.1.2 Deep Learning-based Unsupervised Algorithms	17
2.1.3 Fully Supervised Algorithms	18
2.1.4 Self-supervised Algorithms	21
2.1.5 Semi-supervised Algorithms	23
2.2 Hyperspectral Change Detection Datasets	24
2.3 Evaluation Criteria.....	27
2.4 Summary.....	28
Chapter 3 ABBD: Accumulated Band-wise Binary Distancing for Unsupervised Parameter-Free Hyperspectral Change Detection	31

3.1 Proposed Method	32
3.1.1. Conventional Pixel Differencing.....	32
3.1.2. Band-wise Binary Distancing	33
3.1.3. Adaptively Thresholding in Binary Distancing for Parameter-free Implementation	34
3.1.4. Determining the Overall Change Map.....	35
3.1.5. Applying k-means for Binary Decision-making	36
3.2 Experiments and Results	36
3.2.1 Results and analysis	36
3.2.2 Further discussion of the quantitative results	41
3.3 Further Discussions.....	42
3.3.1 Computational complexity analysis	43
3.3.2 Effect of binary k-means vs. thresholding	44
3.3.3 Effect of the adaptively determined optimal N	45
3.3.4 Comparison with supervised deep learning-based algorithms	46
3.4 Summary.....	48
Chapter 4 Supervised Deep Learning-based Methods for Hyperspectral Change Detection.....	49
4.1 Introduction.....	49
4.2 CBANet: An End-to-End Cross-Band 2-D Attention Network for Hyperspectral Change Detection in Remote Sensing	50
4.2.1 Methodology	51
4.2.2 Experiments and Results	54
4.2.3 Summary.....	64
4.3 SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection	65
4.3.1 Methodology	66
4.3.2 Experiments and Results	71
4.3.3 Ablation Study	76

4.3.4 Discussion	80
4.3.5 Summary	82
Chapter 5 GASSM: Global Attention and State Space Model-Based End-to-End Hyperspectral Change Detection	84
5.1 Introduction	84
5.2 Proposed Method.....	85
5.2.1 Global Attention Mechanism	86
5.2.2 Feature Extraction.....	86
5.2.3 SSM-based Mamba Block.....	87
5.2.4 Binary Decision Making for HCD	89
5.3 Experiments and Results	90
5.3.1 Results analysis for the River dataset.....	91
5.3.2 Results analysis for the Hermiston dataset	93
5.3.3 Model Efficiency	95
5.3.4 Patch size.....	95
5.3.5 Kernel Number of Feature Extracted Module	96
5.3.6 Number of <i>Mambaout</i> channels.....	97
5.3.7 Key Stage Analysis.....	97
5.3.8 Further Discussion.....	99
5.4 Summary	100
Chapter 6 Hyperspectral Anomaly Detection	101
6.1 Introduction	101
6.2 Related Work on Hyperspectral Anomaly Detection	102
6.2.1 Statistics-based Algorithms.....	103
6.2.2 Representation-based Algorithms	104
6.2.3 Deep Learning-based HAD Algorithms	107
6.2.4 Datasets for Hyperspectral Anomaly Detection	109
6.2.5 Evaluation Criteria.....	111

6.2.6 Summary	113
6.3 Proposed Method	115
6.3.1 Methodology	115
6.3.2 Experiments and Results	118
6.4 Summary.....	123
Chapter 7 Conclusions and Future Work.....	124
7.1 Conclusions.....	124
7.2 Future work.....	126
References.....	128
Appendix.....	143
A.1: Results analysis of GASSM on Yancheng dataset.....	143
A.2: Publicly Available Remote Sensing datasets in RGB	145
A.3: Self-Annotated Remote Sensing Change Detection Dataset in RGB.....	147
A.4: Heterogeneous Remote Sensing Change Detection Datasets.....	148

List of Figures

Figure 1. 1 From grayscale image to hyperspectral image 1

Figure 1. 2 Schematic diagram of satellite-based hyperspectral imaging 2

Figure 1. 3 The logical framework of our research topic 9

Figure 2. 1 Basic change detection process of bi-temporal hyperspectral images12

Figure 2. 2 The relationships of five categories of existing HCD methods.13

Figure 2. 3 Statistics of all hyperspectral change detection algorithms from 2021 to June 202430

Figure 3. 1 The architecture of the proposed ABBD algorithm33

Figure 3. 2 Intermediate results of three datasets with ϵ equals to 300, 600, 900 on River dataset (a-c), Yancheng dataset (d-e), Hermiston dataset (g-i), where false alarms and missing pixels are marked in red and green, respectively34

Figure 3. 3 Extracted change maps35

Figure 3. 4 Extracted change maps on the River dataset from different methods.38

Figure 3. 5 Extracted change maps on the Yancheng dataset from different methods.39

Figure 3. 6 Extracted change maps on the Hermiston dataset from different methods.40

Figure 3. 7 The variation comparison under different values ϵ44

Figure 3. 8 Results of the OA and KP vs. an increasing N46

Figure 4. 1 The architecture of the proposed CBANet model51

Figure 4. 2 Extracted change maps on the River Dataset from different methods.56

Figure 4. 3 Extracted change maps on the Yancheng Dataset from different methods58

Figure 4. 4 Extracted change maps on the Hermiston Dataset from different methods60

Figure 4. 5 Comparing the OA and KP results of all DL methods on the River dataset under different training ratios.61

Figure 4. 6 Ablation experiment of the CBANet with different patch sizes62

Figure 4. 7 Ablation experiment of the CBANet with different numbers of spatial-spectral feature extraction kernel.62

Figure 4. 8 Ablation experiment of the CBANet with different numbers of 2D self-attention kernel63

Figure 4. 9 The architecture of the proposed end-to-end SSA-LHCD network66

Figure 4. 10 Examples showing a pair of unchanged pixels from the River dataset.68

Figure 4. 11	Extracted change maps on the River dataset from different methods.	72
Figure 4. 12	Extracted change maps of the Yancheng dataset from the different methods.	74
Figure 4. 13	Extracted change maps of the Hermiston dataset from the different methods.	76
Figure 4. 14	Ablation experiment of the SSA-LHCD model with different patch sizes ...	78
Figure 4. 15	Ablation experiment of the SSA-LHCD model with different kernel numbers of spectral feature extraction module.	79
Figure 4. 16	Ablation experiment of the SSA-LHCD model with different kernel numbers of spatial-spectral feature extraction module.....	79
Figure 4. 17	Ablation experiment of the SSA-LHCD model with different training ratios of all DL-based benchmarks on the River dataset.	80
Figure 5. 1	The architecture of proposed GASSM network	86
Figure 5. 2	Extracted change maps on the River Dataset from different methods.	92
Figure 5. 3	Extracted change maps on the Hermiston Dataset from different methods. ...	94
Figure 5. 4	Experiments and results of proposed GASSM under different patch sizes on two datasets of River and Hermiston.....	96
Figure 5. 5	Experiments and results of proposed method under different kernel numbers of feature extraction module on two datasets of River and Hermiston.	96
Figure 5. 6	Experiments and results of proposed method under different channels of Mambaout on two datasets of River and Hermiston.....	97
Figure 5. 7	Key stage analysis of proposed network on the River and Hermiston datasets	98
Figure 6. 1	Basic process of Hyperspectral anomaly detection.	102
Figure 6. 2	An example of box plot.....	112
Figure 6. 3	The total number of algorithms proposed each year until July 2024.	114
Figure 6. 4	The number of different types HAD models each year by July 2024.....	114
Figure 6. 6	Architecture of the proposed HAD method.....	116
Figure 6. 7	Visual comparison of the detected anomaly maps of the benchmark methods on 6 tested datasets.....	118
Figure 6. 8	Comparison of ROC curves	121
Figure 6. 9	Comparison of separability maps.....	122

Figure A. 1	Extracted change maps on the Yancheng dataset from different methods....	144
Figure A. 2	Example samples from the SYSU-CD dataset.....	145
Figure A. 3	Examples from the LEVIR-CD dataset	146
Figure A. 4	Annotated remote sensing change detection dataset	147
Figure A. 5	Illustrated visual difference between heterogeneous bitemporal image pairs	148

List of Tables

Table 1.1 Current space-borne and airborne HSI sensors (in the alphabet order of their names).....	3
Table 2.1 Hyperspectral change detection datasets.....	25
Table 2.2 Advantages and disadvantages of five categories of HCD algorithms	29
Table 3. 1 Comparing ABBD and unsupervised methods on the River dataset	38
Table 3. 2 Comparing ABBD and unsupervised methods on the Yancheng dataset.....	39
Table 3. 3 Comparig ABBD and unsupervised methods on the Hermiston dataset	41
Table 3. 4 Computational complexity of ABBD and other unsupervised HCD methods..	43
Table 3. 5 Results of using a fixed tolerance threshold	45
Table 3. 6 Results under manually and automatically determined values of N.....	46
Table 3. 7 Comparing ABBD with supervised deep learning methods	48
Table 4. 1 Architecture details for proposed CBANet model	51
Table 4. 2 Quantitative assessment of different methods on the River dataset	56
Table 4. 3 Quantitative assessment of different methods on the Yancheng dataset.....	57
Table 4. 4 Quantitative assessment of different methods on Hermiston dataset	59
Table 4. 5 Comparing the parameters of different DL-based methods on River dataset...	61
Table 4. 6 Comparison of OA and KP from three datasets with various kernel sizes	63
Table 4. 7 Architecture details of each layer in SSA-LHCD	71
Table 4. 8 Quantitative assessment of different methods on the River dataset	73
Table 4. 9 Quantitative assessment of different methods on the Yancheng dataset.....	74
Table 4. 10 Quantitative assessment of different methods on the Hermiston dataset.....	76
Table 4. 11 Complexity comparison of all DL-based methods on the River dataset.....	77
Table 5. 1 Quantitative assessment of different methods on the River dataset	91
Table 5. 2 Quantitative assessment of different methods on the Hermiston dataset.....	93
Table 5. 3 Comparing parameters and performance of different DL-based methods.....	95
Table 6.1 Hyperspectral anomaly detection datasets	110
Table 6.2 Advantages and disadvantages of different types of HAD algorithms	113
Table 6. 4 Parameter setting details of sparse AE	118

Table 6. 5 Comparing AUC values of the different methods on the six datasets.119

Table A. 1 Quantitative analysis results of different methods on Yancheng dataset.....143

Acronyms

1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
AD	Absolute Distance
AE	Autoencoder
AUC	Area Under the Curve
CR	Collaborative Representation
CRL	Contrastive Representation Learning
CVA	Change Vector Analysis
DL	Deep Learning
DTW	Dynamic Time Warping
F1	F1-score
FrFT	Fractional Fourier transform
GA	Global Attention
GT	Ground Truth
HAD	Hyperspectral Anomaly Detection
HCD	Hyperspectral Change Detection
HSI	Hyperspectral Images
ICA	Independent Components Analysis
KP	Kappa Coefficient
KRX	Kernel RX
LRaSMD	Low Rank and Sparse Matrix Decomposition
LRR	Low Rank Representation
LRX	Local RX
LSD	Local Sparse Dispersion
MSE	Mean Square Error
MSI	Multispectral Images
OA	Overall Accuracy
Re	Recall
ROC	Receiver Operating Characteristic
RS	Remote Sensing
RX	Reed-Xiaoli
SAM	Spectral Angle Mapper
SR	Sparse Representation
SSL	Self-supervised Learning
SSM	State Space Model
SWIR	Shortwave Infrared
TIR	Thermal Infrared
VNIR	Visible and Near Infrared
WRX	Weighted RX

Chapter 1 Introduction

1.1 Hyperspectral Remote Sensing

Over the last decades, the field of remote sensing (RS) has experienced transformative advancements, significantly enhancing our capacity to observe and understand the rich activities on the surface of the Earth, covering land and ocean. This evolution includes advanced methods for detecting and classifying land cover, monitoring environmental conditions, and managing Earth's natural resources [7]. Central to these advances are the various types of images used in RS, each providing unique insights through different imaging modalities. Figure 1.1 presents examples of different optical RS images, using the Indian Pines dataset as an example, with each image depicting the same scene. Colour images in RS are generated by combining grayscale images that capture the light in the red, green, and blue (RGB) channels, simulating the human vision and rendering the natural colours of a scene. Beyond basic colour imaging, RS employs multispectral and hyperspectral technologies, which provide more detailed data in the spectral domain. Multispectral imaging (MSI) typically utilises tens of spectral bands, generally fewer than 20 [8], facilitating a good trade-off between the rich spectral information and reduced data volume. These images are essential for applications where specific spectral bands are known to particular tasks for characterisation of the scene. As shown in Figure 1.1, hyperspectral imaging (HSI) represents a substantial leap in spectral imaging technology. Unlike MSI and colour images, hyperspectral sensors can capture hundreds of very narrow, continuous bands across the electromagnetic spectrum, from the visible to the shortwave infrared, which allows for a much more detailed analysis of the spectral properties of the scene being observed.

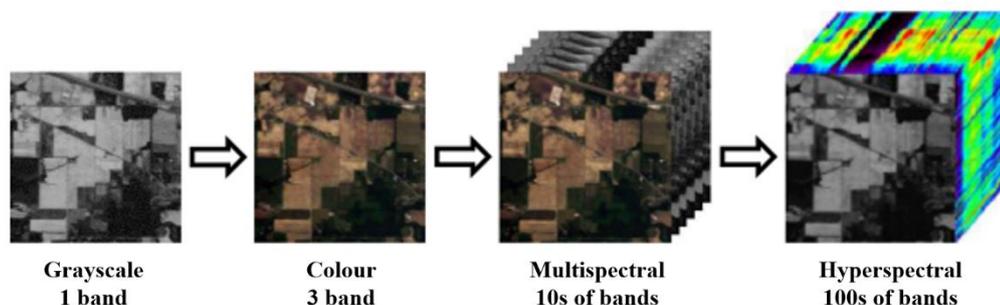


Figure 1. 1 From grayscale image to hyperspectral image

Hyperspectral RS technology employs advanced sensors on satellites or airborne platforms to capture the spectral reflectance or radiance of object surfaces across a wide range of wavelengths. By uniquely combining the high-resolution imaging and sophisticated spectral analysis, HSI RS enables not only for the observation of spatial characteristics but also for dissecting of each pixel into hundreds to thousands of narrow spectral bands. The spectrum range covered spans from approximately 400 nanometres (near the edge of the ultraviolet-visible) to 2500 nanometres (near the end of the shortwave infrared, SWIR) [9]. Consequently, Hyperspectral RS technology produces a detailed 3-D hyperspectral image (HSI) cube, or hypercube, which includes both the 2-D spatial features and the 1-D spectral attributes of the targets. This cutting-edge technology delivers precise details about the objects and enables ongoing monitoring over extensive geographic regions. Figure 1.2 illustrates the operation of a satellite-based hyperspectral imaging system designed for RS applications [10]. The detailed spectral information provided by the hypercube can be used to identify and differentiate various surface covers such as soil, water and vegetation, offering unique signatures of different objects.

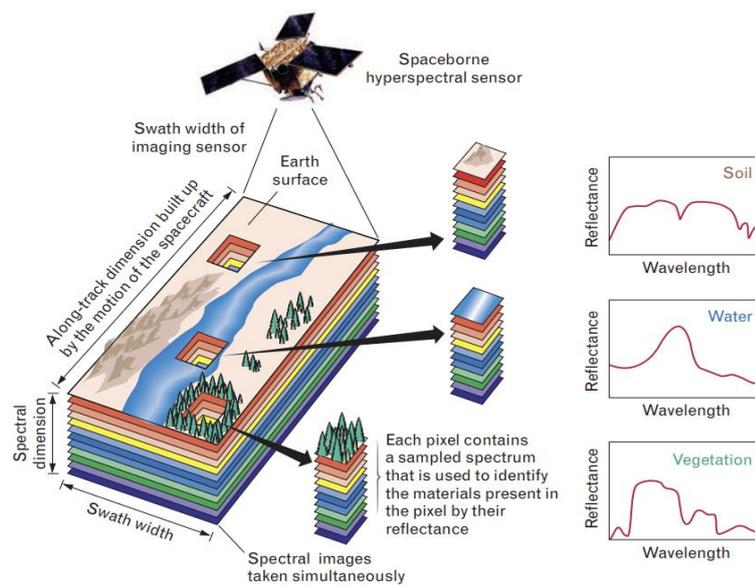


Figure 1. 2 Schematic diagram of satellite-based hyperspectral imaging [10]

In recent years, Hyperspectral RS technology has garnered a wide range of attention, due mainly to its significant effects and impacts in applied earth observation. Table 1.1 lists the current space-borne and airborne HSI sensors along with their key parameters,

including the country and year of satellite launch, the optical subsystems of the sensors, the number of spectral bands, the spectrum range and the spatial resolution. Spatial resolution [11] is defined as the capacity to identify the smallest detail in an image, representing the size of the smallest object that can be distinguished as a separate entity within the image.

Table 1.1 Current space-borne and airborne HSI sensors (sort by year)

Year	Name of Sensor	Optical Subsystem	Spectral Bands	Spectrum Range (μm)	Spatial Resolution (m)	Applications	Country
1992	CASI	VNIR	288	0.38-1.00	4	Precision agriculture, Forest monitoring	UK
1994	DAIS	VNIR	14	0.4-1.1	15	Agriculture and forestry applications	Germany
1995	RODIS	VNIR	115	0.42-0.87	2	Urban heat island effect, environmental assessment	Germany
1997	AISA	VNIR	286	0.45-0.9	3	Land cover	Finland
1998	AVRIS	VNIR, SWIR	224	0.40-2.50	20	Geological exploration, Water quality analysis	USA
1999	ASTER	VNIR, SWIR	14	0.52-1.65	15	Volcanic activity, Hot spring monitoring	USA
1999	MODIS	VNIR, TIR	36	0.40-14.40	250	Climate change; Oceanography	USA
2000	Hyperion	VNIR, SWIR	242	0.40-2.50	30	Mineral mapping, Vegetation types	USA
2000	MightSat-II	VNIR, SWIR	256	0.47-1.05	30	Ecosystem monitoring	USA
2001	CHIRS	VNIR, SWIR	63	0.41-0.98	35	Vegetation health, Water quality monitoring	Europe
2001	PROBA	VNIR, SWIR	63	4.05-1.05	20	Land cover	Europe
2002	MERIS	VNIR	576	0.39-1.04	300	Ocean and lake monitoring	Europe
2002	Shenzhou-3	VNIR	34	0.4-1.25	500	Atmospheric monitoring	China
2006	CRISM	VNIR, SWIR	544	0.38-1.07	20	Atmospheric composition	USA
2007	Chang'e	VNIR	32	0.48-0.96	200	Lunar and planetary exploration	China
2008	HJ-1A/B	VNIR	115	0.45-0.95	100	Environmental monitoring and disaster response	China
2008	HYSI	VNIR, SWIR	64	0.4-0.95	506	Agriculture, Forestry	India
2009	HICO	VNIR	128	0.38-0.95	90	Marine biology	USA
2009	HySpex	VNIR, SWIR	160	0.40-2.50	3	Surface features, environmental monitoring	Norway
2010	APEX	VNIR, SWIR	288	0.37-2.50	2	High precision terrain, vegetation mapping	Switzerland
2011	Tiangong-1	VNIR, SWIR	130	0.4-2.5	5-20	Space-based Earth observation	China
2014	OCO-2	VNIR, SWIR	1016	0.76-2.08	10	CO2 monitoring	USA

2015	SHALOM	VNIR, SWIR	250	0.4-2.5	10	Environmental and security monitoring	Italy
2017	AVIRIS-NG	VNIR, SWIR	426	0.37-2.6	3	Atmospheric, ocean monitoring	USA
2018	ISSDESI	VNIR, SWIR	235	0.4-1.0	30	Space station-based Earth observation	Germany
2018	Zhuhai-1	VNIR	32	0.4-1.0	10	Urban and rural planning	China
2019	GF-5	VNIR, SWIR	330	0.4-2.5	20	Pollution and environmental monitoring	China
2020	PRISM	VNIR, SWIR	185	0.4-2.5	30	Land and infrastructure mapping	Japan
2020	ALOS-3	VNIR, SWIR	185	0.4-2.5	30	Land and infrastructure mapping	Japan
2020	ARTEMIS	VNIR	224	0.4-2.5	5	Climate change, temperature monitoring	USA
2020	EnMAP	VNIR, SWIR	242	0.42-2.45	30	Surface minerals, Soil types	Germany
2019	PRISMA	VNIR, SWIR	236	0.38-2.0	30	Environmental monitoring, geological mapping	Italy
2023	HypSI	VNIR, SWIR	212	0.4-2.5	60	Ecosystem health monitoring	USA

Additionally, Table 1.1 provides a detailed overview of the specific application scenarios for these sensors in the field of RS. These sensors have been extensively applied in various domains such as environmental monitoring, atmospheric research, resource surveys, natural disaster response, and astronomical observations, making invaluable contributions to the advancement of earth observation. Therefore, the development of advanced HSI analysis techniques can not only enhance the capabilities of HSI itself but also extend its applicability across various critical sectors, contributing significantly to more informed decision-making and efficient resource management.

1.2 Aims and Objectives

1.2.1 Significance of Hyperspectral Change Detection

Hyperspectral change detection (HCD) is a powerful method in RS for analysing changes on the earth's surface by comparing bi-temporal HSIs taken at different periods but over the same area. The high spectral resolution of HSIs significantly enhances the accuracy of distinguishing changes and has been applied in various RS applications.

1) Environmental monitoring and management: HCD is invaluable for detecting subtle environmental changes, such as shifts in vegetation cover, forest degradation, and wetland destruction. Early detection allows for timely interventions to minimise the environmental impacts. Some examples of HCD applications in environmental monitoring and management can be found in [12][13].

2) Urban planning: HCD aids in monitoring urban expansion and land use changes, which supports sustainable development for making informed decisions. Some examples of HCD applications in urban planning are given in [14][15].

3) Disaster assessment and response: In the aftermath of natural disasters like earthquakes, floods, and fires, HCD can quickly provide detailed assessments of the affected areas, aiding in the efficient planning of rescue operations and helping mitigate the related impact. Some examples of HCD applications in disaster assessment and response are detailed in [16][17].

1.2.2 Significance of Hyperspectral Anomaly Detection

As an advanced technique in the RS field, hyperspectral anomaly detection (HAD) utilises a single HSI to identify anomalies in a geographical area without any prior knowledge of the targets and has been applied effectively across many RS applications.

1) Security and crisis response: HAD is a critical tool for identifying spectral anomalies of hazardous substances that differ from the environment, which is essential for environmental protection and rapid crisis response. Some examples of HAD applications in security and crisis response are in [18][19].

2) Agricultural management: HAD can be employed non-destructively to monitor the water status of crops and to identify anomalies in the spectral properties, which may signal the indicated diseases, pest infestations, or nutrient deficiencies. Some examples of HAD applications in agricultural management are in [20][21].

3) Military applications: HAD is vital for military reconnaissance, enabling the identification and classification of different military targets by their unique spectral signatures. It also detects camouflaged or concealed military equipment, leveraging the detailed spectral data from a single HSI without any prior information. Some examples of HAD applications in military are in [22][23].

1.2.3 Challenges and Motivations

Despite its potential, HCD and HAD have been fraught with challenges that hinder their full exploitation. These challenges can be broadly categorised into issues related to HSI itself and those associated with the HCD and HAD algorithms, as briefed below.

- 1) **Challenges in hyperspectral data:** The high dimensionality of HSI data increases the computational demands and storage requirements, introducing complexities such as the curse of dimensionality [24]. Additionally, the issues of pixel mixing, where a single pixel captures spectral signatures from multiple materials, thereby making the extraction of accurate features from the data more difficult [25].
- 2) **Challenges in HCD algorithms:** The challenges faced by HCD algorithms in detecting land surface changes are multifaceted. The types of land surface changes are diverse and different types of changes have significantly different spectral characteristics [26]. Moreover, the scale of land surface changes can be very small (such as changes in a single pixel) or very large (such as changes across an entire region) [27]. These two challenges increase the complexity of designing HCD algorithms that can effectively handle both minor and extensive changes. However, traditional unsupervised methods, relying on image algebra and transformation, often result in low detection accuracy, while supervised machine learning algorithms demand substantial computational resources and unrealistic requirements of large datasets of labelled samples. Consequently, finding a balance of these methods between the detection accuracy and running efficiency remains a critical challenge [28].
- 3) **Challenges in HAD algorithms:** The anomalous targets typically occupy a very small portion of the HSIs, often consisting of only a few pixels [29]. This has inevitably led to a severe imbalance of training samples for supervised machine learning, making it extremely challenging for the applicability of HAD tasks [30]. In addition, the complexity of the background in HSIs and the lack of prior characteristics about the anomalous targets further complicate the detection of real anomalies [31]. Moreover, most current HAD algorithms are designed for specific datasets, resulting in a lack of robustness and generalisability [32]. Therefore, designing a HAD algorithm that is both robust and capable of detecting anomalies with high accuracy remains an unresolved challenge.

In response to these challenges, the primary objective of this research is to develop advanced processing techniques for HSI-based change detection and anomaly detection in RS. The specific objectives are defined as follows:

- 1) To develop innovative unsupervised HCD algorithms that can improve the detection accuracy upon existing traditional unsupervised HCD methods.
- 2) To develop lightweight deep learning models based on supervised learning for HCD, capable of training with a smaller proportion of training set with fewer hyperparameters, while balancing detection accuracy and algorithmic efficiency.
- 3) To develop robust HAD methods with a higher detection accuracy, which can improve upon the existing HAD methods.

1.3 Thesis Contributions

In this thesis, several contributions have been made to address the main objectives as defined in Section 1.2.3, where several novel techniques have been developed for improving the detection accuracy of HCD and HAD tasks. The major contributions presented in the thesis are highlighted as follows.

- 1) To keep abreast of the latest developments in HCD and HAD algorithms and the emerging trends in algorithm development, a detailed literature review on HCD and HAD algorithms have been compiled and categorised. The review sections summarise the advantages and disadvantages of each type of method, especially those published in the past three years.
- 2) A new image-algebra-based unsupervised parameter-free method is proposed for HCD, called ABBD, based on the accumulated band-wise binary distancing instead of relying on the absolute pixel difference, which identifies band-wise changes between corresponding pixels and mitigates noise-induced inconsistencies. In addition, ABBD adaptively determines tolerance levels in binary distancing without parameters, ensuring robustness and ease of deployment. Compared with nine state-of-the-art unsupervised HCD methods, higher detection accuracy is achieved using the proposed methodology [2].
- 3) Two 2-D self-attention-based lightweight deep learning networks are proposed for HCD. The 2-D self-attention module is designed to capture local spectral-spatial features with fewer hyperparameters compared to the traditional mechanisms. Firstly,

a lightweight network called CBANet is introduced which integrates a cross-band feature extraction module and a 2-D self-attention module [1]. CBANet demonstrates higher efficiency and fewer parameters compared with four self-attention-based models. Subsequently, another lightweight network, called SSA-LHCD is proposed [3], which combines the singular spectrum analysis (SSA) as a preprocessing step with a 2-D self-attention module, improving the detection accuracy while reducing the number of the hyperparameters of the model. Compared with six advanced supervised deep learning-based networks, these two proposed 2-D self-attention-based networks can provide more accurate detection results with fewer training samples and less computational cost.

- 4) A global attention and state space model-based end-to-end HCD model, namely GASSM, is proposed. In this network, the global attention mechanism is employed to enhance the feature interaction and minimise the information loss across different dimensions. Additionally, the Mamba block is introduced to capture the long-term dependencies of both the local and global features, effectively in modelling the temporal dynamics and enhancing the model's ability to learn complex patterns. To the best of our knowledge, this study is the first to explore the use of the Mamba architecture for HCD. Compared with eight state-of-the-art benchmarks, comprehensive experiments have validated its efficacy and demonstrated its superiority in terms of the detection accuracy and stability.
- 5) A novel HAD approach is proposed, based on a sparse autoencoder combined with the singular spectrum analysis (SSA) for spectral denoising. First, 1-D SSA is employed to eliminate outliers in the spectral domain. Second, the SSA-smoothed hypercube undergoes processing by a sparse autoencoder for background reconstruction where the reconstruction error is used to identify anomalous pixels. Comprehensive experiments conducted on six public datasets demonstrate the superior performance of the method in effectively enhancing the separability between anomaly pixels and their respective backgrounds, outperforming six state-of-the-art methods, particularly in terms of the detection accuracy [6].

1.4 Thesis Structure

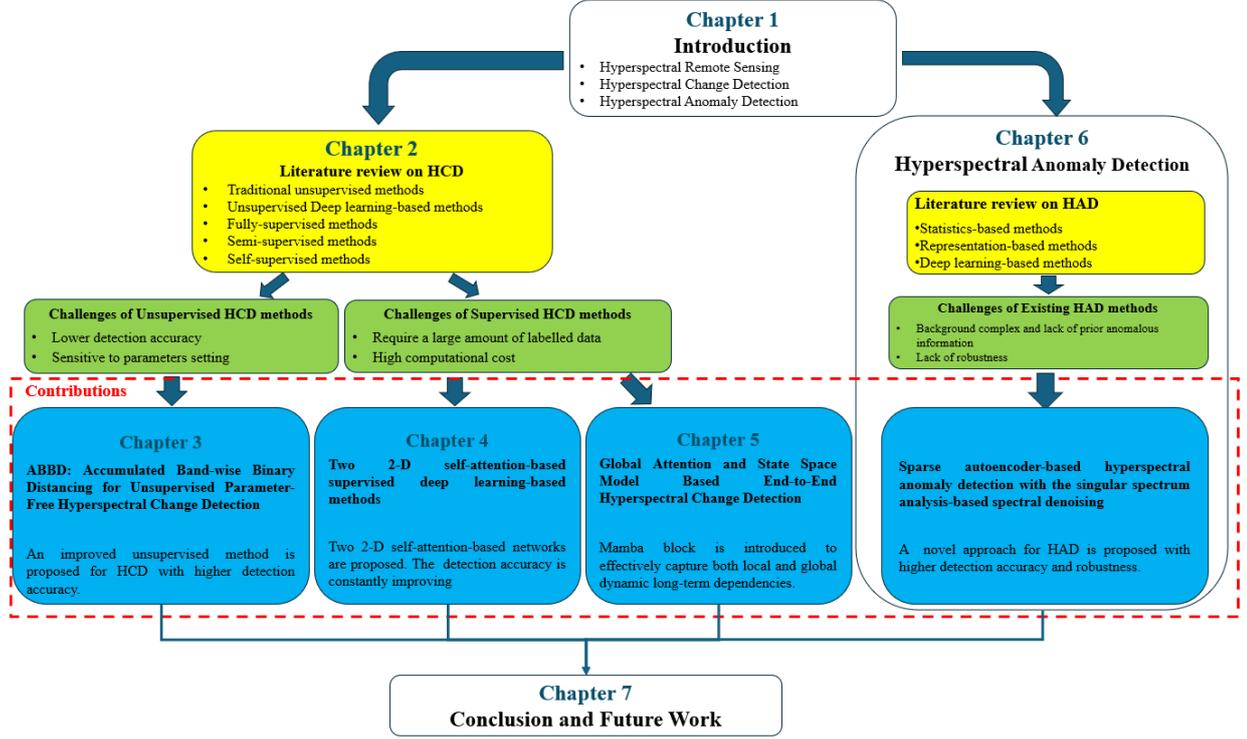


Figure 1.3 The logical framework of our research topic

Figure 1.3 summarises the logical collections of the work presented in this thesis. As can be seen, HCD is focused on Chapters 2-5, followed by the novel HAD methods presented in Chapter 6. The remainder of this thesis is organised as follows.

Chapter 2 provides a comprehensive literature review of existing HCD methods. It reviews representative algorithms within traditional unsupervised, supervised frameworks and discusses cutting-edge methods in unsupervised deep learning, self-supervised and semi-supervised techniques. This chapter methodically discusses the strengths and weaknesses of each category. Additionally, it compiles all HCD algorithms proposed from 2021 to 2024. This chapter also introduces the most commonly used HCD datasets and the metrics for evaluating the detection accuracy.

Chapter 3 presents a novel unsupervised parameter-free method for HCD, namely ABBD. ABBD employs an innovative band-wise binary distancing approach, which is proposed instead of the traditional reliance on absolute pixel differences with thresholding.

By effectively considering variations across the spectral domain, this technique not only improves the detection accuracy but also addresses the challenge associated with high parameter sensitivity found in existing unsupervised HCD methods. Remarkably, the detection accuracy of ABBD even surpasses the performance of some supervised learning algorithms.

Chapter 4 introduces two proposed deep learning-based networks for HCD, CBANet and SSA-LHCD. Initially, the 2-D self-attention module is introduced in the proposed CBANet, achieving higher detection accuracy and stability with fewer parameters. Following this, the SSA-LHCD network is developed, incorporating an SSA preprocessing step that drives a new deep-learning network based on the 2-D self-attention module. SSA-LHCD not only offers higher detection accuracy but also operates with fewer parameters compared to CBANet. This chapter will provide detailed introductions to these two 2-D self-attention-based deep learning networks and assess the stability of these networks through a series of ablation experiments.

Chapter 5 details a proposed supervised method, namely GASSM, based on the global attention and the state space models as in the Mamba structure, which integrates linear transformations, 1-D convolutional layers and a state space model. The Mamba block is introduced to effectively capture both local and global dependencies. To the best of our knowledge, this is the first exploration of combining the Mamba structure with the deep learning-based network for application in HCD tasks. Comprehensive experiments on two publicly available datasets, compared with eight state-of-the-art benchmarks, have validated the efficacy and efficiency of GASSM, demonstrating its superiority in the detection accuracy and stability.

Chapter 6 presents the proposed sparse autoencoder-based model for HAD. This chapter begins with an extensive literature review of existing HAD algorithms, revisiting classical unsupervised methods based on statistics and representation, and exploring advanced unsupervised deep learning and self-supervised approaches. It focuses mainly on the HAD methods proposed from 2021 to 2024, the publicly available HAD datasets and the metrics used to evaluate the detection accuracy. Following this, the chapter introduces the proposed unsupervised HAD method, which employs the 1-D Singular Spectrum Analysis (SSA) and a sparse autoencoder to denoise the HSI dataset in the spectral domain and reconstruct the complex background in the spatial domain. Finally, the proposed HAD algorithm is compared with six advanced unsupervised methods to demonstrate its superiority.

Chapter 7 presents a series of concluding remarks, along with the directions for future work to further advance the investigation in HSI-based HCD and HAD.

Chapter 2 Related Work in Hyperspectral Change Detection

Generally, the basic process of bi-temporal HCD is illustrated in Figure 2.1, where T^1 and T^2 represent the two images in the bi-temporal image set. HCD typically includes four stages: image pre-processing, change feature extraction, difference map segmentation and accuracy evaluation.

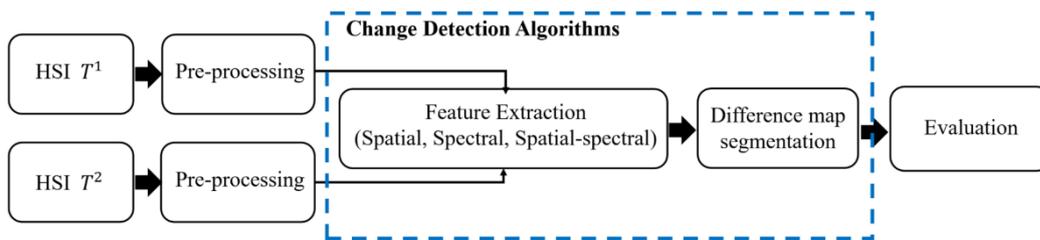


Figure 2. 1 Basic change detection process of bi-temporal hyperspectral images

- 1) **Image pre-processing:** This initial stage is essential to prepare the HSIs for further analysis. Due to inconsistencies in atmospheric conditions, solar elevation, and other conditions during bi-temporal HSI acquisition, pre-processing operations such as spatial registration and spectral correction are required to ensure the bi-temporal images are properly aligned and comparable [33].
- 2) **Feature extraction:** This phase employs either unsupervised or supervised methods to identify relevant features from the corrected HSIs. This step analyses both HSIs to extract features, such as spectral feature extraction based on algebraic methods, spatial feature extraction through image transformation, and spatial-spectral feature extraction using convolution techniques. The feature extraction step is not only crucial in the overall detection process but also a focal point for all researchers involved.
- 3) **Difference map segment:** Following the feature extraction, a difference map is produced which highlights the degree of changes. By applying thresholding or clustering methods, the change map can be segmented into a binary one, including changed and unchanged areas. Machine learning techniques such as support vector machines or deep neural networks may be employed to perform a binary classification. The final binary map can delineate changed pixels from those unchanged areas.

- 4) **Evaluation:** Ultimately, the binary map is rigorously assessed for accuracy using specialised evaluation techniques such as the confusion matrix [34] and its secondary indicators.

In HCD tasks, the key procedures involve the extraction of change features and the segmentation of the difference images. These two critical steps are the primary focus of the majority of current research in this area. Due to the scarce availability of additional bi-temporal HSI acquisition [35], algorithm development and validation are tested on publicly available datasets which are pre-corrected and registered. Overall, existing HCD algorithms can be divided into two main categories: unsupervised and supervised. This classification depends on whether they use manually annotated true labels. The relationship of five categories of HCD approaches, unsupervised and supervised, is illustrated in Figure 2.2, where the details are explained and discussed in the following sections.

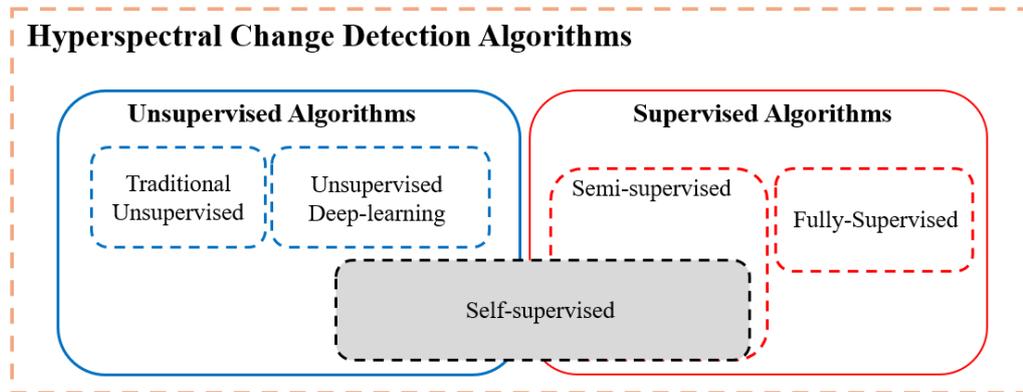


Figure 2. 2 The relationships of five categories of existing HCD methods.

Unsupervised Algorithms

Unsupervised algorithms do not rely on labelled data in the learning or decision-making process. They detect changes through the characteristics of the data itself and are mainly divided into two categories: traditional unsupervised methods and unsupervised deep learning-based methods.

- 1) **Traditional unsupervised methods:** These methods primarily rely on the intrinsic pixel values of the data and do not require pre-labelled training samples or annotations to guide the algorithm in identifying regions of change, and automatically extract

change information from bi-temporal HSIs using statistical analysis, clustering, or other image transformation techniques, without relying on labelled data.

- 2) **Unsupervised Deep learning methods:** These methods can use deep learning model architectures such as autoencoders, generative adversarial networks (GANs) or self-organizing maps to capture and represent the intrinsic properties of the input data, allowing the model to detect underlying regularities or patterns without prior manual labelled guidance.

Supervised Algorithms

Supervised learning is a machine learning method in which the model learns the relationship between input features and output true labels through operations such as convolution and pooling. This allows the model to make predictions on new, unlabelled data. Supervised learning methods can be divided into two categories: Fully supervised and semi-supervised. The definitions of these two categories are as follows.

- 1) **Fully supervised methods:** These methods rely on a large set of manually labelled training data, where regions of change and unchanged are clearly marked. Common architectures used in fully supervised methods include convolutional neural networks, recurrent neural networks and other advanced deep learning frameworks. These models tend to deliver high accuracy when ample labelled data is available but can be limited by the cost and difficulty of acquiring labelled data.
- 2) **Semi-supervised methods:** These methods utilise both manually labelled and unlabelled data for training, which aim to improve model performance by utilising the labelled data to guide the learning process while simultaneously using the unlabelled data to uncover the structure or distribution of the data. This helps the model generalize better and make more accurate predictions, even when only a small portion of the data has been labelled. Common techniques include self-training, consistency regularization and semi-supervised variants of deep learning models like CNNs and GANs.

Self-supervised methods

Self-supervised learning methods utilise labels generated automatically from the unlabelled data, lying between fully unsupervised learning and supervised learning. It can be regarded as a special form of unsupervised learning because it generates pseudo-labels

through traditional unsupervised algorithms and then trains the model in a supervised manner. Self-supervised algorithms can be also viewed as a specific form of semi-supervised algorithms as pseudo-labels are utilised to provide supervisory information for training the model, leveraging the label information to some extent. The difference here is that these labels are generated from the data itself rather than being clearly defined or annotated.

Over the past few decades, numerous HCD methods within traditional unsupervised, unsupervised deep learning, fully supervised, semi-supervised and self-supervised methods have been developed. The following sections will provide an overview of the composition and principles of the representative algorithms from each category. The summary section will discuss the advantages and disadvantages of each category. Additionally, an analysis of relevant algorithms proposed from 2021 to June 2024 will explore current research trends in HCD. Finally, publicly available HCD datasets and the associated evaluation metrics for HCD are detailed.

2.1 Hyperspectral change detection algorithms

2.1.1 Traditional Unsupervised Algorithms

In early research, the unsupervised HCD methods can be categorised into algebra-based and transformation-based ones. The algebra-based methods mainly include the image difference, image ratio, absolute distance (AD) [36], etc. The most representative method is Change Vector Analysis (CVA) [37], which subtracts the bi-temporal images to derive the spectral change vectors. The magnitude and direction of change vectors show the degree of variation. These change vectors are typically classified according to an adaptive threshold to identify significant changes. Compressed CVA (C^2VA) [38] is subsequently proposed to obtain a 2-D compressed representation of the multi-dimensional change vectors. In [39], spectral angle mapper (SAM) is proposed to compare the angle between corresponding spectral vectors to determine their similarity. Dynamic time warping (DTW) [40] is utilised to measure the spectral similarity by computing the minimum cumulative distance of corresponding pixels across all bands. These early algebra-based algorithms often combine adaptive threshold segmentation [41] or the k -means clustering method [42] to distinguish changed and unchanged pixels before generating a binary change map. The primary advantage of the algebra-based methods is their intuitiveness and ease of

implementation, allowing for rapid extraction of the change information from large-scale hyperspectral data [43]. However, these methods also have drawbacks. Due to atmospheric effects, sensor noise, or other factors, the same object may exhibit large variations, while different objects may show similar spectrums [44]. Additionally, the subjectivity of threshold selection can lead to instability in binary classification results [27], and clustering methods may result in inconsistent outcomes depending on the strategies used [45].

Instead of relying on the raw spectral features, image transformation-based methods compute the pixel differences in the transformed spatial feature domain to highlight the changed features whilst reducing the redundancy and dimension of the data. One typical method is the Principal Component Analysis (PCA) with k -means (PCA k M) [46], which projects linearly the high-dimensional data to a low-dimensional space whilst preserving most of the energy or information, followed by k -means clustering on the reduced features to obtain the change results. Multivariate Alteration Detection (MAD) method used the Canonical Correlation Analysis (CCA) [47] to maximise the correlation between the features of bi-temporal images. Subsequently, iteratively reweighted MAD (IRMAD) [48] is proposed to conduct the weighted iteration according to the chi-square distance. Slow feature analysis (SFA) [49] extracts invariant and slowly varying features from the difference image, which is obtained by subtracting bi-temporal images and transforming the results into a new low-dimensional spatial feature space. Although image transformation-based methods can effectively utilise the spatial information and reduce the dimensionality and redundancy of the high-dimensional data, which may perform well on occasions, they suffer from poor explainability, due to the complex mathematical transformations involved, making it difficult to directly relate the transformed results to the physical meaning of the original data [50]. Additionally, these methods may fail to preserve the band-wise correlations and spectral characteristics of the original data during the transformation process [51], which is a significant drawback for applications that require precise spectral information. Moreover, transformation-based methods are often computationally expensive [52], particularly when processing large datasets, which has inevitably constrained their wide applicability, especially in scenarios that require rapid data processing in real time.

Some advanced unsupervised HCD methods have been proposed based on techniques of statistics [53][54], band selection [55][56] and spectral unmixing [57][58], etc. Chen *et al.* [54] proposed an automated image analysis method that relies on the utilisation of different images and histogram statistics for HCD. This method distinguishes the changes based on both positive and negative values within the difference images and establishes location-

specific thresholds by identifying the minimum points within the histogram in accordance with variations of image brightness. Lv *et al.* [56] proposed a spatial-contextual feature extraction method that incorporates the band selection to reduce the spectral redundancy, along with an iterative spatial-adaptive filter used for noise reduction. The change magnitude is evaluated by coupling CVA and the adaptive regions around each pixel, resulting in a change magnitude image, followed by binary thresholding using the OTSU method to derive the changed pixels. Guo *et al.* [58] proposed a joint unmixing and spatial information co-guidance approach to extract the endmembers and estimate abundance, incorporating spectral perturbed regularisation to enhance the robustness against the spectral variability. Hou *et al.* [59] proposed a novel unsupervised method based on three-order tucker decomposition and reconstruction detector. Firstly, tucker decomposition and reconstruction strategies are utilised to eliminate the influence of different environment conditions during bi-temporal image acquisition. Specifically, a singular value accumulation strategy is used to determine the principal components in the factor matrices. Meanwhile, a spectral angle is applied to analyse spectral changes after tensor processing in different domains. Finally, a new detector based on SAM is designed to achieve binary classification. Marinelli *et al.* [60] proposed a new unsupervised HCD method based on a discrete representation of change information, focusing on the represented change information in each spectral band. The band-wise radiation information is analysed to generate a quantised discrete representation of the change vector, followed by a tree representation to distinguish between different types of changes. In summary, these unsupervised HCD methods have generally significant advantages over the algebra-based and transformation-based algorithms, particularly in terms of improved detection accuracy [61]. However, they also have notable limitations. Most algorithms require manual intervention to adjust parameters, which affects their usability and lacks the generalisation capabilities [62]. Additionally, the introduction of numerous parameters increases the complexity of these models [63].

2.1.2 Deep Learning-based Unsupervised Algorithms

Unsupervised deep learning (DL) methods, which learn features and patterns from unlabelled data without requiring the training samples, are primarily employed to identify latent patterns and reconstruct the structures within the data [64]. They have gained significant interest in recent years, particularly in the field of HCD. The most representative unsupervised DL-based network is the autoencoder [65]. By minimising the reconstruction error, autoencoders can capture latent structures and features within HSIs, enabling feature

extraction and dimensionality reduction of the original HSIs. Hu *et al.* [66] proposed a novel unsupervised deep learning model based on a two-stream coupled autoencoder. The network consists of two symmetric encoders and a decoder, which can jointly decompose bi-temporal images into abundance coefficients corresponding to the same set of spectral bases. Then, the pixel-wise difference is calculated based on the Euclidean distance of their abundance features, and the binary result is obtained by the OTSU threshold algorithm. Chakraborty and Ghosh [67] proposed an autoencoder-based method that exploits multiresolution deep feature maps derived by a convolutional autoencoder and automatically learns the spatial features without requiring any labelled data. Furthermore, Saha *et al.* [68] proposed a feature extraction method that employs an untrained VGG16 model to extract feature maps from bi-temporal HSIs. Subsequently, CVA is utilised to fuse and analyse these feature maps before producing the binary change map. Lei *et al.* [69] proposed a novel unsupervised DL framework which combines a spectral mapping module based on a generative adversarial network (GAN), and optimised discriminant analysis into AE-based framework for spatial attributes. In this network, the generator in the GAN structure is employed to capture the mapping spectral features, producing latent feature representations from the input data. The spectral mapping constraint loss is applied in the reconstruction space, while adversarial loss is incorporated into the latent space to improve the quality of the features extracted by the spectral mapping network. Then, spatial attribute optimization uses the spatial correlation to further improve the performance. Compared to the traditional unsupervised methods, the above-mentioned unsupervised DL-based methods, while capable of eliminating the dependence on labelled data and achieving good detection results for specific datasets with simple and distinct change features, have relatively less desirable detection accuracy and show lack of robustness when applied to more complex datasets [70]. This is due mainly to the absence of explicit label guidance for different scales of change features, as these unsupervised DL networks are sensitive to initial hyperparameter settings hence showing different performance when applied to various datasets [71].

2.1.3 Fully Supervised Algorithms

Fully supervised HCD algorithms refer to methods in machine learning and deep learning where the model is trained using labelled data, and each training sample is paired with its corresponding label, providing explicit guidance for the model to learn the relationship between input data and the annotated GT [72]. In conventional shallow machine learning, support vector machine (SVM) and random forest (RF) are commonly

used. SVM [73] treats each pixel, from either the raw data or after dimension reduction using transforms or band selection alike, as an independent sample vector. With these feature vectors, the SVM trains a classifier to determine an optimal hyperplane that separates the pixels of different categories on a per-pixel basis. RF [74] performs classification by constructing multiple decision trees and determining the final category of each pixel through a majority voting mechanism. These shallow classifiers treat each pixel as an independent sample, which allows for parallel processing to improve the detection efficiency [75]. However, they overlook the spatial features between neighbouring pixels [76]. With the advancement in computing power and the development of DL technology, deep neural networks [77] have gradually become an effective alternative to conventional machine learning methods. The multi-layered structures can model tasks in finer detail and improve the discriminative capabilities. In the field of HSI processing, many existing algorithms rely on the convolutional neural networks (CNNs) [78] and recurrent neural networks (RNNs) [79] to extract multiscale spatial-spectral features and discriminative temporal features, respectively. CNNs can automatically extract spatial features whilst incorporating spectral information. By employing the convolutional and pooling operations, CNNs can effectively capture the local patterns and spatial relationships within the images [80]. On the other hand, RNNs can capture temporal dependencies and sequential information by employing recurrent connections and hidden states. These features enable the RNNs to process and retain information from the previous iteration or steps, allowing them to identify patterns and relationships within the sequential data more effectively [81]. An advanced RNN-based approach, the Long Short-Term Memory (LSTM) network [82], further enhances this capability by mitigating issues such as the vanishing gradients, thereby maintaining long-term dependencies and improving the detection accuracy in the modelling of temporal sequences. Finally, the feature maps extracted from CNNs and RNNs can take as feature vectors and fed into the fully connected layers for non-linear transformations, ultimately achieving the binary classification. Zhan *et al.* [83] proposed a DL-based network which combines multiple 1-D and 2-D CNN layers to extract spectral and spatial features, respectively. Wang *et al.* [84] proposed a general end-to-end 2-D CNN-based method, which performs spectral unmixing on the input HSIs to obtain a mixed affinity matrix, followed by applying multiple 2-D CNN layers for feature mining. Lin *et al.* [85] proposed a bilinear CNN model that employs two symmetric CNNs to extract spatial-spectral features of the pair of HSI patches, respectively. Then, the outer product of the matrices is applied to the output feature maps to generate combined bilinear features, which are then fed into a softmax binary classifier for classification. Song *et al.* [86]

proposed a recurrent 3-D CNN fully convolutional network to extract spatial-spectral features along with a combined LSTM module for extracting bi-temporal change features to achieve binary classification. Shi *et al.* [87] designed a multipath convolutional LSTM and CNN architecture to extract multilevel temporal dependencies and multiscale temporal-spatial-spectral features. Luo *et al.* [88] proposed a multiscale diff-changed feature fusion network based on multi-scale 2-D CNN layers to enhance the feature representation by learning the refined changing components between bi-temporal HSIs at different scales. Finally, binary classification is achieved by adaptively fusing multi-scale features. Qu *et al.* [89] proposed a two-stream encoder-decoder model that integrates hierarchical features from convolutional layers of bitemporal images. It employs a context-guided attention module for enhanced feature transfer and an LSTM subnetwork to capture temporal dependencies.

The self-attention mechanism, which can focus on key information within an image with powerful modelling capabilities, has been widely used for effective feature extraction in HCD, often enhancing the feature representation ability of CNN-based networks. For example, a deep multiscale pyramid network [90] is proposed, enhanced by a spatial-spectral residual attention module. The multiscale pyramid convolution can effectively excavate multilevel and multiscale spatial-spectral features, which are further enhanced by a spatial-spectral residual attention module of each scale. Song *et al.* [91] proposed a cross-temporal interaction symmetric attention (CSA) network, where a traditional self-attention module is employed for supporting the extraction and integration of joint spatial-spectral-temporal features to enhance feature representation. Wang *et al.* [92] proposed a joint spectral, spatial, and temporal transformer for integration and extraction of relevant change detection features from bi-temporal HSIs based on the self-attention mechanism. Yu *et al.* [93] proposed a new gate spectral-spatial-temporal attention network with a spectral similarity filtering module to reduce the spectral redundancy while capturing intra-image spatial features and extracting inter-image temporal changes. Then, fed into a lightweight gated spectral-spatial-temporal attention module to integrate time-series spectral feature from bi-temporal-phase patches. Furthermore, a domain adaptive and interactive differential attention network [94] is proposed that incorporates domain adaptive constraints to mitigate pseudo-variation interface by mapping bi-temporal images to shared deep feature space for alignment. The interactive differential attention module enhances feature representation by integrating differential information, effectively filtering out irrelevant data and improving discriminative capability.

Although fully supervised DL-based models generally produce good detection accuracy, they often rely on large portion of the data, usually more than 50% of the total data, to be manual annotated as ground truth maps for training. Additionally, fully supervised DL-based methods typically entail high computational costs and have a large number of hyperparameters. Achieving high detection accuracy while substantiating computational load remains the major challenge within the existing fully supervised DL-based HCD models.

2.1.4 Self-supervised Algorithms

In recent years, the rapid advancement of self-supervised learning (SSL), which focuses on learning feature representations from unlabelled data to predict pseudo labels, has sparked numerous research initiatives in the field of HCD. Unlike fully supervised DL-based methods that rely on the labelled data for training, SSL leverages the inherent structural features of the data to predict its labels, allowing the networks to extract effective feature representations by learning these self-generated labels. Another approach involves using traditional unsupervised methods to generate pseudo labels, which can then serve as the ground truth for DL-based networks, enabling SSL-based HCD.

Contrastive representation learning (CRL) is another group of representative and powerful techniques within SSL. CRL focuses on learning useful representations of the data by contrasting positive pairs (similar or related samples) with negative pairs (dissimilar or unrelated samples). The goal is to encode the data in such a way that similar items are closer together in the representation space, while dissimilar items are further apart. For example, a deep subspace-based SSL method is proposed [95] that extracts the spatial-spectral features using the CRL. Firstly, adjacent spectral vectors are selected as positive sample pairs in an HSI, spectral vectors from different HSIs are regarded as negative sample pairs. This construction of positive and negative sample pairs is to enable the model to recognize similar and different data patterns, thereby forming clear inter-class boundaries in the feature space. Then, the trained model is transferred to the change detection task to distinguish between change and background pixels. Cao *et al.* [96] proposed a spatial-spectral contrastive clustering model that combines the CRL and autoencoder to reconstruct the spatial-spectral features to improve the feature learning ability of the network. Moreover, Li *et al.* [97] proposed a CRL-based framework that uses a cross-domain CNN to learn multi-scale representations with varying spatial-spectral features. Cai *et al.* [98] proposed a novel network, which combines unsupervised contrastive losses using a multi-channel contrastive learning strategy with multiple image

transformation methods to extract spatial-spectral features, further improving the detection accuracy. These CRL-based SSL methods are effective for extracting the changing features from complex backgrounds [99] and can be trained without the need for annotated data, significantly reducing reliance on manually annotated labels [100]. However, CRL-based methods typically require substantial computational resources, which may lead to longer training times and higher computational costs [101]. Moreover, the selection of appropriate positive and negative sample pairs is crucial in CRL. If the randomly selected positive samples cannot represent all scale changes or the selected features are severely affected by noise, the generalisation capability of the model can be compromised [102].

Graph self-supervised learning (GSSL), a special category of SSL methods, learns the graph structure and context information within the data to predict its label. GSSL offers a potential solution for estimating both the attribute and structure uncertainty due to its ability to learn the graph-based embedding. For instance, Jian *et al.* [103] proposed an uncertainty-aware GSSL network, in which the spectral and spatial correlations in bi-temporal HSIs are characterised using the graph models. Based on the constructed graph models, novel node- and edge-data augmentation methods are devised to enhance contrastive sample pairs, followed by the dual-branch GSSL contrastive network to maximise the mutual information between low-dimensional feature embeddings. In addition, a novel loss function was introduced to prioritise reliable features in feature representation. Ding *et al.* [104] proposed a graph convolutional embedding-based GSSL method that employs a spectral-spatial transformation pre-processing mechanism to learn super pixel-level spectral-spatial features from HSIs, reducing the number of graph nodes for subsequent processing. In addition, locality-preserving low-pass graph convolutional embedding autoencoder is designed, in which the low-pass graph convolution and layer-wise graph attention are proposed for extracting smoother features and supervising the clustering process to predict appropriate hidden representations labels. In summary, GSSL-based networks can model complex spatial relationships through the graph embedding structure, providing a more robust feature representation. However, GSSL-based methods are still relatively new and under development when applied to HCD algorithms. The high dimensionality and complexity of HSIs require the design of specific graph structures and learning tasks to achieve enhanced detection accuracy [105].

In addition, there are other SSL-based frameworks proposed to utilise generated pseudo-ground truth through traditional unsupervised algorithms to train DL-based networks, thereby avoiding manually annotated training samples. For instance, a PCA-guided [106] SSL network is proposed, in which the binary pseudo-labels are generated via PCAKM

before being used to train a CNN-based DL network. Li *et al.* [107] proposed a credible pseudo-label generation method, combining two unsupervised techniques: structural similarity (SSIM) and CVA. Then, select credible pseudo labels to train the CNNs, and let the uncertain pixels be predicted unambiguously. Zhao *et al.* [108] proposed a novel SSL-based network, in which pseudo-labels are obtained using a modified threshold segmentation method based on the changing intensity. Then, trained an RNN-based models to achieve binary classification. Hu *et al.* [109] proposed that the pseudo-labels can be obtained using CVA and expectation to boost the binary change detection result. Furthermore, Song *et al.* [110] utilised the pseudo-labels which are generated by PCA and spectral correction angle. These SSL-based methods generate pseudo-labels to train the deep learning models, completely avoiding the step of manual annotation of the data, thereby reducing the cost and time associated with data annotation [111]. However, the accuracy of the pseudo-labels directly impacts the performance of the pre-trained model [112]. Moreover, combining multiple unsupervised methods with a deep model in a two-stage training process, as opposed to traditional single-stage training, may significantly increase the complexity and computational cost of these algorithms [113].

2.1.5 Semi-supervised Algorithms

Semi-supervised learning leverages a limited number of labelled samples to guide the training process, along with a large number of unlabelled samples to expand the training set or retrain the network, thereby improving the confidence of the network. In a semi-supervised network proposed by Liu *et al.* [114], an SVM classifier is pre-trained using a limited number of labelled samples. Subsequently, the spatial neighbourhood of labelled training samples is combined with an active learning algorithm to select the most informative unlabelled samples of the same category for retraining the SVM classifier. Yuan *et al.* [115] employed the Laplacian regularised metric learning to extract feature representation from limited labelled samples. By assigning high-confidence pseudo-labels to unlabelled samples, it further expands the training set to train an SVM classifier. Jiang *et al.* [116] trained the generative adversarial network (GAN) by all samples without any prior information, which combines two identical multiple connected layers to build a dual-pipeline joint classifier to classify the GAN results. Finally, the classifier is fine-tuned by a very small number of labelled samples to enhance the confidence. Luo *et al.* [117] proposed to utilise multi-scale Kullback-Leibler (KL) divergence [118] and feature-enhanced probabilistic contrast learning to constrain two Siamese branches, aiming to generate and predict pseudo-label for unlabelled samples. When training the network with limited

labelled samples, the contrast loss and predicted unlabelled samples are added to the network to supervise the tuning of parameters. Moreover, in a semi-supervised method from Huang *et al.* [119], the training process begins with using a limited set of training samples as a teacher model. The student model is trained using both labelled and unlabelled samples, where the unlabelled samples are subjected to weak and strong perturbations to ensure weak-to-strong consistency and calculate unsupervised loss. Simultaneously, weak perturbations are applied to labelled samples, and the supervision loss is computed by comparing the weakly perturbed prediction with the ground truth. The final step involves a fine-tuning process on the student model, retraining the labelled samples without perturbations to improve detection accuracy.

Semi-supervised methods aim to overcome the challenges posed by limited labelled data and enhance the generalisation ability of models. However, semi-supervised learning methods are often constrained by the size of labelled samples. On the other hand, some semi-supervised methods rely on traditional preprocessing methods to generate pseudo-labels for unlabelled samples, not fully leveraging the properties of the unlabelled data itself. In addition, most semi-supervised methods require training a pre-trained model with unlabelled samples and then fine-tuning it with a small number of labelled samples, which seems less efficient and fail to achieve end-to-end learning.

2.2 Hyperspectral Change Detection Datasets

Currently, the River [120], Yancheng [121], Hermiston-1 and Hermiston-2 [121] datasets are the most commonly used public HCD datasets. These three datasets are obtained from the Hyperion sensor installed on the Earth Observing-1 (EO-1) satellite, which provides a total of 242 bands ranging from 0.4 to 2.5 μm , with a spatial resolution of 10 m. The bi-temporal HSIs and their corresponding binary change ground truth maps are summarised in Table 2.1. T^1 and T^2 represent the two images in the bi-temporal image set, and the ground truth is obtained through manual annotation.

The River dataset, shown in Table 2.1 (a-c), was collected over the Jiangsu Province, China on May 3, 2013, and December 31, 2013, respectively. After water band removal and image registration, this dataset contains 463×241 spatial pixels and 198 spectral bands, where the major changed regions are the substance in the river and the structure of the riverbank. River dataset is used in chapters 3-5.

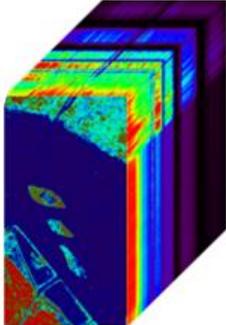
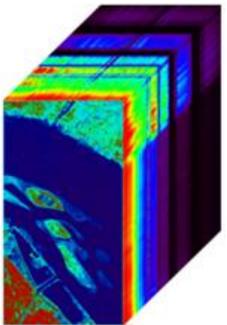
The Yancheng dataset, shown in Table 2.1(d-f), was collected over the Yancheng city, China on May 3, 2006, and April 23, 2007, respectively. After water band removal and

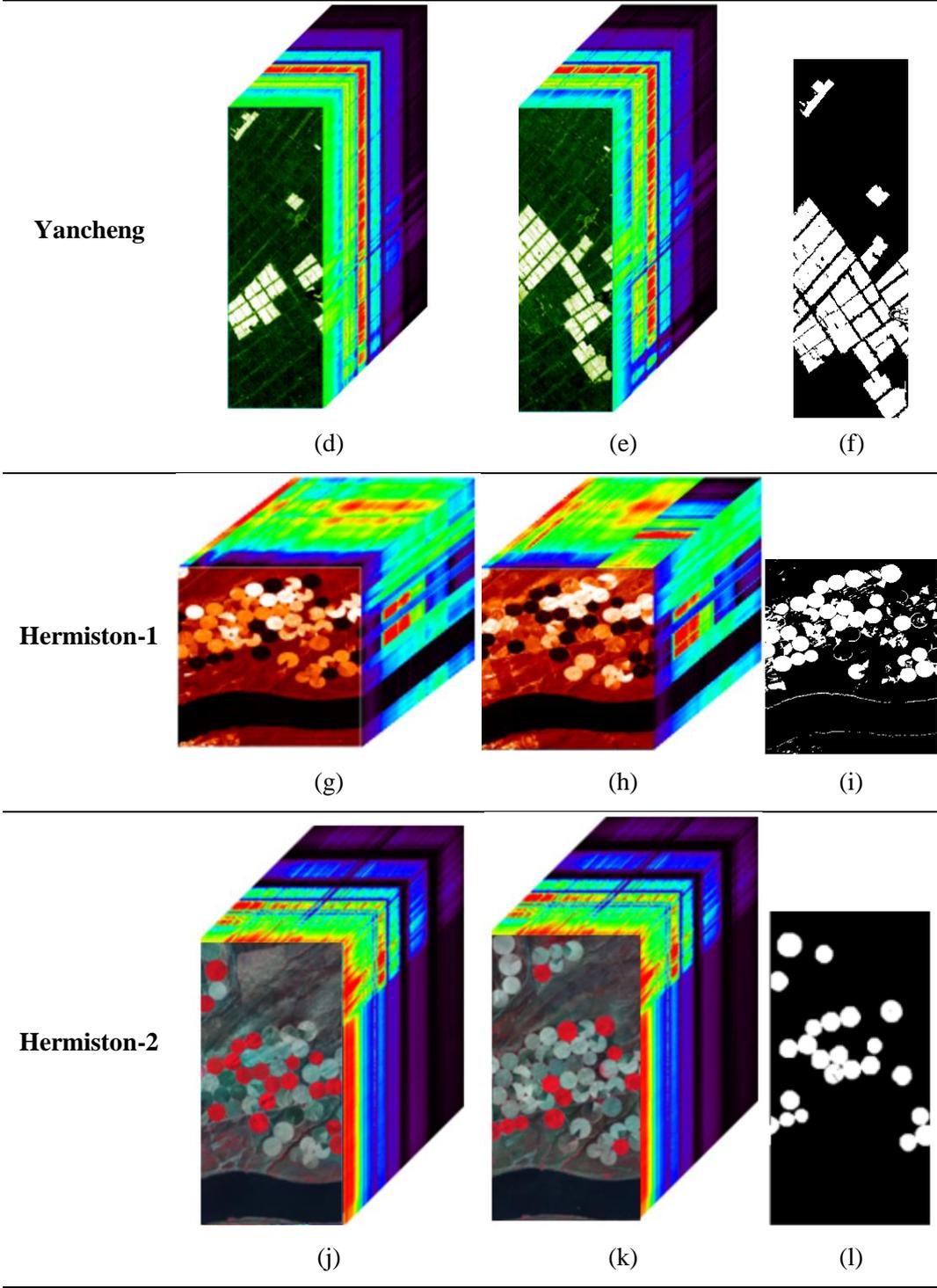
image registration, this dataset contains 420×140 spatial pixels and 154 spectral bands, where the major change is the land cover on wetlands. Yancheng dataset is used in chapters 3-4.

The Hermiston-1 dataset, shown in Table 2.1 (g-i), was collected over the Hermiston city, Oregon, United States on May 1, 2004, and May 8, 2007, respectively. After water band removal and image registration, this dataset contains 307×241 spatial pixels and 154 spectral bands, where the changing factors are crop growth situation and the water content of crops that were affected by irrigation conditions in the farmland. Hermiston-1 dataset is used in chapter 4.

The Hermiston-2 dataset, shown in Table 2.1 (j-l), was collected over the Hermiston city, Oregon, United States on May 1, 2004, and May 8, 2007, the same as Hermitson-1 dataset. However, the Hermiston-2 dataset is raw and retains all 242 bands, without any water band removal. The spatial size of this dataset is 390×200 . In terms of data labelling, some subtle variations were left unannotated compared to the Hermiston-1 dataset. Hermiston-2 dataset is used in chapters 3 and 6.

Table 2. 1 Hyperspectral change detection datasets

	T^1	T^2	Ground Truth
River			
	(a)	(b)	(c)



2.3 Evaluation Criteria

The change detection task is treated as a binary classification problem, with changed and unchanged pixels are denoted as 1 (positive) and 0 (negative), respectively. The confusion matrix [34] and its secondary indicators such as the overall accuracy (OA) and Kappa coefficient (KP) are used for quantitative performance evaluations. OA here indicates the percentage of correctly classified pixels as defined below:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

where TP , TN , FP , and FN denote the correctly detected changed pixels, correctly detected unchanged pixels, incorrectly detected changed pixels, and incorrectly detected unchanged pixels, respectively.

The KP is used to measure the inter-rater reliability as the degree of similarity between the change map and the ground truth:

$$KP = \frac{OA - PRE}{1 - PRE} \quad (2.2)$$

$$PRE = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \quad (2.3)$$

Recall (Re) represents the ratio of the number of TP observations to the total number of actual positives.

$$Re = \frac{TP}{TP + FN} \quad (2.4)$$

The F1 score ($F1$) defines a balanced index that can be considered as the harmonic mean of precision (Pre) and Re , where Pre is defined as the ratio of the number of TP observations to the total number of predicted positive observations.

$$Pre = \frac{TP}{TP + FP} \quad (2.5)$$

$$F1 = 2 * \frac{Pre * Re}{Pre + Re} \quad (2.6)$$

For more intuitive comparison, CA and NCA are used to represent the detection accuracy of the changed cluster and the non-changed cluster, respectively, as given below.

$$CA = \frac{M_C}{N_1} \quad (2.7)$$

$$NCA = \frac{M_G}{N_0} \quad (2.8)$$

where M_C and M_G denote the number of the corrected detected changed and non-changed pixels in the change map, respectively; N_1 and N_0 denote the number of changed and non-changed pixels in the ground truth, respectively.

2.4 Summary

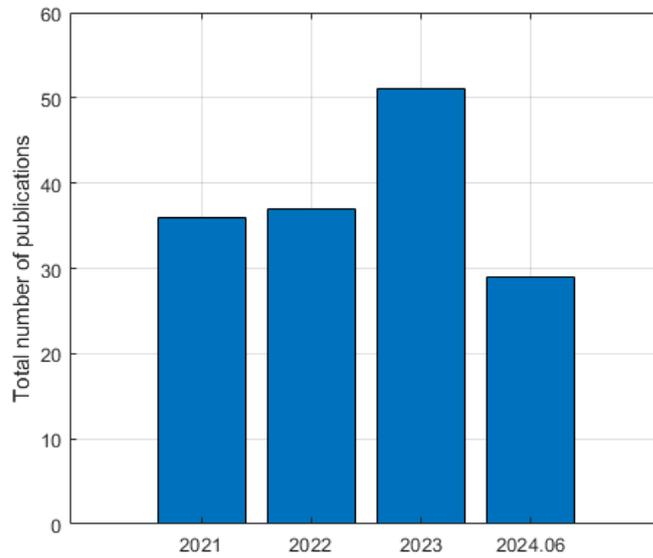
Based on the analysis of each category of HCD algorithm, including traditional unsupervised, unsupervised deep learning, fully supervised, semi-supervised and self-supervised methods, their advantages and disadvantages are summarised in Table 2.2. Traditional unsupervised methods have the advantage of requiring no labelled data, making them relatively easy to implement and interpret. However, their accuracy is often lower compared to supervised models, and the unsupervised methods are sensitive to parameter settings and noise. Unsupervised deep learning methods also do not need the labelled data and can extract complex patterns within the data, but they only perform well on specific datasets and have lower detection accuracy, interpretability, and robustness. Supervised methods are known for their high detection accuracy, especially when there is ample labelled data available. They can capture detailed spatial, spectral, and temporal features. However, they require a large amount of labelled data, which is time-consuming to label and involves high computational costs. Semi-supervised methods improve the model confidence by leveraging a small amount of labelled data alongside a large amount of unlabelled data, retaining the capability to capture detailed features similar to the supervised methods. Nevertheless, balancing labelled and unlabelled data introduces higher complexity, risks overfitting with limited data, and can make the training process inefficient if selecting low-quality unlabelled datasets. Self-supervised methods do not rely on labelled data and share the advantages of supervised methods in capturing detailed spatial, spectral, and temporal features. However, their detection accuracy depends on the quality of generated pseudo-labels, and the training process is complex. Therefore, proper selection of the appropriate method requires to balance the specific application scenario, data availability, and the available computational resources.

All articles on HCD algorithms from the past three years were compiled, with the number of publications per year and the quantity of each category. As shown in Figure 2.2 (a), from 2021 to June 2024, the total number of published articles on new HCD algorithm development has steadily increased year on year. Specifically, there were 36 articles published in 2021, 37 in 2022, and 51 in 2023. By June 2024, 29 articles had already been published. From Figure 2.2 (b), it is evident that supervised methods have remained a key focus of research in the past three years. The trends suggest a potential continuation of the upward trend in 2024, especially in supervised and semi-supervised deep learning methods. In addition, there are two review papers on HCD algorithms development, which were published in 2018 [26] and 2019 [28], respectively.

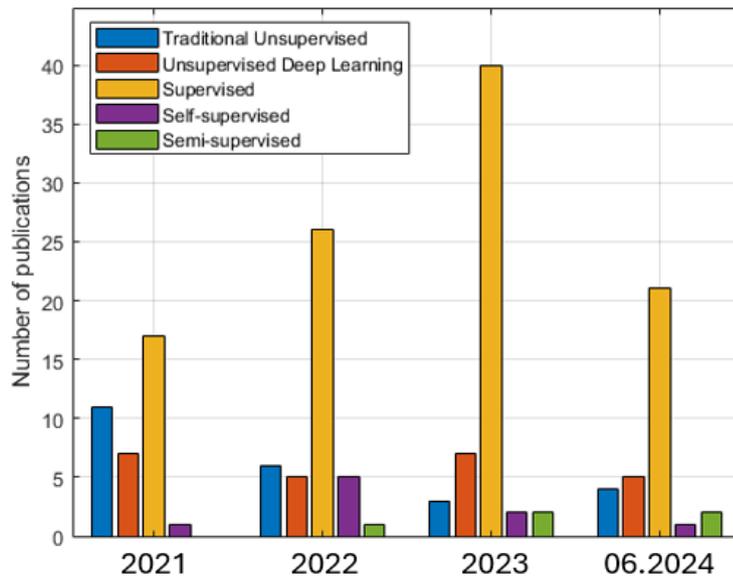
Table 2. 2 Summary of the advantages and disadvantages of five categories of HCD algorithms

Advantages		Disadvantages
Traditional Unsupervised	<ul style="list-style-type: none"> • No need for labelled data • Relatively simple implementation • Strong interpretability 	<ul style="list-style-type: none"> • Lower accuracy compared to supervised methods • Sensitive to parameters settings • Noise sensitivity
Unsupervised Deep Learning	<ul style="list-style-type: none"> • No need for labelled data • Can extract complex patterns • Good detection results for the specific dataset 	<ul style="list-style-type: none"> • Less desirable detection accuracy and robustness • Poor Interpretability
Fully supervised	<ul style="list-style-type: none"> • High accuracy with sufficient labelled data • Capture detailed spatial, spectral, and temporal features 	<ul style="list-style-type: none"> • Requires a large amount of labelled data. • Time-consuming data labelling • High computational costs • Poor Interpretability
Semi-supervised	<ul style="list-style-type: none"> • Requires fewer labelled data than supervised methods • Leverage unlabelled data to improve model confidence • Same advantages as supervised methods that can capture detailed spatial, spectral, and temporal features 	<ul style="list-style-type: none"> • Higher complexity in balancing labelled and unlabelled data • Potential overfit • Inefficient two-stage training
Self-supervised	<ul style="list-style-type: none"> • No need for labelled data • Same advantages as supervised methods that can capture detailed spatial, spectral, and temporal features 	<ul style="list-style-type: none"> • Detection accuracy depends on the quality of pseudo-labels • Complex two-stage training process

In summary, this comprehensive review lays a solid foundation for future research. Firstly, by reviewing the representative articles from the past three years, we have identified the technical gap for motivating our study in the thesis. Specifically, to overcome the limitations of existing methods, novel solutions have been proposed to tackle these challenges, which span from unsupervised change detection, deep learning based supervised change detection to anomaly detection using the most recent Mamba model.



(a) The total number of algorithms proposed each year



(b) The number of different types proposed each year

Figure 2. 3 Statistics of all hyperspectral change detection algorithms from 2021 to June 2024

Chapter 3 ABBD: Accumulated Band-wise Binary Distancing for Unsupervised Parameter-Free Hyperspectral Change Detection

Unsupervised algorithms are crucial in HCD tasks because they eliminate the need for pre-acquired manually labelled samples or the training stage, which enables the rapid and automated identification and extraction of areas where changes have occurred within the images. Consequently, these methods demonstrate high adaptability and broad applicability, particularly when handling large-scale HSIs in real time. However, as summarised in Section 2.2, existing unsupervised HCD algorithms have certain drawbacks: 1) lower detection accuracy, 2) sensitivity to noise, 3) the need for manual parameter tuning, which reduces the robustness of the model.

To tackle these challenges, a novel Accumulated Band-wise Binary Distancing (ABBD) model is proposed for unsupervised parameter-free HCD. Rather than relying on the absolute pixel difference with thresholding as in conventional approaches, the binary distancing only indicated whether a pixel was changed or not in a certain band, which could alleviate the adverse effects of noise-induced inconsistency of measurement. The band-wise binary distance map is then accumulated to form a grayscale change map, on which the simple k -means was applied for final binary decision-making. Experiments on three publicly available datasets have validated the superiority of ABBD, which has yielded comparable or slightly better results in comparison to twelve state-of-the-art methods including several deep learning models. The major contributions are summarised below.

1) We propose Accumulated Band-wise Binary Distancing (ABBD), a novel unsupervised method for HCD. By leveraging binary distancing to indicate if there is a band-wise change between corresponding pixels, ABBD can effectively mitigate the adverse effects of noise-induced measurement inconsistency.

2) To adaptively determine the tolerance level when reaching the binary distancing, a parameter-free solution is derived in ABBD for robustness and ease of deployment whilst maintaining a high accuracy.

3) By applying the k -means to the accumulated N grayscale change map, the conventional thresholding is avoided for robustness in binary decision-making. Experiments on three publicly available datasets have validated its superior efficacy and

efficiency when compared to several state-of-the-art unsupervised and even deep learning-based methods.

The remaining part of this chapter is organised as follows. In section 3.2, technical details of the proposed ABBD approach. Section 3.3 presents the experimental results of ABBD compared to nine classical and state-of-the-art unsupervised HCD algorithms on three publicly available datasets. Section 3.4 discusses the ablation analysis of the proposed ABBD, including complexity analysis, key stage analysis and comparisons with classical deep learning algorithms. Finally, Section 3.5 summarises the chapter with certain concluding remarks.

3.1 Proposed Method

3.1.1. Conventional Pixel Differencing

In bi-temporal HCD tasks, let $T^{1,2} \in \mathfrak{R}^{W \times H \times B}$ represent two HSIs captured at different times after spatial registration and spectral correction, where W , H and B denote the numbers of rows, columns and spectral bands, respectively. Let $T_{(i,j)}^1 = [x_1, x_2, \dots, x_B]$ and $T_{(i,j)}^2 = [y_1, y_2, \dots, y_B]$ denote a pair of spectral vectors at (i, j) in $T^{(1)}$ and $T^{(2)}$ ($i \in [1, W]$, $j \in [1, H]$), where x_b and y_b ($b \in [1, B]$) represent the corresponding intensity values at band b . The pixel-based band-wise distance between T^1 and T^2 can be calculated by:

$$\begin{aligned} T_{(i,j)}^{diff} &= |T_{(i,j)}^2 - T_{(i,j)}^1| = [d_{(i,j)}^1, d_{(i,j)}^2, \dots, d_{(i,j)}^B], \\ d_{(i,j)}^{(b)} &= |y_b - x_b|, b \in [1, B]. \end{aligned} \tag{3.1}$$

where $d_{(i,j)}^{(b)}$ is the difference of the corresponding pixel (i, j) at band b on two spectral vectors $T_{(i,j)}^1$ and $T_{(i,j)}^2$.

In unsupervised wisdom, changed pixels can be determined based on the magnitude of $T_{(i,j)}^{diff}$, the accumulated $d_{(i,j)}^{(b)}$ or the vector distance between $T_{(i,j)}^1$ and $T_{(i,j)}^2$ [122]. However, due to the inconsistent illumination [123], varying environmental and weather conditions and noise [124], $d_{(i,j)}^{(b)}$ often appears noise-sensitive, leading to quite unreliable results of HCD. In particular, the noise caused significant changes in certain bands may

affect the overall decision-making even though the changes are minor in the majority of the bands. Therefore, more effective spectral matching is needed for more robust HCD.

To tackle the aforementioned issues, we propose to consider the band-wise matching in HCD and define $M_{(i,j)} \in \mathcal{R}^{1*B}$ to indicate whether the pixel pair at (i,j) are changed or not. Four strategies are further proposed for improving the robustness of HCD as follows. The flowchart of the proposed ABBD algorithm is shown in Figure 3.1.

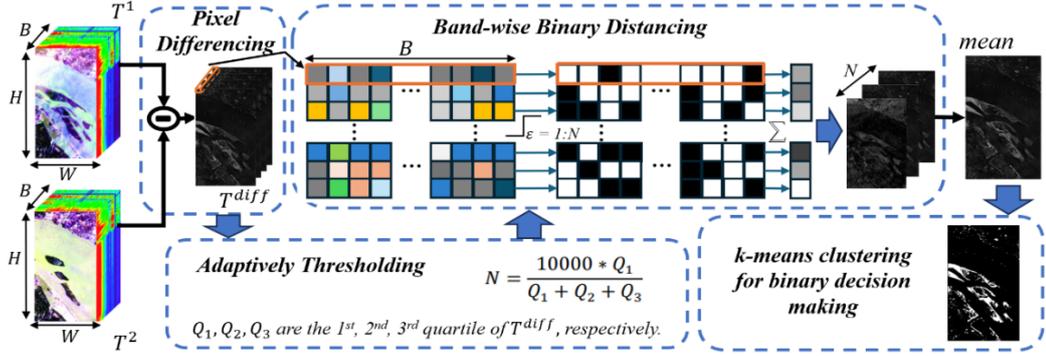


Figure 3. 1 The architecture of the proposed ABBD algorithm

3.1.2. Band-wise Binary Distancing

First, we apply a tolerance threshold ε to $d_{(i,j)}^{(b)}$, and the pair of pixels will be considered as changed at band b only if we have $d_{(i,j)}^{(b)} \geq \varepsilon$. This can help to determine a band-wise binary change map M as follows.

$$M_{(i,j)}^{(b)} = \begin{cases} 1 & \text{if } d_{(i,j)}^{(b)} \geq \varepsilon \\ 0 & \text{Otherwise.} \end{cases} \quad (3.2)$$

The strategy here has two advantages. First, the tolerance applied can filter the insignificant difference that has widely occurred in natural HSI scenes. More importantly, binary distancing can help to suppress the effect of noise caused by big changes in certain bands for more robust decision-making for robustness. By adopting band-wise binary distancing, the effect of abnormal values caused by noise can be mitigated in comparison to the absolute differences used in conventional approaches. This has been further validated in the experiment section.

3.1.3. Adaptively Thresholding in Binary Distancing for Parameter-free Implementation

Based on $M^{(b)}$, as our second strategy, the overall change for each pixel pair in $T^{(1)}$ and $T^{(2)}$ is decided by accumulating the band-wise binary change map by

$$C_{(i,j)} = \sum_{b=1}^B M_{(i,j)}^{(b)}, C_{(i,j)} \in [0, B]. \quad (3.3)$$

where C is a grayscale image to indicate the overall degree of changes in all the bands. Note the intensity of C is within $[0, B]$, where B is the number of bands.

As the $M^{(b)}$ is a binary indicator to show whether there is a noticeable change at the corresponding pixel pairs, against a pre-determined threshold ε , the extracted overall change map C is also sensitive to ε . For the three original datasets, the extracted change maps under different values of ε are shown in Figure 3.2, where the values of ε are set to 300, 600 and 900, respectively.

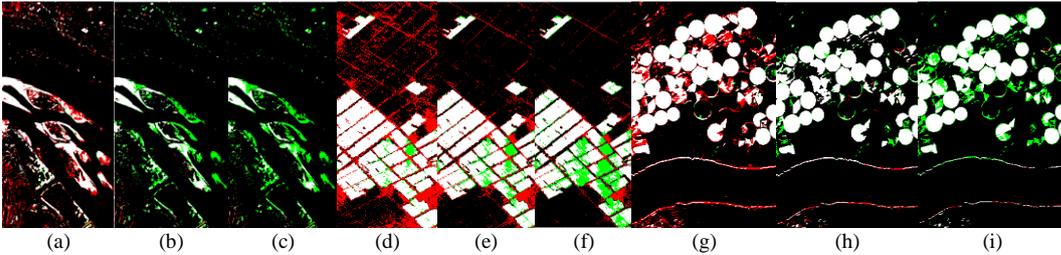


Figure 3. 2 Intermediate results of three datasets with ε equals to 300, 600, 900 on River dataset (a-c), Yancheng dataset (d-e), Hermiston dataset (g-i), where false alarms and missing pixels are marked in red and green, respectively.

As seen in Figure 3.2, the change maps C are very sensitive to ε . Setting the fixed threshold too low can result in a higher number of false alarms as it incorrectly classifies pixels with subtle difference as changed ones. With an increasing ε , more and more false alarm pixels are suppressed, yet real changed pixel may also be filtered off if the corresponding difference is not high enough, leading to more missing pixels in the resulting binary image.

3.1.4. Determining the Overall Change Map

Due to the inherent challenge of determining the optimal ε for each dataset, we propose as the third strategy an iterative process to retrieve the best change map as follows. Herein, we automatically determine a new parameter N to bypass the fixed ε as follows.

$$N = \frac{V * Q_1}{Q_1 + Q_2 + Q_3} \quad (3.4)$$

where, v is the coefficient used to ensure that the threshold aligns with the original data, which is set to 10000 in this experiment as it helps to produce the best results. Q_1, Q_2, Q_3 are the first, second and third quartile of T^{diff} , respectively.

Note the statistics here will reflect the distribution characteristics of pixel-wise difference in all spectral bands that statistical analysis-based adaptive thresholding is employed for optimal decision making, which helps to achieve a parameter-free implementation rather than relying on certain unadjusted parameters.

Subsequently, the final change map C_{acc} can be determined as the mean of the accumulated band-wise binary distancing results below, where $C^{(n)}$ denotes the change map with the parameter ε set to $n, n \in [1, N]$.

$$C_{acc} = \frac{1}{N} \sum_{n=1}^N C^{(n)} \quad (3.5)$$

The extracted final grayscale change maps of three datasets are shown in Figure 3.3, where the changed pixels are clearly distinguishable when compared to the GTs. This has verified the value of the third strategy for bypassing the threshold ε and reach a parameter-free solution. In addition, the consistent results in Figure 3.3 have validated the efficacy of the introduced strategy.

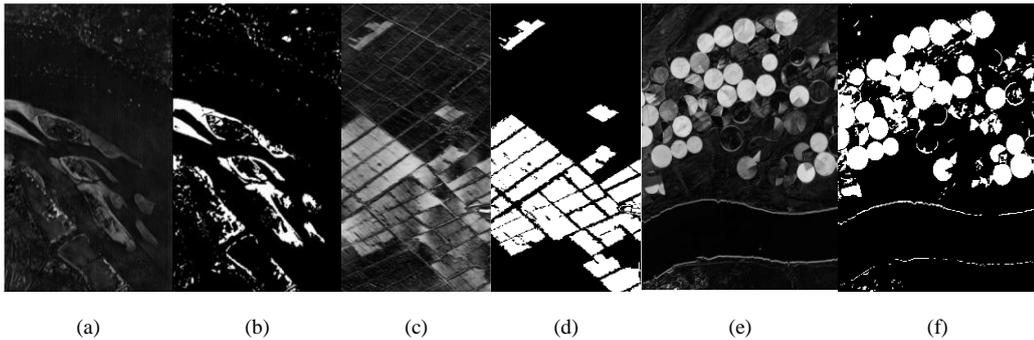


Figure 3. 3 Extracted change maps for the River (a), Yancheng (c) and Hermiston (e) datasets in comparison to the GTs in (b), (d), and (f).

3.1.5. Applying k-means for Binary Decision-making

Rather than apply the thresholding to the refined change map C_{acc} , the fourth strategy is to apply the k -means clustering method [125] to determine the final binary change map (Ω) for robustness as follows.

$$\Omega = kmeans(C_{acc}, k) \quad (3.6)$$

where $k = 2$ for binary classification to classify the change map C_{acc} into the changed foreground and the unchanged background accordingly before evaluation.

3.2 Experiments and Results

3.2.1 Results and analysis

To evaluate the efficacy of the proposed ABBD, we compare it with some classic unsupervised HCD methods, including image algebraic-based ones such as absolute distance (AD) [36], CVA [37], SAM [39], and DTW-KM [40], image transform based e.g. PCAKM [46], and IR-MAD [48], and hybrid methods e.g. patch tensor-based HCD (PTCD) [126], spectral angle weighted local AD (SALA) [127], and three-order tucker decomposition and reconstruction detector (TDRD) [59]. A brief of these selected approaches is given as follows.

- 1) AD [36]: the absolute difference between spectral values is accumulated as the change map, followed by k -means binary classification.
- 2) CVA [37]: the Euclidean distance between two spectral pixels is used to decide the change map, followed by k -means binary classification.
- 3) SAM [39]: the angle between the two spectral vectors is employed to determine the change map.
- 4) DTW-KM [40]: It gauges the similarity between two spectral vectors by calculating the minimum cumulative distance of corresponding pixels across all spectral bands.
- 5) PCAKM [46]: With PCA to reduce the data dimension and redundancy, followed by k -means clustering for binary classification of changed pixels.
- 6) IR-MAD [48]: By extracting altered objects via canonical correlation analysis to amplify the variance of independently transformed variables, it iterates the weights of distinct observations, assigning larger weights to those with fewer changes and vice versa.

- 7) PTCDD [126]: Unsupervised tensor-based method, it utilises tensor decomposition and reconstruction to mitigate the impact of various factors in bi-temporal images, incorporating a patch-based approach to leverage spatial structural information by considering non-overlapping local similarities.
- 8) SALA [127]: Spectral angle weighted local AD is used to reconstruct a discriminative feature for robustness.
- 9) TDRD [59]: After addressing the impact of diverse factors via tucker decomposition and reconstruction, it employs the singular value accumulation to identify key components in factor matrices and utilises spectral angle analysis to assess spectral changes across different domains, thereby capturing both the spectral and spatial information.

For quantitative assessments, the OA , AA , KP , Pre , Re , and $F1$ on the three datasets are compared in Tables 3.1-3.3, which have clearly shown the superiority of ABBD in all three metrics. The highest results averaged on all datasets have validated the superiority of ABBD, thanks to the binary distancing the parameter-free implementation. Detailed analysis and visual comparison of the detected change maps for each dataset are detailed below. In the change maps shown in Figures 3.4 - 3.6, white and black areas denote the correctly detected changed and unchanged pixels, whilst the false alarms and missing pixels are marked in red and green, respectively.

1) Results on the River Dataset: For the River dataset, Table 3.1 presents the indices describing the binary classification accuracy of all unsupervised methods. ABBD has produced the highest OA and KP values within the group, although the AA is slightly lower than the highest one derived from PCAKM. As seen in the detected change maps in Figure 3.4, ABBD has missed some small objects to the bottom-right side of the image than the PCAKM, though the latter seems to have more false detection. In addition, algebraic-based methods have the worse results, which has been significantly improved by the transformation-based approaches, especially for detection of large changed areas, due mainly to their inclusion of global spatial information. However, image-transformation based approaches fail to perform well in detecting the subtle sporadic changing pixels, leading to a high false alarm rate. For ABBD, it demonstrates a well-balanced detection of the changed and unchanged regions. Even without the spatial information, ABBD still outperforms all others in terms of OA and KP , the achieved KP at 0.7928 surpassing the next one by 0.0602.

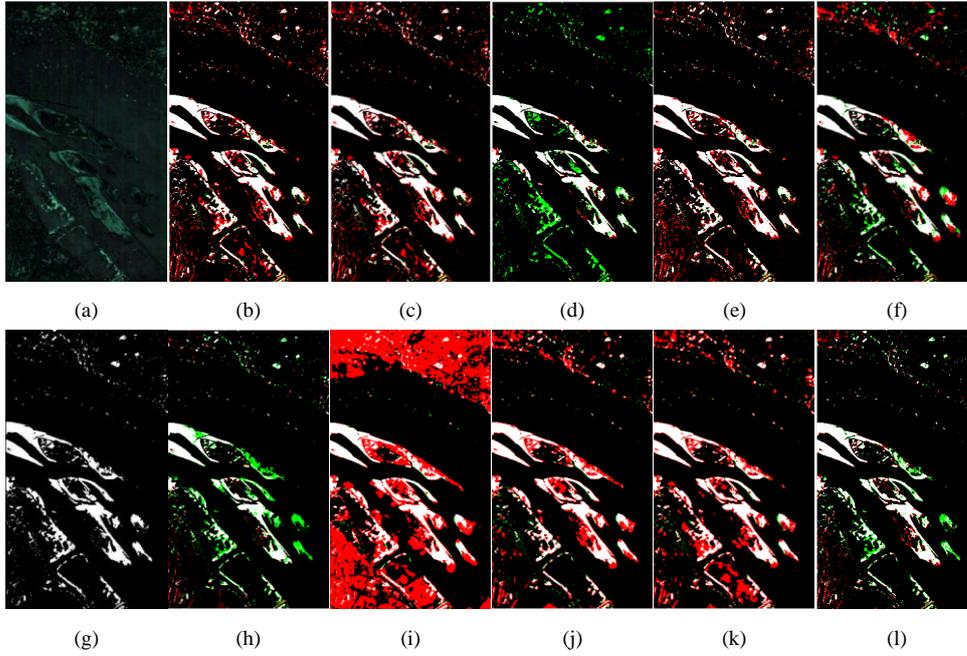


Figure 3. 4 Extracted change maps on the River dataset from different methods of AD (b), CVA (c), SAM (d), PCAKM (e), IR-MAD (f), DTW-KM (h), PTCD (i), TDRD (j), SALA (k) and our ABBD (l) in comparison to the pseudo-colour image of the difference between bi-temporal images (a) and the ground-truth map (g), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Table 3. 1 Comparisons between ABBD and unsupervised methods on the River dataset

Method	OA (%)	KP	AA (%)	Pre	Re	F1
AD	94.3092	0.7137	94.6912	0.6108	0.9515	0.7440
CVA	92.5293	0.6528	94.2577	0.5393	0.9635	0.6915
SAM	96.3041	0.7326	81.8914	0.9024	0.6445	0.7519
PCAKM	95.1686	0.7478	95.1199	0.6524	0.9506	0.7738
IR-MAD	94.4391	0.6745	86.2955	0.7644	0.6541	0.7050
DTW-KM	96.0066	0.7095	80.7302	0.8837	0.6224	0.7304
PTCD	71.1103	0.2638	82.6174	0.2269	0.9655	0.3675
TDRD	92.3142	0.6336	92.4186	0.5333	0.9254	0.6767
SALA	91.4629	0.6146	92.8807	0.5047	0.9460	0.6582
ABBD	96.3758	0.7928	93.6957	0.7377	0.9045	0.8126

2) Results on the Yancheng Dataset: For the Yancheng dataset, the extracted change maps are compared in Figure 3.5 and the quantitative assessment results are shown in Table 3.2. As seen, all algebra-based methods, including CVA, AD, SAM, and DTW-KM, yield poor outcomes with many missing detected pixels, leading to low values of *OA* (less than 88%) and *KP* (less than 0.71). In contrast, PTCD achieved the highest *KP* among unsupervised methods, as the spatial information used has improved the detection of intermediate regions. However, some non-changing pixels situated in the middle of the visual map are

misclassified, leading to an increased false alarm. This drawback prevents the accurate separation of the boundaries of each changing area, which could potentially affect the precision of the results. Similar to other methods, ABBD has the second highest values of OA and KP , lower than PTCD by 0.0129 in KP , but the missing pixels in the middle part are the least. The values of AA and Re are the highest among all benchmarks. Overall, these highlight again the efficacy of ABBD.

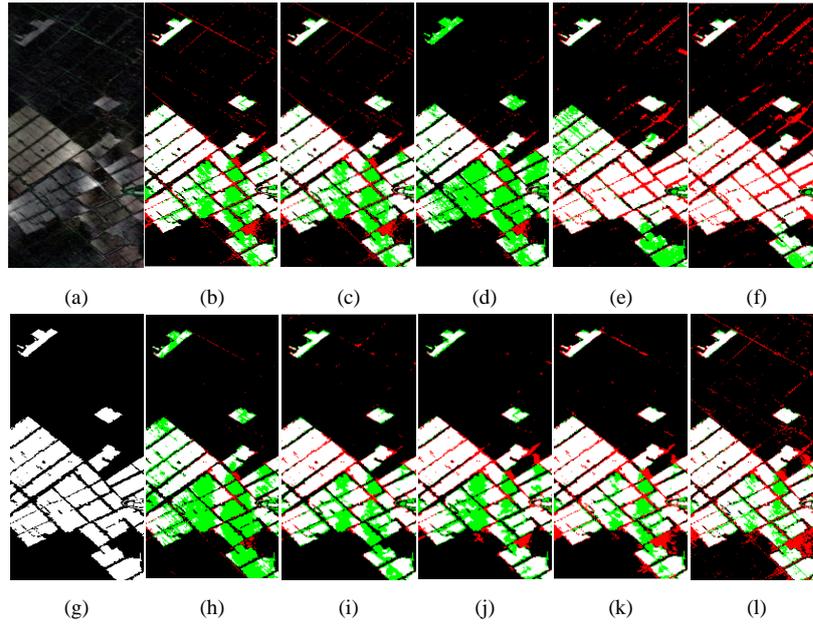


Figure 3. 5 Extracted change maps on the Yancheng dataset from different methods of AD (b), CVA (c), SAM (d), PCAKM (e), IR-MAD (f), DTW-KM (h), PTCD (i), TDRD (j), SALA (k) and our ABBD (l) in comparison to the pseudo-colour image of the difference between bi-temporal images (a) and the ground-truth map (g), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Table 3. 2 Comparisons between ABBD and unsupervised methods on the Yancheng dataset

Method	OA (%)	KP	AA (%)	Pre	Re	F1
AD	87.8027	0.7074	84.2977	0.8430	0.7494	0.7935
CVA	87.5459	0.7025	84.2058	0.8327	0.7529	0.7908
SAM	82.9694	0.5558	74.7914	0.8767	0.5297	0.6604
PCAKM	88.2789	0.7180	84.7108	0.8557	0.7519	0.8004
IR-MAD	88.7007	0.7388	87.1766	0.8119	0.8311	0.8214
DTW-KM	83.1992	0.5559	74.4774	0.9113	0.5122	0.6558
PTCD	89.9031	0.7556	86.3193	0.8945	0.7676	0.8261
TDRD	88.1735	0.7163	84.7379	0.8494	0.7557	0.7998
SALA	88.6412	0.7378	87.1896	0.8092	0.8332	0.8210
ABBD	88.7908	0.7427	87.6470	0.8053	0.8459	0.8251

3) Results on the Hermiston Dataset: For the Hermiston dataset, the visualized results are shown and compared in Figure 3.6 and the quantitative assessment results are shown in Table 3.3. The four algebraic-based methods still perform the worst, especially the SAM. Among image-transform based methods, PCAKM detected a considerable number of missing pixels, whereas IR-MAD had more false alarms. Three advanced unsupervised algorithms, PTCO, TDRD, and SALA, have shown significantly superior performance on this dataset compared to image-algebra and image-transform based methods. Notably, the SALA outperformed all others except ABBD with an OA of 96.0115% and a KP value of 0.8842. For our ABBD, it was the best among all compared methods, with an OA of 97.4874%, a KP of 0.9281 and the AA of 96.4372%, all the highest, to confirm again the superiority of our ABBD in HCD.

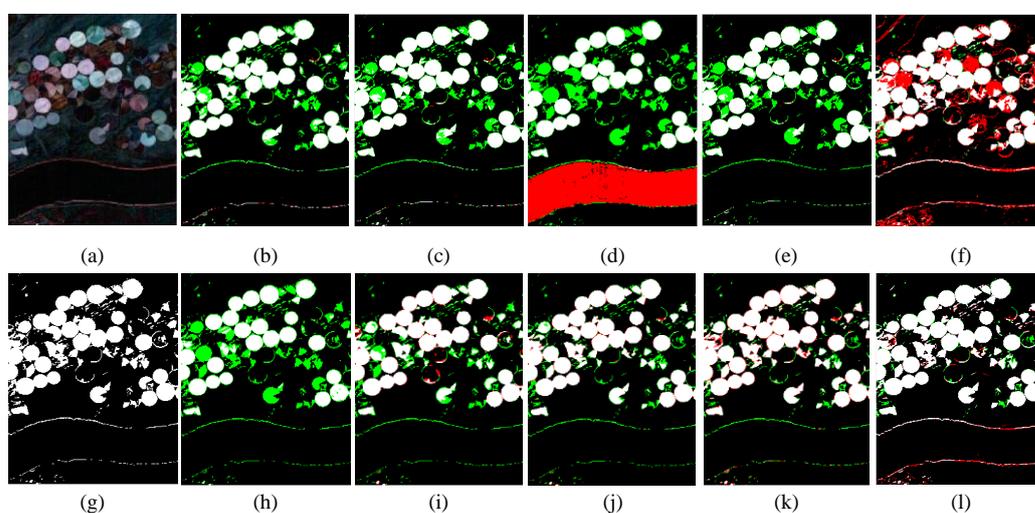


Figure 3. 6 Extracted change maps on the Hermiston dataset from different methods of AD (b), CVA (c), SAM (d), PCAKM I, IR-MAD (f), DTW-KM (h), PTCO (i), TDRD (j), SALA (k) and our ABBD (l) in comparison to the pseudo-colour image of the difference between bi-temporal images (a) and the ground-truth map (g), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Table 3. 3 Comparisons between ABBD and unsupervised methods on the Hermiston dataset

Method	OA (%)	KP	AA (%)	Pre	Re	F1
AD	93.4146	0.7904	85.5461	0.9941	0.7122	0.8298
CVA	92.8663	0.7705	84.2897	0.9953	0.6867	0.8127
SAM	72.0397	0.2888	66.2565	0.4112	0.5573	0.4733
PCAKM	92.2378	0.7472	82.8444	0.9973	0.6574	0.7924
IR-MAD	89.7509	0.6533	77.6549	0.9806	0.5563	0.7099
DTW-KM	90.2834	0.6716	78.4514	0.9997	0.5691	0.7253
PTCD	94.2030	0.8226	88.4859	0.9536	0.7808	0.8586
TDRD	95.8817	0.8765	91.8463	0.9683	0.8450	0.9024
SALA	96.0128	0.8843	93.6210	0.9277	0.8927	0.9098
ABBD	97.4874	0.9281	96.4372	0.9434	0.9453	0.9443

3.2.2 Further discussion of the quantitative results

Overall, the quantitative assessment results across three datasets shed light on the strengths and weaknesses of various unsupervised methods for HCD. One notable observation is that isolated pixels pose a significant challenge for image-transformation and tensor-based methods. Relying heavily on spatial features, it makes these approaches less effective in dealing with isolated pixels that exhibit changes, leading to high false alarms when the neighbouring pixels are considered. Consequently, the detection accuracy of PCAKM and IR-MAD is not as high as that of image-algebraic methods on datasets with a substantial number of isolated changing pixels.

On the contrary, the relatively poorer results of ABBD on the Yancheng dataset is primarily due to the following reasons: 1) As seen from the GT map, the changed areas in the Yancheng dataset are all connected large blocks, rather than loosely distributed in the other two images, the accurate classification of unchanged pixels within the blocks and the pixels along the edges of the blocks poses a challenge for this dataset; 2) As shown in the pseudo-colour image of the difference image, many dark areas in the blocks mean that there are subtle changes marked as changes, where the edges of each block in the upper right corner are marked as unchanged even they appear very bright. These are the challenges of the dataset itself that present the detection accuracy of unsupervised algorithms. From the comparison of visual maps, like other image algebra-based methods, the ABBD method generates a significant number of false positives in the top right corner. This is because these methods do not consider spatial features, resulting in less effective handling of edge regions compared to methods based on image transformations, such as PCAKM and PTCD

methods. However, compared to image algebra-based methods, ABBD handles subtle changes in pixels within connected blocks quite well, accurately detecting subtle changes. It is evident that ABBD's missing pixel count is much lower than other unsupervised algorithms. This is why ABBD's detection accuracy is significantly better than other algebraic-based methods, although it is slightly inferior to image transformation-based methods.

As for other advanced methods e.g. PTCD, TDRD, and SALA, they significantly outperform conventional unsupervised algorithms on the Yancheng and Hermiston datasets though the robustness is relatively poor. Moreover, their performance on the River dataset is notably deficient, falling below the performance of all benchmark tests. As an unsupervised method based on image algebra without considering the spatial information, ABBD has produced significantly improved results in quantitative assessment across all three datasets. This is mainly due to the proposed binary distancing and parameter-free decision-making to mitigate the measurement inconsistency in handling both subtle and obvious spectral changes. The highest averaged OA , KP and AA have fully validated its efficacy in HCD, especially when processing datasets with a large number of isolated changing pixels.

As shown in Figures 3.4 – 3.6 (b-c), image algebra-based methods such as AD and CVA that use the absolute difference to measure the degrees of changes tend to produce a much higher level of false alarms than ABBD. This is mainly owing to the proposed binary distancing, which has successfully mitigated the adverse effect of noise-induced inconsistency of measurement and resulted in a significant reduction of false alarms. Along with the k -means-based binary classification on the cumulated band-wise binary change map, more accurate and robust HCD has been achieved on all these three datasets.

3.3 Further Discussions

To comprehensively validate the effectiveness of our proposed method, we conduct a series of experiments covering computational complexity, threshold adaptive iteration, classifier selection, and comparison with results obtained from advanced DL-based algorithms.

3.3.1 Computational complexity analysis

Table 3. 4 Computational complexity of our ABBD and the classic unsupervised HCD methods

	AD	CVA	SAM	PCAKM	IR-MAD
Parameters	0	0	0	1	3
Complexity	$O(L*B)$	$O(L*B)$	$O(L*B)$	$O(L*B^2 + B^3 + LBp)$	$O(L*B^2 + L*B^3)$
Computation time (s)	1.0463	1.0650	1.7159	1.9782	3.5046
	DTW-KM	PTCD	TDRD	SALA	ABBD
Parameters	0	2	1	0	0
Complexity	$O(L*B^2+L)$	$O(L*S^2*B + L^3)$	$O(L*B^3 + L*B)$	$O(L*B)$	$O(N*L*B)$
Computation time (s)	1.7601	56.4090	29.6332	1.2889	1.1912 ($N = 1$)

Herein we analyse the computational complexity of all unsupervised methods briefly in Table 3.4, where L denotes the total number of pixels ($L=W*H$), p is the number of the principal components for dimension reduction, S is the patch size and N represents the number of iterations.

In general, algebraic operations are much simpler than image-transformation-based ones. AD, CVA, SAM, and SALA have the least computational complexity because they rely on the pixel-wise difference between the spectral vectors. Due to the need for sequence alignment in the spectral domain, DTW-KM has a much higher computational complexity.

For the image-transformation-based methods, such as PCAKM and IR-MAD, their computational complexity is much larger than the algebraic operations, due to the need to calculate the difference or correlation in the transformed domain. PTCD involves the segmentation of the HSI into different patches before applying feature extraction on the constructed tensor. TDRD requires a three-order Tucker decomposition of the high-dimensional HSI, followed by feature extraction from the core tensor or mode matrices. Additionally, it needs to reconstruct the detected change regions from the feature space back to the original image space. Consequently, these two methods are most complex within the group. For ABBD, it has the same lowest computational complexity as the image-algebraic methods when ε is set to a fixed value ($N=1$). With the increasing of iterations, the computational cost of ABBD will grow linearly.

We have also compared the computation times of all algorithms on the River dataset, which are summarized in Table 3.4. All algorithms were executed based on MATLAB on an NVIDIA RTX A2000, and their parameters were set according to the specifications provided in the original paper. To ensure experimental fairness, the computation time reported only includes the runtime of the algorithms and excludes data loading time. From the comparison result, it can be observed that methods based on algebraic operations

generally require significantly less time compared to those based on image transformations. For ABBD, it also has a similar runtime compared to other algebraic-based methods when ε is set to a fixed value ($N=1$). However, as N increases, the runtime will grow linearly. How to reduce the number of loops based on the determined N will be further investigated.

In addition, the number of control parameters in these approaches is analysed for comparison. Image algebra-based methods, including AD, CVA, SAM, and SALA, inherently extract and compare spectral differences, obviating the need for control parameters. On the contrary, image transformation-based methods exhibit varying sensitivities to the parameter settings. Specifically, PCAKM's detection outcomes depend on the number of principal components used, while IR-MAD is impacted by three key parameters: the maximum number of iterations, thresholds, and penalty terms. PTCDD's detection accuracy hinges on factors such as the window size and inter-block stride, while TDRD's performance is affected by the rate parameter between bi-temporal images. Thanks to the capability embedded within the ABBD that can automatically determine the associated parameters. This has enabled ABBD as a parameter-free solution for extra robustness and ease of deployment whilst producing high accuracy results of HCD.

3.3.2 Effect of binary k-means vs. thresholding

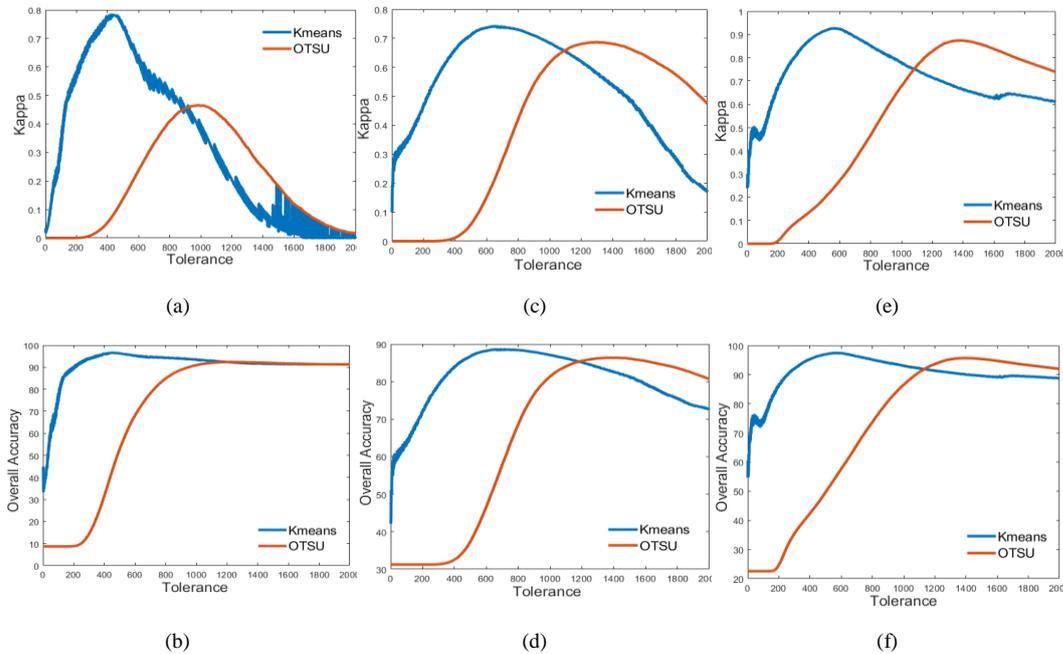


Figure 3. 7 The variation comparison under different values ε : *KP* on the River (a), Yancheng (c), and Hermiston dataset (e); and *OA* comparison on the River (b), Yancheng (d), and Hermiston (f).

Based on the accumulated change map, there are two ways to make the final binary decision in HCD, i.e., k -means clustering and thresholding, while maintaining the parameter ε spanning within [1, 2000] to explore the full range of the variations. The corresponding results on the three datasets are given in Figure 3.7, where OTSU was used to determine the optimal threshold for the accumulated change map at each ε . As seen, all the KP and OA curves exhibited an initial ascent followed by a decline as ε increases. This is due to the fact that a larger ε will put a higher threshold to detect changed pixels. In other words, it will result in more missing detection hence the declined KP towards zero along with a stabilized OA .

It's highlighting that k -means consistently outperformed OTSU thresholding in both the KP and OA . The optimal ε for k -means turned out to be considerably smaller than that for OTSU. This finding underscores the efficiency advantage of k -means clustering over OTSU thresholding while delivering superior results of HCD. Hence, in the ultimate selection of the binary classifier, we opted for k -means clustering. In addition, the quantitative assessment is summarized in Table 3.5. As seen, the maximum KP achieved by thresholding with a fixed ε falls short of the k -means binary classification, indicating the efficacy of k -means vs. thresholding.

Table 3.5 Results of using a fixed tolerance threshold

	River		Yancheng		Hermiston	
	k -means	OTSU	k -means	OTSU	k -means	OTSU
OA (%)	96.5398	92.3621	88.6497	86.3452	97.4631	95.7128
KP	0.7791	0.4713	0.7383	0.6868	0.9269	0.8747
Optimal ε	453	981	679	1300	569	1385

3.3.3 Effect of the adaptively determined optimal N

From Eq. (3.2), ABBD can automatically derive the optimal values of N as 1055, 1319, and 978 for the three datasets of the River, Yancheng, and Hermiston, respectively. In addition, to further validate the efficacy of the automatically determined parameter N , we compare the detection results from this N with those manually determined optimal values of N according to the best detection accuracy, as shown in Figure 3.8. Note that the manually determined values of N are 1999, 1296, and 1053, in comparison to the

automatically derived values of 1055, 1319, and 978 from the three datasets of River, Yancheng, and Hermiston, respectively. The OA and KP yielded on the three datasets, using both the manually and automatically determined N , are compared in Table 3.6. Although the N values can be much different, the produced OA and KP are very close to each other, which has validated the efficacy of the adaptive solution in determining the N .

Interestingly, the comparison of results revealed that the OA and KP values attained at the optimal N values were slightly improved when compared to the adaptively determined N , although the differences were relatively small. This outcome further validates the efficacy of our ABBD in determining the appropriate N values for the iterative process and achieve truly parameter-free unsupervised HCD that is applicable to different datasets.

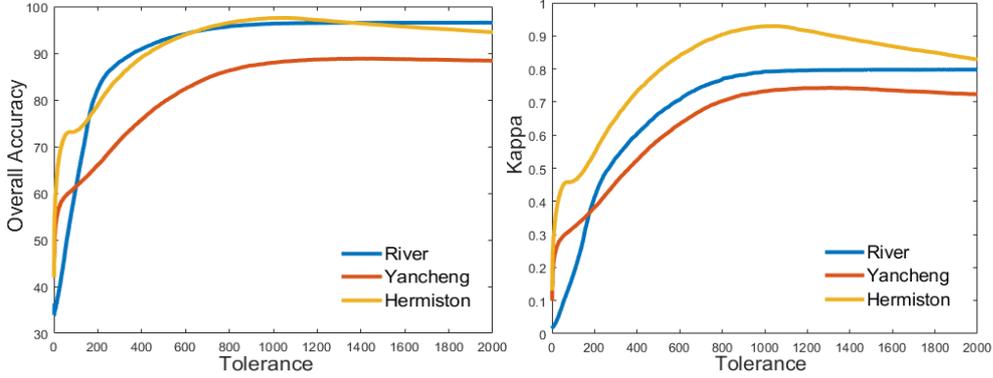


Figure 3. 8 Results of the OA and KP vs. an increasing N

Table 3. 6 Results under manually and automatically determined values of N

	River		Yancheng		Hermiston	
	Manual	Auto	Manual	Auto	Manual	Auto
N	1999	1055	1296	1319	1053	978
OA	0.9654	0.9638	0.8884	0.8879	0.9757	0.9749
KP	0.7981	0.7928	0.7433	0.7427	0.9296	0.9281

3.3.4 Comparison with supervised deep learning-based algorithms

For extended performance assessment, we compare ABBD with several supervised deep learning-based models, including 2-D CNN [128], 3-D CNN [129], CSANet [91] and

CBANet [1]. We conducted the training using two distinct subsets, i.e., 5% and 10% of the labelled pixels, respectively, while using the remaining for testing. The training samples were randomly selected each time, and the averaged *OA* and *KP* in three runs were used for comparison, as shown in Table 3.7.

For the River dataset, with a training ratio of 5%, ABBD outperformed all the compared DL models in both *OA* and *KP*. When increasing the training ratio to 10%, ABBD was beaten by CBANet and CSANet, though it has comparable or even slightly better results than 3-D CNN and 2-D CNN, especially in *KP*. For the Yancheng dataset, irrespective of using 5% or 10% of pixels for training, all DL-based models surpass ABBD and other unsupervised methods by a large margin. This was due to the removal of a large number of noisy bands from the dataset, resulting in a weakened representation of the handcrafted features hence the low detection accuracy of the unsupervised algorithms [130]. For the Hermiston dataset, ABBD outperforms all DL-based methods in both *OA* and *KP* when the training ratio is up to 10%.

ABBD has shown much worse results on the Yancheng dataset than the River and Hermiston, which can be explained as follows. In both the River and Hermiston datasets, the changed areas include a complex landscape of interconnected regions and numerous isolated pixels. Addressing the distinctions among these isolated pixels in the spectral domain is the strength of ABBD. Thanks to the proposed binary distancing and adaptive solution, ABBD outperforms those that solely rely on spatial characteristics and neglect the pixel-level spectral features, such as 2-D CNN, 3-D CNN, and CSANet, as they focus on image-level operations to exploit the spatial correlation [131]. Therefore, they show remarkable detection accuracy when handling changed pixels within connected regions on both datasets, but not the isolated small regions. This also explains why these models exhibited clear advantages of better results when applied to the Yancheng dataset, which contained a substantial number of interconnected areas of changes. On the contrary, CBANet's proficiency in addressing isolated pixels of changes becomes more pronounced as it leverages 1×1 convolutional layers to consider pixel-level spectral features, which amplifies with an increasing training ratio. As changed regions can be of various sizes, this has shown the value and importance of both spectral and spatial features in HCD when applying different approaches including DL-based methods.

Table 3. 7 Comparing ABBD with supervised deep learning methods

Models	Train ratio	River		Yancheng		Hermiston	
		<i>OA</i>	<i>KP</i>	<i>OA</i>	<i>KP</i>	<i>OA</i>	<i>KP</i>
3-D CNN	5%	0.9601	0.7364	0.9558	0.8977	0.9505	0.8588
2-D CNN		0.9579	0.7261	0.9497	0.8839	0.9443	0.8375
CSANet		0.9670	0.7786	0.9619	0.9112	0.9547	0.8638
CBANet		0.9609	0.7518	0.9601	0.9068	0.9673	0.9045
ABBD	0%	0.9638	0.7920	0.8878	0.7418	0.9749	0.9288
3-D CNN	10%	0.9647	0.7706	0.9649	0.9183	0.9575	0.8803
2-D CNN		0.9644	0.7661	0.9591	0.9050	0.9486	0.8490
CSANet		0.9706	0.8128	0.9692	0.9290	0.9512	0.8661
CBANet		0.9769	0.8344	0.9657	0.9213	0.9727	0.9218

3.4 Summary

A novel accumulated band-wise binary distancing (ABBD) for unsupervised parameter-free hyperspectral change detection is proposed in this chapter. The proposed four strategies of ABBD have helped to significantly improve the efficiency and robustness of ABBD in HCD. First, band-wise binary distancing can successfully mitigate the measurement inconsistency. Second, it is found that the k -means used for binary decision-making have surpassed thresholding using OTSU. Third, the adaptive solution is found particularly useful in automatically determining the parameter N for achieving a fully parameter-free approach in HCD. With these strategies, ABBD has outperformed a number of state-of-the-art approaches including several deep learning models when the training ratio is relatively low.

Chapter 4 Supervised Deep Learning-based Methods for Hyperspectral Change Detection

4.1 Introduction

In recent years, advancements in DL technology have provided innovative approaches for addressing the complexities of HCD tasks. DL-based models are known for their powerful feature learning capabilities, which can automatically extract feature representations to effectively capture complex spatial and spectral information. This capability significantly enhanced the detection accuracy and robustness of HCD methods. However, as summarised in Chapter 2, most current DL-based networks face several challenges. Firstly, these networks typically require large amounts of manually labelled data for training. Secondly, they often have an excessive number of hyperparameters, making the optimization process complex and computationally expensive. Therefore, developing a lightweight DL-based network that can achieve higher detection accuracy with fewer hyperparameters, and less labelled training data remains a key focus for researchers.

In this chapter, two DL-based networks are proposed to address these challenges in the HCD task. Initially, an end-to-end cross-band 2-D attention network (CBANet), incorporates a 2-D self-attention module designed to extract local spatial-spectral features. The 2-D self-attention module is integrated with cross-band feature extraction module, which yields higher detection accuracy, and requires fewer hyperparameters. The effectiveness and efficiency of CBANet have been thoroughly validated through comprehensive experiments, showcasing its superiority over several state-of-the-art approaches based on self-attention mechanisms. However, experimental analysis revealed that CBANet has limitations in accurately detecting edge pixels of changed areas. To address this limitation, a singular spectrum analysis-driven lightweight network (SSA-LHCD) is developed, building upon the 2-D self-attention module. SSA-LHCD integrates the singular spectrum analysis, residual blocks, and an additional 1×1 convolutional layer. These enhancements improve the spatial feature extraction, enhance the pixel-wise detection accuracy, and further reduce the number of the network's hyperparameters. Sections 4.2 and 4.3 provide detailed descriptions of the proposed CBANet and SSA-LHCD network, respectively, along with comparisons to other advanced HCD algorithms,

highlighting the efficiency and robustness of these two proposed 2-D self-attention-based networks.

4.2 CBANet: An End-to-End Cross-Band 2-D Attention Network for Hyperspectral Change Detection in Remote Sensing

In this section, a lightweight deep learning network, namely CBANet, is proposed, which fuses the cross-band module for extracting spectral domain features pixel-by-pixel and designs a new 2-D attention module based on traditional self-attention mechanisms for improved extraction of local spatial-spectral features whilst keeping the network compact for efficiency. The major contributions are summarised as follows.

- 1) A cross-band feature extraction module is proposed to extract the mutual and representative features from bi-temporal hypercubes, where a 1×1 convolutional layer is introduced to greatly increase the non-linear characteristics (using the subsequent activation function) of the feature map while keeping the scale of the feature map unchanged.
- 2) A 2-D self-attention module is proposed for focused extraction of local spatial-spectral features and improved feature representation and discrimination capability, resulting in enhanced network reliability.
- 3) A novel end-to-end lightweight CBANet is proposed which can produce higher detection accuracy but has fewer hyperparameters. Its efficacy and efficiency have been fully validated in comprehensive experiments when compared with a few state-of-the-art approaches.

The remainder of this section is organized as follows. Section 4.2.2 describes the details of the proposed CBANet. Section 4.2.3 presents the experimental results and assessments. Finally, some remarkable conclusions of CBANet are summarized in Section 4.2.4.

4.2.1 Methodology

The diagram of the proposed CBANet is presented in Figure 4.1, which is composed of three main modules, i.e., 1) cross-band spectral feature extraction; 2) spectral-spatial feature extraction, and 3) 2-D self-attention based deep feature extraction. The specific details of the network are shown in Table 4.1.

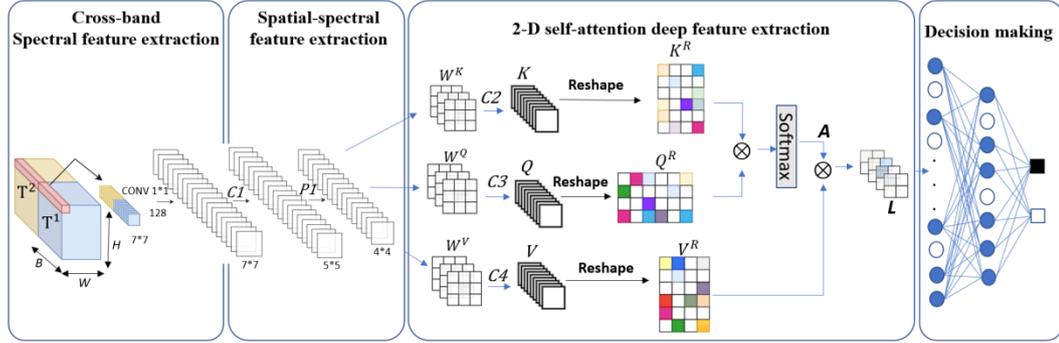


Figure 4. 1 The architecture of the proposed CBANet model

Table 4. 1 Architecture details for proposed CBANet model

Layers	Type	No. Kernel	Size
Input	-	$B*2$	-
$Conv1$	Conv2D+BN	128	1×1
$C1$	Conv2D+BN	128	3×3
$P1$	Average Pooling	-	2×2
$C2$	Conv2D+BN	32	3×3
$C3$	Conv2D+BN	32	3×3
$C4$	Conv2D+BN	32	3×3
Flatten	Flatten	512	-
$FC1$	Linear (Dropout=0.4)	64	-
$FC2$	Linear (Dropout=0.4)	16	-
$FC3$	Linear (Dropout=0.4)	2	-

A. Cross-band spectral feature extraction

Given a pair of spatially aligned bi-temporal hypercubes $T^1 \in \mathfrak{R}^{W*H*B}$ and $T^2 \in \mathfrak{R}^{W*H*B}$, where W and H denote the width and height of the spatial size, and B represents the number of spectral bands. T^1 and T^2 are first concatenated to form a new hypercube $Q \in \mathfrak{R}^{W*H*2B}$, which will be divided into a group of overlapped 3-D neighboring patches denoted as $Z_{(\alpha,\beta)} \in \mathfrak{R}^{S*S*2B}$, where S is the spatial size of Z , (α, β) denote the coordinates of the patch centre in the spatial domain where $\alpha \in [1, W], \beta \in [1, H]$. Firstly, padding is added to the T^1 and T^2 to increase the spatial dimensions, ensuring that the window S fits even on the edges. The total number of 3-D patches from Q will be $(W - S + 1) \times (H - S + 1)$. For each patch $Z_{(\cdot)}$, the whole spectral vector may contain highly redundant information and cause huge computational cost in training the deep learning model. Thus, reducing the data dimension whilst keeping the discriminative information in the spectral domain becomes the key issue here. For this purpose, a 1×1 convolutional layer [132] with a proper setting of k_{Conv1} is applied to the dual spectral bands $Z_{(\cdot)}$. the weighted fusion across the whole spectral vector can help to compose a new feature fusion space with a much lower spectral dimensionality. Meanwhile, the input patch size S of the proposed methods is set to 7×7 and the number of kernels k_{Conv1} in the cross-band feature extraction module is set to 128, as it can achieve a good balance between the computational efficiency and the retained the principal components.

B. Spectral-spatial feature extraction

The low-dimensional feature cube constructed after extracting the spectral features by passing through the cross-band fusion module, which is a 1×1 convolutional layer to preserve the characteristics of the bi-temporal cubes and remove redundant information. In the next step, a 2-D convolutional kernel is employed for global feature extraction in the spatial domain. The convolution sums up the dot product between the input feature map and the kernel. The 2-D kernels are stride over the input feature map to cover the entire spatial domain. The convolutional results with an adding on additional bias will pass through a ReLU function. In 2-D convolution, the j^{th} feature map in the i^{th} layer at position (x, y) is denoted as $F_{i,j}^{x,y}$, which is calculated as follows:

$$F_{i,j}^{x,y} = \text{ReLu} \left(\text{BN}(b_{i,j}) + \sum_r \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{i,j,r}^{p,q} F_{i-1,r}^{x+p,y+q} \right) \quad (4.1)$$

where P_i , Q_i are the height and width of the 2-D kernel, $b_{i,j}$ is the bias, $w_{i,j,r}^{p,q}$ is the weight parameter at the position (p, q) of the kernel connected to the r^{th} feature map, where r represents the set of feature maps in the $(i-1)^{\text{th}}$ layer connected to the i^{th} layer [131]. $\text{ReLu}(\cdot)$ is the Rectified Linear Unit [133] as an activation function to introduce the nonlinearity, reduce parameter interdependence and alleviate overfitting $\text{BN}(\cdot)$ represents the batch normalization function. In this module, 2-D convolution with a kernel size of 3×3 is used in order to reduce the network parameters as well as extract more representative local information. Afterwards, 2×2 sub-sampling average pooling is adopted to prevent feature from rotation and scale during convolution [134]. The extracted spectral-spatial features are represented as $X \in \mathfrak{R}^{h \times h \times k_{C1}}$, where $h = 4$ after pooling. The number of kernels k_{C1} for spectral-spatial feature extraction is set to 128, as it reaches a good trade-off between the classification accuracy and robustness.

C. 2-D self-attention based deep feature extraction

Previous studies have found that the self-attention mechanism is beneficial to conventional change detection tasks [135] and HSI classification [136][137]. However, these self-attention models use the 1×1 convolutional kernel and focus on pixel-wise band features to assign the pixel-wise weights and only pay attention to the spectral information, leading to insufficient detection performance especially when dealing with the changing areas in various sizes. Motivated by this issue and inspired by the work in [138], we propose a 2-D self-attention module to build adjacent pixels dependency in local space as well as enhancing the spatial-spectral features from middle-level towards deeper-level. The feature X is taken as the input and fed into three 3×3 2-D convolutional layers ($C2, C3, C4$) to generate three new spatial feature maps, denoted as Query (Q), Key (K), and Value (V), where $(K, Q, V) \in \mathfrak{R}^{m \times m \times k_{C2}}$, we set $k_{C2} = k_{C3} = k_{C4} = 32$ in this study. Each feature map will be converted to 2-D attention matrices denoted as K^R, Q^R, V^R , respectively, where $(K^R, V^R) \in \mathfrak{R}^{m^2 \times k_{C2}}$ and $Q^R \in \mathfrak{R}^{k_{C2} \times m^2}$. Then the correlation can be obtained by the dot product of the attention matrix K^R and Q^R , from the properties of the dot product, the higher similarity between the two matrices, the value of the dot product will be larger which represents the more obvious the local change feature and will be assigned a greater weight.

The spatial attention matrix A is calculated by multiplication between K^R and Q^R followed by the *Softmax* operations.

$$A = \text{Softmax}(K^R * Q^R) \quad (4.2)$$

Finally, the 2-D attention feature map $L \in \mathfrak{R}^{m*m*kc2}$ can be obtained by multiplying A by V^R . In this process, all local features are involved in the calculation, therefore, 2-D self-attention not only can capture the global feature distribution, but also focus on the key changing features. The larger the weight value in feature map L , the more prominent the feature.

Since change detection can be considered as a binary classification problem of distinguishing the change and non-change pixels, the cross entropy, which is commonly used for classification, is adopted as the loss function.

$$Loss_{(pred, label)} = -\frac{1}{u} \sum_{i=1}^n (l * \log(p) + (1 - l) * \log(1 - p)) \quad (4.3)$$

where u denotes the number of samples, l represents the ground truth value where 0 and 1 represent unchanged and changed regions. p represents the probability predicted by the Linear function. The selected optimizer is the adaptive momentum (Adam) [139] with the initial learning rate of 0.0001.

4.2.2 Experiments and Results

In this section, we evaluate the effectiveness of the proposed CBANet by comparing it with three classical unsupervised methods, which include the absolute distance (AD) [30], change vector analysis (CVA) [31], and the principal component analysis (PCAKM) [40] as well as several deep-learning (DL) based methods such as 2-D-CNN [128], 3-D-CNN [129], HybridSN [35], CSANet [91] and Long-short-term-memory (LSTM) [48]. It is worth noting that the compared methods except CSANet will need to take the difference of the given HSI pairs as input, which may thus break the continuity of the spectral features. Thanks to the cross-band fusion module used, such image differencing is not needed for our proposed end-to-end network. The proposed CBANet and all other DL-based methods are trained based on the PyTorch on an NVIDIA RTX A2000, with the batch size set to 32 and the number of training epoch as 500. We randomly select 20% pixels in the changed and unchanged pixels as the training set, and the remaining for testing. To make a fairer

and more reliable comparison, all DL algorithms are repeated ten times in each experiment, and the averaged results with the standard deviations are reported. In the produced change maps, false alarms and missing pixels are marked in red and green respectively for ease of comparison, white areas represent correctly detected and black area for true negatives.

Experiments on the River Dataset

The experimental results from the River dataset are shown in Figure 4.2 and Table 4.2. As seen in the ground-truth map in Figure 4.2 (j), the most obvious differences are the differently shaped sediment accumulations in the river and the land-cover changes on the riverbank, in addition to many others. In Figure 4.2 (a-c), most of the non-changed pixels are detected as false alarms, distributed in the upper and lower left corners of the maps, are wrongly detected as changed areas by all unsupervised algorithms. However, most false alarms can be correctly classified by the DL-based algorithms. In the regions in the upper left corner of the maps, although most changing pixels can be distinguished by the 3-D-CNN and 2-D-CNN in Figure 4.2 (f-g), some sporadic changing pixels are still not identified, due possibly to that both these DL-based models only extract the relationship between local and global spatial-spectral features but ignoring the changing features of the independent pixels in the spatial domain. The CSANet has produced the second highest OA, CA, AA, and KP values among all compared DL-based models, only slightly worse than our CBANet, owing to the joint spatial-spectral-temporal features extracted by the introduced self-attention module. Also, our CBANet has a much higher CA than the CSANet in effective detection of the changed pixels whilst maintaining the same or even slightly lower level of false alarms as measured by NCA. Thanks to our cross-band fusion module and the 2-D self-attention module, both sporadic changing pixels and large regions can be accurately detected.

As shown in Table 4.2, not surprisingly, the DL based supervised methods all have higher OA and KP and outperform the unsupervised ones. As for CA and AA, however, DL methods seem inferior to unsupervised ones, due mainly to the fact that the changing pixels have the characteristics of wide distribution and various scales. Note that CVA, AD and PCAKM are all pixel-wise methods, which do not consider the influence of adjacent pixels thus are more sensitive to noise. Therefore, they tend to misclassify the changing pixels, resulting in a low NCA. On the contrary, DL algorithms are more accurate in distinguishing small changes. In the benchmarked DL methods, LSTM has the worst performance with an average KP of 0.7261 and OA of 95.69%. Our proposed CBANet has produced the highest

OA, KP and NCA over all compared methods, achieving the highest CA value over all DL methods, which indicates the correct detection of changing areas in various sizes.

Table 4. 2 Quantitative assessment of different methods on the River dataset

	OA	CA	NCA	AA	KP
AD	0.9431	0.9423	0.9515	0.9469	0.7137
CVA	0.9253	0.9217	0.9635	0.9425	0.6528
PCAKM	0.9517	0.9518	0.9505	0.9511	0.7476
LSTM	0.9569±0.0011	0.7671±0.0074	0.9746±0.0019	0.8704±0.0038	0.7216±0.0070
HybridSN	0.9671±0.0019	0.7605±0.0298	0.9867±0.0043	0.8736±0.0130	0.7826±0.0087
3-D-CNN	0.9700±0.0008	0.7888±0.0299	0.9871±0.0036	0.8879±0.0124	0.8045±0.0053
2-D-CNN	0.9682±0.0007	0.8346±0.0118	0.9806±0.0021	0.9083±0.0063	0.7946±0.0033
CSANet	0.9743±0.0012	0.8623±0.0049	0.9847±0.0009	0.9235±0.0094	0.8360±0.0049
CBANet	0.9765±0.0012	0.8800±0.0110	0.9865±0.0018	0.9308±0.0065	0.8526±0.0036

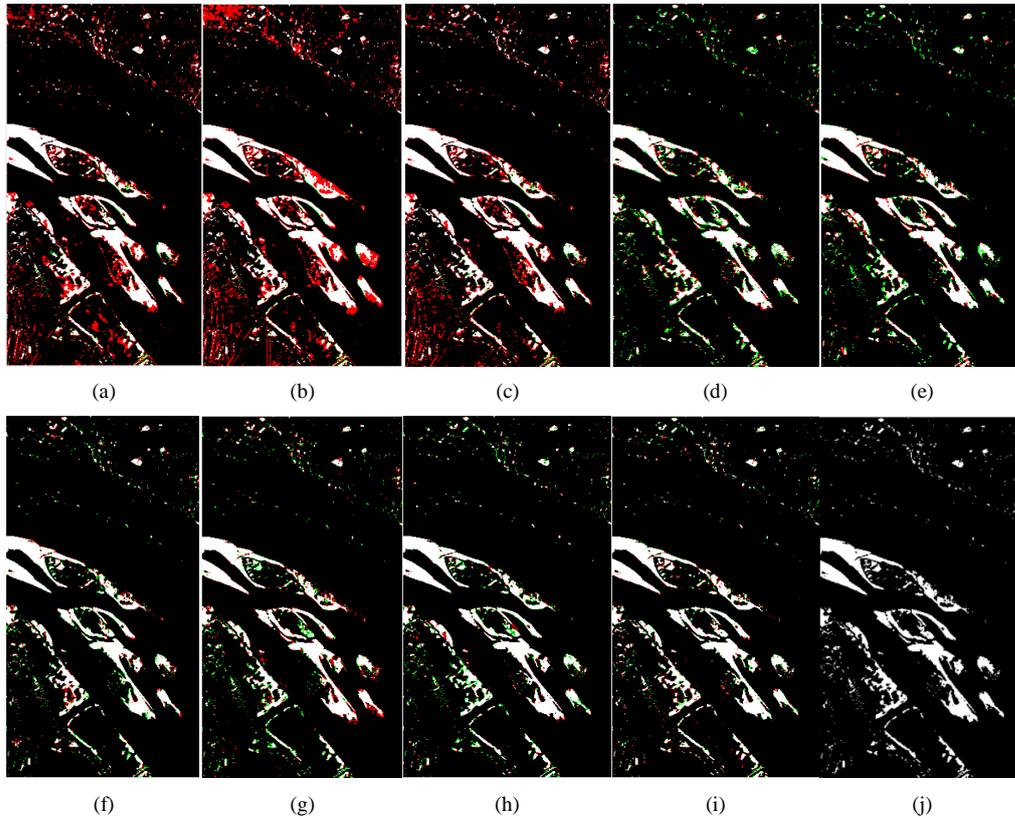


Figure 4. 2 Extracted change maps on the River Dataset from different methods of AD (a), CVA (b), PCAKM (c), LSTM (d), HybridSN (e), 3-DCNN (f), 2-DCNN (g), CSANet (h) and CBANet (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Experiments on the Yancheng Dataset

According to the HCD results shown in Figure 4.3 and Table 4.3, the primarily changing area in this dataset is land-cover on wetlands, see in Figure 4.3 (j). Again, all three unsupervised methods have quite poor results as shown in Figure 4.3 (a-c), where many changing pixels are missed along with false alarms in striped lines and other occasions, resulting in low values of KP at around 0.71 and OA less than 90% (Table 4.3). Obviously, all the DL-based methods have outperformed the unsupervised ones, as these are region-wise classification methods and more robust to spatial noise than pixel-wise unsupervised approaches. Although the OA from LSTM and HybridSN is relatively high, their CA is even lower than that of the unsupervised methods, leading to poor detection results in Figure 4.3 (d-e), especially the boundaries of the changing areas. For 2-D-CNN and 3-D-CNN, they have produced the similar OA and KP as LSTM and HybridSN, though the visual results seem slightly better, although the connected changing region in the middle of the maps cannot be well distinguished. CSANet has yielded almost the same OA, AA, and KP as our proposed CBANet, where both of them are the top-performed models. However, our CBANet has a higher CA than the CSANet in the detection of the changed pixels, whilst the false alarms as indicated by NCA remains very comparable.

Table 4.3 Quantitative assessment of different methods on the Yancheng dataset

	OA	CA	NCA	AA	KP
AD	0.8780	0.7494	0.9365	0.8429	0.7074
CVA	0.8755	0.7529	0.9312	0.8421	0.7025
PCAKM	0.8828	0.7519	0.9424	0.8471	0.7180
LSTM	0.9555±0.0010	0.9246±0.0042	0.9702±0.0011	0.9472±0.0016	0.8967±0.0023
HybridSN	0.9641±0.0021	0.9350±0.0191	0.9790±0.0042	0.9555±0.0052	0.9160±0.0055
3-D-CNN	0.9665±0.0015	0.9427±0.0058	0.9774±0.0016	0.9601±0.0025	0.9221±0.0035
2-D-CNN	0.9667±0.0014	0.9413±0.0037	0.9781±0.0016	0.9603±0.0026	0.9223±0.0030
CSANet	0.9715±0.0009	0.9584±0.0015	0.9774±0.0020	0.9677±0.0003	0.9335±0.0023
CBANet	0.9713±0.0006	0.9605±0.0070	0.9768±0.0041	0.9679±0.0019	0.9332±0.0014

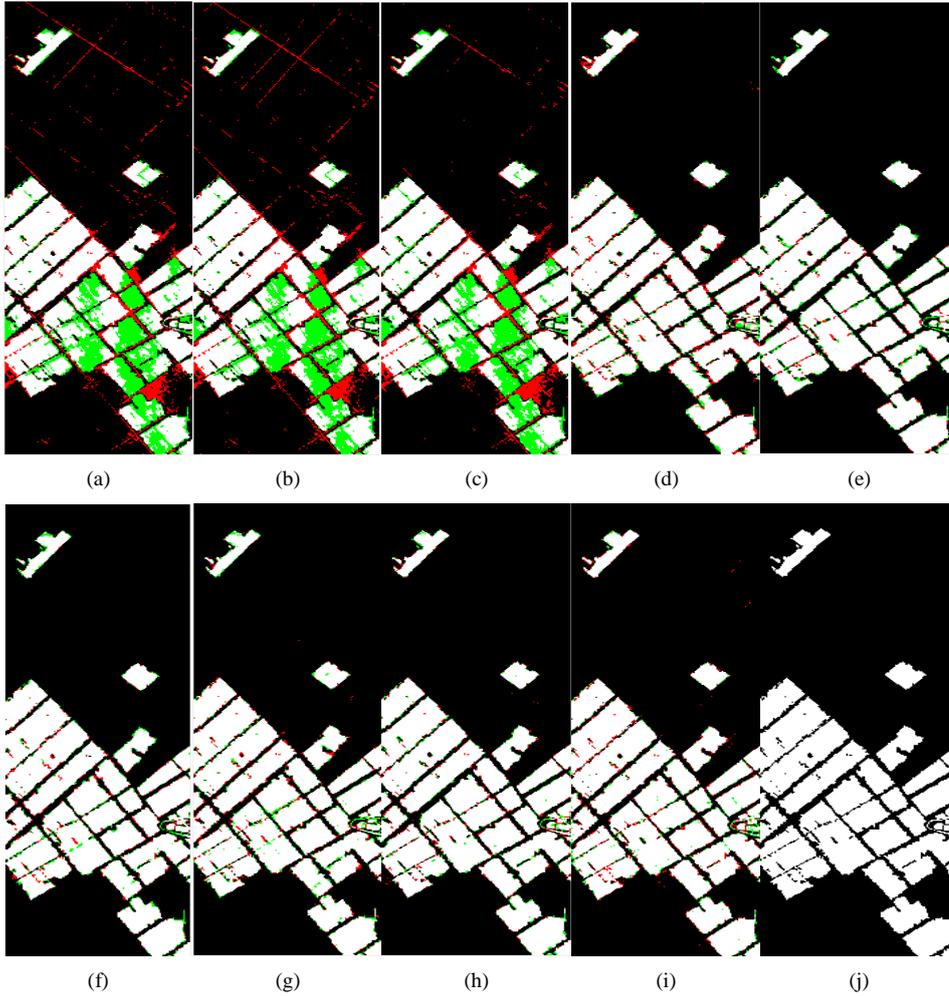


Figure 4.3 Extracted change maps on the Yancheng Dataset from different methods of AD (a), CVA (b), PCAKM (c), LSTM (d), HybridSN (e), 3-DCNN (f), 2-DCNN (g), CSANet (h) and CBANet (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Experiments on the Hermiston Dataset

For the Hermiston dataset, the HCD results are shown and compared with the ground truth in Figure 4.4 and Table 4.4, where the changing areas are mainly crop regions with simple round shapes. The results of the quantitative analysis are presented in Table 4.4. OA of all methods have achieved at least 97%, or over 99% for DL-based methods. However, unsupervised methods have still produced quite a few false alarms, leading to lower UCA and KP values than all DL methods. For DL-based methods, as highlighted in Figure 4.4, LSTM and HybridSN fail to accurately segment the edges of the changing area, where the boundaries of crop regions are connected together. Though 2-D-CNN and 3-D-CNN have slightly better results than LSTM and HybridSN, it is still difficult for them to detect the

crop regions with smooth edges. On the contrary, our CBANet can much more accurately detect the changing areas, with the KP 0.40-0.72% higher than that of 3-D-CNN and 2-D-CNN yet a much-reduced STD by 0.0008-0.0013, again validating the high efficacy of the proposed approach in HCD. In this dataset, although DL methods outperform all unsupervised ones with fewer false alarms and missing detection, the difference in terms of quantitative assessments is small, due mainly to the relatively simple background hence less noise caused false alarms. Within the DL methods, LSTM has the poorest performance, whilst the results from 2-D-CNN and 3-D-CNN are quite similar. As the combination of 2-D-CNN and 3-D-CNN, HybridSN can extract spectral-spatial features of local regions, yet it fails to feature changing pixels. Also, it may suffer from overfitting due to too many convolutional layers contained. In addition, these three CNN-based models suffer from dealing with sporadic and isolated changing pixels because the large perceptual field in their convolutional layers can help to extract the global features but neglect small details. Again, CSANet and our CBANet have about the same results in terms of OA, AA, and KP, though it has a slightly higher CA than CBANet. In addition, it is worth noting that in all three datasets, the proposed CBANet has a (slightly) higher AA than the CSANet, indicating its strong capability in characterizing both large and small features for their accurate detection.

Table 4. 4 Quantitative assessment of different methods on Hermiston dataset

	OA	CA	NCA	AA	KP
AD	0.9728	0.9781	0.9367	0.9574	0.8824
CVA	0.9781	0.9843	0.9351	0.9597	0.9035
PCAKM	0.9789	0.9858	0.9322	0.9590	0.9068
LSTM	0.9901±0.0010	0.9602±0.0074	0.9945±0.0009	0.9773±0.0036	0.9555±0.0046
HybridSN	0.9893±0.0006	0.9580±0.0014	0.9939±0.0011	0.9759±0.0047	0.9519±0.0030
3-D-CNN	0.9919±0.0003	0.9728±0.0081	0.9948±0.0014	0.9834±0.0033	0.9638±0.0016
2-D-CNN	0.9912±0.0004	0.9662±0.0077	0.9949±0.0012	0.9806±0.0033	0.9606±0.0021
CSANet	0.9923±0.0006	0.9747±0.0075	0.9950±0.0003	0.9848±0.0037	0.9659±0.0031
CBANet	0.9928±0.0010	0.9745±0.0057	0.9955±0.0007	0.9850±0.0024	0.9678±0.0008

For our CBANet model, however, the cross-band feature extraction module can extract the representative spectral feature whilst reducing the spectral dimension. The 2-D self-attention module can further fuse the spatial and spectral features for distinguishing both sporadic changing pixels and large changing areas. As a result, the proposed CBANet can

consistently produce the best results than other benchmarking methods on all these three datasets.

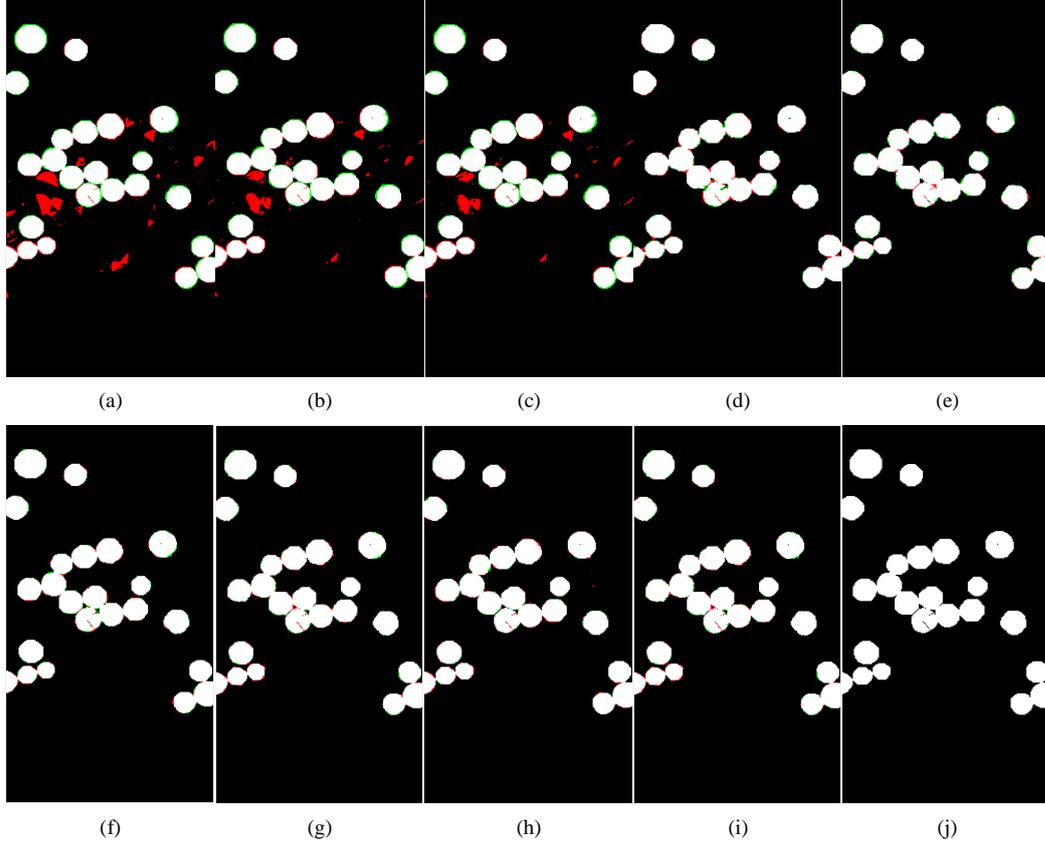


Figure 4. 4 Extracted change maps on the Hermiston Dataset from different methods of AD (a), CVA (b), PCAKM (c), LSTM (d), HybridSN (e), 3-DCNN (f), 2-DCNN (g), CSANet (h) and CBANet (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Hyperparameter analysis

To further validate the efficiency of our proposed CBANet, we compare the hyperparameters, the number of floating-point operations (FLOPs), and the overall running time in minutes (m) on the River dataset in Table 4.5, including both the training time and testing time. It can be observed that HybridSN, 3-D-CNN and 2-D-CNN, CSANet have much more hyperparameters, resulting longer running time than our proposed method. Although LSTM has less running time and less hyperparameters, it has the worst detection accuracy on three datasets. Thanks to the 1×1 convolutional kernel in the cross-band feature extraction module and 2-D self-attention module, our proposed CBANet model can be

considered as a lightweight model which has fewer hyperparameters but performs better than other benchmarking methods.

Table 4. 5 Comparing the hyperparameters of different DL-based methods on River dataset

	LSTM	HybridSN	3-D-CNN	2-D-CNN	CSANet	CBANet
Hyperparameters (k)	213.79	5128.74	1613.03	607.43	2452.88	319.36
FLOPs (M)	3.51	1597.24	215.35	368.21	144.44	6.66
Running Time (m)	10.21	130.22	55.90	32.42	101.43	20.71

Training ratio analysis

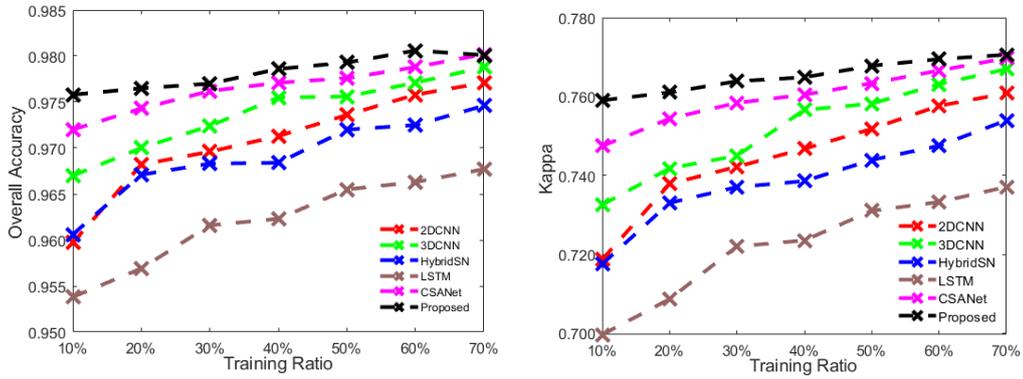


Figure 4. 5 Comparing the OA and KP results of all DL methods on the River dataset under different training ratios.

To fully validate the effectiveness of our proposed model, we evaluate the results of all aforementioned DL-based methods on the River dataset when the training pixels vary from 10% to 70%. As seen in Figure 4.5, more training pixels will make the deep learning methods achieve better detection accuracy. Meanwhile, our CBANet can consistently achieve the highest OA and KP, where the best OA and KP on the River dataset can reach 98.01% and 0.8765, respectively.

Patch size

We tested four patch sizes of $\{5 \times 5, 7 \times 7, 9 \times 9, \text{ and } 11 \times 11\}$ to analyse their impact on the CBANet. As shown in Figure 4.6, an increase in patch size has a very limited effect on the

KP and OA achieved when other module parameter settings are unchanged, though the largest patch size of 11×11 seems to have slightly improved OA and KP value on Yancheng dataset. That is why we choose the patch size of 7×7 in our experiments for all three datasets to balance between the detection accuracy and computational efficiency.

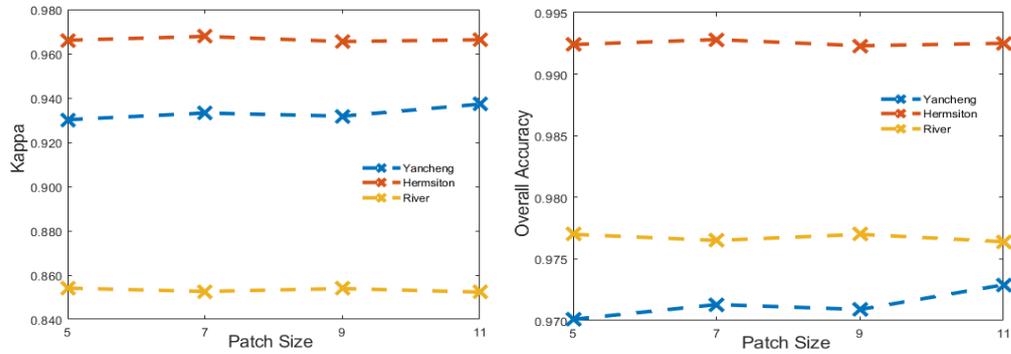


Figure 4. 6 Ablation experiment of the CBANet with different patch sizes

Number of spatial-spectral feature extraction kernels

To find the optimal number of kernels in the spatial-spectral module, six different settings of $\{8, 16, 32, 64, 128, 256\}$ are tested. As shown in Figure 4.7, all three datasets have the highest OA and KP value when the number of kernels is 128. Therefore, we set the kernel number for spatial-spectral feature extraction as 128.

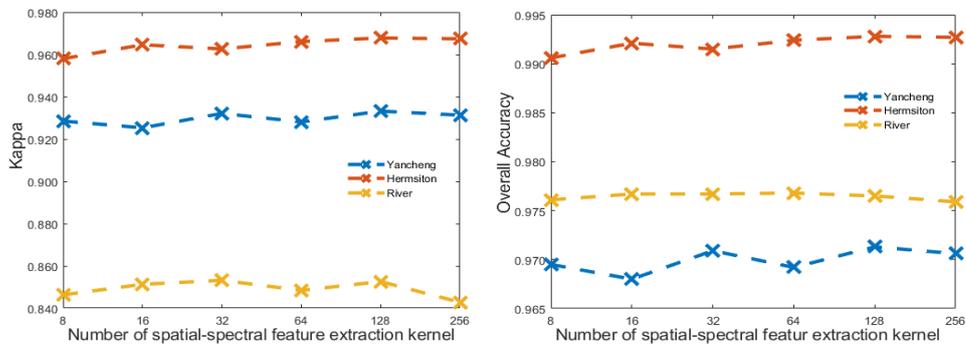


Figure 4. 7 Ablation experiment of the CBANet with different numbers of spatial-spectral feature extraction kernel.

Number of 2-D self-attention kernels

We have also evaluated the selection of the number of 2-D self-attention kernels, where the experiments on five different settings of $\{8, 16, 32, 64, 128\}$ are conducted. As shown in Figure 4.8, the variation trends of OA and KP value on the three datasets increase first and then decrease with the increasing number of kernels, and the classification result has the highest OA and KP value when kernel number of the 2-D self-attention module is 32.

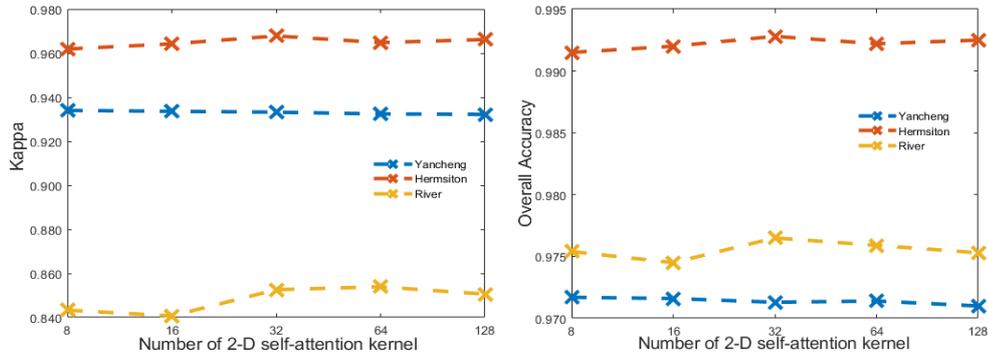


Figure 4. 8 Ablation experiment of the CBANet with different numbers of 2D self-attention kernel

Key stage analysis

In this section, compared with the traditional self-attention mechanism with 1×1 kernel, the efficacy of 1-D convolution module with 1×3 or 3×1 kernel, 2-D self-attention module with 5×5 kernel, and the proposed 2-D self-attention module with 3×3 kernel is compared. As seen in Table 4.6, the 3×3 kernel outperforms other 1-D and 2-D kernels in the self-attention module. Specifically, for the River dataset, the OA and KP values from the 3×3 kernel are 0.15% and 1% higher than those from the 1×1 kernel, respectively. Meanwhile, comparing with the 1×1 kernel, the standard deviation of the OA and KP in the 3×3 kernel has been reduced by 53.8% and 56.6%, respectively.

Table 4. 6 Comparison of OA and KP from three datasets with various kernel sizes

		Kernel size	1×1	1×3	3×1	3×3	5×5
River	OA		0.9750 ± 0.0026	0.9759 ± 0.0029	0.9756 ± 0.0009	0.9765 ± 0.0012	0.9757 ± 0.0008
	KP		0.8426 ± 0.0083	0.8467 ± 0.0005	0.8471 ± 0.0031	0.8526 ± 0.0036	0.8452 ± 0.0051
Yancheng	OA		0.9707 ± 0.0005	0.9700 ± 0.0009	0.9711 ± 0.0007	0.9713 ± 0.0006	0.9689 ± 0.0013
	KP		0.9319 ± 0.0011	0.9301 ± 0.0024	0.9325 ± 0.0018	0.9332 ± 0.0014	0.9276 ± 0.0029
Hermiston	OA		0.9916 ± 0.0014	0.9914 ± 0.0009	0.9923 ± 0.0006	0.9928 ± 0.0010	0.9910 ± 0.0006
	KP		0.9628 ± 0.0017	0.9616 ± 0.0039	0.9656 ± 0.0016	0.9678 ± 0.0008	0.9596 ± 0.0027

For the Yancheng and Hermiston datasets, an interesting finding is that the 5×5 kernel produces the worst results than all others. The possible reason is that this kernel is too large to the connected changed regions contained in these two datasets. In addition, a larger kernel in the 2-D self-attention module will inevitably lead to higher FLOPs and significantly more hyperparameters. In summary, the 2-D self-attention module with the 3×3 kernel can provide more accurate and robust results than other kernel sizes we have compared for HCD.

4.2.3 Summary

A novel lightweight end-to-end deep learning-based network, namely CBANet is proposed for hyperspectral change detection in this section. With the CBANet, the proposed cross-band feature extraction module has shown very good performance to fully extract and fuse the spectral information from bi-temporal HSI data whilst using the 1×1 kernels in the convolutional layer for efficiency. In addition, the proposed 2-D self-attention module has helped to capture deep spatial-spectral features for improving the feature representation and discrimination capabilities. The experiments on three publicly available HCD datasets have shown that the proposed CBANet outperforms other benchmarking models and has better stability and lighter weight than benchmarking deep learning models. This has fully validated the effectiveness and efficiency of the proposed model for the HCD task.

There are still some limitations to our proposed method. Through the analysis of experimental results, it was found that the NCA of CBANet is inferior to other DL-based methods. To further improve the NCA, enhancing the detection accuracy of edge pixels in changed areas would be crucial. To address this limitation, we made further attempts and proposed a new DL-based network called SSA-LHCD, which built upon the 2-D self-attention module. SSA-LHCD introduces singular spectrum analysis (SSA) to denoise spectral signals. Additionally, a residual block based on 1×1 convolutional layers is incorporated to deeply extract spectral features. This improvement not only increased detection accuracy but also reduced the number of the model's hyperparameters. Detailed information is provided in Section 4.3.

4.3 SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection

To improve the detection accuracy of edge pixels in change areas and to further reduce the number of model’s hyperparameters, we propose a singular spectrum analysis (SSA)-driven-lightweight network for HCD, where three crucial components are incorporated to tackle these challenges. Firstly, SSA is applied to alleviate the effect of noise. Next, a residual block-based module is designed to effectively extract the spectral features for efficiency. Finally, the 2-D self-attention module is employed to effectively handle multi-scale changes. The major contributions of this work are summarised as follows:

- 1) To apply the 1-D SSA for spectral domain denoising and mitigating the effect of noise on the tasks of feature extraction and change detection.
- 2) To propose an efficient spectral feature-extraction module, which utilises a residual block and an extra 1×1 convolutional layer to restrict the gradient propagation range via skip connections, and to adeptly capture the spectral features with instance normalization, further benefiting the greatly increased non-linear characteristics with fewer hyperparameters and computational costs.
- 3) To employ a 2-D self-attention module to capture local spatial–spectral features. By dynamically adjusting the attention across diverse positions with multi-scale changing areas, feature representation and discrimination capability are improved through strategic weight allocation, resulting in significantly enhanced module reliability.

This section is organized as follows. Section 4.3.2 elucidates the particulars of the proposed SSA-LHCD model. Section 4.3.3 discusses the experimental results for three publicly available datasets. Section 4.3.4 discusses the ablation experiment results regarding the parameter setting of the SSA-LHCD model. Then, a comprehensive discussion on the benchmark methods and all experiment results is summarized in Section 4.3.5. Finally, a short summary with some concluding remarks is given in Section 4.3.6.

4.3.1 Methodology

The SSA-LHCD network is designed in four main steps: 1) SSA-based preprocessing for noise removal, 2) spectral feature extraction module, 3) 2-D self-attention-based local spatial-spectral feature extraction module, and 4) decision making. The details of the SSA-LHCD network are presented in Figure 4.9 and are further discussed in the following subsections.

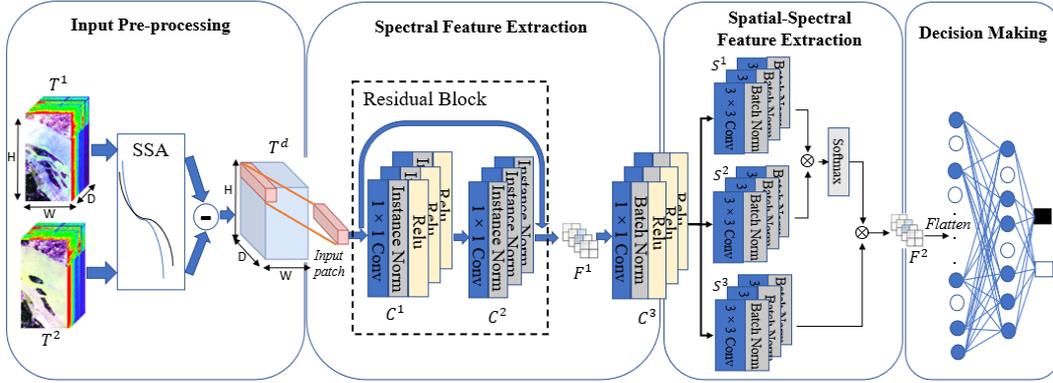


Figure 4. 9 The architecture of the proposed end-to-end SSA-LHCD network

SSA-based pre-processing

In the conventional task of land-mapping, SSA was used to extract the representative spectral information from the HSI data. For this purpose, each spectral profile was decomposed into several independent components, including the trend, oscillations, and noise, followed by spectral reconstruction using selected components whilst discarding the noisy ones. In HCD, a pair of bitemporal hypercubes, $T^1 \in \mathcal{R}^{W \times H \times B}$ and $T^2 \in \mathcal{R}^{W \times H \times B}$, are presented, where W and H denote the width and height in the spatial domain, respectively, and B is the number of spectral bands. SSA is applied to reduce the inheritable noise in each hypercube, aiming to mitigate the noise caused by outliers in the differentiation process, as detailed below.

Embedding

Let $x = [x_1, x_2, \dots, x_B]$ denote a pixel-wise spectral vector; it will be firstly embedded to form a trajectory matrix, X , by an embedding window, L , $L \in [1, B]$.

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_B \end{pmatrix} \quad (4.4)$$

where $K = B - L + 1$, and each column of X is a lagged vector that can be considered as a Hankel matrix as it has equal values along the antidiagonals.

Eigen Decomposition

The singular value decomposition (SVD) is applied for eigen decomposition on the matrix x , where the eigenvalues and eigenvectors of XX^T are denoted as $(\lambda_1, \lambda_2, \dots, \lambda_L)$ and (U_1, U_2, \dots, U_L) , respectively. The trajectory matrix can be reconstructed as the sum of elementary matrices as follows.

$$X = X_1 + \dots + X_i + \dots + X_L, \quad X_i = \sqrt{\lambda_i} U_i V_i^T, V_i = X^T U_i / \sqrt{\lambda_i} \quad (4.5)$$

Grouping and Projection

Dividing the total set of L components into M disjoint subsets (I_1, I_2, \dots, I_M) , where $\sum |I_m| = L$ and $m \in [1, M]$. Let $I = [i_1, i_2, \dots, i_p]$ represent a divided subset as $X_I = X_{I_1} + X_{I_2} + \dots + X_{I_p}$. Then, the trajectory matrix is represented by:

$$x = X_{I_1} + X_{I_2} + \dots + X_{I_M} \quad (4.6)$$

Let $Z_m = [Z_{m1}, Z_{m2}, \dots, Z_{mN}] \in \mathcal{R}^N$ denote the 1-D signal projected from X_{I_m} , which can be obtained via the diagonal averaging of each X_{I_m} . Finally, the original 1-D signal, x , can be reconstructed using its eigenvalues in one or more principal groups, with highly noisy and less significant components discarded by:

$$SSA_{(x)} = z_1 + z_2 + \dots + z_M = \sum_{m=1}^M Z_m \quad (4.7)$$

An example of a 1-D SSA application is shown in Figure 4.10, which shows a pair of corresponding non-changed pixels from the bitemporal images on the River dataset, including the original spectral vectors and their difference, as well as the SSA-smoothed results and the new difference. As can be seen, both the original corresponding spectral signals and the differential spectral signal preserve the basic trend of the profile whilst smoothing out the noise and thus the outliers in the difference signal for more robust change detection.

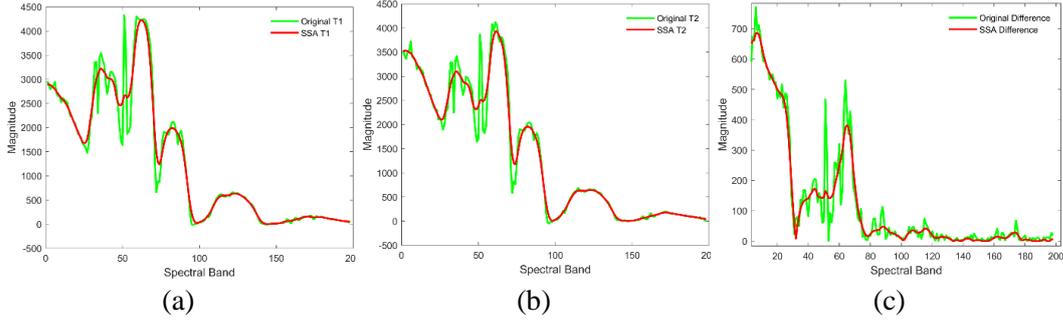


Figure 4.10 Examples showing a pair of unchanged pixels from the River dataset, where the outliers in the difference signal have been mitigated via SSA-based noise removal from the original profiles. (a) Original T1/ SSA T1, (b) original T2/ SSA T2, and (c) original difference/ SSA difference.

Spectral Feature Extraction

From the SSA-smoothed hypercubes of T^1 and T^2 , their absolute difference can be obtained as a new hypercube, T^{diff} :

$$T^{diff} = |SSA_{(T^2)} - SSA_{(T^1)}| \quad (4.8)$$

where $T^{diff} \in \mathfrak{R}^{W \times H \times B}$. To produce more training samples, T^d is divided into 3-D overlapped patches, $P_{(\alpha, \beta)} \in \mathfrak{R}^{O \times O \times B}$, with a window size, O ; (α, β) denote the coordinates of the patch center in the spatial domain, where $\alpha \in [1, W]$, $\beta \in [1, H]$, and the truth label is decided by the centered pixel. Padding is added to the T^{diff} to increase the spatial dimensions, ensuring that the window O fits even on the edges. In our experiments, 20% of pixels from both the changed and unchanged regions were randomly selected for training, while the remaining were used for testing.

The spectral feature-extraction module is structured into two main components. The initial part is a residual block composed of $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$, aiming to extract features in the spectral domain. $\mathcal{C}^{(1)}$ is constructed with a 1×1 convolutional layer, serving as an initial extraction for spectral features, followed by an instance normalization (IN) layer and the rectified linear unit (ReLU) activation function. The incorporated IN layer [140] independently normalizes each pixel rather than the entire batch, thereby ensuring that the features of each sample have similar means and variances, which aids in speeding up the convergence and enhancing the model's generalization capability. Additionally, the non-linear properties of the ReLU [141] activation function foster the acquisition of more intricate functions and alleviate the issue of vanishing gradients, consequently amplifying the model's capacity for non-linear representation.

For $C^{(2)}$, it also has a 1×1 convolutional layer and an IN layer for deep spectral feature extraction, enhancing the network's representation capacity. Unlike $C^{(1)}$, the difference lies in the absence of a ReLU activation function, aiming to preserve the feature information for compatibility with the residual connection. The incorporated skip connections facilitate the addition of input features to the output features. This design enables the network to effectively capture the residual information between the input, $P_{(\alpha,\beta)}$, and output, F^1 , as follows.

$$F^1 = C^{(2)} \left(C^{(1)}(P_{(\alpha,\beta)}) \right) + P_{(\alpha,\beta)} \quad (4.9)$$

Utilising the 1×1 convolution operation on the input layer enables the linear combination of features across different channels, resulting in the generation of novel feature representations. This process enhances the network's representational capacity and the overall performance by extracting more expressive features. The subsequent component is a profound spectral feature-extraction layer, denoted as the 1×1 convolutional layer, $C^{(3)}$, accompanied by the batch normalization (BN) [142] and ReLU functions. $C^{(3)}$ is adept at adjusting the number of convolutional kernels, thereby effectively reducing the number of spectral channels. This dimensionality reduction serves a dual purpose, i.e., trimming down the number of parameters and computational complexity, whilst simultaneously preserving crucial spectral features. The outcomes are the improved computational efficiency of the network and dimension reduction-mitigated challenges associated with gradient vanishing, fostering improved information propagation within the network [143]. In other words, the combination of these features has not only refined the network's efficiency, but also addressed key issues related to gradient flow and parameter optimisation.

Spatial-spectral feature extraction

Inspired by the work in CBANet, here we integrate a 2-D self-attention module into the proposed SSA-LHCD model, serving local spatial-spectral feature extraction, aiming to boost the stability of the feature extractor within the model. Taking the output of the spectral feature extractor as the input, after traversing through three successive 2-D convolutional layers of $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$, it can generate a novel spatial-spectral feature map. By converting the feature map into a 2-D attention matrix, it can facilitate the creation of a refined spatial-spectral feature representation. This multi-step process enriches the model's

ability to capture intricate relationships and latent dependencies within the input. This comprehensive representation encapsulates both spatial and spectral information, offering a robust foundation for subsequent stages of the model. Upon completing the deep spatial–spectral feature extraction, the final extracted feature map is derived as $F^{(2)}$ as follows.

$$F^{(2)} = (S^{(3)}(Z_{(\alpha,\beta)}))^T * Softmax(S^{(1)}(Z_{(\alpha,\beta)}))^T * S^{(2)}(Z_{(\alpha,\beta)}) \quad (4.10)$$

Incorporating the self-attention module empowers the SSA-LHCD model to capture intricate spatial–spectral dependencies, fostering enhanced stability and robustness in feature extraction for a diverse range of applications.

Decision making

Change detection can be regarded as a binary classification problem for distinguishing changed and unchanged pixels. Firstly, the spatial–spectral features, $F^{(2)}$, obtained from the previous spatial–spectral feature-extraction stage are flattened into a one-dimensional vector. This transformation prepares the features for input into a fully connected neural network suitable for decision making. The flattened feature vector is then fed into a series of fully connected layers. Each layer performs linear transformations, followed by nonlinear activations to learn complex patterns and relationships within the input features, $F^{(2)}$. Subsequently, the final layer of the fully connected network employs a SoftMax activation function, which converts the network’s outputs into a probability distribution over two classes. The final classification decision is made by selecting the class with the highest probability, thus achieving a binary classification. The selected optimizer is adaptive momentum (Adam) [139], and the selected loss function is cross-entropy [144], with the initial learning rate of 0.0001. The specific details of each layer in the end-to-end SSA-LHCD model are summarized in Table 4.7.

Table 4. 7 Architecture details of each layer in SSA-LHCD

Layers	Type	Channels	Kernel
T^1	SSA Preprocessing	B	-
T^2	SSA Preprocessing	B	-
T^d	Difference	B	-
C^1	Conv2D+IN+Relu	B	1×1
C^2	Conv2D+IN	B	1×1
C^3	Conv2D+BN+Relu	64	1×1
S^1	Conv2D+BN	32	3×3
S^2	Conv2D+BN	32	3×3
S^3	Conv2D+BN	32	3×3
Flatten	Flatten	288	-
FC^1	Linear (Dropout=0.4)	64	-
FC^2	Linear (Dropout=0.4)	8	-
FC^3	Linear (Dropout=0.4)	2	-

4.3.2 Experiments and Results

In this section, we evaluate the effectiveness of the proposed SSA-LHCD by comparing it with three classical unsupervised methods, which include the AD [30], CVA [31], and PCAKM [40] as well as several DL-based methods such as SVM [73], 2-D CNN [128], CSANet [83], ML-EDAN [89], and CBANet [2]. The benchmarks are established according to the specified parameters in the default settings, where DL-based methods are trained using PyTorch on NVIDIA RTX A2000, with a batch size of 32 and 500 training epochs. For training, 20% of pixels from both changed and unchanged regions are randomly selected, while the remaining pixels are used for testing. To ensure fairness and reliability, each supervised method is repeated ten times in our experiments, and the averaged results of OA and KP are reported for comparison. In addition, comparisons of the resulting change maps as well as quantitative evaluations using Pre , Re , and $F1$ for all methods are conducted. In the resulting change maps, false alarms and missing pixels are highlighted in red and green, respectively, while correctly detected changed areas are presented in white, and true negatives are depicted in black for an easy visual comparison.

Results on the River dataset

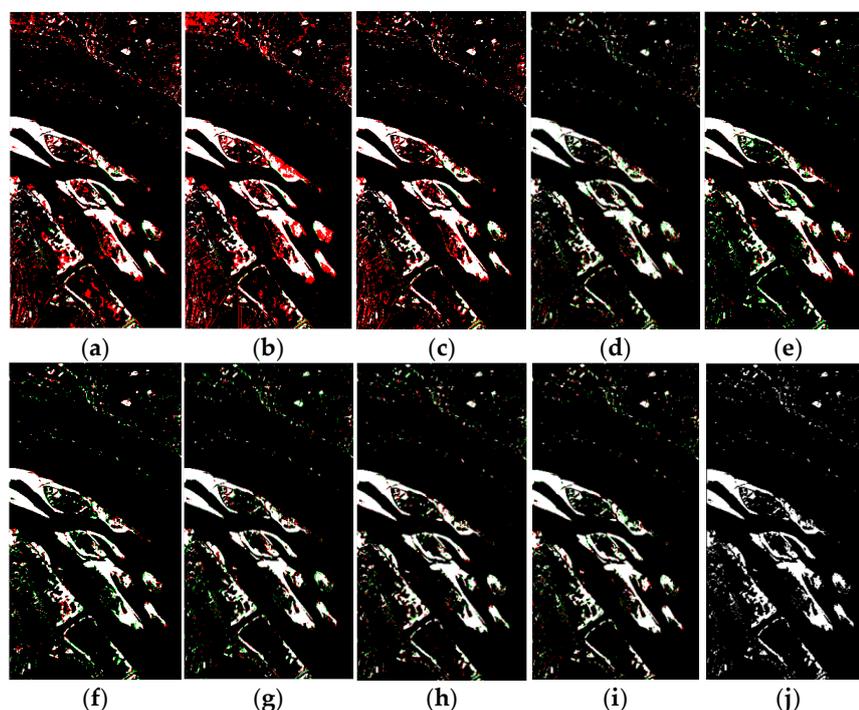


Figure 4.11 Extracted change maps on the River dataset from different methods of AD (a), CVA (b), PCAKM (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

The extracted change maps and quantitative results from the River dataset for all benchmarks are shown in Figure 4.11 and Table 4.8, respectively. Although all three unsupervised algorithms achieve Re values no less than 99.5%, they exhibit quite a low Pre , i.e., excessive false alarms. These false alarms are visibly concentrated in the upper and lower left corners of the change maps, as seen in Figure 4.11 (a–c), due to the misclassification of subtle sporadic change pixels in the River dataset. Consequently, the Pre of all unsupervised algorithms drops to below 66%, where all KP values fall below 0.75. For the supervised methods, however, the extracted maps exhibit much less false alarms, yet there is a prevalent issue of missing detections, especially for 2-D CNN and ML-EDAN approaches. These results demonstrate a relatively low detection accuracy, as indicated by OA values below 97% and KP values hovering around 0.80. Interestingly, the SVM performs marginally better than the 2-D CNN and ML-EDAN, with the OA boosted to 97.02% and KP to 0.8109. However, the SVM has the highest standard deviation of

0.0078 in *OA* among all supervised methods. Not surprisingly, thanks to the SSA pre-processing and proposed feature-extraction modules, our SSA-LHCD model outperforms all benchmarks on the River dataset, surpassing the CBANet by 0.24% in the *OA* and 0.0144 in the *KP*.

Table 4. 8 Quantitative assessment of different methods on the River dataset

	<i>OA</i> (%)	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	94.31	0.7137	0.6108	0.9515	0.7440
CVA	92.53	0.6528	0.5393	0.9635	0.6915
PCAKM	95.17	0.7478	0.6524	0.9512	0.7738
SVM	97.02±0.0078	0.8109±0.0049	0.8358	0.8417	0.8387
2-D CNN	96.82±0.0007	0.7946±0.0033	0.9073	0.8888	0.8978
CSANet	97.43±0.0012	0.8360±0.0049	0.9130	0.9175	0.9152
ML-EDAN	96.96±0.0014	0.8009±0.0049	0.9220	0.8975	0.9093
CBANet	97.65±0.0036	0.8526±0.0036	0.9405	0.9119	0.9256
SSA-LHCD	97.89±0.0007	0.8670±0.0026	0.9322	0.9343	0.9332

Results on the Yancheng dataset

Similar to the results in the River dataset, the inadequate performance of all three unsupervised methods is evident, as shown in Figure 4.12 (a–c). The quantitative results on Yancheng dataset are shown in Table 5.9. These methods have a notable number of missing detection of pixels, coupled with the presence of false alarms, particularly in striped lines and other field regions. As a result, the *KP* values remain consistently low, hovering around 0.71, with the *OA* dropping below 90%; both *Pre* and *Re* are below 90%. Here, the SVM becomes the poorest performer among all supervised algorithms, with *OA* and *KP* values of only 94.87% and 0.8806, respectively, due mainly to the SVM’s limitation in pixel-wise learning without considering the spatial features. In contrast, deep learning-based approaches have an *OA* exceeding 96% and a *KP* over 0.92. Nevertheless, our SSA-LHCD model remains the best, showcasing the highest average values for both the *KP* and *OA*. Furthermore, the standard deviation of the *KP* is only 0.0012, the lowest among all supervised methods, along with the highest *F1* score compared with all benchmark tests. These outcomes serve as compelling evidence, substantiating the effectiveness and robustness of our proposed SSA-LHCD model.

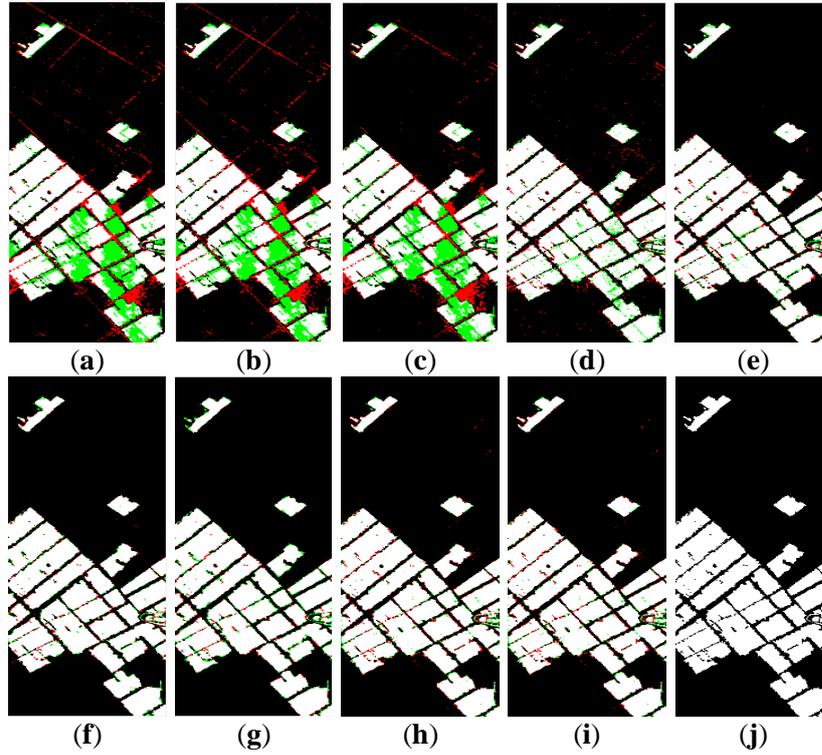


Figure 4.12 Extracted change maps of the Yancheng dataset from the different methods of AD (a), CVA (b), PCAKM (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Table 4.9 Quantitative assessment of different methods on the Yancheng dataset

	<i>OA</i> (%)	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	87.80	0.7074	0.8430	0.7494	0.7935
CVA	87.55	0.7025	0.8327	0.7529	0.7908
PCAKM	88.28	0.7180	0.8557	0.7519	0.8004
SVM	94.87 ± 0.0013	0.8806 ± 0.0029	0.9063	0.9110	0.9086
2-D CNN	96.67 ± 0.0014	0.9223 ± 0.0030	0.9608	0.9557	0.9582
CSANet	97.15 ± 0.0009	0.9335 ± 0.0023	0.9658	0.9641	0.9650
ML-EDAN	97.15 ± 0.0012	0.9316 ± 0.0034	0.9685	0.9517	0.9598
CBANet	97.13 ± 0.0006	0.9332 ± 0.0014	0.9645	0.9633	0.9639
SSA-LHCD	97.16 ± 0.0011	0.9365 ± 0.0012	0.9680	0.9701	0.9691

Results on the Hermiston dataset

For the Hermiston dataset, the extracted change maps and quantitative assessment are shown in Figure 4.13 and Table 4.10, respectively. Due to the absence of scattered variation pixels and the distinct visibility of all changed features, the *OA* values of all benchmarks surpassed 97%, or are over 99% for all supervised methods, though the SVM remains the worst supervised model due to the lack of spatial features. Here, our SSA-LHCD model emerges as the second-highest performer among all deep learning models, while the *OA* is only 0.11% lower than the top-performing ML-EDAN, and the *KP* is merely 0.0067 lower than the leading CBANet. This is attributed to the relatively homogeneous change type in this dataset, limiting the prominence of deep spectral feature extraction. The primary disparity lies in the detected edges of changed regions. CBANet, with its incorporation of deep spatial feature learning and small kernels, accurately identifies pixels along the edges of each change region through spatial feature extraction. In contrast, our SSA-LHCD model focuses solely on extracting spectral features by utilising the 1×1 convolutional layer and residual block and does not explicitly learn deep spatial features like the CBANet. As a result, the *OA* and *KP* of our model on the Hermiston dataset are slightly lower than those of the CBANet and ML-EDAN.

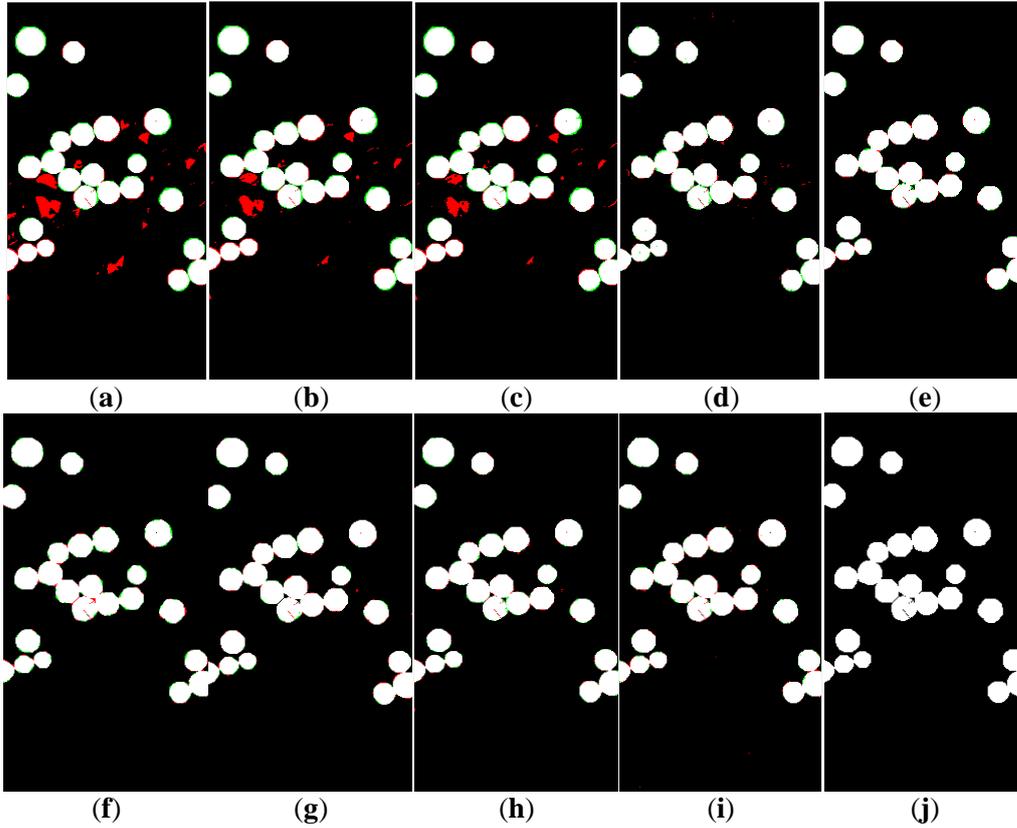


Figure 4. 13 Extracted change maps of the Hermiston dataset from the different methods of AD (a), CVA (b), PCAKM (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

Table 4. 10 Quantitative assessment of different methods on the Hermiston dataset

	<i>OA</i> (%)	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	97.28	0.8824	0.8625	0.9367	0.8981
CVA	98.43	0.9035	0.8978	0.9351	0.9161
PCAKM	97.89	0.9068	0.9060	0.9322	0.9189
SVM	99.07± 0.0002	0.9581±0.0012	0.9519	0.9759	0.9638
2-D CNN	99.12± 0.0004	0.9662±0.0077	0.9819	0.9779	0.9799
CSANet	99.23± 0.0006	0.9659±0.0031	0.9822	0.9705	0.9763
ML-EDAN	99.32± 0.0001	0.9669±0.0008	0.9806	0.9820	0.9813
CBANet	99.28± 0.0010	0.9745± 0.0030	0.9808	0.9883	0.9845
SSA-LHCD	99.21± 0.0009	0.9678± 0.0008	0.9781	0.9909	0.9844

4.3.3 Ablation Study

To comprehensively validate the effectiveness of our proposed SSA-LHCD model, we conducted a series of experiments covering computational hyperparameters, the effect of

the modular block, and different numbers of spectral or spatial–spectral feature-extraction kernels.

Hyperparameter Analysis

In Table 4.11, we compare the numbers of hyperparameters and floating-point operations (FLOPs) and the overall running time in minutes (m), including both training time and testing time, for all the DL-based models, including ours, on the River dataset. For those using multi-layer CNNs, such as the 2-D CNN, CSANet, and ML-EDAN, the numbers are much higher, i.e., over two magnitudes, than ours. The inclusion of the 1×1 convolutional kernel in the spectral feature module and the residual block contribute to the lightweight nature of our SSA-LHCD model, which also outperforms other benchmarking methods.

Table 4. 11 Complexity comparison of all DL-based methods on the River dataset

	2-D CNN	CSANet	ML-EDAN	CBANet	SSA-LHCD
Parameters (k)	607.43	2452.88	88933.34	319.36	167.24
FLOPs (M)	368.21	144.44	590.22	6.66	2.80
Running Time (m)	35.42	53.43	76.27	18.53	14.21

Effect of Modular Blocks and Patch Size

In this section, we conducted three sets of experiments: (i) SSA-LHCD without the SSA based pre-processing, (ii) SSA-LHCD without the residual block, and (iii) SSA-LHCD with both module blocks with different patch sizes on all three datasets. We tested five patch sizes of $\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, \text{ and } 11 \times 11\}$, and the results are presented in Figure 4.14, where the training ratio is set as 20%. First, the performance of the SSA-LHCD model degrades when either the SSA or the residual block is absent, showing their unique value to the proposed model. Second, with the increasing patch size, the KP values on the Hermiston dataset increase and reach 0.9697 at the patch size of 11×11 . However, the varying trends of the KP on the River and Yancheng datasets are different, which show an initial increase followed by a decrease when the patch size exceeds 5×5 . The observed phenomenon can be attributed to the abundance of sparsely distributed change or non-change pixels in the River and Yancheng datasets, whilst the Hermiston dataset only

contains large, regular regions. Smaller patch sizes are better suited for extracting these scattered pixels effectively, and when using a large patch size, may lead to false alarms. However, in the case of the Hermiston dataset with distinct spectral features, the designed deep spectral feature module loses its advantage. Conversely, larger patch sizes encompass more spatial edge information, leading to improved edge detection accuracy for small changed areas. For balancing the detection accuracy and computational efficiency, we chose a patch size of 5×5 for our SSA-LHCD model.

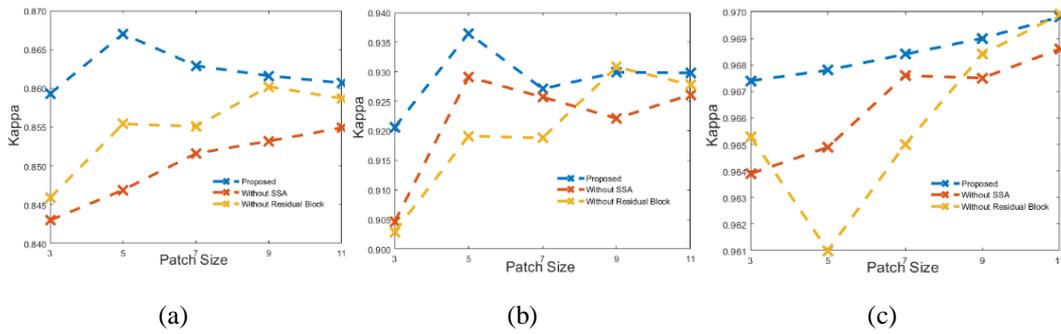


Figure 4. 14 Ablation experiment of the SSA-LHCD model with different patch sizes on River (a), Yancheng (b), and Hermiston datasets (c).

Number of Spectral Feature-Extraction Kernels

To determine the optimal number of kernels in the spectral feature-extraction module, five different settings of 16, 32, 64, 128, and 256 were tested. As shown in Figure 4.15, the varying trends of the KP on the three datasets appear similar to those from increasing the patch size. When the kernel number of the spectral feature-extraction module is set to 64, the highest KP values can be achieved on the River and Yancheng datasets. It is worth noting that, when the kernel number is set to 128, the average KP value on the Hermiston dataset is 0.9735, which is very close to the KP value of the CBANet. However, for the overall performance of the proposed network, we decided to set the kernel number of the spectral feature-extraction module to 64 for all datasets.

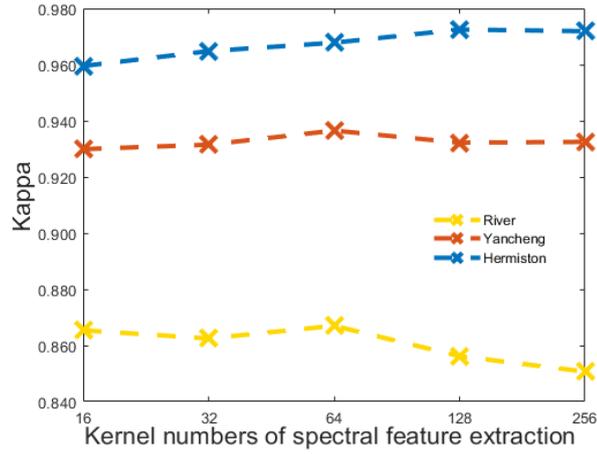


Figure 4. 15 Ablation experiment of the SSA-LHCD model with different kernel numbers of spectral feature extraction module.

Number of Spatial–Spectral Feature-Extraction Kernels

We also evaluated the selection of the number of 2-D self-attention kernels by conducting experiments using five different settings, including 8, 16, 32, 64, and 128. The variation trend of KP on the three datasets is shown in Figure 4.16. Similarly, the kernel number is set to 32 to balance the model’s parameters and robustness.

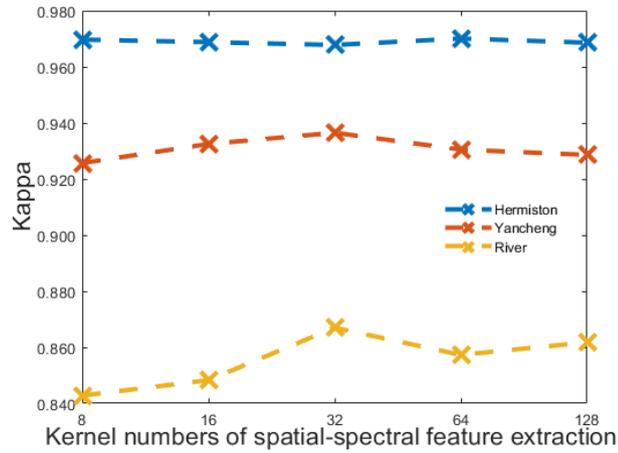


Figure 4. 16 Ablation experiment of the SSA-LHCD model with different kernel numbers of spatial-spectral feature extraction module

Training Ratios

To further validate the efficacy of our SSA-LHCD model, its performance is assessed on the River dataset, considering varying percentages of training ratios from 10% to 50%. As shown in Figure 4.17, a larger training ratio generally leads to an improved detection accuracy, where our model consistently achieves the highest KP . Specifically, when the training ratio is 50%, our model can achieve a KP of 0.8843, surpassing the second best, CBANet, by a margin of 1.39%.

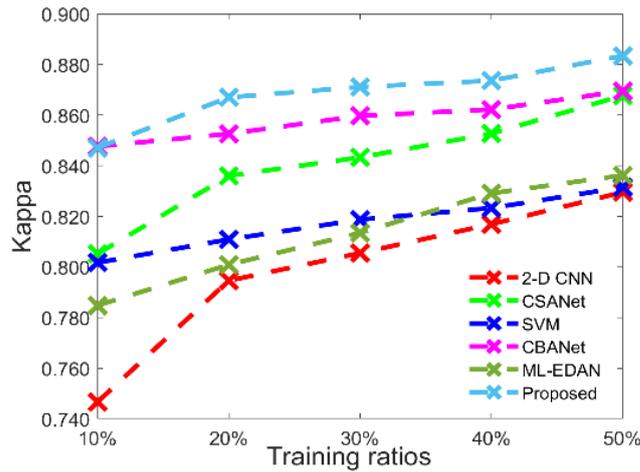


Figure 4. 17 Ablation experiment of the SSA-LHCD model with different training ratios of all DL-based benchmarks on the River dataset.

4.3.4 Discussion

The proposed SSA-LHCD network demonstrates significant advantages in terms of higher detection accuracy and fewer hyperparameters compared to the benchmarked state of the arts. These are mainly due to the residual block-based spectral feature-extraction module and the 2-D self-attention-based spatial-spectral feature-extraction module, as well as SSA-based pre-processing to effectively reduce noise whilst preserving valuable features, enabling our lightweight DL network to extract spectral and spatial-spectral features more effectively.

As shown in the compared results for the three datasets, image algebra-based, CVA and AD and image transformation-based PCAKM are all noise sensitive. Furthermore, the threshold segmentation or clustering processes in these methods fail to accurately classify subtle changes, leading to numerous false alarms or missing pixels. The SVM, as a classical supervised binary classifier, is trained using pixel-wise spectral vectors. Consequently, its

detection accuracy is significantly lower compared to methods based on spatial features and spatial–spectral feature extraction using deep learning models. This is due to the SVM’s inability to capture spatial features, which are crucial for precise classification.

In comparison to DL-based state-of-the-art approaches, our proposed SSA-LHCD method outperforms almost all of them, offering higher detection accuracy with reduced hyperparameters. The 2-D CNN method uses multi-layer 2-D convolutions with large kernels to extract local spatial features from input patches, yet it fails to account for spectral features. This limitation results in the lowest detection accuracy for the three datasets, especially the River dataset, which contains many sporadic pixels. The influence of neighbouring pixels due to the large kernel used leads to the misclassification of many sporadic changed pixels as unchanged ones, resulting in a high number of missing pixels. Conversely, on the Yancheng and Hermiston datasets, which consist of connected regions, a large number of false alarms are detected at the edges of the connected areas. By using multi-level spatial-spectral feature extraction via encoder-decoder and LSTM subnetworks, the ML-EDAN becomes the most complex network among all the compared models, with the number of parameters and FLOPs being approximately 531 times and 210 times greater than those of our proposed SSA-LHCD model, respectively. Based on the Siamese 2-D CNN structure, the CSANet extracts the joint spatial-spectral–temporal features of corresponding patches, along with the cross-temporal self-attention module utilised to integrate the jointed features oriented from each temporal embedding. Similarly, also as a self-attention-based network, the CBANet can effectively extract spectral and spatial–spectral features. The detection accuracy of the CSANet and CBANet, two self-attention mechanism-based algorithms, rank second and third, respectively, in all benchmark tests on the River and Yancheng datasets. However, it is still inferior to our proposed SSA-LHCD model, due to the inability of the SSA module to mitigate noise for the effective extraction of spectral features.

From the ablation experiments, as shown in Figure 4.14, it can be found that the SSA pre-processing step and residual block-based spectral feature-extraction module significantly improve the detection accuracy under different patch sizes. For the River and Yancheng datasets with varying scales of change, a smaller patch size results in higher detection accuracy. However, for the Hermiston dataset, characterized by a regular change scale and single change type, larger patch sizes increase the detection accuracy. Regarding the kernel numbers of the spectral feature-extraction module and the spatial–spectral feature-extraction module, the SSA-LHCD model achieves the best detection accuracy for the Hermiston dataset when set to 128 and 64, respectively. For a balanced performance

across different datasets, we set the kernel numbers of these two modules to 64 and 32, which yielded the best results for both the River and Yancheng datasets. Furthermore, the SSA-LHCD model can achieve the best detection performance across various training settings, indicating that a higher detection accuracy can be achieved with fewer training samples.

In summary, DL-based methods tend to outperform SVM and unsupervised approaches in HCD. As shown in all the quantitative results, the three self-attention-based models, CSANet, CBANet, and our SSA-LHCD model, outperform the two models that only use multi-scale 2-D convolutional layers and the Siamese autoencoder-based network, as shown in both the visual map comparison results and quantitative results. Notably, when examining the change maps, the three self-attention-based models demonstrate a superior detection performance, particularly for sparsely distributed change regions. For the River dataset with many sporadic pixels, the SSA-LHAD model achieves the best detection accuracy. Overall, our approach exhibits significant advantages over other existing models, especially for detecting different scales of changes.

There remain certain limitations to our proposed method. Currently, the difference of the presented HSI pairs after SSA pre-processing is taken as the input, followed by a single channel 1×1 convolutional layer for the deep extraction of the spectral features. In current implementations, only the trend signal of the SSA is used. Considering that the other components can also be potentially useful, their effects will be explored further in our future work. Although the proposed SSA-LHCD network has surpassed the state-of-the-art benchmarks in overall accuracy when using fewer training samples, it still requires manually labelled data due to its supervised nature. This dependency on manual labelling is a significant limitation in practical applications.

4.3.5 Summary

In section 4.3, a novel, lightweight end-to-end DL-based network (SSA-LHCD) is proposed for HCD. First, bitemporal HSIs were pre-processed using SSA for noise reduction. Initial change features are then extracted through subtraction. Following this, a residual block-based spectral feature-extraction module is employed to refine these initial change features by effectively capturing spectral information. Subsequently, a 2-D self-attention mechanism is integrated to capture local spatial-spectral features, enhancing both feature representation and discrimination capabilities. Finally, a fully connected layer serves as the classifier, facilitating binary HCD decision-making.

SSA-based noise reduction, 1×1 convolutional layer, and the residual block significantly improve the model's overall performance of change detection by enabling efficient spectral feature learning. Moreover, the inclusion of the 2-D self-attention module is crucial for capturing complex spatial-spectral features, further enhancing the model's ability to discriminate changed regions, thus improving HCD accuracy. Comprehensive experiments demonstrate SSA-LHCD's superiority over eight state-of-the-art methods on three publicly available datasets, highlighting its capability to produce higher detection accuracy with fewer hyperparameters. This innovative approach offers significant advancements in HCD by enhancing noise reduction, multi-scale change handling, and computational efficiency, setting a new benchmark in supervised HCD.

Chapter 5 **GASSM: Global Attention and State Space Model-Based End-to-End Hyperspectral Change Detection**

5.1 Introduction

Traditional deep learning networks primarily based on convolution neural networks (CNNs), optimise the parameters through gradient descent but face limitations in modelling long-range spatial-spectral feature dependencies, which restricts their contextual reasoning capabilities in understanding of the global context. Furthermore, methods based on transformers, while capable of modelling long-range dependencies, suffer from quadratic growth of the computational complexity. Recently, there is a trend to integrate the state space models (SSMs) into deep learning networks for sequence modelling [145]. A notable advancement involves introducing a selection mechanism into SSM, allowing them to selectively propagate or discard information based on the current token in the sequence or scan path for effective modelling of long-range dependencies with linear computational complexity [146]. Subsequently, selective SSMs have been integrated with simplified linear transformation and CNN architecture, namely Mamba. Mamba combines the modelling power of transformers with linear scalability by considering sequence length to analyse the long-range dependencies of features, positioning it as a promising foundation model for various sequence modelling tasks. The versatility of Mamba has been demonstrated across multiple RS applications, including semantic segmentation [147], HSIs classification [148] and HSIs dehazing [149].

In this chapter, we present a pioneering attempt to combine the SSM-based Mamba with the global attention (GA) and spatial-spectral feature extraction modules in the proposed end-to-end deep learning network, GASSM for HCD tasks. To enable the Mamba block to integrate spatial and spectral understanding, the introduced GA module can capture and consolidate the complex relationships in spatial and spectral domains, enhancing spatial-spectral feature interactions. To ensure efficient feature extraction of Mamba block, we designed a spatial-spectral feature extraction module, aiming to reduce feature redundancy while extracting high-quality features. This has not only preserved key information but also reduced computational complexity. Finally, the Mamba block extracts long-range features and models global spatial-spectral features, followed by a fully connected layer to

determine the HCD outcomes. Experiments on two publicly available datasets, the River and Hermiston datasets, have validated its superior performance when benchmarked with eight state-of-the-art methods. In addition, we have also tested the GASSM on the Yancheng dataset to further evaluate its effectiveness, by comparing it with a few state-of-the-art techniques. The visualization maps and quantitative analysis of the experimental results are given in Appendix A.1 for comparison.

5.2 Proposed Method

The diagram of the proposed GASSM is presented in Figure 5. 1, which is composed of four main modules, i.e., 1) global attention mechanism based on spectral and spatial attention; 2) feature extraction module for minimizing redundancy; 3) SSM-based Mamba block for capturing long-range dependencies; and 4) decision making of detected HCD results.

Given a pair of spatially aligned bi-temporal hypercubes $T^1 \in \mathfrak{R}^{W*H*B}$ and $T^2 \in \mathfrak{R}^{W*H*B}$, where W and H denote the width and height of the spatial size, and B represents the number of spectral bands. Then, their absolute difference can be obtained as a new hypercube T^d :

$$T^{diff} = |T^2 - T^1| \quad (5.1)$$

where $T^{diff} \in \mathfrak{R}^{W*H*B}$. To produce more training samples, T^{diff} is divided into 3-D overlapped patches $P_{(\alpha,\beta)} \in \mathfrak{R}^{S*S*B}$ with a window size S (set $S=7$ in this network); (α, β) denote the coordinates of the patch centre in the spatial domain, where $\alpha \in [1, W]$, $\beta \in [1, H]$ and the truth label is decided by the centre pixel. Firstly, padding is added to the T^{diff} to increase the spatial dimensions, ensuring that the window S fits even on the edges. In our experiments, 20% of pixels from both the changed and unchanged regions are randomly selected for training, while the remaining are used for testing.

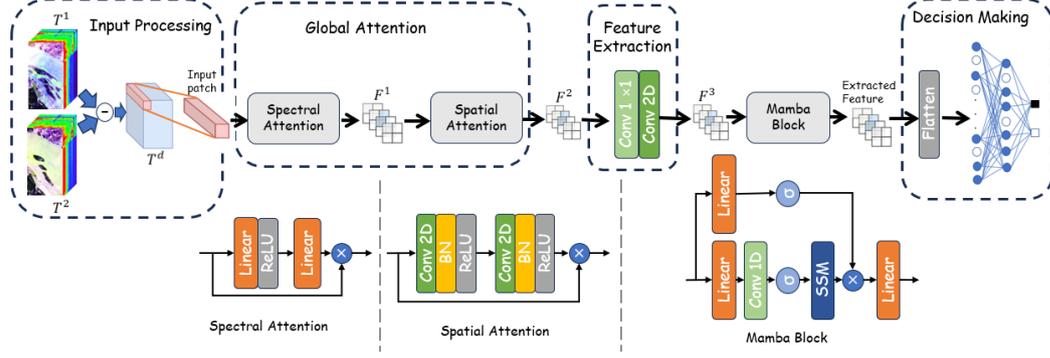


Figure 5. 1 The architecture of proposed GASSM network

5.2.1 Global Attention Mechanism

The global attention mechanism (GAM) [150] integrates spectral and spatial attention to minimize information loss and enhance dimension-interactive features. It leverages 3-D permutation with a multilayer perceptron (MLP) for spectral attention and incorporates a convolutional-based spatial attention submodule. The spectral attention submodule preserves information across three dimensions using 3-D permutation and employs a two-layer MLP, designed as an encoder-decoder, to strengthen cross-dimension spectral-spatial dependencies.

The spatial attention (*SA*) submodule enhances the spatial features by using two convolutional layers for spatial fusion and eliminates pooling to avoid information loss, thus preserving feature maps. This combined approach effectively retains more information and improves the interaction of features across dimensions, ensuring a comprehensive understanding and utilisation of complex data structures. The output of GAM, denoted as F^2 can be represented as

$$F^2 = SA(Spectral Attention (Input patch)) \quad (5.2)$$

here, the output F^2 and the input patch have the same size, with $F^2 \in \mathfrak{R}^{S*S*B}$.

5.2.2 Feature Extraction

In the feature extraction stage, following the global attention, the refined feature F^2 are subjected to further spectral and spatial feature extraction through convolutional operations.

This process incorporates a combination of 1×1 and 2-D convolutional layer to effectively process F^2 and generate F^3 .

The 1×1 convolution layer is primarily used to extract the spectral features and decrease the number of channels, thereby enhancing computational efficiency. By processing each pixel across the channels, the network can learn interactions between different channels without changing the spatial dimensions. By selecting an appropriate number of kernels (set to 128 in this network), This not only improves computational efficiency but also allows for increasing the network depth, which helps in learning more complex features. Furthermore, the 1×1 convolution introduces more nonlinearity, thereby enhancing the model's expressive power.

On the other hand, the 2-D convolution captures spatial features by employing a larger convolution kernel (set to 5×5 in the network). This allows it to extract features from a wider spatial area, identifying relationships and patterns among pixels in local regions, thus generating feature maps enriched with spatial structure information. In addition, by setting fewer number of kernels (set to 64 in this network), the channels of F^3 are further reduced.

By combining 1×1 convolution and 2-D convolution, this feature extraction stage not only reduces the spatial and spectral dimensions of the features but also preserves and enhances critical spectral and spatial characteristics. As a result, the generated feature map F^3 contains both refined spectral information and essential spatial structure details, providing high-quality input for the subsequent Mamba block. By this point, the size of feature map F^3 is $F^3 \in \mathfrak{R}^{3 \times 3 \times 64}$.

5.2.3 SSM-based Mamba Block

The Mamba block integrates linear transformations, a 1-D convolutional layer and a SSM to further extract the long-term dynamic dependencies of F^3 . This integration is designed to both local and global dependencies within the feature maps, enhancing the model's ability to learn more complex patterns. The detailed operations of the Mamba block are given below.

The input feature maps F^3 is first processed by a linear transformation and a 1-D convolutional layer. The operations can be expressed as:

$$\hat{F} = \text{SiLU} (\text{Conv1D} (\text{Linear} (F^3))) \quad (5.3)$$

where the Sigmoid-Weight linear units (*SiLU*) [151] is a smooth, non-linear activation function, which can retain input information and improve the gradient flow during backpropagation. The 1-D convolutional layer, represented as *Conv1D*, applies the convolutional filters across the transformed feature map. The output of this operation, denoted as \hat{F} , serves as the input for the subsequent SSM operations.

The SSM is employed to model the temporal dynamics within the transformed feature maps. The SSM consists of state update and observation equations that capture the dependencies across different time steps. The observation equation is defined as:

$$y_t = C_t * h_{t+1} \quad (5.4)$$

here, y_t represents the observation output at time step t , and C_t is a time-varying matrix that transforms the state vector h_{t+1} into the observation space. This transformation is crucial for linking the hidden states to the observable outputs. The state update equation is given by:

$$h_{t+1} = dA * h_t + \hat{F} * dB \quad (5.5)$$

where, h_t is the state vector at time step t , and h_{t+1} is the state vector at the next time step h_{t+1} . The matrices dA and dB are dynamically adjusted state transition and input matrices, respectively, which modulate the influence of the previous state and the current input on the next state.

The matrices B_t and C_t , which are essential for the state update and observation processes, are computed as follows:

$$B_t = fc^1(\hat{F}) \quad (5.6)$$

$$C_t = fc^2(\hat{F}) \quad (5.7)$$

here, fc^1 and fc^2 are distinct linear transformation functions in SSM that applied to the transformed feature map \hat{F} . These functions typically consist of fully connected layers that transform \hat{F} into the respective matrices B_t and C_t . These matrices play crucial roles in adjusting the state transitions and linking hidden states to observations.

The dynamic factor δ_t and the dynamic matrices dA and dB are computed as:

$$\delta_t = \text{softplus}(B_t) \quad (5.8)$$

$$dA = \exp(\delta_t * A) \quad (5.9)$$

$$dB = \delta_t * B_t \quad (5.10)$$

where, the *softplus* function is used as an activation function to ensure that δ_t is positive, providing a smooth and continuous dynamic adjustment factor. The matrix A is the initial state transition matrix, which is modulated by δ_t to produce the dynamic state transition matrix dA . Similarly, δ_t adjusts the matrix B_t to produce the dynamic input matrix dB . These dynamic matrices allow the model to adaptively adjust the influence of past states and current inputs, enhancing its ability to capture temporal dependencies.

The Mamba block uses another set of linear transformation combined with *SiLU* activation to capture and extract the static features of F^3 . By combining these static features with the dynamic features obtained from SSM, the model can better capture long-term dependencies and short-term variations in the data. This enhances the overall representational power and robustness of the model, improving its performance and accuracy. This process can be illustrated as:

$$Mamba_{out} = Linear(y_t * SiLU(Linear(F^3))) \quad (5.11)$$

$Mamba_{out}$ denotes the 2D feature matrix output by the Mamba block, which is converted into a feature map with spatial characteristics, consisting of m channels, which is empirically set to 64. The extracted feature map has the same spatial size and number of channels as F^3 . Finally, this extracted feature is passed to the final stage for binary decision-making.

5.2.4 Binary Decision Making for HCD

In the final stage, the extracted feature is flattened and fed into a fully connected layer. This layer is responsible for the binary decision-making process, interpreting the refined features and producing the final binary classification output. By leveraging the comprehensive feature representation obtained from the preceding stages, this stage aims to make accurate and reliable predictions.

5.3 Experiments and Results

In this section, we evaluate the effectiveness of the proposed method by comparing it with three start-of-the-art unsupervised methods, which include the CVA [37], PCAKM [46] and ABBD [2] as well as five DL-based methods such as ML-EDAN [89], HyGSTAN [152], CSANet [79], CBANet [1] and SSTFormer [92], which represent either classical or the state-of-the-art techniques in HCD.

- 1) **CVA [37]**: This method relies on the computed Euclidean distance between spectral vectors of bitemporal images. OTSU thresholding is applied to generate a binary change map.
- 2) **PCAKM [46]**: Principal Component Analysis is used to reduce data dimensionality and redundancy, k-means is subsequently applied to classify pixels into changed or unchanged categories.
- 3) **ABBD [2]**: This method evaluates binary distance for each band to indicate pixel changes. The band-wise binary distance maps are accumulated into a grayscale change map, which is then processed using k-means clustering for binary classification.
- 4) **ML-EDAN [89]**: A two-stream encoder-decoder model that integrates hierarchical features from convolutional layers of bitemporal images. It employs a context-guided attention module for enhanced feature transfer and an LSTM subnetwork to capture temporal dependencies.
- 5) **HyGSTAN [153]**: This self-attention-based network uses cosine similarity to reduce spectral redundancy. It employs gated spectral-spatial attention with single-head weak self-attention for spatial feature extraction and gated spectral-spatial-temporal attention for capturing temporal changes.
- 6) **CSANet [79]**: A traditional self-attention-based method that enhances joint spatial, spectral, and temporal feature representation within each band.
- 7) **CBANet [1]**: A 2-D self-attention-based method that combined with a cross-band feature extraction module, enhancing feature representation and discrimination.
- 8) **SSTFormer [84]**: This method employs a joint spectral, spatial, and temporal transformer to integrate and extract relevant change features using self-attention mechanisms.

The benchmarks are established according to the specified parameters in the default settings, where DL-based methods are trained using PyTorch on an NVIDIA RTX A2000, with a batch size of 32 and training epochs of 200. For training, 20% of pixels from both

changed and unchanged regions are randomly selected, while the remaining pixels are used for testing. To ensure fairness and reliability, each supervised method is repeated ten times in our experiments, and the averaged results of OA , KP , Pre , Re , and $F1$ are reported for comparison. In addition, in the resulting change maps comparison, false alarms (FP) and missing pixels (FN) are highlighted in red and green, respectively, while correctly detected changed pixels (TP) are put in white, and true negatives (TN) are depicted in black for ease of visual comparison.

5.3.1 Results analysis for the River dataset

Table 5. 1 Quantitative assessment of different methods on the River dataset

	$OA(\%)$	KP	Pre	Re	$F1$
CVA	92.53	0.6528	0.5393	0.9635	0.6915
PCAKM	95.17	0.7478	0.6524	0.9506	0.7738
ABBD	96.38	0.7928	0.7377	0.9045	0.8126
ML-EDAN	96.96±0.0014	0.8009±0.0049	0.9078±0.0012	0.8961±0.0088	0.9016±0.0013
HyGSTAN	97.22±0.0012	0.8202±0.0048	0.9211±0.0109	0.9002±0.0075	0.9101±0.0024
CSANet	97.43±0.0012	0.8360±0.0049	0.9296±0.0022	0.9081±0.0030	0.9185±0.0012
CBANet	97.65±0.0036	0.8526±0.0036	0.9346±0.0105	0.9235±0.0065	0.9262±0.0008
SSTFormer	97.46±0.0015	0.8383±0.0023	0.9266±0.0216	0.9137±0.0187	0.9191±0.0011
Proposed	97.91±0.0004	0.8662±0.0026	0.9402±0.0027	0.9261±0.0038	0.9330±0.0014

The quantitative assessment and extracted change maps on the River dataset for all benchmarks are shown in Table 5.1 and Figure 5.2, respectively. Firstly, from the comparison of extracted change maps, the binary maps produced by these three unsupervised algorithms contain a significant number of false alarms or missing pixels. In contrast, all DL-based algorithms detect only a few false alarms or missing pixels. This observation is also supported by quantitative analysis that the Pre of these three unsupervised algorithms does not exceed 0.74, while all DL-based algorithms achieve a Pre of over 90%. Similarly, the $F1$ for all unsupervised algorithms are below 0.82, whereas the $F1$ for all DL-based algorithms are all above 0.90. However, due to the CVA method misclassifying many unchanged pixels as changed category, it results in very few missing pixels (FN). Consequently, CVA has the highest Re among all benchmarks, but the lowest

F1 score, due to the highest number of false alarms (*FP*). Among all DL-based methods, ML-EDAN achieved the worst results, with an average *KP* of only 0.8009 and an average *F1* of 0.9016. HyGSTAN, CSANet and SSTFormer have *KP* values ranging from 0.81 to 0.84 and *F1* between 0.90 and 0.92. CBANet consistently ranks second across all benchmarks, with average *KP* and *F1* of 0.8526 and 0.9296, respectively. The proposed method outperforms all others, achieving the highest scores across *OA*, *KP*, *Pre* and *F1*. Notably, it shows an improvement of approximately 0.0136 in *KP* and 0.0156 in *F1* compared to the second-best method, CBANet. Additionally, the proposed method has the highest *Pre* among all benchmarks, indicating a high accuracy in detecting *TP* with only a few false alarms. These quantitative analysis results demonstrate the proposed method's superior capability in accurately detecting changes with minimal false alarms and missing pixels, confirming its effectiveness and reliability in HCD tasks.

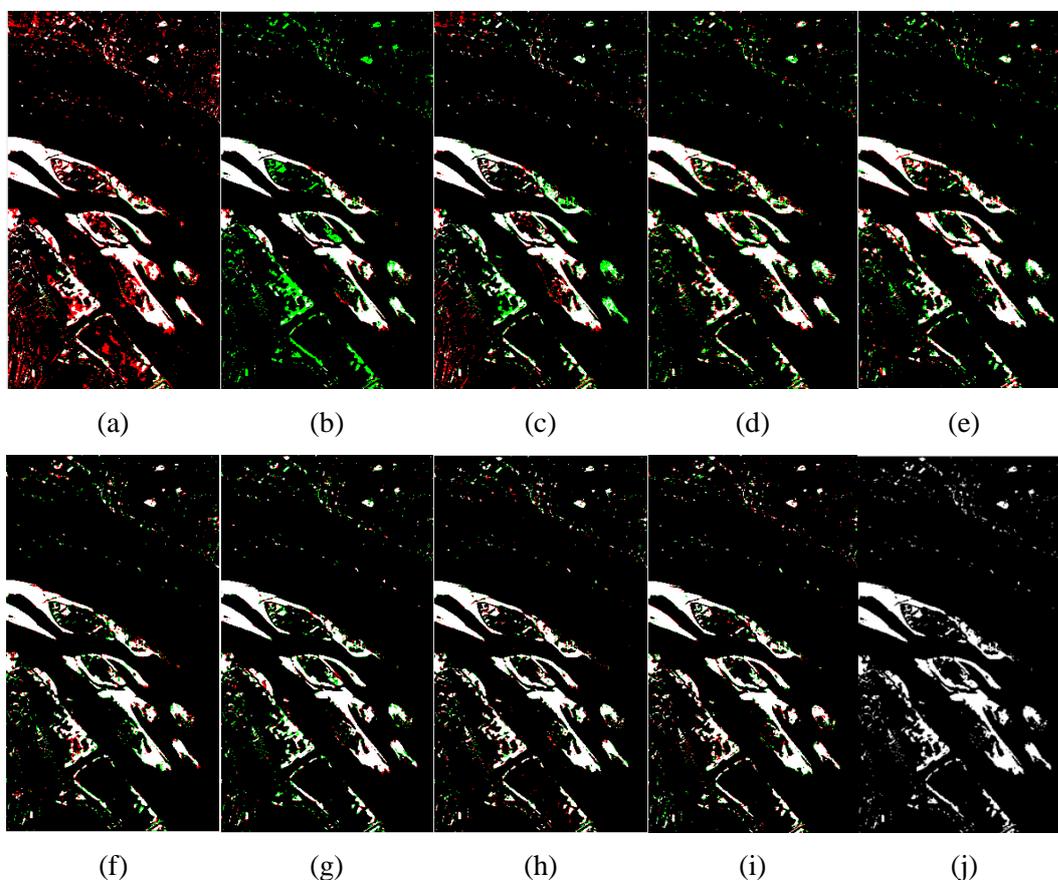


Figure 5. 2 Extracted change maps on the River Dataset from different methods of CVA (a), PCAKM (b), ABBD (c), ML-EDAN (d), HyGSTAN (e), CSANet (f), CBANet (g), SSTFormer (h) and proposed (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

5.3.2 Results analysis for the Hermiston dataset

The extracted change maps and quantitative analysis results for all benchmarks on the Hermiston dataset are presented in Figure 5.3 and Table 5.2, respectively. Similar to the findings on the River dataset, CVA and PCAKM exhibit a significant number of missing pixels (*FN*) in their binary maps and detect only a few false alarms (*FP*). This leads to very high *Pre* values, both exceeding 99.5%, but their *F1* are only 0.8127 and 0.7924, respectively. Consequently, they both achieve very high precision (*Pre*) values, exceeding 99.5%, but their *F1* scores are significantly lower at 0.8127 and 0.7924, respectively, making them the lowest among all benchmarks. ABBD stands out among the unsupervised algorithms for its superior detection accuracy, even surpassing ML-EDAN, HyGSTAN, and CSANet in the *KP*.

Table 5. 2 Quantitative assessment of different methods on the Hermiston dataset

	<i>OA</i>(%)	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
CVA	92.87	0.7705	0.9953	0.6867	0.8127
PCAKM	92.24	0.7472	0.9973	0.6574	0.7924
ABBD	97.49	0.9281	0.9434	0.9453	0.9443
ML-EDAN	97.05±0.0020	0.9151±0.0058	0.9595±0.0040	0.9559±0.0054	0.9576±0.0029
HyGSTAN	96.74±0.0009	0.9054±0.0024	0.9598±0.0029	0.9461±0.0018	0.9527±0.0012
CSANet	96.87±0.0005	0.9097±0.0016	0.9585±0.0024	0.9514±0.0030	0.9596±0.0076
CBANet	97.50±0.0021	0.9285±0.0051	0.9648±0.0084	0.9640±0.0039	0.9642±0.0025
SSTFormer	97.34±0.0026	0.9234±0.0075	0.9663±0.0062	0.9563±0.0044	0.9617±0.0037
Proposed	98.00±0.0003	0.9426±0.0008	0.9725±0.0016	0.9702±0.0012	0.9713±0.0004

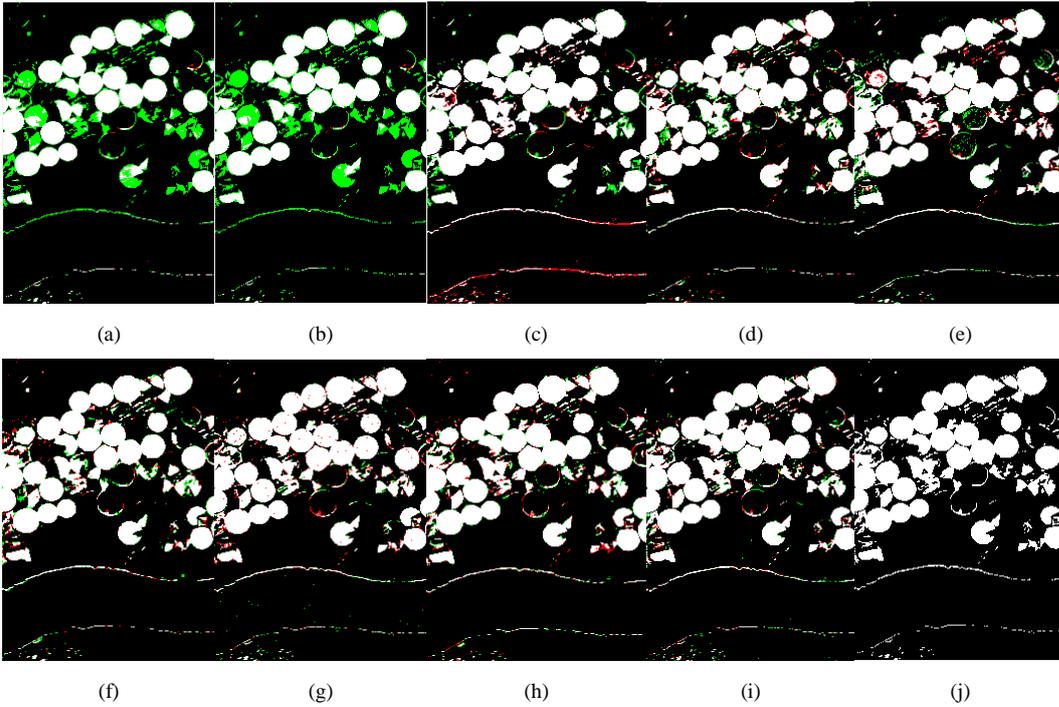


Figure 5.3 Extracted change maps on the Hermiston Dataset from different methods of CVA (a), PCAKM (b), ABBD (c), ML-EDAN (d), HyGSTAN (e), CSANet (f), CBANet (g), SSTFormer (h) and proposed (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively..

Among all DL-based methods, HyGSTAN performs the worst, with a higher number of missing pixels compared to other DL-based algorithms, as illustrated in Figure 5.3 (e). CBANet and SSTFormer demonstrate similar detection capabilities, both achieving average KP and FI scores around 0.92 and 0.96. However, our proposed method significantly outperforms these algorithms, achieving an OA of over 98%. The average KP and FI scores of the Proposed method surpass those of the second-best algorithm by 0.0141 and 0.0071, respectively. Again, these results provide compelling evidence of the superior performance and robustness of our proposed method. Furthermore, the standard deviations for the five evaluation metrics in the quantitative analysis are all below 0.0017, the smallest among all benchmarks, indicating the stability of our proposed method.

5.3.3 Model Efficiency

Table 5.3 Comparing parameters and performance of different DL-based methods

	Parameters	FLOPs	Training Time
	(M)	(M)	(s/per epoch)
ML-EDAN	88.53	569.94	131.36
HyGSTAN	0.03	0.83	2.26
CSANet	2.45	144.33	29.53
CBANet	0.45	11.05	18.49
SSTFormer	2.53	443.28	44.19
Proposed	1.63	136.71	24.95

Table 5.3 provides a comprehensive comparison of the number of parameters, FLOPs, and the training time of single epoch for all DL-based benchmarks on the River dataset, which contains 198 bands, with a batch size set to 1. From Table 5.3, it is evident that our proposed algorithm ranks third in terms of parameter count, it boasts a significantly lower parameter count compared to ML-EDAN and SSTFormer. This reduction in parameters translates to lower computational resource requirements. Moreover, despite the lower parameter count, our method does not compromise on performance. It demonstrates outstanding detection capabilities, achieving high accuracy and robustness. This exceptional detection performance, coupled with the reduced computational load, highlights the strength of our approach. It strikes a perfect balance between efficiency and effectiveness, making it an optimal choice for scenarios where both resource constraints and high performance are critical.

5.3.4 Patch size

Experiments were conducted using five patch sizes: $\{5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11, 13 \times 13\}$, the variations of KP are illustrated in Figure 5.4. It can be observed that the average of KP for test accuracy initially increased and then declined, with both metrics peaking at a patch size of 7×7 . Moreover, it is evident that as the patch size increases, the training time progressively lengthens. Observing these trends, it becomes clear that the patch size of 7×7 can balance performance and efficiency.

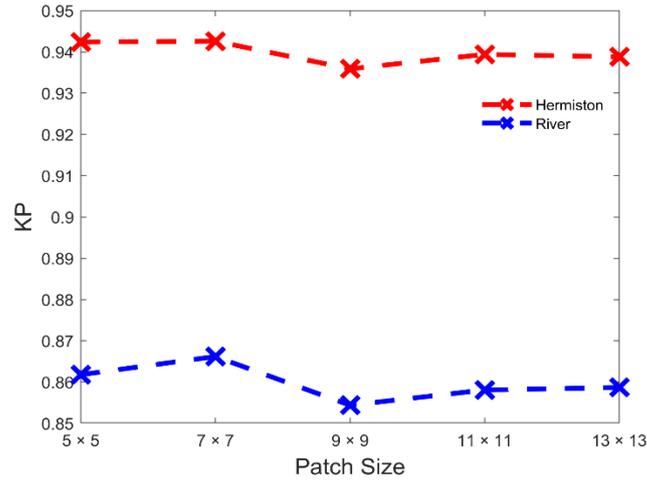


Figure 5. 4 Experiments and results of proposed GASSM under different patch sizes on two datasets of River and Hermiston.

5.3.5 Kernel Number of Feature Extracted Module

The kernel number of the feature extraction module determines the number of channels in the feature map F^3 , which subsequently serves as the input for the Mamba block. To investigate the impact of different kernel numbers, experiments were conducted with the following values: {8, 16, 32, 64, 128}. The KP curves of two datasets are depicted in Figure 5.5 reveals that setting the kernel number to 64 yields the optimal performance across both datasets.

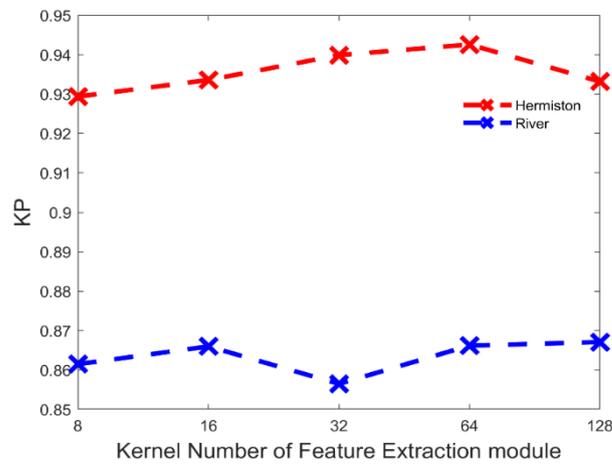


Figure 5. 5 Experiments and results of proposed method under different kernel numbers of feature extraction module on two datasets of River and Hermiston.

5.3.6 Number of $Mamba_{out}$ channels

The number of $Mamba_{out}$ channels determine the number of output channels in the Mamba block, which subsequently affects the size of the flattened dimensions in the subsequent fully connected layer. To investigate the impact of different $Mamba_{out}$ channel numbers, experiments were conducted using the following values: {8, 16, 32, 64, 128}. As shown in Figure 5.6, the KP curves for two datasets are depicted in the figure. The results reveal that the KP values for the Hermiston dataset remain consistently high, ranging from 0.94 to 0.95, regardless of the number of $Mamba_{out}$ channels. This indicates that the Hermiston dataset maintains high performance with minimal impact from varying the number of $Mamba_{out}$ channels. In contrast, the KP values for the River dataset are ranging from 0.85 to 0.87. However, within this range, the optimal performance is observed when the number of $Mamba_{out}$ channels is set to 64. Thus, for both datasets, a $Mamba_{out}$ channel number of 64 is recommended to achieve the best model performance.

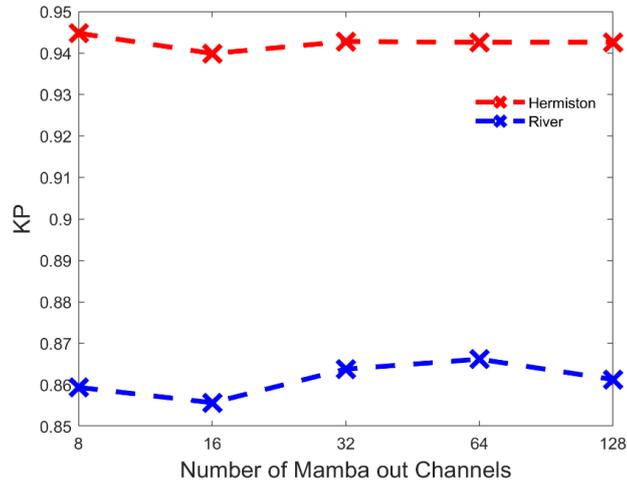


Figure 5. 6 Experiments and results of proposed method under different channels of $Mamba_{out}$ on two datasets of River and Hermiston.

5.3.7 Key Stage Analysis

This section primarily analyses the role of each module within the model. By separately removing the three main modules (spectral attention module, spatial attention module, and feature extraction module), we evaluate their impact on the model's performance. Figure 5.7 presents the quantitative assessment results on the River and Hermiston datasets.

Although removing certain modules can reduce the model's parameters and FLOPs, the overall performance analysis indicates that removing any module negatively affects the model's detection accuracy. The feature extraction module has the most significant impact on both datasets, particularly on the Hermiston dataset, where its removal leads to a substantial performance drop. Removing the spatial attention module also results in notable declines across various metrics, especially the KP value, highlighting the importance of spatial attention in enhancing performance on both datasets. The spectral attention module, when removed, causes slight decreases in performance metrics, particularly for the Hermiston dataset, indicating its helpful role in improving model performance. The proposed combination of these modules achieves optimal performance, ensuring that the model performs exceptionally well across different datasets.

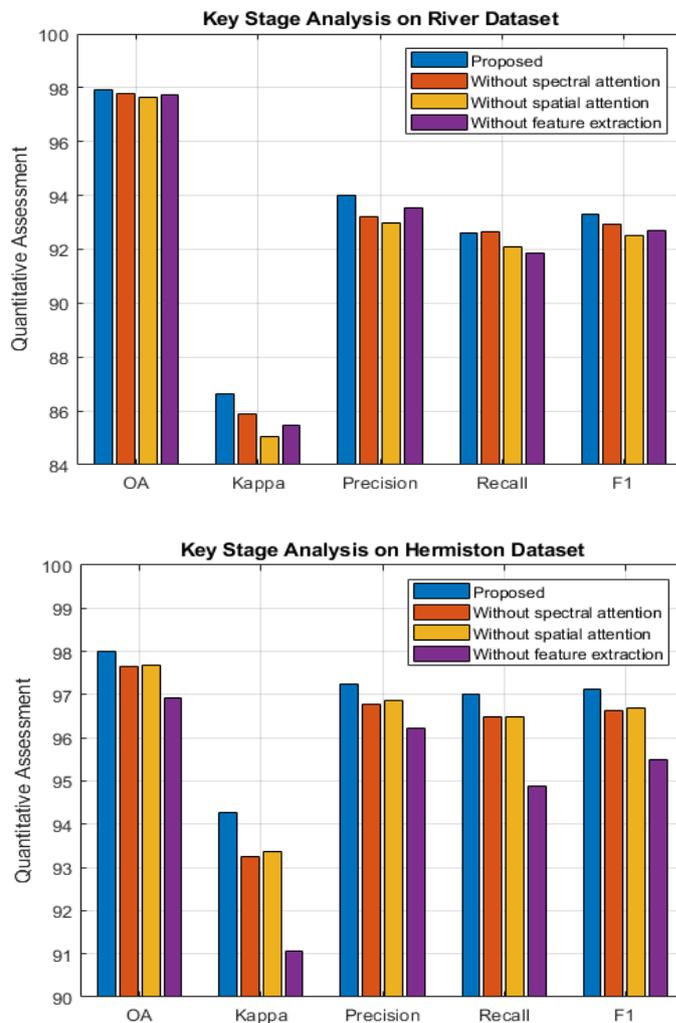


Figure 5. 7 Key stage analysis of proposed network on the River and Hermiston datasets

5.3.8 Further Discussion

The proposed method demonstrates significant advantage in terms of higher detection accuracy compared to state-of-the-art benchmarks across two datasets. Firstly, traditional attention mechanisms typically extract information within a local scope, whereas the global attention mechanism can fuse information over the whole image. This global perspective ensures that even subtle changes are precisely detected, significantly reducing false alarms and missed detections. By comprehensively considering spatial and spectral features, our method achieves a notable improvement in detection accuracy compared to other methods.

The comparison with unsupervised algorithms, such as CVA, PCAKM and ABBD, while computationally efficient and relatively simple to implement, often struggle with false alarms or missing pixels, leading to lower *Pre* and *F1* score, this is primarily due to their limited capability to exploit spatial context and noise sensitively. PCAKM and ABBD, although better than CVA in certain metrics, still fall short in comparison to DL-based methods. These findings reinforce the limitations of traditional unsupervised methods in handling the complexity of change detection tasks, particularly in diverse and intricate datasets like River and Hermiston. In contrast, DL-based methods exhibit remarkable capabilities in extracting and integrating relevant features, achieving higher *Pre*, *Re* and *F1* scores. ML-EDAN integrates hierarchical features from convolutional layers and employs a context-guided attention module and an LSTM subnetwork for temporal dependencies. Despite its advanced architecture, ML-EDAN's performance is hindered by the complexity of the model, which may lead to overfitting and increased computational burden. In addition, the integration of hierarchical features may not fully capture the global spatial and spectral context. Although HyGSTAN has minimal parameters, its single-head weak self-attention mechanism seems insufficient for complex change detection tasks that require capturing global dependencies. CSANet enhances joint spatial, spectral, and temporal feature representation using a traditional self-attention method. While it effectively captures detailed features within each band, CSANet may struggle with integrating these features into a coherent global context, resulting in lower accuracy. CBANet enhances feature representation and discrimination by introducing 2-D self-attention. Despite its strong performance, CBANet may not fully leverage long-range dependencies relationships, which are crucial for precise change detection. Transformer-based SSTFormer network also leverages extensive linear transformations and self-attention mechanisms, making it powerful in capturing complex features, it only extracts static features and does not account for long-term feature dependencies. While its detection

accuracy is somewhat better than other benchmarks, both the model's efficiency and detection accuracy fall short compared to our proposed method, which is based on dynamic feature extraction.

In summary, the superior performance of our proposed GASSM is attributed to its advanced feature extraction and state space model, which effectively captures dynamic, spatial, and spectral characteristics. This comprehensive approach ensures high accuracy in distinguishing the characteristics of changes, making our method a reliable and efficient solution for HCD tasks.

5.4 Summary

In this chapter, a novel end-to-end DL-based network, GASSM, is proposed for HCD task. The detection accuracy of GASSM surpasses existing state-of-the-art benchmarks and demonstrates robustness across various testing scenarios. Its innovation architecture includes a global attention and a dedicated feature extraction module, effectively enhancing the spatial-spectral feature interactions and reducing feature redundancy. The Mamba block incorporates a state space model to adeptly capture long-range dependencies, highlighting the network's ability to comprehensively model global spatial-spectral features. Additionally, GASSM has lower computational costs compared to most deep learning methods, particularly those based on transformers.

Due to the necessity of manually labelled training sets, our GASSM model still faces a common limitation inherent to all supervised learning algorithms. To address this challenge, our next step is to leverage the powerful feature representation capabilities of GASSM by utilising it as the backbone and incorporating pseudo ground truth generated by unsupervised algorithms. Specifically, we will experiment with different combinations of unsupervised algorithms and enhance the accuracy of the pseudo ground truth through a majority voting mechanism. Our strategy is to implement a self-supervised learning approach that does not require any manually annotated training samples.

Chapter 6 Hyperspectral Anomaly Detection

6.1 Introduction

The hyperspectral anomaly detection (HAD) task focuses on identifying abnormal pixels within a single HSI that exhibit significantly different spectral and spatial characteristics from their surroundings, without any prior information about the anomalies. Developing highly accurate detection algorithms is crucial for various applications, such as military reconnaissance and environmental monitoring. However, current research faces the following three main challenges: 1) the spectrum is complex and changeable, usually affected by various factors such as changes in environment, atmosphere and temporal conditions; 2) The correlation between adjacent bands is strong, and the information is redundant; 3) The spatial resolution of the image is limited, resulting in the widespread phenomenon of complex background and mixed pixels. Low detection rates and high false alarms are thus involved due to the mixed pixels. In response to these challenges, there has been a concerted effort within the research community to develop more robust HAD algorithms that can achieve higher accuracy with robustness.

In Section 6.2, this section reviews prominent HAD methodologies that utilise statistical-based techniques, representation-based methods, and DL-based approaches, detailing the strengths and weaknesses of each. It also outlines publicly available HAD datasets and the metrics used to evaluate detection accuracy.

Following this review, Section 6.3 introduces a novel unsupervised HAD algorithm that combines 1-D SSA with a sparse autoencoder. SSA is applied within the spectral domain to denoise data and tackle the issues arising from strong inter-band correlations and redundant information. The sparse autoencoder operates in the spatial domain, employing sparsity constraints to enhance the learning and reconstruction of crucial background features. This method efficiently isolates anomalous pixels from the background, thereby improving the precision of anomaly detection. Comprehensive experiments on six publicly available datasets, which include anomalies such as airplane, vehicle, and building scenarios, have demonstrated that this proposed approach significantly improves the distinction between anomalies and their backgrounds. The proposed unsupervised HAD algorithm surpasses six advanced methods in terms of both detection accuracy and robustness.

6.2 Related Work on Hyperspectral Anomaly Detection

HAD involves analysing a single HSI to detect anomalous pixels in a geographical region without prior knowledge of the target, which is a sophisticated process utilised in RS to identify objects or phenomena that differ significantly from their surroundings [153]. The basic workflow of HAD is illustrated in Figure 6.1 and primarily consists of three steps. The primary goal of the feature extraction step is to derive spatial, spectral, and spatial-spectral features that highlight the characteristics of anomalous targets from backgrounds. Since the single HSI is analysed without prior information about the anomalous target, it becomes more difficult to separate abnormal pixels in a complex background. Currently, most existing feature extraction methods are based on background reconstruction or low-rank and sparse representation. Following this, the detection and segmentation of anomalous pixels identify those that statistically deviate significantly from the background spectral behaviour. Finally, the accuracy and effectiveness of the detection results are evaluated by comparing them with GT during the algorithm development phase to verify the performance of the model. The accuracy of anomaly detection is critically evaluated, often employing metrics like receiver operating characteristic (ROC) curves [154] and area under the curve (AUC) [155].

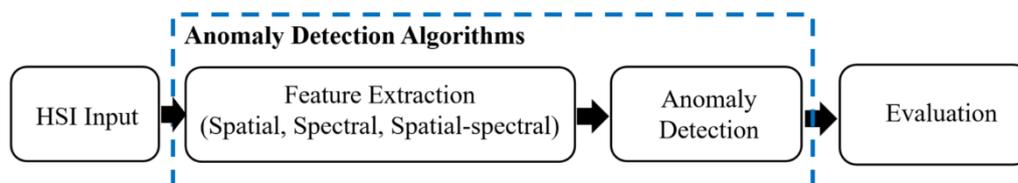


Figure 6. 1 Basic process of Hyperspectral anomaly detection.

Over the past few decades, numerous HAD methods have been developed. The following sections will provide an overview of the composition and principles of representative algorithms from traditional unsupervised methods, including statistics-based, representation-based and deep learning-based methods, including unsupervised DL and self-supervised DL. In the summary section, the advantages and disadvantages of each category will be discussed. Additionally, all the proposed HAD algorithms from 2021 to July 2024 will be categorised and a statistic analysed will be presented at the end of this chapter. Finally, public datasets and evaluation metrics for HAD are introduced.

6.2.1 Statistics-based Algorithms

Statistics-based methods are among the earliest and most widely used techniques for HAD. These methods require the assumption that the background pixels of a hyperspectral image conform to a specific statistical model, with pixels that deviate from this model marked as anomalies. The Reed-Xiaoli (RX) algorithm [156], proposed by Reed and Yu in 1990, is a representative statistical method for HAD. RX algorithm assumes that the background pixels adhere to a multivariate Gaussian distribution, calculates the mean and variance of background pixels and uses the Mahalanobis distance to identify anomalies. Due to its efficiency and simplicity in straightforward scenarios, the RX algorithm is considered a foundational technique in HAD. Subsequent advancements have built upon the RX algorithm to enhance its performance. For example, the Kernel-RX (KRX) [157] uses kernel functions to project hyperspectral data into a higher-dimensional space, which improves the separation between anomalies and background pixels, thus boosting detection accuracy. The Local-RX (LRX) [158] algorithm employs locally adaptive kernel density estimation to model background data more effectively, minimizing noise interference. Other methods like Weighted-RX (WRX) [159] and Linear Filter-Based RX (LF-RX) [160] have also been developed to tackle HAD. WRX increases the weights of background pixels while decreasing the weights of noise pixels, enhancing the evaluation of background information and improving detection precision. Conversely, LF-RX filters noise pixels to achieve a more accurate estimate of the background covariance matrix, leading to more reliable detection results. Another improved method involves obtaining a pure background by removing potential anomalous pixels. For example, the subspace RX [161] reduces the impact of anomaly contamination on background modelling by eliminating several background dimensions with large variances. The local adaptive iterative RX [162] algorithm iteratively removes potential anomalies until the detected anomalous targets are the same as those removed in the previous iteration. These RX-based statistical methods are part of the early HAD methods. They are advantageous due to their straightforward approach. This can lead to effective anomaly detection in simple scenarios. However, these methods often assume that the background data follows a multivariate Gaussian distribution, an assumption that is rarely met with real-world data. As a result, their detection accuracy tends to be quite limited when applied to the complex and varied nature of actual hyperspectral data. In addition, RX-based methods require calculating the inverse of the covariance matrix to describe the statistical properties of the background. Due to the

high dimensionality of HSIs, this operation has a high computational complexity, which limits the RX-based algorithm's scalability for large-scale applications.

In recent years, various non-RX-based statistical methods have also been widely applied in HAD tasks. For instance, Schweizer and Moura [163] proposed an adaptive detection algorithm, which based on a three-dimensional Gaussian Markov random field uses a joint spatial-spectral random field to represent background statistical properties. By considering the correlation between spatial and spectral information, it provides an explicit expression for the inverse of the covariance matrix, effectively mitigating the computational cost associated with the inversion process in RX algorithms. Moreover, some subspace-based methods have shown high detection capabilities by projecting the original data into a lower-dimensional subspace where the differences between background and anomalies are more distinct. Xiang *et al.* [164] proposed a method that combines Mahalanobis distance detection with spectral angle distance detection to construct a feature projection space. Chang *et al.* [165] employ an isolation discriminative forest model to perform effective subspace searches during the binary tree splitting process. Song *et al.* [166] proposed a statistics-based method that integrates independent components analysis (ICA) with orthogonal subspace projection iteratively extracts anomalous components using ICA and effectively suppresses the background by employing orthogonal subspaces for the background and anomalies.

6.2.2 Representation-based Algorithms

Representation-based HAD methods do not require assumptions about the statistical distribution of the background, thereby overcoming the limitations of statistical methods. The main idea is that the background pixels in hyperspectral data can be approximated by spatial neighbourhood pixels or a background dictionary, whereas anomalous pixels cannot. The basic process of these methods involves first constructing a background dictionary and a sparse coefficient matrix. Then, the reconstruction error of the test pixels is calculated based on the background dictionary and coefficient matrix to determine whether a pixel is an anomaly. Based on the constraints on the dictionary representation coefficients, representation-based methods can be further categorized into sparse representation (SR) methods, collaborative representation (CR) methods and low-rank representation (LRR) methods.

SR-based methods assume that background pixels can be sparsely represented by several atoms in an over-complete dictionary, whereas anomaly pixels cannot. Yuan *et al.*

[167] proposed a novel method based on local sparsity divergence without any distribution hypothesis. Firstly, a sliding double window strategy is employed to construct a local spectral and spatial dictionary, enabling the extraction of sparse coefficients for each pixel in HSIs. Next, a consistent sparse dispersion index is introduced to compute the local sparse dispersion map (LSD) for each band individually. Finally, the LSD maps for various bands are combined and segmented to produce the final anomaly map. Li *et al.* [168] developed a technique that employs adaptive orthogonal background complementary subspace, estimated through joint sparse representation. This method adaptively selects the most representative background bases tailored for each local region, enhancing the accuracy of the background modelling process. Zhao *et al.* [169] employ archetypal analysis scheme as spectral unmixing method to extract the representative background endmember signatures. Then, the archetypal analysis unmixing reconstruction error is integrated with the structured sparse representation reconstruction error to distinguish the anomalous pixels from the background. Zhu *et al.* [170] utilised the endmember abundance matrix to derive the background dictionary, and an adaptive weight sparse matrix is incorporated into the residual matrix to effectively suppress the background.

The SR-based method leverages the competition among atoms in an overcomplete dictionary to achieve optimal representation of the original data using only a few non-zero atoms, i.e., the sparse representation coefficients of the background. In contrast, CR prioritizes cooperation over competition among dictionary atoms. It involves all atoms in the dictionary in the linear representation to achieve optimal results. Li *et al.* [171] proposed a CR-based HAD method for the first time, suggesting that background pixels can be linearly represented by their spatially neighbouring pixels, whereas anomalous pixels, being relatively rare, do not exhibit this characteristic. To adjust the contribution of each neighbouring pixel to the representation, a distance-weighted regularization matrix is introduced. Additionally, imposing a sum-to-one constraint on the weight vector enhances the stability of the solution. The basic CR assumes that the importance of each band is equal, which is not pragmatic in practical application. To alleviate this problem. Wang *et al.* [172] proposed a self-weighted CR-based detector, which combines the weight learning and CR into a joint objective function. Lu *et al.* [173] proposed a novel ensemble and random CR-based detector that the random sub-sampling is processed to gain several detection results instead of the sliding dual window strategy, which significantly reduces the computational complexity. Then, ensemble learning is employed to refine these multiple random sub-sampling results, providing abundant complementary information to better detect different anomalies. Zhao *et al.* [174] proposed a novel CR-based method, which Fractional Fourier

transform (FrFT) is associated with CR-based detector. FrFT can transfer HSI pixels into a FrFT domain, which can suppress noise and improve the discrimination between background and anomalies.

The SR and CR based HAD methods operate at pixel-level, reconstructing the image pixel-by-pixel. In contrast, LRR describes the global structure of the entire image. In HSIs, background pixels exhibit high similarity in both spatial and spectral domains, making them highly correlated and resulting in low rank. Anomalous pixels occur infrequently and occupy a small proportion of the data, giving them sparse characteristics. Consequently, the LRR-based method can decompose HSIs into a low-rank background component and a sparse anomaly component. Sun *et al.* [175] proposed an anomaly detection method based on low rank and sparse matrix decomposition (LRaSMD). It posits that the background and anomaly matrices exhibit low rank and sparse characteristics, respectively. Utilising convex optimization theory, rapid matrix decomposition is achieved. Then, the Euclidean distance is directly used to determine the anomaly degree of each pixel in sparse anomaly component. Zhang *et al.* [176] utilised LRaSMD technology to estimate background statistical characteristics. Anomalies are then determined by calculating the Mahalanobis distance between the pixels to be detected and the background to ensure the correlation in the anomaly matrix. Qu *et al.* [177] proposed an LRR-based method based on performing low rank decomposition on the abundance vectors from spectral unmixing and background coefficients from dictionary construction. Finally, the sparse part is summed pixel-by-pixel to obtain the anomaly score map, which indicates the degree of anomaly. The traditional low rank and sparse decomposition-based methods assume that the anomalies and noise reside modelled by one signal distribution, which potentially confuses weak anomalies and noise in the sparse component. To address this potential problem, Li *et al.* [178] proposed a modified low rank sparse decomposition model that incorporates a mixture noise model with a low rank background to characterize complex distributions. In this framework, variational Bayes algorithm is utilised to infer a posterior mixture of Gaussian model, facilitating the separation of anomalies from noise components. Finally, a Manhattan distance-based detector is employed for effective anomaly detection under complex distributions.

In summary, representation-based methods can overcome the limitation of statistical-based methods that rely on Gaussian and other assumptions. By determining the optimal coefficient matrix of the background dictionary, these representation-based methods can effectively isolate anomalous pixels. Consequently, representation-based methods can be applied in more complex scenarios and have shown excellent detection accuracy on specific

datasets. However, the construction of background and anomalous target dictionaries is prone to noise interference, and the steps involved in dictionary construction and reconstruction add to the representation-based method's complexity. Then, some coefficients are uncertain and need to be manually set based on experience. Moreover, due to the lack of prior information on anomaly targets, it is challenging to establish the accurately dictionary for the anomalous targets under different backgrounds. Finally, the detection results may vary each time due to the randomness of iterative convergence [179]. As such, developing robust representation-based methods for HAD remains a significant challenge.

6.2.3 Deep Learning-based HAD Algorithms

In many complex scenarios, obtaining the prior information such as the spectra and texture of anomalous targets is challenging. Thus, HAD algorithms are required to employ a 'blind detection' approach to identify these targets [180]. Additionally, the imbalance between anomalous pixels and background further limits the applicability of supervised deep learning methods. To address these limitations, researchers have shifted their focus to unsupervised deep learning methods, which operate under the assumption that anomalies are rare and significantly distinct from normal instances [181]. Typical unsupervised deep learning networks include autoencoder (AE) and generative adversarial network (GAN), with many HAD methods based on AEs and GANs being proposed in recent years.

AE-based methods are commonly unsupervised models and achieved remarkable results in HAD tasks. Zhao *et al.* [182] proposed a HAD method based on stacked denoising AE that learn the nonlinear characteristics and enhance feature extraction capabilities. This approach incorporates two models: the spectral feature extraction model which analyses individual pixel spectra, and the fused feature by clustering model which clusters similar pixels and employs stacked denoising AEs to learn deep nonlinear features. Zhao *et al.* [183] proposed a spectral-spatial stacked AE model, integrating low rank and sparse matrix decomposition to effectively reduce the influence of mixed anomaly targets and background after AE reconstruction. Lu *et al.* [184] proposed a novel manifold constrained AE network, which the latent representations are learned by an AE network with the learned embedding manifold constraints. The reconstruction errors are calculated to distinguish anomalies. Wang *et al.* [185] proposed an autonomous HAD methods, employing a fully convolutional AE with the skip connections to reconstruct the background, an adaptive-

weight loss function is utilised, reducing the weights of pixels with large reconstruction errors, thereby focusing on normal background features. Li *et al.* [186] proposed a spectral difference guided graph attention AE network that utilised a graph attention encoder with a spectral sharpening constraint to captures and emphasizes spectral differences between central nodes and their neighbours to enhance feature extraction. Then, a corresponding decoder reconstructs node attributes and calculates node reconstruction errors to extract anomaly features.

GAN-based methods have also been popularly utilised to reconstruct background within the spatial domain for HAD tasks. Due to the high ratio of the background to anomalies, the GAN-based generator usually demonstrates better learning performance for background characteristics, while the anomaly pixels can be identified by a higher error value compared to background pixels [187]. Jiang *et al.* [188] proposed an integrated approach combining the AE with a GAN framework. Firstly, the AE is used to reconstruct background features from the original image, serving as the GT for the GAN's generator, which allows the discriminator to learn the differences between the reconstructed image and the real original data. The discriminator classifiers pixels with large errors as anomaly through the feedback function loss. Based on this breakthrough. Furthermore, Li *et al.* [32] employed a sparse coding inspired regularized network instead of AE to generator background features. Then, integrated into an end-to-end GAN for unsupervised HAD, which can learn a discriminative latent reconstruction with small errors for background and large errors for anomalous region. Arisoy *et al.* [189] trained a traditional GAN-based model to generate a synthetic background image. By subtracting the synthetic image from the original one, the background can be removed from the original image. Wang *et al.* [190] proposed a GAN-based module to estimate the background distribution, overcoming the deficiency of prior information. During the generation process, a differentiable data augmentation strategy is used to enhance real and fake samples. This augmentation method can maintain the effective propagation of gradients during training, thereby optimizing the training processing of the generator and discriminator.

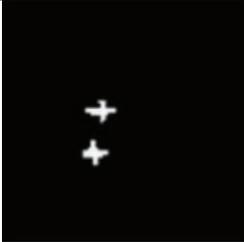
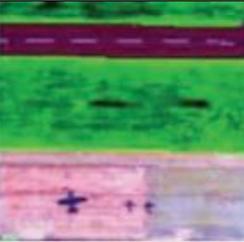
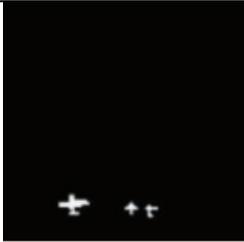
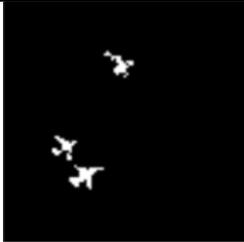
More recently, researchers started to apply self-supervised methods to the HAD tasks. Wang *et al.* [191] proposed a transformer-based self-supervised learning network to reconstruct background images. The proposed transformer-based network is trained by taking pseudo RGB images as input and using HSI cube, which selected through the optimal clustering framework [192] to select the most representative bands as the pseudo HSI cube. The scores by integrating weight maps from different layers along with the root mean square error are calculated, thereby highlighting anomalies with high scores. Liu *et*

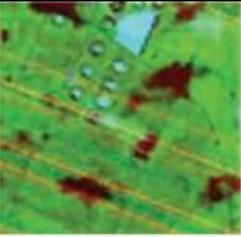
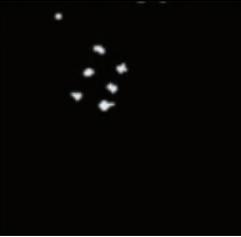
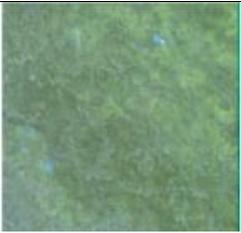
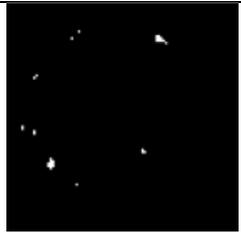
al. [193] proposed a novel self-supervised HAD method. Firstly, a ResNet34-based classification model is trained in a self-supervised manner, which on an expanded dataset that introduces pseudo-anomalies, aims to identify and extract features from these pseudo-changes, thus generating feature representation vectors from the feature maps within the network. These vectors capture the key characteristics of anomalies and the background in the data. Then, the network processes a new dataset to obtain feature representations of both anomalies and background. These representation vectors serve as prior dictionary information for low-rank and sparse representations, reconstruct the test data to effective separation of anomalies from the background. Gao *et al.* [31] proposed a novel self-supervised learning method based on blind-spot prediction. This method involves masking the centre pixel of each image patch and using the surrounding pixels as input to predict the masked pixel's value, focusing on reconstructing the background. The difference between the original and reconstructed values serves as pseudo labels, optimising the network by minimising this reconstruction error. During testing, pixels with high reconstruction errors are flagged as potential anomalies. While these self-supervised methods demonstrate enhanced detection accuracy on specific datasets compared to unsupervised algorithms, they incur higher computational costs due to the necessities of network pre-training and pseudo label generation. Furthermore, a significant limitation of these methods is their lack of robustness across diverse datasets. Therefore, the specific pipeline of self-supervised learning approaches applied to HAD task still requires further exploration and refinement to enhance their adaptability and efficiency.

6.2.4 Datasets for Hyperspectral Anomaly Detection

This subsection sorts real published dataset that are widely used for HAD algorithms development. The San Diego dataset [194], and airport-urban dataset [195] are all collected by Airborne Visible/Infrared Imaging Spectrometer (AVIRS), which cover a spectrum range of 0.4 to 2.5 μm and provide 224 bands. Some other publicly available datasets, such as MUUFL Gulfport [196] is acquired by CASI sensor. Hyperion dataset [197] is collected by EO-1 Hyperion sensor. The Detailed characteristics of each dataset are summarized in Table 6.1.

Table 6. 1 Hyperspectral anomaly detection datasets

Dataset	Pseudo Image	Ground Truth	Size	Captured data	Anomalies
Airport-1			100×100× 205	Nov. 2011	Airplanes
Airport-2			100×100× 205	Nov. 2011	Airplanes
Airport-3			100×100× 205	Nov. 2011	Airplanes
Airport-4			100×100× 205	Nov. 2011	Airplanes
San Diego			100×100× 186	Aug. 1995	Airplanes
Bench			150×150× 102	Jul. 2010	Vehicles

Urban			100×100× 205	Aug. 2010	Buildings
MUUFLL Gulfport			325×220× 64	Nov. 2010	Cloth panels
Hyperion			150×150× 149	Oct. 2008	Storage silo

6.2.5 Evaluation Criteria

Generally, the Receiver operating characteristic (ROC) curve [154] and area under the ROC curve (AUC) [155] are used for the qualitative analysis of the HAD model. The abscissa of the ROC curve is the false positive rate (FPR), and the ordinate is the true positive rate (TPR). ROC curve demonstrates the trade-off between the model's ability to capture positive samples and its rate of mistakenly labelling negative samples by altering the decision threshold. TPR is the proportion of positive samples correctly identified as positive and FPR is the proportion of negative samples that are incorrectly classified as positive. TPR and FPR can be expressed as:

$$TPR = \frac{TP}{TP + FN} \quad (6.1)$$

$$FPR = \frac{FP}{FP + TN} \quad (6.2)$$

where TP , TN , FP , and FN denote the correctly detected changed pixels, correctly detected unchanged pixels, incorrectly detected changed pixels, and incorrectly detected unchanged pixels, respectively.

The AUC value [155], representing the area under the ROC curve, quantifies the overall performance of a classifier and ranges from 0 to 1. The larger the AUC value, the better the detection accuracy of the algorithm.

Box plot [198] is an effective tool for evaluating an algorithm's ability to suppress background noise and distinguish the target. The box plot displays the interquartile range (IQR), with the top and bottom of the box indicating the upper and lower quartiles, which encompass 50% of the data. The median, representing the central tendency of the data, is marked within the box. The whiskers of the plot extend to the maximum and minimum values in the dataset. An example of the box plot [199] is shown in Figure 6.2, the height of the red and blue boxes illustrates the background suppression and abnormalities across different algorithms. The lower the blue box, the more severe the background is suppressed; the gap between the red box and the blue box represents the algorithm's separation of the background and the anomaly. The larger the interval, the more conducive to the separation of the target and the anomaly which means the separation of the algorithm is better

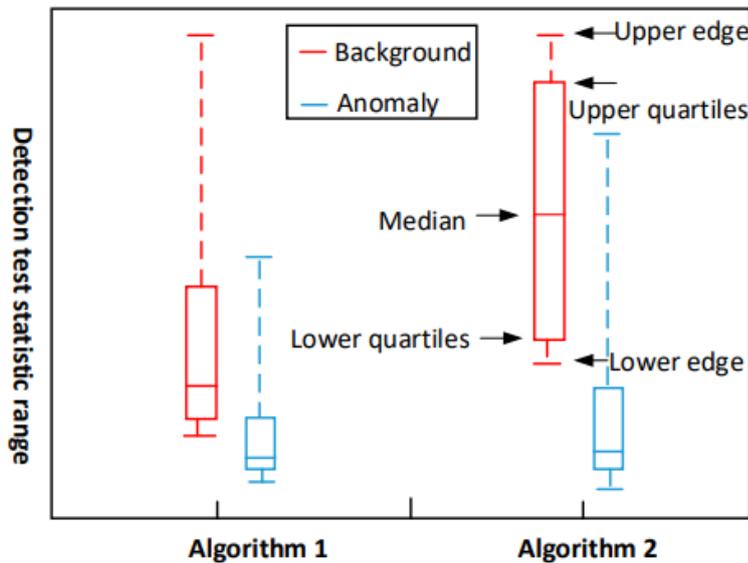


Figure 6. 2 An example of box plot

6.2.6 Summary

In summary, statistics-based methods have several key advantages, including high interpretability and low computational complexity. These methods can analyse data using simple statistical models, making them easy to understand and implement. However, they are highly sensitive to noise and have poor adaptability, making them less effective in handling complex anomalies. Representation-based methods do not require assumptions about the statistical distribution of data and are suitable for complex scenarios. Nevertheless, these methods lack prior knowledge of anomalous targets when constructing and reconstructing dictionaries, making them susceptible to noise interference. Unsupervised deep learning methods offer the advantage of high-quality background reconstruction and excellent performance in complex backgrounds. However, they require long training times and may suffer from overfitting, which can negatively impact their performance on new data. Self-supervised deep learning methods effectively utilise unlabelled data and do not require prior knowledge. They perform well on certain datasets. However, their detection accuracy can be affected by noise or inaccurate label generation, the pre-training process is complex, and they lack robustness. Based on the analysis of each type of algorithm, their advantages and disadvantages are summarized in Table 6.2.

Table 6. 2 Summary of the advantages and disadvantages of different types of HAD algorithms

Method	Advantages	Disadvantages
Statistic-based	<ul style="list-style-type: none"> • High interpretability • Low computational complexity 	<ul style="list-style-type: none"> • Sensitive to noise • Poor adaptability
Representation-based	<ul style="list-style-type: none"> • No statistical distribution assumptions required • Suitable for complex scenarios 	<ul style="list-style-type: none"> • Constructing and reconstructing dictionaries are susceptible to noise interface • Some coefficients need to be set manually
Unsupervised deep learning-based	<ul style="list-style-type: none"> • High-quality background reconstruction • Excellent performance under complex background 	<ul style="list-style-type: none"> • Long training time • Possible overfitting
Self-supervised deep learning-based	<ul style="list-style-type: none"> • Effective use of unlabelled data • No prior knowledge required • Performs well on certain datasets 	<ul style="list-style-type: none"> • Detection accuracy is affected by noise or inaccurate label generation • Pre-training is complex • Lack of robustness

As shown in Figure 6.3, articles we reviewed on HAD algorithm development from 2021 to July 2024 have been compiled, revealing a steady year on year increase in the total number of publications. Specifically, there were 72 articles published in 2021, 85 in 2022, and 91 in 2023. By July 2024, 59 articles had already been published, surpassing half the total number of articles published in 2023.

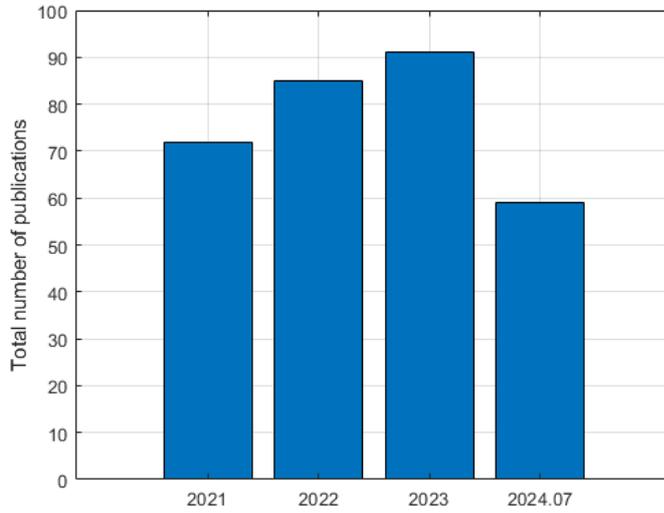


Figure 6. 3 The total number of algorithms proposed each year until July 2024.

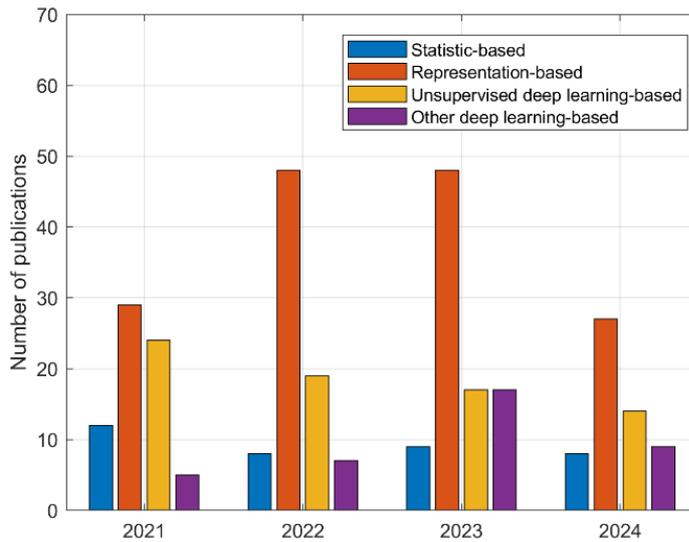


Figure 6. 4 The number of different types HAD models each year by July 2024.

As can be seen from Figure 6.4, the representation-based HAD algorithm has remained the focus of researchers' research in the past three years and deep learning methods based on self-supervised are rising in popularity. These self-supervised methods also offer new ideas for the future work plan. In addition, the number of review paper published each year is as follows: 2 in 2021 [153][30], 3 in 2022 [200,199,201], and 1 so far in by July 2024 [29].

6.3 Proposed Method

As summarised in section 6.2, there are still many challenges in HAD algorithm development. For instance, HSIs often have high dimensions, leading to high processing costs and susceptibility to noise. Additionally, the spectral characteristics can vary greatly across different complex background scenarios, demanding models with robust generalization capabilities to adapt to various conditions. To tackle these difficulties, a novel unsupervised HAD method is proposed. First, 1-D SSA is employed to eliminate outliers in the spectral domain. Second, the SSA-smoothed hypercube undergoes a sparse autoencoder for background reconstruction, where the reconstruction error is used to extract anomalous pixels. Finally, the RX algorithm is employed to segment anomalous pixels from the background. Comprehensive experiments on six publicly available datasets, including abnormal objects in airplane, vehicle and buildings, have validated the superior performance of our method in effectively enhancing the separability between anomaly pixels and their respective backgrounds, outperforming a few state-of-the-art methods, particularly in terms of the detection accuracy and robustness.

6.3.1 Methodology

The flowchart of the proposed algorithm is illustrated in Figure 6.5, which contains three main parts. Initially, we employ the first component of SSA to derive the smoothed hypercube so as to mitigate the noise and outliers in the spectral domain. The denoised hypercube is subsequently fed into a sparse AE model for background reconstruction. We extract anomalous targets based on the reconstruction error. Ultimately, the conventional RX algorithm is used to segment the anomalous pixels from the background. These are detailed as follows.

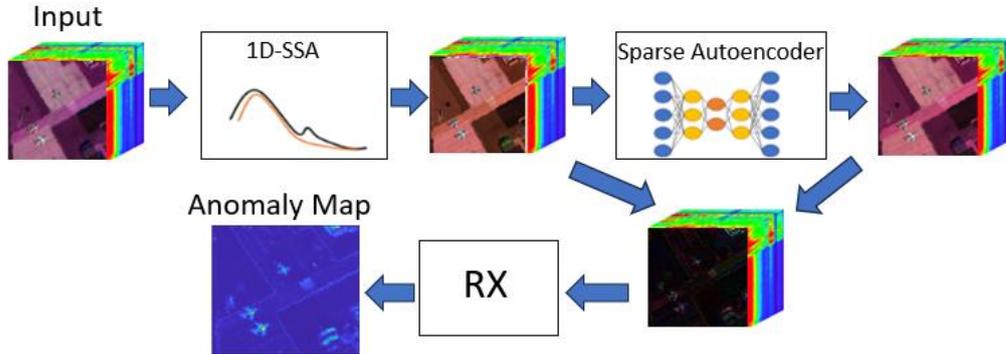


Figure 6. 5 Architecture of the proposed HAD method.

In the traditional classification task, the main objective of SSA is to extract the representative spectral information from the HSI data. For this purpose, each spectral profile will be decomposed into several independent components including trend, oscillations, or noise. Then, several components will be used to reconstruct the HSI data. In HAD task, given a hypercube $T \in \mathcal{R}^{W \times H \times B}$, where W and H denote the size of the spatial domain, and B represents the number of spectral bands. The SSA algorithm will be used to reduce the noise in spectral domain corresponding to each pixel, followed by a differentiation process, as described in the equations (4.4 – 4.7) in section 4.3.1.

Sparse Autoencoder

Sparse Autoencoder is a specialized type of autoencoder that incorporates sparsity constraints to learn more meaning feature representations [202]. Unlike traditional autoencoder, which primarily focuses on accurately reconstructing input, sparse autoencoder adds a constraint that limits the activation of neurons in the hidden layer. This constraint ensures that only a few neurons are activated for each input, resulting in a sparse representation [203]. By balancing reconstruction with the need to maintain sparse hidden layer activations, sparse autoencoder is more effective at identifying important features [204]. The loss function of a sparse autoencoder consists of three components: reconstruction error, sparsity constraint and weight regularization, detailed below.

$$\begin{aligned}
J_{(W,b)} = & \frac{1}{2m} \sum_{i=1}^m \|h(Wx^{(i)} + b) - x^{(i)}\|^2 \\
& + \lambda \sum_{j=1}^n \left(\rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \right) \\
& + \frac{\beta}{2} \sum_{l=1}^L \|W^{(l)}\|^2
\end{aligned} \tag{6.3}$$

reconstruction error term: $\frac{1}{2m} \sum_{i=1}^m \|h(Wx^{(i)} + b) - x^{(i)}\|^2$. This is the first part of the loss function, which measures the error in reconstructing the input data by the autoencoder. $x^{(i)}$ represents the i^{th} input sample. $h(Wx^{(i)} + b)$ is the output obtained through the activation function of the hidden layer in the autoencoder, which corresponds to the reconstruction of the input sample. $\| \cdot \|^2$ denotes the squared Euclidean distance, used to quantify the difference between the original input $x^{(i)}$ and the reconstructed output.

sparsity constraint term: $\lambda \sum_{j=1}^n \left(\rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \right)$. This part imposes a sparsity constraint on the activations of the hidden layer using KL divergence. ρ is the desired average activation value for the hidden units. $\hat{\rho}_j$ represents the actual average activation value of the j^{th} hidden unit over the training data. By controlling $\hat{\rho}_j$ to be close to ρ , the model forces the hidden layer neurons to remain sparsely activated, meaning that most neurons are inactive for a given input. λ is the weight of the sparsity constraint, determining its importance in the overall loss function.

weight regularization term: $\frac{\beta}{2} \sum_{l=1}^L \|W^{(l)}\|^2$. This part of the loss function is the regularization term, which helps to prevent overfitting. β is the regularization coefficient, controlling the contribution of the regularization term to the loss function. $\sum_{l=1}^L \|W^{(l)}\|^2$ is the sum of the squared L2 norms of the weight matrices, where L is the number of layers in the network. The regularization term helps to constrain the size of the weight matrices W , thereby improving the model's generalization ability.

Finally, taking the hypercube obtained from sparse autoencoder above as the input, the RX algorithm is applied to extract the anomaly map. Specifically, RX was employed to calculate the global mean vector and covariance matrix of all the pixels within the image. For each pixel, its anomaly score can be determined by the Mahalanobis distance of its reconstruction error from the global mean and the covariance. These scores can be used to build an anomaly map, as a grayscale image, where small and large values indicate the low

and high anomaly levels of the corresponding pixels, respectively. In this work, specific key parameters' settings of sparse autoencoder model are detailed in Table 6.3.

Table 6.3 Parameter setting details of sparse AE

Hidden size	128
Max epochs	200
β	0.01
ρ	0.01
λ	0.01

6.3.2 Experiments and Results

The proposed method is compared with six state-of-the-art unsupervised benchmarks including the RX [156], LRASR [205], LSMAD [176], VABS [206], robust PCA with RX (RPCA-RX) [207] and LRSNCR [208]. The detection intensity maps of all methods on the six datasets are shown in Figure 6.6. Corresponding AUC value for each method on individual images, as well as the average AUC values across all methods for the six datasets, are presented in Table 6.4 for comparison.

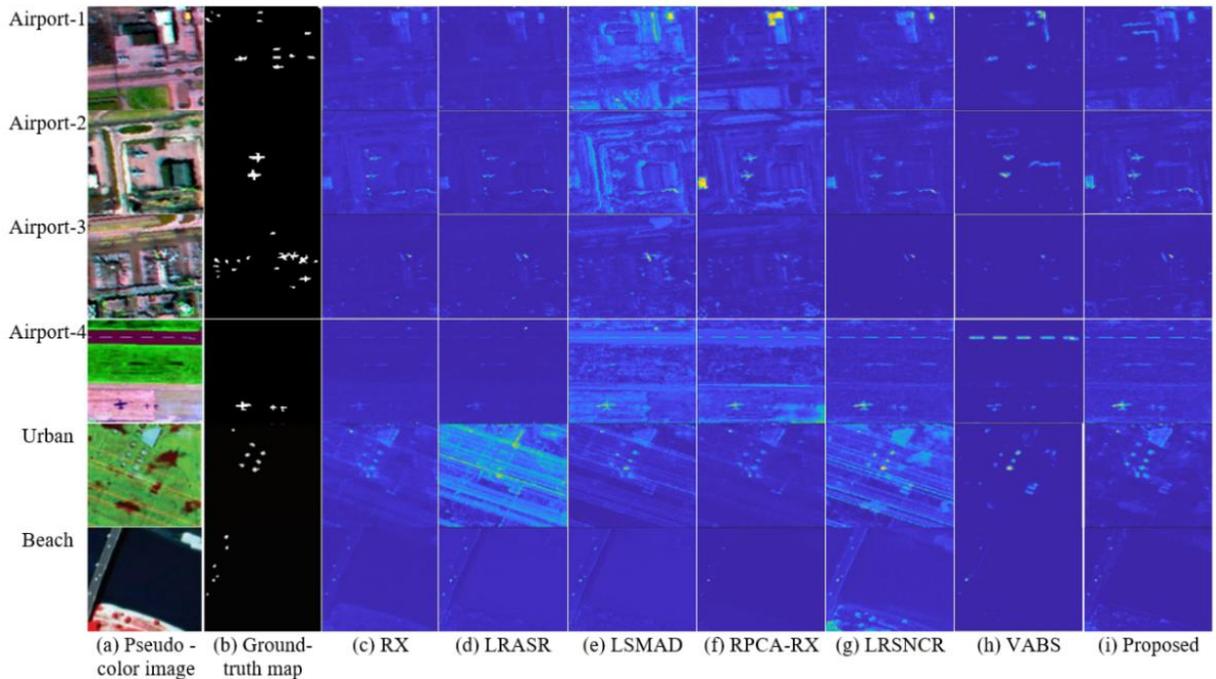


Figure 6.6 Visual comparison of the detected anomaly maps of the benchmark methods on six tested datasets.

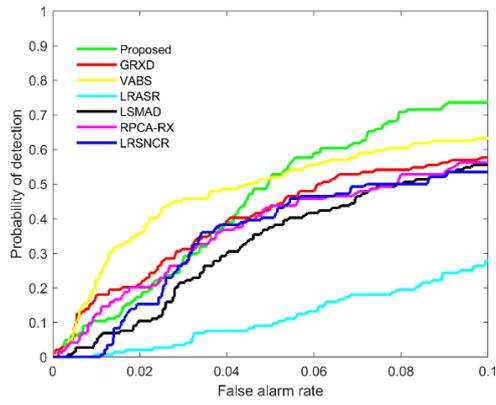
1) Comparisons of detection maps: By analysing the visualisation results from Figure 6.6, it becomes apparent that each method exhibits unique characteristics in detection anomaly targets. On the Airport 1-4 datasets, which include complex background and larger anomalous airplane objects, the detection results from the RX and LRASR methods show insufficient brightness levels for these anomalies. Although other benchmarks can also detect these airplane targets more conspicuously, some background pixels are incorrectly identified as anomalies, leading to false alarms. Particularly in the visualisation maps from LSMAD and LRSNCR, the brightness of buildings in the background is more pronounced than that of airplane targets. On the Urban dataset, due to the background noise interference, the low-rank matrix decomposition-based methods, such as the LRASR, LSMAD, and LRSNCR, have produced a significant number of false alarms in the detection maps. Regarding the comparison results of Beach dataset, which has a simpler background with small anomalous vehicle targets, all benchmarks produce satisfactory visualization results. RPCA-RX and VABS, while the suppression of background is effective, the detected number of anomalous vehicle targets is somewhat incomplete compared to the GT. In the compared detection maps, our proposed method not only achieves high precision in detection all types of anomalous target but also maintains a low rate of false alarms. Moreover, compared to other benchmarks, it consistently exhibits the highest level of brightness and preserves the integrity of target edges. Consequently, the proposed algorithm demonstrates outstanding detection accuracy and robustness across all advanced benchmarks from the comparison of visual detection maps.

Table 6. 4 Comparing AUC values of the different methods on the six datasets.

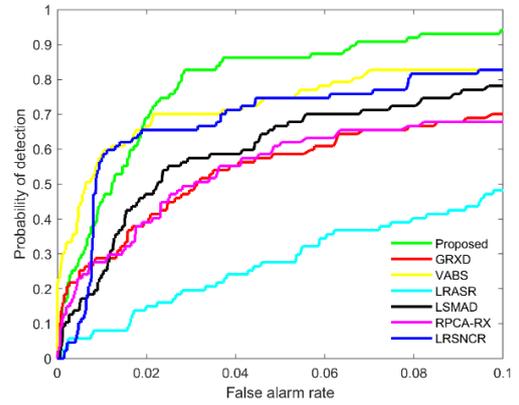
	RX	LRASR	LSMAD	RPCA-RX	LRSNCR	VABS	Proposed
Airport-1	0.8221	0.7775	0.8317	0.8088	0.8677	0.8224	0.9072
Airport-2	0.8403	0.8664	0.9187	0.8426	0.9507	0.9191	0.9725
Airport-3	0.9288	0.8891	0.9383	0.9274	0.9526	0.9216	0.9594
Airport-4	0.9526	0.9846	0.9868	0.9628	0.9501	0.9331	0.9937
Urban	0.9907	0.8257	0.9830	0.9922	0.9432	0.9166	0.9931
Beach	0.9538	0.9393	0.9704	0.9600	0.9327	0.9628	0.9759
Average	0.9147	0.8804	0.9382	0.9156	0.9328	0.9126	0.9670

2) Comparisons of AUC values: As shown in Table 6.4, the AUC values are compared to measure these seven methods. RPCA-RX achieves an AUC value of 0.9922 on the Urban dataset, but only 0.8088 on the Airport-1 dataset. LRASR scores an AUC value of 0.9846 on the Airport-4 dataset but drops to 0.7775 when applied to the airprot-1. Similar discrepancies are observed with other benchmarks that certain algorithms perform exceptionally well on specific datasets, indicating a lack of robustness in handling varied and complex backgrounds. In contrast, our proposed method outperforms other benchmarks, achieving an average precision of 0.9670, which represents significant improvements over the average AUC values of RX, LRASR, LSMAD, RPCA-RX, LRSNCR and VABS by 5.72%, 9.86%, 3.07%, 5.61%, 3.67%, and 5.95%, respectively. Overall, it is verified that our proposed method exhibits superior stability and versatility compared to other benchmarks.

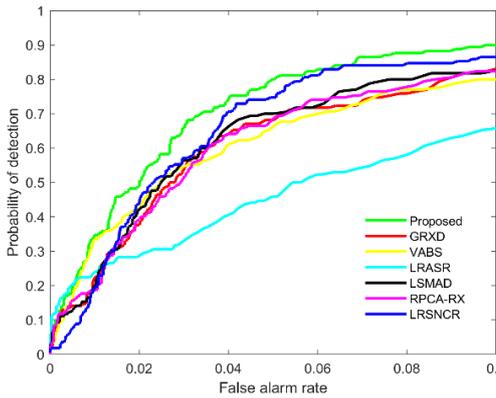
3) Comparisons of ROC curves and separability maps: The ROC curves for all benchmarks across the six different datasets are shown in Figure 6.7. While GRX and LRASR performed well on certain datasets, their overall performance was somewhat inconsistent, displaying significant variability across different datasets. In contrast, LRSNCR consistently ranked as the weakest algorithm. Although it showed some improvement on certain datasets, its overall detection capability remained inferior to the other algorithms. The proposed method proved to be the most robust overall, particularly excelling at lower false alarm rates, which demonstrated stable and efficient detection capabilities of our proposed method. For the comparison of the separability maps in Figure 6.8, our proposed method once again stands out, particularly on the datasets with more complex backgrounds, such as Airports datasets. The statistical ranges for the background and anomalies are distinctly separated, demonstrating our proposed method's strong discriminative capability.



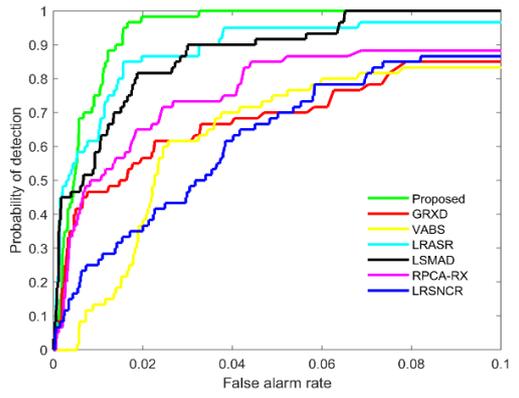
(a) Airport-1



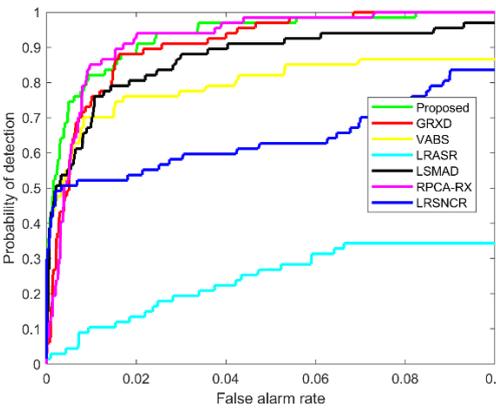
(b) Airport-2



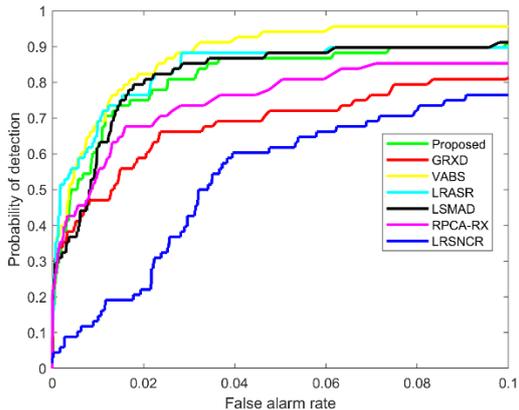
(c) Airport-3



(d) Airport-4

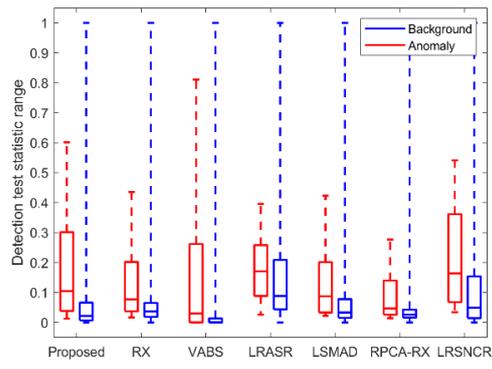


(e) Urban

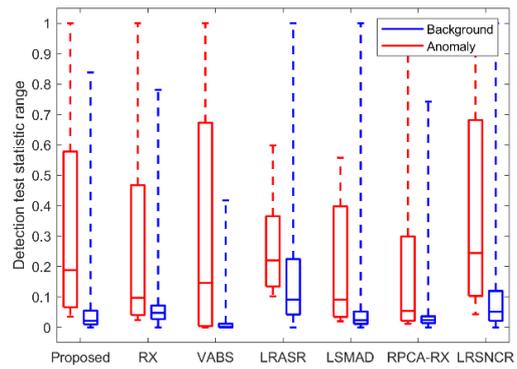


(f) Beach

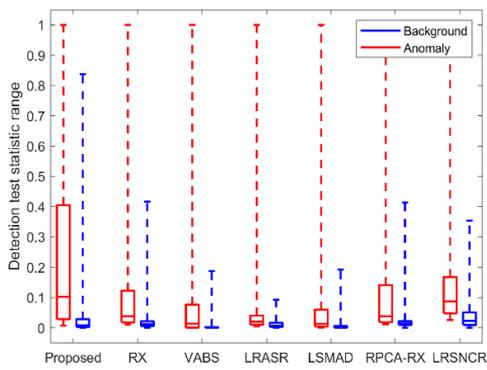
Figure 6. 7 Comparison of ROC curves (false alarm range 0-0.1)



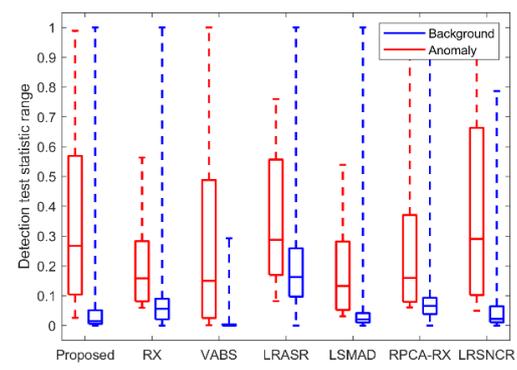
(a) Airport-1



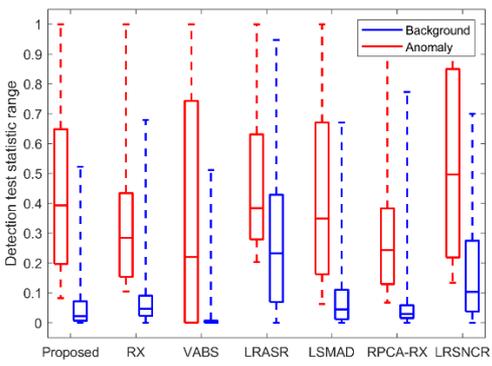
(b) Airport-2



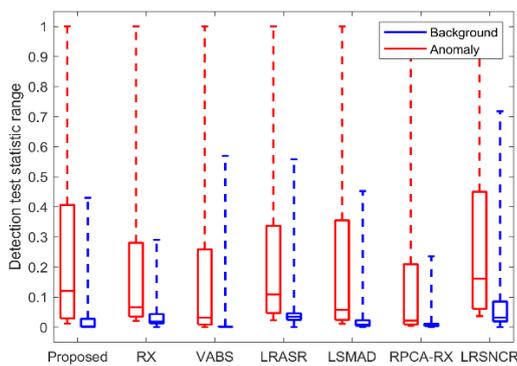
(c) Airport-3



(d) Airport-4



(e) Urban



(f) Bench

Figure 6. 8 Comparison of separability maps

6.4 Summary

In this chapter, efforts have been focused on developing effective algorithms for HAD task. Firstly, this chapter commences with a comprehensive review of existing HAD algorithms, including statistical-based, representation-based and DL-based methods. Subsequently, a novel representative-based method is introduced, leveraging 1-D SSA for denoising in the spectral domain. This is followed by employing a sparse autoencoder for the reconstruction of the background from the denoised hypercube. Finally, the anomaly strength of each pixel is determined by calculating the reconstruction errors, and the RX method is used to identify the anomalous regions. Comparative results indicate that the proposed method surpasses six advanced HAD methods across six datasets including anomalies of various types and sizes such as airplanes, vehicles and buildings, thereby showcasing its robustness and effectiveness.

Building upon the foundations laid in this chapter, future research will continue to advance the field of HAD algorithm development. Specifically, efforts will be directed towards integrating the robust spectral reconstruction with semantic segmentation strategies. The objective is to enhance anomaly detection capabilities in a more refined and unsupervised manner, even in more generic remote sensing images rather than only HSIs.

Chapter 7 Conclusions and Future Work

7.1 Conclusions

The primary objective of this thesis is to develop advanced processing techniques for change and anomaly detection in hyperspectral images. The comprehensive review of existing hyperspectral change detection algorithms provided a solid foundation for understanding the current landscape and identifying gaps in the research. In response to these gaps, four novel and innovation HCD methods for HCD are developed, including one unsupervised HCD method, two deep learning and self-attention based supervised networks as well as a Mamba based deep learning network. These methods were thoroughly evaluated and demonstrated significant improvements over existing approaches. In the last main chapter, the current research trends in hyperspectral anomaly detection are reviewed, followed by a novel background reconstruction based hyperspectral anomaly detection method proposed for HAD. The contributions and limitations of the thesis are summarised in detail as follows.

Chapter 2 provides a comprehensive overview of the HCD algorithms, covering recent advancements in unsupervised, fully supervised, semi-supervised and self-supervised approaches mainly in the last three years. The review concluded that unsupervised methods are straightforward but often limited by lower accuracy and noise sensitivity. In contrast, fully supervised and semi-supervised methods achieve higher accuracy and capture detailed features but require substantial labelled data and are computationally expensive. Self-supervised methods offer a balance by leveraging pseudo-labels but depend heavily on their quality and involve complex training processes.

Chapter 3 introduces the proposed accumulated band-wise binary distancing (ABBD) method for unsupervised, parameter-free HCD. Unlike traditional methods that depend on absolute pixel differences with thresholding, ABBD utilises binary distancing to merely indicate whether a pixel has undergone a change in a specific band, which helps mitigate the adverse effects of measurement inconsistencies. An adaptive, parameter-free tolerance setting within the framework enhances both robustness and usability, while maintaining high accuracy. Comprehensive results demonstrate that the proposed algorithm exhibits robustness across various datasets. However, ABBD is an algebra-based method, lacking

consideration for spatial characteristics. Furthermore, the tolerance value needs to be optimized to achieve the optimal solution.

In Chapter 4, two 2-D self-attention based lightweight deep learning networks are proposed for HCD. Initially, the 2-D self-attention module is introduced to capture local spectral-spatial features with fewer hyperparameters and enhanced detection accuracy compared to traditional self-attention mechanisms. Subsequently, an end-to-end cross-band 2-D attention network (CBANet) is proposed for HCD. To enhance the detection accuracy of edge pixels in changed areas, a singular spectrum analysis-driven lightweight network with 2-D self-attention (SSA-LHCD) is developed, building on the foundational 2-D self-attention module. The SSA-LHCD network incorporates singular spectrum analysis, residual block. These enhancements significantly improve spatial feature extraction, boost pixel-wise detection accuracy, and further reduce the number of hyperparameters.

Chapter 5 introduces the proposed global attention and state space model-based end-to-end model (GASSM) for HCD, which effectively integrates the global attention mechanism and the state space model with the Mamba block to address the limitations of traditional HCD methods based on CNN. The GASSM network leverages the SSM-based Mamba block to capture global spatial-spectral features, overcoming the challenge of limited receptive fields in CNNs and the computational complexity of transformer-based methods. Comprehensive experiments conducted on two publicly available datasets demonstrated that our GASSM model outperforms eight state-of-the-art benchmarks in terms of accuracy and stability, establishing its effectiveness in HCD tasks. This work represents the first exploration of combining the SSM-based Mamba model with global attention for HCD, offering a promising direction for future research in this field.

However, either using a 2-D self-attention-based network or an SSM-based network, both approaches, though capable of achieving high detection accuracy with a small proportion of the training set, still face some major challenges, especially the manual annotated GTs.

Chapter 6 presents a useful exploration of HAD, started by a comprehensive review of key methodologies including statistical-based, representation-based and DL-based methods. Insights into the strengths and weaknesses of each category of these methods are highlighted. In addition, a novel unsupervised HAD algorithm is proposed, which combines the 1-D SSA with a sparse autoencoder. The 1-D SSA is utilised in the spectral domain for denoising and addressing inter-band correlations, while the sparse autoencoder focuses on enhancing the background reconstruction in the spatial domain. This approach

effectively isolates anomalies from the background, significantly improving the detection accuracy. Extensive experiments on six publicly available datasets have fully validated the algorithm's superior performance. However, challenges remain, notably the high training time due to autoencoder-based background reconstruction, along with the need to manually tune the three key parameters.

7.2 Future work

Following by the conclusions of those innovative techniques proposed for hyperspectral change and anomaly detection in this thesis, the potential directions for future research are summarised as follows.

- 1) For the development of hyperspectral change detection algorithms, the plan is to leverage existing and proposed unsupervised method to generate pseudo-ground truth map using the Majority vote approach [209]. The pseudo-ground truth map will then be used to train the developed deep learning networks including CBANet, SSA-LHCD or GASSM. Additionally, there is a plan to design one more lightweight network, aiming to implement a self-supervised learning approach that eliminates the need for manual annotated data. The goal is to enhance the efficiency and accuracy of hyperspectral change detection while reducing the reliance on labour-intensive data annotation processes.
- 2) Regarding hyperspectral anomaly detection, future research plans include to further advance the field of HAD algorithm development. Specifically, efforts will be focused on integrating robust spectral reconstruction with semantic segmentation strategies. The objective is to enhance anomaly detection capabilities in a more refined and unsupervised manner. Even combining the spectral characteristics of prior known targets, with some deep learning networks to achieve self-supervised HAD.
- 3) In addition, the development of change detection algorithms will be extended to more generic and large RS images rather than HSI only, even using homologous and heterogeneous modalities of images. Some representatively homologous bi-temporal images are shown in Appendix A.1. Besides, we have manually annotated a new RGB change detection dataset, which contains 80 pairs of images. The dataset includes 690 airplanes, 60 cars and 755 ship targets, which will be used for algorithm validation. Some of the annotated bi-temporal images are illustrated in Appendix A.2.

- 4) Heterogeneous change detection involves bi-temporal images from different modalities, covering the same geographical area but captured by different sensors, resulting in varying resolutions and physical properties [210]. The public available heterogeneous change detection datasets are shown in Appendix A.3 that illustrates how different modality images appear visually distinct for the same geographical area. Future research will also focus on designing of effective multimodal fusion methods to address the challenges of heterogeneous change detection tasks.

References

- [1] Li, Y.; Ren, J.; Yan, Y.; Liu, Q.; Ma, P.; Petrovski, A.; Sun, H. CBANet: An End-to-End Cross Band 2-D Attention Network for Hyperspectral Change Detection in Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*. **2023**, *61*, 1-11.
- [2] Li, Y.; Ren, J.; Yan, Y.; Ma, P.; Assaad, M.; Gao, Z. ABBD: Accumulated Band-Wise Binary Distancing for Unsupervised Parameter-Free Hyperspectral Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **2024**, *17*, 9880-9893.
- [3] Li, Y.; Ren, J.; Yan, Y.; Sun, G.; Ma, P. SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection. *Remote sensing* **2024**, *16*, 2353.
- [4] Li, Y.; Ren, J.; Yan, Y.; Liu, Q.; Petrovski, A.; McCall, J. Unsupervised Change Detection in Hyperspectral Images using Principal Components Space Data Clustering. In *Journal of Physics: Conference Series*; **2022**, pp. 012021.
- [5] Li, Y.; Yan, Y.; Ren, J.; Liu, Q.; Sun, H. MLM-LSTM: Multi-Layer Memory Learning Framework Based on LSTM for Hyperspectral Change Detection. In *International Conference on Brain Inspired Cognitive Systems*; **2023**, 51–61.
- [6] Li, Y.; Ren, J.; Gao, Z.; Sun, G. Sparse Autoencoder Based Hyperspectral Anomaly Detection with the Singular Spectrum Analysis Based Spectral Denoising. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*; pp. 9210–9213.
- [7] Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geoscience and remote sensing magazine* **2013**, *1*, 6–36.
- [8] TREATY, N.A. Survey of Hyperspectral and Multispectral Imaging Technologies. **2007**.
- [9] Chutia, D.; Bhattacharyya, D.K.; Sarma, K.K.; Kalita, R.; Sudhakar, S. Hyperspectral Remote Sensing Classifications: A Perspective Survey. *Transactions in GIS* **2016**, *20*, 463–490.
- [10] Khan, M.J.; Khan, H.S.; Yousaf, A.; Khurshid, K.; Abbas, A. Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access* **2018**, *6*, 14118–14129.
- [11] Li, F.; Kustas, W.P.; Anderson, M.C.; Prueger, J.H.; Scott, R.L. Effect of Remote Sensing Spatial Resolution on Interpreting Tower-Based Flux Observations. *Remote Sens. Environ.* **2008**, *112*, 337–349.
- [12] Borghys, D.; Shimoni, M.; Perneel, C. Change Detection in Urban Scenes by Fusion of SAR and Hyperspectral Data. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VII*, **2007**, pp. 187–198.
- [13] Navin, M.S.; Agilandeewari, L. Multispectral and Hyperspectral Images Based Land use/Land Cover Change Prediction Analysis: An Extensive Review. *Multimedia Tools Appl* **2020**, *79*, 29751–29774.

- [14] Heiden, U.; Heldens, W.; Roessner, S.; Segl, K.; Esch, T.; Mueller, A. Urban Structure Type Characterization using Hyperspectral Remote Sensing and Height Information. *Landscape Urban Plann.* **2012**, *105*, 361–375.
- [15] Weber, C.; Aguejdad, R.; Briottet, X.; Avala, J.; Fabre, S.; Demuynck, J.; Zenou, E.; Deville, Y.; Karoui, M.S.; Benhalouche, F.Z. Hyperspectral Imagery for Environmental Urban Planning. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, **2018**, pp. 1628–1631.
- [16] Bovolo, F.; Bruzzone, L. A Split-Based Approach to Unsupervised Change Detection in Large-Size Multitemporal Images: Application to Tsunami-Damage Assessment. *IEEE Transactions on Geoscience and Remote Sensing.* **2007**, *45*, 1658–1670.
- [17] Liu, S.; Chi, M.; Zou, Y.; Samat, A.; Benediktsson, J.A.; Plaza, A. Oil Spill Detection Via Multitemporal Optical Remote Sensing Images: A Change Detection Perspective. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 324–328.
- [18] S. Matteoli; M. Diani; G. Corsini. A Tutorial Overview of Anomaly Detection in Hyperspectral Images. *IEEE Aerospace and Electronic Systems Magazine* **2010**, *25*, 5–28.
- [19] Chang, C.; Chen, S.; Zhong, S.; Shi, Y. Exploration of Data Scene Characterization and 3D ROC Evaluation for Hyperspectral Anomaly Detection. *Remote Sensing* **2023**, *16*, 135.
- [20] McCann, C.; Repasky, K.S.; Lawrence, R.; Powell, S. Multi-Temporal Mesoscale Hyperspectral Data of Mixed Agricultural and Grassland Regions for Anomaly Detection. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *131*, 121–133.
- [21] Liu, Y.; Zhou, S.; Wu, H.; Han, W.; Li, C.; Chen, H. Joint Optimization of Autoencoder and Self-Supervised Classifier: Anomaly Detection of Strawberries using Hyperspectral Imaging. *Comput. Electron. Agric.* **2022**, *198*, 107007.
- [22] Briottet, X.; Boucher, Y.; Dimmeler, A.; Malaplate, A.; Cini, A.; Diani, M.; Bekman, H.; Schwing, P.; Skauli, T.; Kasen, I. Military Applications of Hyperspectral Imagery. In *Targets and Backgrounds XII: Characterization and Representation*, **2006**, pp. 82–89.
- [23] Hupel, T.; Stütz, P. Adopting Hyperspectral Anomaly Detection for Near Real-Time Camouflage Detection in Multispectral Imagery. *Remote Sensing* **2022**, *14*, 3755.
- [24] Signoroni, A.; Savardi, M.; Baronio, A.; Benini, S. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. *Journal of imaging* **2019**, *5*, 52.
- [25] Ibarrola-Ulzurrun, E.; Marcello, J.; Gonzalo-Martin, C. Assessment of Component Selection Strategies in Hyperspectral Imagery. *Entropy* **2017**, *19*, 666.
- [26] Hasanlou, M.; Seydi, S.T. Hyperspectral Change Detection: An Experimental Comparative Study. *Int. J. Remote Sens.* **2018**, *39*, 7029–7083.
- [27] Wilson, H.G.; Nelson, T.; Boots, B.; Wulder, M.A. Using a Spatial Statistic Threshold Technique for Defining Change in a Forested Environment. In *Canadian Symposium on Remote Sensing*, October **2003**, pp. 14–17.

- [28] Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges. *IEEE Geoscience and Remote Sensing Magazine* **2019**, *7*, 140–158.
- [29] Bo, Q.U.; Xiangtao, Z.; Xueming, Q.; Xiaoqiang, L.U. Research Progress on Hyperspectral Anomaly Detection. *National Remote Sensing Bulletin* **2024**, *28*, 42–54.
- [30] Racetin, I.; Krtalić, A. Systematic Review of Anomaly Detection in Hyperspectral Remote Sensing Applications. *Applied Sciences* **2021**, *11*, 4878.
- [31] Gao, L.; Wang, D.; Zhuang, L.; Sun, X.; Huang, M.; Plaza, A. BS 3 LNet: A New Blind-Spot Self-Supervised Learning Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2023**, *61*, 1–18.
- [32] Li, Y.; Jiang, T.; Xie, W.; Lei, J.; Du, Q. Sparse Coding-Inspired GAN for Hyperspectral Anomaly Detection in Weakly Supervised Learning. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *60*, 1–11.
- [33] Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote sensing* **2017**, *9*, 1110.
- [34] Luque, A.; Carrasco, A.; Martín, A.; de Las Heras, A. The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix. *Pattern Recognition*. **2019**, *91*, 216–231.
- [35] Wasowski, J.; Bovenga, F. Investigating Landslides and Unstable Slopes with Satellite Multi Temporal Interferometry: Current Issues and Future Perspectives. *Eng. Geol.* **2014**, *174*, 103–138.
- [36] Du, P.; Liu, S.; Gamba, P.; Tan, K.; Xia, J. Fusion of Difference Images for Change Detection Over Urban Areas. *IEEE journal of selected topics in applied earth observations and remote sensing* **2012**, *5*, 1076–1086.
- [37] Bovolo, F.; Bruzzone, L. A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain. *IEEE Transactions on Geoscience and Remote Sensing*. **2006**, *45*, 218–236.
- [38] Bovolo, F.; Marchesi, S.; Bruzzone, L. A Framework for Automatic and Unsupervised Detection of Multiple Changes in Multitemporal Images. *IEEE Transactions on Geoscience and Remote Sensing*. **2011**, *50*, 2196–2212.
- [39] Kuching, S. The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis. *Journal of Computer Science* **2007**, *3*, 419–423.
- [40] Csillik, O.; Belgiu, M.; Asner, G.P.; Kelly, M. Object-Based Time-Constrained Dynamic Time Warping Classification of Crops using Sentinel-2. *Remote sensing* **2019**, *11*, 1257.
- [41] Schwarz, J.; Staenz, K. Adaptive Threshold for Spectral Matching of Hyperspectral Data. *Canadian Journal of Remote Sensing* **2001**, *27*, 216–224.
- [42] Ahmed, M.; Seraj, R.; Islam, S.M.S. The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* **2020**, *9*, 1295.

- [43] Renard, N.; Bourennane, S.; Blanc-Talon, J. Denoising and Dimensionality Reduction using Multilinear Tools for Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters* **2008**, *5*, 138–142.
- [44] Kerekes, J.P.; Baum, J.E. Hyperspectral Imaging System Modeling. *Lincoln Laboratory Journal* **2003**, *14*, 117–130.
- [45] Gross, W.; Borchardt, S.; Middelman, W. Evaluation of Spectral Unmixing using Nonnegative Matrix Factorization on Stationary Hyperspectral Sensor Data of Specifically Prepared Rock and Mineral Mixtures. *Proc. OCM* **2013**, 169–178.
- [46] Celik, T. Unsupervised Change Detection in Satellite Images using Principal Component Analysis and k -Means Clustering. *IEEE Geoscience and Remote Sensing Letters*, **2009**, *6*, 772–776.
- [47] Nielsen, A.A.; Conradsen, K.; Simpson, J.J. Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies. *Remote Sens. Environ.* **1998**, *64*, 1–19.
- [48] Nielsen, A.A. The Regularized Iteratively Reweighted MAD Method for Change Detection in Multi-and Hyperspectral Data. *IEEE Transactions on Image Process.* **2007**, *16*, 463–478.
- [49] Wu, C.; Du, B.; Zhang, L. Slow Feature Analysis for Hyperspectral Change Detection. In 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS); pp. 1–4.
- [50] Wargnier-Dauchelle, V.; Grenier, T.; Durand-Dubief, F.; Cotton, F.; Sdika, M. Constrained Non-Negative Networks for a More Explainable and Interpretable Classification. In *Medical Imaging with Deep Learning*. Jul **2024**, Paris, France.
- [51] Shi, J.; Zhang, Z.; Tan, C.; Liu, X.; Lei, Y. Unsupervised Multiple Change Detection in Remote Sensing Images Via Generative Representation Learning Network. *IEEE Geoscience and Remote Sensing Letters* **2021**, *19*, 1–5.
- [52] He, Z.; Huang, M.; Luo, L.; Yang, X.; Zhu, C. Towards Real-Time Practical Image Compression with Lightweight Attention. *Expert Syst. Appl.* **2024**, 124142.
- [53] Bruzzone, L.; Prieto, D.F. Automatic Analysis of the Difference Image for Unsupervised Change Detection. *IEEE Transactions on Geoscience and Remote Sensing.* **2000**, *38*, 1171–1182.
- [54] Chen, H.; Ye, S.; Zhang, D.; Areshkina, L.; Ablameyko, S. Change Detection Based on Difference Image and Energy Moments in Remote Sensing Image Monitoring. *Pattern Recognition and Image Analysis* **2018**, *28*, 273–281.
- [55] Liu, S.; Du, Q.; Tong, X.; Samat, A.; Pan, H.; Ma, X. Band Selection-Based Dimensionality Reduction for Change Detection in Multi-Temporal Hyperspectral Images. *Remote Sensing* **2017**, *9*, 1008.
- [56] Lv, Z.; Zhang, M.; Sun, W.; Benediktsson, J.A.; Lei, T.; Falco, N. Spatial-Contextual Information Utilization Framework for Land Cover Change Detection with Hyperspectral Remote Sensed Images. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *61*, 1–11.

- [57] Ertürk, A.; Iordache, M.; Plaza, A. Sparse Unmixing-Based Change Detection for Multitemporal Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **2015**, *9*, 708–719.
- [58] Guo, Q.; Zhang, J.; Zhang, Y. Multitemporal Hyperspectral Images Change Detection Based on Joint Unmixing and Information Coguidance Strategy. *IEEE Transactions on Geoscience and Remote Sensing*, **2021**, *59*, 9633–9645.
- [59] Hou, Z.; Li, W.; Tao, R.; Du, Q. Three-Order Tucker Decomposition and Reconstruction Detector for Unsupervised Hyperspectral Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 6194–6205.
- [60] Marinelli, D.; Bovolo, F.; Bruzzone, L. A Novel Change Detection Method for Multitemporal Hyperspectral Images Based on Binary Hyperspectral Change Vectors. *IEEE Transactions on Geoscience and Remote Sensing*, **2019**, *57*, 4913–4928.
- [61] Sun, Y.; Lei, L.; Guan, D.; Kuang, G.; Liu, L. Graph Signal Processing for Heterogeneous Change Detection. *Transactions on Geoscience and Remote Sensing*, **2022**, *60*, 1–23.
- [62] Seydi, S.T.; Hasanlou, M. A New Structure for Binary and Multiple Hyperspectral Change Detection Based on Spectral Unmixing and Convolutional Neural Network. *Measurement* **2021**, *186*, 110137.
- [63] A. Shafique; S. T. Seydi; T. Alipour-Fard; G. Cao; D. Yang. SSViT-HCD: A Spatial–Spectral Convolutional Vision Transformer for Hyperspectral Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *16*, 6487–6504.
- [64] Olaode, A.; Naghdy, G.; Todd, C. Unsupervised Classification of Images: A Review. *International Journal of Image Processing* **2014**, *8*, 325–342.
- [65] Ng, A. Sparse Autoencoder. *CS294A Lecture notes* **2011**, *72*, 1–19.
- [66] Sun, J.; Liu, J.; Xiao, L. Intrinsic Decomposition Model-Guided Two-Stream Coupled Autoencoder for Unsupervised Hyperspectral Image Change Detection. *IEEE Geoscience and Remote Sensing Letters* **2023**.
- [67] Chakraborty, D.; Ghosh, A. Unsupervised Change Detection in Hyperspectral Images using Feature Fusion Deep Convolutional Autoencoders. *arXiv preprint arXiv:2109.04990* **2021**.
- [68] Saha, S.; Kondmann, L.; Zhu, X.X. Deep no Learning Approach for Unsupervised Change Detection in Hyperspectral Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2021**, *3*, 311–316.
- [69] Lei, J.; Li, M.; Xie, W.; Li, Y.; Jia, X. Spectral Mapping with Adversarial Learning for Unsupervised Hyperspectral Change Detection. *Neurocomputing* **2021**, *465*, 71–83.
- [70] J. Qu; P. Yang; W. Dong; X. Zhang; Y. Li. A Semi-Supervised Multiscale Convolutional Sparse Coding-Guided Deep Interpretable Network for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* **2024**, *62*, 1–14.
- [71] Li, Q.; Mu, T.; Tuniyazi, A.; Yang, Q.; Dai, H. Progressive Pseudo-Label Framework for Unsupervised Hyperspectral Change Detection. *International Journal of Applied Earth Observation and Geoinformation* **2024**, *127*, 103663.

- [72] Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, **2016**, *54*, 6232–6251.
- [73] Suthaharan, S.; Suthaharan, S. Support Vector Machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning* **2016**, 207–235.
- [74] Rigatti, S.J. Random Forest. *Journal of Insurance Medicine* **2017**, *47*, 31–39.
- [75] Yin, X.; Yang, C.; Pei, W.; Hao, H. Shallow Classification or Deep Learning: An Experimental Study. In *2014 22nd International Conference on Pattern Recognition*; pp. 1904–1909.
- [76] LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444.
- [77] Sze, V.; Chen, Y.; Yang, T.; Emer, J.S. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proc IEEE* **2017**, *105*, 2295–2329.
- [78] Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J. Recent Advances in Convolutional Neural Networks. *Pattern Recognition* **2018**, *77*, 354–377.
- [79] Medsker, L.; Jain, L.C. *Recurrent Neural Networks: Design and Applications*.; CRC press, 1999.
- [80] Malik, K.; Robertson, C.; Roberts, S.A.; Remmel, T.K.; Long, J.A. Computer Vision Models for Comparing Spatial Patterns: Understanding Spatial Scale. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 1–35.
- [81] Salehinejad, H.; Sankar, S.; Barfett, J.; Colak, E.; Valaee, S. Recent Advances in Recurrent Neural Networks. *arXiv preprint arXiv:1801.01078* **2017**.
- [82] Graves, A.; Graves, A. Long Short-Term Memory. *Supervised sequence labelling with recurrent neural networks* **2012**, 37–45.
- [83] Zhan, T.; Song, B.; Xu, Y.; Wan, M.; Wang, X.; Yang, G.; Wu, Z. SSCNN-S: A Spectral-Spatial Convolution Neural Network with Siamese Architecture for Change Detection. *Remote Sensing* **2021**, *13*, 895.
- [84] Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 3–13.
- [85] Lin, Y.; Li, S.; Fang, L.; Ghamisi, P. Multispectral Change Detection with Bilinear Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters* **2019**, *17*, 1757–1761.
- [86] Song, A.; Choi, J.; Han, Y.; Kim, Y. Change Detection in Hyperspectral Images using Recurrent 3D Fully Convolutional Networks. *Remote Sensing* **2018**, *10*, 1827.
- [87] Shi, C.; Zhang, Z.; Zhang, W.; Zhang, C.; Xu, Q. Learning Multiscale Temporal–spatial–spectral Features Via a Multipath Convolutional LSTM Neural Network for Change Detection with Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, **2022**, *60*, 1–16.
- [88] Luo, F.; Zhou, T.; Liu, J.; Guo, T.; Gong, X.; Ren, J. Multiscale Diff-Changed Feature Fusion Network for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **2023**, *61*, 1–13.
- [89] Qu, J.; Hou, S.; Dong, W.; Li, Y.; Xie, W. A Multilevel Encoder–decoder Attention Network for Change Detection in Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, **2021**, *60*, 1–13.

- [90] Yang, Y.; Qu, J.; Xiao, S.; Dong, W.; Li, Y.; Du, Q. A Deep Multiscale Pyramid Network Enhanced with Spatial–spectral Residual Attention for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **2022**, *60*, 1–13.
- [91] Song, R.; Ni, W.; Cheng, W.; Wang, X. CSANet: Cross-Temporal Interaction Symmetric Attention Network for Hyperspectral Image Change Detection. *IEEE Geoscience and Remote Sensing Letters* **2022**.
- [92] Y. Wang; D. Hong; J. Sha; L. Gao; L. Liu; Y. Zhang; X. Rong. Spectral–Spatial–Temporal Transformers for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–14.
- [93] Yu, H.; Yang, H.; Gao, L.; Hu, J.; Plaza, A.; Zhang, B. Hyperspectral Image Change Detection Based on Gated Spectral–Spatial–Temporal Attention Network with Spectral Similarity Filtering. *IEEE Transactions on Geoscience and Remote Sensing*, **2024**.
- [94] Ji, Y.; Sun, W.; Wang, Y.; Lv, Z.; Yang, G.; Zhan, Y.; Li, C. Domain Adaptive and Interactive Differential Attention Network for Remote Sensing Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, early access, **2024**.
- [95] Lee, H.; Kwon, H. Self-Supervised Contrastive Learning for Cross-Domain Hyperspectral Image Representation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; pp. 3239–3243.
- [96] Cao, Z.; Li, X.; Feng, Y.; Chen, S.; Xia, C.; Zhao, L. ContrastNet: Unsupervised Feature Learning by Autoencoder and Prototypical Contrastive Learning for Hyperspectral Imagery Classification. *Neurocomputing* **2021**, *460*, 71–83.
- [97] Li, J.; Wei, Y.; Wang, C.; Hu, Q.; Liu, Y.; Xu, L. 3-D CNN-Based Multichannel Contrastive Learning for Alzheimer’s Disease Automatic Diagnosis. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–11.
- [98] Cai, Y.; Zhang, Z.; Liu, Y.; Ghamisi, P.; Li, K.; Liu, X.; Cai, Z. Large-Scale Hyperspectral Image Clustering using Contrastive Learning. *arXiv preprint arXiv:2111.07945* **2021**.
- [99] Qin, J.; Zhao, H. Spatial-Spectral-Associative Contrastive Learning for Satellite Hyperspectral Image Classification with Transformers. *Remote Sensing* **2023**, *15*, 1612.
- [100] Hu, X.; Li, T.; Zhou, T.; Peng, Y. Deep Spatial-Spectral Subspace Clustering for Hyperspectral Images Based on Contrastive Learning. *Remote Sensing* **2021**, *13*, 4418.
- [101] Yu, W.; Wan, S.; Li, G.; Yang, J.; Gong, C. Hyperspectral Image Classification with Contrastive Graph Convolutional Network. *IEEE Transactions on Geoscience and Remote Sensing*. **2023**, *61*, 1–15.
- [102] Liang, X.; Du, S.; Duan, Y.; Chen, Y.; Zhu, K.; Yin, Y.; Jian, L. Kernel-Based Algorithms for Image Classification: A Review. **2023**.
- [103] Jian, P.; Ou, Y.; Chen, K. Uncertainty Aware Graph Self-Supervised Learning for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2024**.

- [104] Ding, Y.; Zhang, Z.; Zhao, X.; Cai, Y.; Li, S.; Deng, B.; Cai, W. Self-Supervised Locality Preserving Low-Pass Graph Convolutional Embedding for Large-Scale Hyperspectral Image Clustering. *IEEE Transactions on Geoscience and Remote Sensing*. **2022**, *60*, 1–16.
- [105] J. Wang; F. Gao; J. Dong; S. Zhang; Q. Du. Change Detection from Synthetic Aperture Radar Images Via Graph-Based Knowledge Supplement Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2022**, *15*, 1823–1836.
- [106] Wang, Z.; Jiang, F.; Liu, T.; Xie, F.; Li, P. Attention-Based Spatial and Spectral Network with PCA-Guided Self-Supervised Feature Extraction for Change Detection in Hyperspectral Images. *Remote Sensing* **2021**, *13*, 4927.
- [107] Li, Q.; Gong, H.; Dai, H.; Li, C.; He, Z.; Wang, W.; Feng, Y.; Han, F.; Tuniyazi, A.; Li, H. Unsupervised Hyperspectral Image Change Detection Via Deep Learning Self-Generated Credible Labels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 9012–9024.
- [108] Zhao, H.; Feng, K.; Wu, Y.; Gong, M. An Efficient Feature Extraction Network for Unsupervised Hyperspectral Change Detection. *Remote Sensing* **2022**, *14*, 4646.
- [109] M. Hu; C. Wu; B. Du; L. Zhang. Binary Change Guided Hyperspectral Multiclass Change Detection. *IEEE Transactions on Image Processing* **2023**, *32*, 791–806.
- [110] Song, A.; Choi, J.; Han, Y.; Kim, Y. Change Detection in Hyperspectral Images using Recurrent 3D Fully Convolutional Networks. *Remote Sensing* **2018**, *10*, 1827.
- [111] Fan, Y.; Dai, D.; Kukleva, A.; Schiele, B. Cossi: Co-Learning of Representation and Classifier for Imbalanced Semi-Supervised Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; pp. 14574–14584.
- [112] Rani, V.; Nabi, S.T.; Kumar, M.; Mittal, A.; Kumar, K. Self-Supervised Learning: A Succinct Review. *Archives of Computational Methods in Engineering* **2023**, *30*, 2761–2775.
- [113] Kim, G.; Choi, J.G.; Ku, M.; Lim, S. Developing a Semi-Supervised Learning and Ordinal Classification Framework for Quality Level Prediction in Manufacturing. *Comput. Ind. Eng.* **2023**, *181*, 109286.
- [114] Liu, L.; Hong, D.; Ni, L.; Gao, L. Multilayer Cascade Screening Strategy for Semi-Supervised Change Detection in Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2022**, *15*, 1926–1940.
- [115] Yuan, Y.; Lv, H.; Lu, X. Semi-Supervised Change Detection Method for Multi-Temporal Hyperspectral Images. *Neurocomputing* **2015**, *148*, 363–375.
- [116] Jiang, F.; Gong, M.; Zhan, T.; Fan, X. A Semi-supervised GAN-Based Multiple Change Detection Framework in Multi-Spectral Images. *IEEE Geoscience and Remote Sensing Letters* **2019**, *17*, 1223–1227.
- [117] Luo, F.; Zhou, T.; Liu, J.; Guo, T.; Gong, X.; Gao, X. DCENet: Diff-Feature Contrast Enhancement Network for Semi-Supervised Hyperspectral Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2024**.

- [118] Pérez-Cruz, F. Kullback-Leibler Divergence Estimation of Continuous Distributions. In 2008 IEEE International Symposium on Information Theory; pp. 1666–1670.
- [119] Huang, Y.; Zhang, L.; Qi, W.; Song, R.; Huang, C.; Cen, Y. Semi-Supervised Hyper Match-Driven Cross Temporal and Spatial Interaction Transformer for Hyperspectral Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2024**.
- [120] Ou, X.; Liu, L.; Tu, B.; Zhang, G.; Xu, Z. A CNN Framework with Slow-Fast Band Selection and Feature Fusion Grouping for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2022**, *60*, 1–16.
- [121] Li, X.; Yuan, Z.; Wang, Q. Unsupervised Deep Noise Modelling for Hyperspectral Image Change Detection. *Remote Sensing* **2019**, *11*, 258.
- [122] Veltkamp, R.C. Shape Matching: Similarity Measures and Algorithms. In Proceedings International Conference on Shape Modelling and Applications; pp. 188–197.
- [123] Guo, Q.; Zhang, J.; Zhang, Y. Multitemporal Hyperspectral Images Change Detection Based on Joint Unmixing and Information Cognisance Strategy. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *59*, 9633–9645.
- [124] Eismann, M.T.; Meola, J.; Hardie, R.C. Hyperspectral Change Detection in the Presence of Diurnal and Seasonal Variations. *IEEE Transactions on Geoscience and Remote Sensing*. **2007**, *46*, 237–249.
- [125] K. P. Sinaga; M. -S. Yang. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727.
- [126] Hou, Z.; Li, W.; Du, Q. A Patch Tensor-Based Change Detection Method for Hyperspectral Images. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS; pp. 4328–4331.
- [127] Hou, Z.; Li, W.; Li, L.; Tao, R.; Du, Q. Hyperspectral Change Detection Based on Multiple Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *60*, 1–12.
- [128] He, N.; Paoletti, M.E.; Haut, J.M.; Fang, L.; Li, S.; Plaza, A.; Plaza, J. Feature Extraction with Multiscale Covariance Maps for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*. **2018**, *57*, 755–769.
- [129] Hamida, A.B.; Benoit, A.; Lambert, P.; Amar, C.B. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*. **2018**, *56*, 4420–4434.
- [130] Qu, J.; Hou, S.; Dong, W.; Li, Y.; Xie, W. A Multilevel Encoder–decoder Attention Network for Change Detection in Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *60*, 1–13.
- [131] Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters* **2019**, *17*, 277–281.
- [132] Humphreys, G.W.; Sui, J. Attentional Control and the Self: The Self-Attention Network (SAN). *Cognitive neuroscience* **2016**, *7*, 5–17.

- [133] Agarap, A.F. Deep Learning using Rectified Linear Units (Relu). arXiv preprint arXiv:1803.08375 **2018**.
- [134] Tong, Z.; Tanaka, G. Hybrid Pooling for Enhancement of Generalization Ability in Deep Convolutional Neural Networks. *Neurocomputing* **2019**, *333*, 76–85.
- [135] Guo, Q.; Zhang, J.; Zhu, S.; Zhong, C.; Zhang, Y. Deep Multiscale Siamese Network with Parallel Convolutional Structure and Self-Attention for Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *60*, 1–12.
- [136] Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-Attention with Relative Position Representations. arXiv preprint arXiv:1803.02155 **2018**.
- [137] Xia, J.; Cui, Y.; Li, W.; Wang, L.; Wang, C. Lightweight Self-Attention Residual Network for Hyperspectral Classification. *IEEE Geoscience and Remote Sensing Letters* **2022**, *19*, 1–5.
- [138] Wu, C.; Du, B.; Zhang, L. Hyperspectral Anomalous Change Detection Based on Joint Sparse Representation. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *146*, 137–150.
- [139] Kingma, D.P. & Ba J.(2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 **2015**.
- [140] Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. arXiv preprint arXiv:1607.08022 **2016**.
- [141] Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*; pp. 807–814.
- [142] Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International Conference on Machine Learning*; pp. 448–456.
- [143] Zhang, H.; Dauphin, Y.N.; Ma, T. Fixup Initialization: Residual Learning without Normalization. arXiv preprint arXiv:1901.09321 **2019**.
- [144] Zhang, Z.; Sabuncu, M. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *Advances in neural information processing systems* **2018**, *31*.
- [145] Wang, X.; Wang, S.; Ding, Y.; Li, Y.; Wu, W.; Rong, Y.; Kong, W.; Huang, J.; Li, S.; Yang, H. State Space Model for New-Generation Network Alternative to Transformers: A Survey. arXiv preprint arXiv:2404.09516 **2024**.
- [146] Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:2312.00752 **2023**.
- [147] Ma, X.; Zhang, X.; Pun, M. RS 3 Mamba: Visual State Space Model for Remote Sensing Image Semantic Segmentation. *IEEE Geoscience and Remote Sensing Letters* **2024**.
- [148] Chen, K.; Chen, B.; Liu, C.; Li, W.; Zou, Z.; Shi, Z. Rsmamba: Remote Sensing Image Classification with State Space Model. *IEEE Geoscience and Remote Sensing Letters* **2024**.
- [149] Fu, H.; Sun, G.; Li, Y.; Ren, J.; Zhang, A.; Jing, C.; Ghamisi, P. HDMba: Hyperspectral Remote Sensing Imagery Dehazing with State Space Model. arXiv preprint arXiv:2406.05700 **2024**.
- [150] Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. arXiv preprint arXiv:2112.05561 **2021**.

- [151] Elfving, S.; Uchibe, E.; Doya, K. Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning. *Neural Networks* **2018**, *107*, 3–11.
- [152] Yu, H.; Yang, H.; Gao, L.; Hu, J.; Plaza, A.; Zhang, B. Hyperspectral Image Change Detection Based on Gated Spectral–Spatial–Temporal Attention Network with Spectral Similarity Filtering. *IEEE Transactions on Geoscience and Remote Sensing*. **2024**.
- [153] Su, H.; Wu, Z.; Zhang, H.; Du, Q. Hyperspectral Anomaly Detection: A Survey. *IEEE Geoscience and Remote Sensing Magazine* **2021**, *10*, 64–90.
- [154] Zweig, M.H.; Campbell, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* **1993**, *39*, 561–577.
- [155] Lobo, J.M.; Jiménez-Valverde, A.; Real, R. AUC: A Misleading Measure of the Performance of Predictive Distribution Models. *Global Ecol. Biogeogr.* **2008**, *17*, 145–151.
- [156] Reed, I.S.; Yu, X. Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution. *IEEE transactions on acoustics, speech, and signal processing* **1990**, *38*, 1760–1770.
- [157] Kwon, H.; Nasrabadi, N.M. Kernel RX-Algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*. **2005**, *43*, 388–397.
- [158] Xiong, Y.; Zuo, R.; Wang, K.; Wang, J. Identification of Geochemical Anomalies Via Local RX Anomaly Detector. *Journal of Geochem. Explore.* **2018**, *189*, 64–71.
- [159] Liu, W.; Feng, X.; Wang, S.; Hu, B.; Gan, Y.; Zhang, X.; Lei, T. Random Selection-Based Adaptive Saliency-Weighted RXD Anomaly Detection for Hyperspectral Imagery. *International Journal of Remote Sensing*. **2018**, *39*, 2139–2158.
- [160] Guo, Q.; Zhang, B.; Ran, Q.; Gao, L.; Li, J.; Plaza, A. Weighted-RXD and Linear Filter-Based RXD: Improving Background Statistics Estimation for Anomaly Detection in Hyperspectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 2351–2366.
- [161] Ranney, K.I.; Soumekh, M. Hyperspectral Anomaly Detection within the Signal Subspace. *IEEE Geoscience and Remote Sensing Letters* **2006**, *3*, 312–316.
- [162] Taitano, Y.P.; Geier, B.A.; Bauer, K.W. A Locally Adaptable Iterative RX Detector. *EURASIP Journal on Advances in Signal Processing* **2010**, *2010*, 1–10.
- [163] Schweizer, S.M.; Moura, J.M. Hyperspectral Imagery: Clutter Adaptation in Anomaly Detection. *IEEE Trans. Inf. Theory* **2000**, *46*, 1855–1871.
- [164] Xiang, P.; Zhou, H.; Li, H.; Song, S.; Tan, W.; Song, J.; Gu, L. Hyperspectral Anomaly Detection by Local Joint Subspace Process and Support Vector Machine. *International Journal of Remote Sensing*. **2020**, *41*, 3798–3819.
- [165] Chang, S.; Du, B.; Zhang, L. A Subspace Selection-Based Discriminative Forest Method for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2020**, *58*, 4033–4046.

- [166] Song, S.; Zhou, H.; Zhou, J.; Qian, K.; Cheng, K.; Zhang, Z. Hyperspectral Anomaly Detection Based on Anomalous Component Extraction Framework. *Infrared Phys. Technol.* **2019**, *96*, 340–350.
- [167] Yuan, Z.; Sun, H.; Ji, K.; Li, Z.; Zou, H. Local Sparsity Divergence for Hyperspectral Anomaly Detection. *IEEE Geoscience and Remote Sensing Letters* **2014**, *11*, 1697–1701.
- [168] Li, J.; Zhang, H.; Zhang, L.; Ma, L. Hyperspectral Anomaly Detection by the use of Background Joint Sparse Representation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2015**, *8*, 2523–2533.
- [169] Zhao, G.; Li, F.; Zhang, X.; Laakso, K.; Chan, J.C. Archetypal Analysis and Structured Sparse Representation for Hyperspectral Anomaly Detection. *Remote Sensing* **2021**, *13*, 4102.
- [170] Zhu, L.; Wen, G. Hyperspectral Anomaly Detection Via Background Estimation and Adaptive Weighted Sparse Representation. *Remote Sensing* **2018**, *10*, 272.
- [171] W. Li; Q. Du. Collaborative Representation for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 1463–1474.
- [172] Wang, R.; Hu, H.; He, F.; Nie, F.; Cai, S.; Ming, Z. Self-Weighted Collaborative Representation for Hyperspectral Anomaly Detection. *Signal Process* **2020**, *177*, 107718.
- [173] Lu, Y.; Zheng, X.; Xin, H.; Tang, H.; Wang, R.; Nie, F. Ensemble and Random Collaborative Representation-Based Anomaly Detector for Hyperspectral Imagery. *Signal Process* **2023**, *204*, 108835.
- [174] Zhao, C.; Li, C.; Feng, S. A Spectral–spatial Method Based on Fractional Fourier Transform and Collaborative Representation for Hyperspectral Anomaly Detection. *IEEE Geoscience and Remote Sensing Letters* **2020**, *18*, 1259–1263.
- [175] Sun, W.; Liu, C.; Li, J.; Lai, Y.M.; Li, W. Low-Rank and Sparse Matrix Decomposition-Based Anomaly Detection for Hyperspectral Imagery. *Journal of Applied Remote Sensing* **2014**, *8*, 083641.
- [176] Zhang, Y.; Du, B.; Zhang, L.; Wang, S. A Low-Rank and Sparse Matrix Decomposition-Based Mahalanobis Distance Method for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *54*, 1376–1389.
- [177] Qu, Y.; Wang, W.; Guo, R.; Ayhan, B.; Kwan, C.; Vance, S.; Qi, H. Hyperspectral Anomaly Detection through Spectral Unmixing and Dictionary-Based Low-Rank Decomposition. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56*, 4391–4405.
- [178] Li, L.; Li, W.; Du, Q.; Tao, R. Low-Rank and Sparse Decomposition with Mixture of Gaussian for Hyperspectral Anomaly Detection. *IEEE Transactions on Cybernetics* **2020**, *51*, 4363–4372.
- [179] Xiang, P.; Ali, S.; Zhang, J.; Jung, S.K.; Zhou, H. Pixel-Associated Autoencoder for Hyperspectral Anomaly Detection. *International Journal of Applied Earth Observation and Geoinformation* **2024**, *129*, 103816.
- [180] Jiang, K.; Xie, W.; Li, Y.; Lei, J.; He, G.; Du, Q. Semi-supervised Spectral Learning with Generative Adversarial Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing* **2020**, *58*, 5224–5236.

- [181] Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep Learning for Anomaly Detection: A Review. *ACM computing surveys (CSUR)* **2021**, *54*, 1–38.
- [182] Zhao, C.; Li, X.; Zhu, H. Hyperspectral Anomaly Detection Based on Stacked Denoising Autoencoders. *Journal of Applied Remote Sensing* **2017**, *11*, 042605.
- [183] Zhao, C.; Zhang, L. Spectral-Spatial Stacked Autoencoders Based on Low-Rank and Sparse Matrix Decomposition for Hyperspectral Anomaly Detection. *Infrared Phys. Technol.* **2018**, *92*, 166–176.
- [184] Lu, X.; Zhang, W.; Huang, J. Exploiting Embedding Manifold of Autoencoders for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing.* **2019**, *58*, 1527–1537.
- [185] Wang, S.; Wang, X.; Zhang, L.; Zhong, Y. Auto-AD: Autonomous Hyperspectral Anomaly Detection Network Based on Fully Convolutional Autoencoder. *IEEE Transactions on Geoscience and Remote Sensing.* **2021**, *60*, 1–14.
- [186] Li, K.; Ling, Q.; Wang, Y.; Cai, Y.; Qin, Y.; Lin, Z.; An, W. Spectral Difference Guided Graph Attention Autoencoder for Hyperspectral Anomaly Detection. *IEEE Transactions on Instrumentation and Measurement* **2022**, *72*, 1–17.
- [187] Xia, X.; Pan, X.; Li, N.; He, X.; Ma, L.; Zhang, X.; Ding, N. GAN-Based Anomaly Detection: A Review. *Neurocomputing* **2022**, *493*, 497–535.
- [188] Jiang, T.; Li, Y.; Xie, W.; Du, Q. Discriminative Reconstruction Constrained Generative Adversarial Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing.* **2020**, *58*, 4666–4679.
- [189] Arisoy, S.; Nasrabadi, N.M.; Kayabol, K. GAN-Based Hyperspectral Anomaly Detection. In *2020 28th European Signal Processing Conference (EUSIPCO)*; pp. 1891–1895.
- [190] Wang, J.; Guo, S.; Hua, Z.; Huang, R.; Hu, J.; Gong, M. CL-CaGAN: Capsule Differential Adversarial Continual Learning for Cross-Domain Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing.* **2024**.
- [191] Wang, Z.; Ma, D.; Yue, G.; Li, B.; Cong, R.; Wu, Z. Self-Supervised Hyperspectral Anomaly Detection Based on Finite Spatial-Wise Attention. *IEEE Transactions on Geoscience and Remote Sensing.* **2023**, *62*, 1-18.
- [192] Wang, Q.; Zhang, F.; Li, X. Optimal Clustering Framework for Hyperspectral Band Selection. *IEEE Transactions on Geoscience and Remote Sensing.* **2018**, *56*, 5910–5922.
- [193] Liu, Y.; Xie, W.; Jiang, K.; Zhang, J.; Li, Y.; Fang, L. Hyperspectral Anomaly Detection with Self-Supervised Anomaly Prior. *arXiv preprint arXiv:2404.13342* **2024**.
- [194] Fan, G.; Ma, Y.; Mei, X.; Fan, F.; Huang, J.; Ma, J. Hyperspectral Anomaly Detection with Robust Graph Autoencoders. *IEEE Transactions on Geoscience and Remote Sensing.* **2021**, *60*, 1–14.
- [195] Kang, X.; Zhang, X.; Li, S.; Li, K.; Li, J.; Benediktsson, J.A. Hyperspectral Anomaly Detection with Attribute and Edge-Preserving Filters. *IEEE Transactions on Geoscience and Remote Sensing.* **2017**, *55*, 5600–5611.

- [196] Du, X.; Zare, A. Scene Label Ground Truth Map for MUUFL Gulfport Data Set. Dept.Elect.Comput.Eng., Univ.Florida, Gainesville, FL, USA, Tech.Rep **2017**, 20170417.
- [197] Li, K.; An, W.; Wang, Y.; Zhang, T.; Qin, Y.; Gao, T. Explicit Background Endmember Learning for Hyperspectral Anomaly Detection. *IEEE Transactions on Instrumentation and Measurement* **2024**.
- [198] McGill, R.; Tukey, J.W.; Larsen, W.A. Variations of Box Plots. *The american statistician* **1978**, 32, 12–16.
- [199] Hu, X.; Xie, C.; Fan, Z.; Duan, Q.; Zhang, D.; Jiang, L.; Wei, X.; Hong, D.; Li, G.; Zeng, X. Hyperspectral Anomaly Detection using Deep Learning: A Review. *Remote Sensing* **2022**, 14, 1973.
- [200] Raza Shah, N.; Maud, A.R.M.; Bhatti, F.A.; Ali, M.K.; Khurshid, K.; Maqsood, M.; Amin, M. Hyperspectral Anomaly Detection: A Performance Comparison of Existing Techniques. *International Journal of Digital Earth* **2022**, 15, 2078–2125.
- [201] Xu, Y.; Zhang, L.; Du, B.; Zhang, L. Hyperspectral Anomaly Detection Based on Machine Learning: An Overview. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2022**, 15, 3351–3364.
- [202] Ng, A. Sparse Autoencoder. *CS294A Lecture notes* **2011**, 72, 1–19.
- [203] Shu, M.; Fyshe, A. Sparse Autoencoders for Word Decoding from Magnetoencephalography. In *Proceedings of the Third NIPS Workshop on Machine Learning and Interpretation in NeuroImaging (MLINI)*.
- [204] Kamimura, R.; Takeuchi, H. Sparse Semi-Autoencoders to Solve the Vanishing Information Problem in Multi-Layered Neural Networks. *Appl. Intell.* **2019**, 49, 2522–2545.
- [205] Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; Wei, Z. Anomaly Detection in Hyperspectral Images Based on Low-Rank and Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*. **2015**, 54, 1990–2000.
- [206] Xiang, P.; Song, J.; Qin, H.; Tan, W.; Li, H.; Zhou, H. Visual Attention and Background Subtraction with Adaptive Weight for Hyperspectral Anomaly Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, 14, 2270–2283.
- [207] Xu, Y.; Wu, Z.; Chanussot, J.; Wei, Z. Joint Reconstruction and Anomaly Detection from Compressive Hyperspectral Images using Mahalanobis Distance-Regularized Tensor RPCA. *IEEE Trans. Geosci. Remote Sens.* **2018**, 56, 2919–2930.
- [208] Yao, W.; Li, L.; Ni, H.; Li, W.; Tao, R. Hyperspectral Anomaly Detection Based on Improved RPCA with Non-Convex Regularization. *Remote Sensing* **2022**, 14, 1343.
- [209] Davani, A.M.; Díaz, M.; Prabhakaran, V. Dealing with Disagreements: Looking Beyond the Majority Vote in Subjective Annotations. *Transactions of the Association for Computational Linguistics* **2022**, 10, 92–110.
- [210] Lv, Z.; Huang, H.; Li, X.; Zhao, M.; Benediktsson, J.A.; Sun, W.; Falco, N. Land Cover Change Detection with Heterogeneous Remote Sensing Images: Review, Progress, and Perspective. *Proc IEEE* **2022**, 110, 1976–1991.

- [211] Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. **2021**, *60*, 1–16.
- [212] Chen, H.; Shi, Z. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing* **2020**, *12*, 1662.
- [213] Ji, F.; Ming, D.; Zeng, B.; Yu, J.; Qing, Y.; Du, T.; Zhang, X. Aircraft Detection in High Spatial Resolution Remote Sensing Images Combining Multi-Angle Features Driven and Majority Voting CNN. *Remote Sensing* **2021**, *13*, 2207.
- [214] Lv, Z.; Huang, H.; Li, X.; Zhao, M.; Benediktsson, J.A.; Sun, W.; Falco, N. Land Cover Change Detection with Heterogeneous Remote Sensing Images: Review, Progress, and Perspective. *Proc IEEE* **2022**, *110*, 1976–1991.

Appendix

A.1: Results analysis of GASSM on Yancheng dataset

The quantitative results of GASSM on the Yancheng dataset and the extracted change maps from different methods are compared in Table A.1 and Figure A.1, respectively. As shown in Table A.1, all deep learning-based methods significantly outperform unsupervised ones in HCD. Models such as CSANet, CBANet and SSTFormer have consistently exceeded in terms of the OA, KP, Pre, Re and F1, indicating their strong capability using the deep learning for modelling complex data features. Our proposed GASSM model has achieved comparable performance to state-of-the-art approaches e.g. CSANet and SSTFormer, with notable advantages in KP and high stability, further validating its superior HCD performance and practical potential. However, as shown in Figure A.1, all deep learning methods exhibit a high number of false alarms when handling large areas of change pixels especially at the boundaries. As discussed in Chapter 4, this could be attributed to inaccurate annotations of the edge pixels in the Yancheng dataset. Future work will focus on utilizing more advanced algorithms to further mitigate the false alarms.

Table A. 1 Quantitative analysis results of different methods on Yancheng dataset

	OA(%)	KP	Pre	Re	F1
CVA	87.55	0.7025	0.8421	0.8327	0.7529
PCAKM	88.28	0.7180	0.8471	0.8557	0.7519
ABBD	88.79	0.7427	0.8765	0.8053	0.8459
ML-EDAN	97.15±0.0012	0.9316±0.0034	0.9665±0.0034	0.9517±0.0021	0.9298±0.0027
HyGSTAN	96.74±0.0009	0.9454±0.0024	0.9698±0.0029	0.9561±0.0018	0.9327±0.0012
CSANet	97.15±0.0009	0.9584±0.0015	0.9774±0.0020	0.9677±0.0003	0.9335±0.0023
CBANet	97.13±0.0006	0.9605±0.0070	0.9768±0.0041	0.9679±0.0019	0.9332±0.0014
SSTFormer	97.34±0.0016	0.9634±0.0005	0.9673±0.0032	0.9563±0.0044	0.9317±0.0017
Proposed	97.33±0.0003	0.9636±0.0008	0.9675±0.0006	0.9592±0.0012	0.9324±0.0024

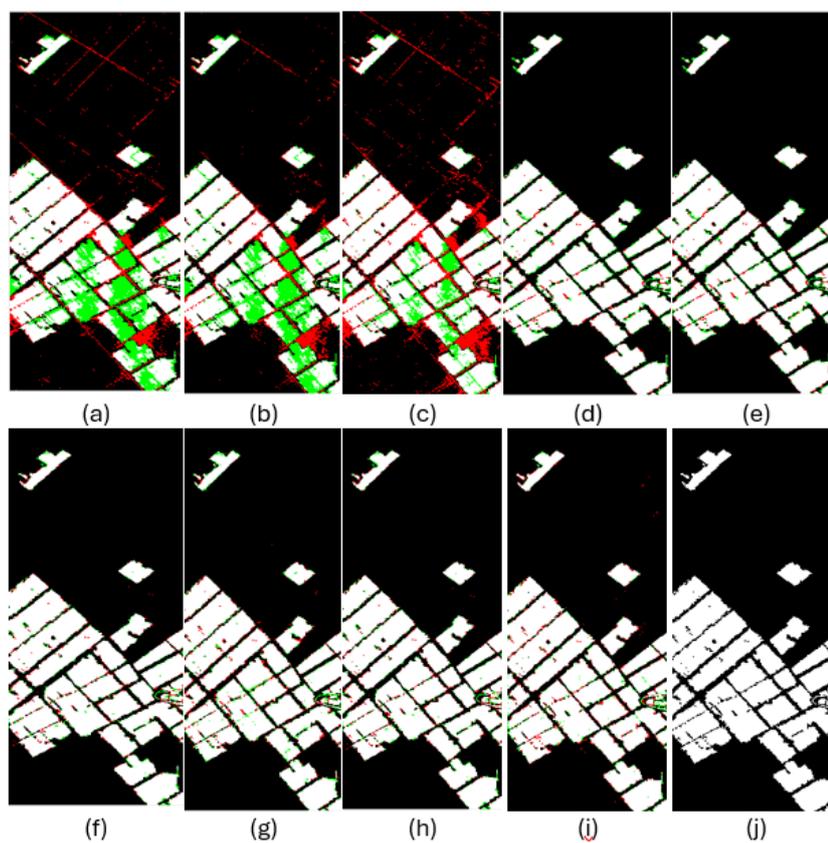


Figure A. 1 Extracted change maps on the Yancheng dataset from different methods of CVA (a), PCAKM (b), ABBD (c), ML-EDAN (d), HyGSTAN (e), CSANet (f), CBANet (g), SSTFormer (h) and proposed (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, whilst white and black denotes true positive and true negative, respectively.

A.2: Publicly Available Remote Sensing datasets in RGB

1) ***SYSU-CD*** [211]: This dataset comprises a set of 20,000 pairs of aerial images with a spatial resolution of 0.5m, captured over Hong Kong from 2007 to 2014. Each pair of the bi-temporal images measures 256×256 pixels. The primary types of changes documented in the dataset, as illustrated in Figure A.2, include the construction of new urban buildings, suburban expansion, initial groundwork for construction projects, alterations in vegetation, road widening, and maritime construction activities.



2) **LEVIR-CD** [212] dataset contains 637 image pairs with a size of 1024×1024 pixels and a spatial resolution of 0.5 m via the Google Earth application programming interface. The dataset mainly focuses on changes in buildings, including newly constructed buildings and destructed buildings, with 31,333 changed buildings were annotated. Figure A. 3 presents the examples from the LEVIR-CD dataset.



Figure A. 3 Examples from the LEVIR-CD dataset

A.3: Self-Annotated Remote Sensing Change Detection Dataset in RGB

The self-annotated RS CD dataset in RGB is derived from a portion of the DIOR dataset [213], each with a spatial size of 800×800 pixels, the spatial resolution is 0.5 m. The self-annotated RS CD dataset contains 80 pair of images, which divided into 42 pairs for the training set and 38 pairs for the testing set. Example samples from this annotated dataset are shown in Figure A.4. Compared to the existing datasets, this annotated dataset features a greater number of changed targets in small sizes. This dataset will be utilised to validate the proposed deep learning algorithms in the future research.

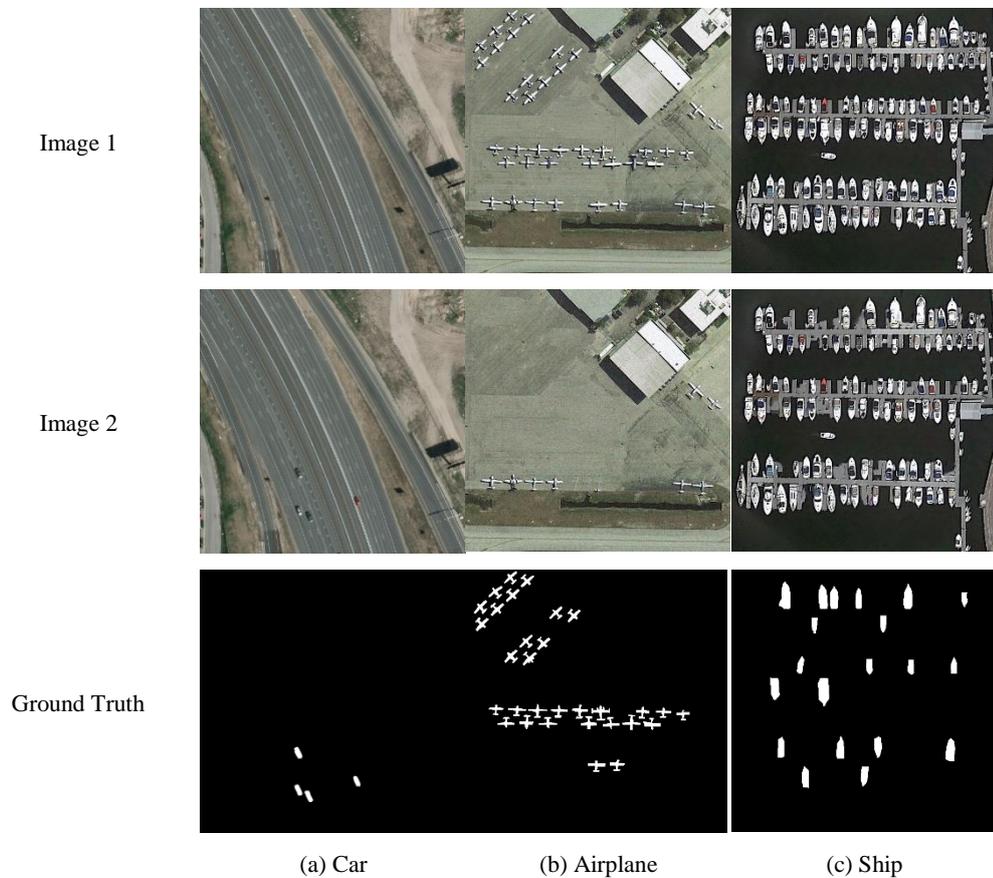


Figure A. 4 Annotated remote sensing change detection dataset

A.4: Heterogeneous Remote Sensing Change Detection Datasets

The multimodal data includes images such as SAR, Optical, LiDAR, LiDAR-depth and Open Street map. The details of each bi-temporal image pair, including the multimodal composition, resolution and acquisition time. The publicly available heterogeneous datasets and further information can be found in the review paper on heterogeneous CD [214]. Four demonstrations of visual differences between bi-temporal image pairs are shown in Figure A.5.

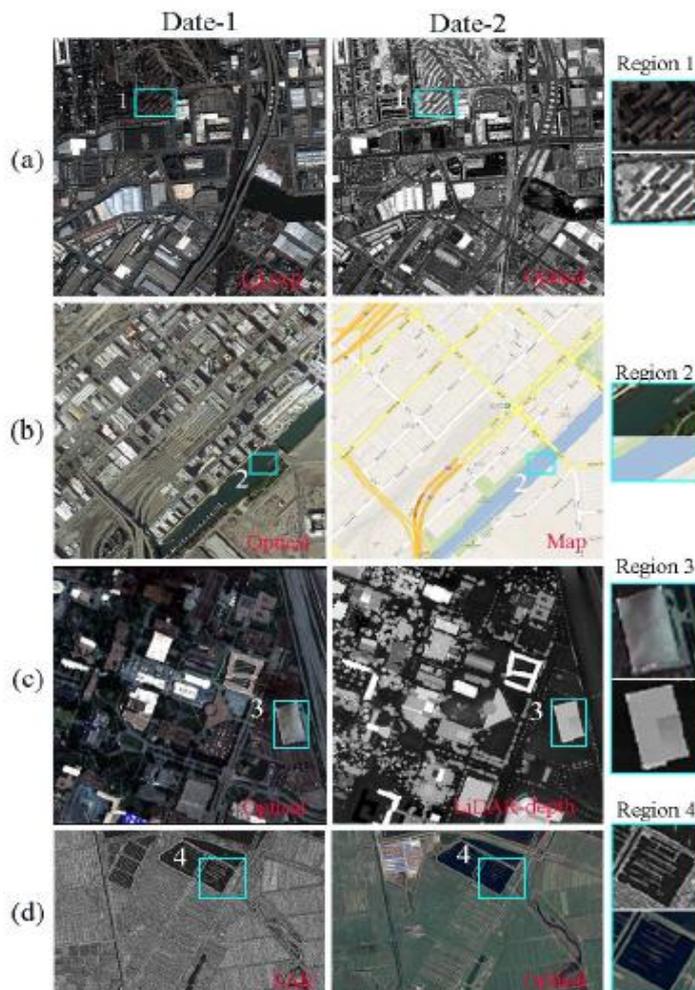


Figure A. 5 Illustrated visual difference between heterogeneous bitemporal image pairs: (a) optical-LiDAR images, (b) optical image-map, (c) optical-LiDAR-depth images, and (d) SAR-optical images