LI, Y., YAN, Y. and REN, C., LIU, Q. and SUN, H. 2023. MLM-LSTM: multi-layer memory learning framework based on LSTM for hyperspectral change detection. In: *Ren, J., Hussain, A., Liao, I.Y. et al. (eds.) Advances in brain inspired cognitive systems: proceedings of the 13th International conference on Brain-inspired cognitive systems 2023 (BICS 2023), 5-6 August 2023, Kuala Lumpur, Malaysia*. Lecture notes in computer sciences, 14374. Cham: Springer [online], pages 51-61. Available from: https://doi.org/10.1007/978-981-97-1417-9_5

MLM-LSTM: multi-layer memory learning framework based on LSTM for hyperspectral change detection.

LI, Y., YAN, Y. and REN, C., LIU, Q. and SUN, H.

2023

MLM-LSTM: Multi-layer Memory Learning Framework Based on LSTM for Hyperspectral Change Detection

Yinhe Li[1], Yijun Yan[1], Jinchang Ren[1](✉), Qiaoyuan Liu[2], and Haijiang Sun[2]

[1] National Subsea Centre, Robert Gordon University, Aberdeen, UK
{y.li24,y.yan2,j.ren}@rgu.ac.uk

[2] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China
liuqy@ciomp.ac.cn

Abstract. Hyperspectral change detection plays a critical role in remote sensing by leveraging spectral and spatial information for accurate land cover variation identification. Long short-term memory (LSTM) has demonstrated its effectiveness in capturing dependencies and handling long sequences in hyperspectral data. Building on these strengths, a multilayer memory learning model based on LSTM for hyperspectral change detection is proposed, called MLM-LSTM for hyperspectral change detection is proposed. It incorporates shallow memory learning and deep memory learning. The deep memory learning module performs deep feature extraction of long-term and short-term memory separately. Then fully connected layers will be used to fuse the features followed by binary classification for change detection. Notably, our model has higher detection accuracy compared to other state-of-the-art deep learning-based models. Through comprehensive experiments on publicly available datasets, we have successfully validated the effectiveness and efficiency of the proposed MLM-LSTM approach.

Keywords: Hyperspectral image · Change detection · Long short-term memory

1 Introduction

Change detection (CD) is a fundamental approach for monitoring land-cover changes, involving the analysis of disparities between bi-temporal remote sensing images of the same geographic area [1]. Hyperspectral images (HSIs) provide a comprehensive representation of objects by combining pixel-wise 1-D spectral data with spatial information in the form of a standard 2-D image. The integration of spectral and spatial data in HSIs enables a more detailed and accurate assessment of changes occurring within the observed area [2]. Therefore, hyperspectral change detection (HCD) has emerged as a prominent research area in recent years.

Despite the abundance of spatial and spectral information, HSIs often plagued by highly redundancy information and various noise that primarily due to sensor limitations and atmospheric effects during the data acquisition step. As a result, processing HSIs data poses significant challenges [3]. Over the past two decades, numerous methods for hyperspectral change detection (HCD) have been proposed to address these challenges, encompassing both unsupervised and supervised approaches [4].

Recently, the application of deep learning-based methods to HCD tasks has emerged as a new trend in extracting more effective and representative spectral, spatial, and spatial-spectral features. In [5], a cross-temporal interaction symmetric attention (CSA) network was proposed that employed a Siamese module to hierarchically extract change information in a symmetric manner. A cross-temporal self-attention module was incorporated to joint spatial-spectral-temporal features and enhance the feature representation ability. In [6], a novel end-to-end 2-D CNN was introduced that utilizes a mixed affinity matrix and subpixel representation to effectively extract cross-band gradients. In addition, some deep learning-based methods based on recurrent neural networks (RNN) and Long Short-Term memory (LSTM) have been proposed, for example, in [7], Re3FCN was proposed that leverages 3-D CNN layers to extract spatial-spectral features, while LSTM is utilized for extracting comprehensive features and screening significantly changed features. In [8], a novel multilevel encoder-decoder attention network is introduced to extract hierarchical spatial-spectral features more effectively. In this approach, the extracted features are transferred to a LSTM module for analysing temporal dependencies. Although deep learning models have yielded impressive results, they often rely on a substantial amount of training data, which is difficult to acquire. Consequently, the computational cost becomes exceptionally high, thus needs further efforts to address this issue.

Although LSTM has shown good performance in hyperspectral change detection (HCD), it has limitations:

1. Difficulty handling long sequence data: HSIs have hundreds of consecutive bands, which can lead to issues like gradient explosion when processing with LSTM models, hindering the capture of long-term dependencies effectively.
2. Information omission: Traditional LSTM methods only utilize the final output result, neglecting the full utilization of memory from the last hidden state.

To address these limitations, we propose an end-to-end multi-layer memory learning framework based on LSTM. It incorporates unsupervised PCA for dimensionality reduction, retaining essential features while reducing dimensionality and mitigating the gradient explosion risk. Multi-layer LSTM modules are employed for initial feature extraction, with long-term and short-term memories integrated separately for fusion in subsequent layers. This facilitates the identification of significant change features.

The remainder of this paper is organized as follows. Section 2 describes the details of the proposed MLM-LSTM. Section 3 presents the experimental results and assessments. Finally, some remarkable conclusions are summarized in Sect. 4.

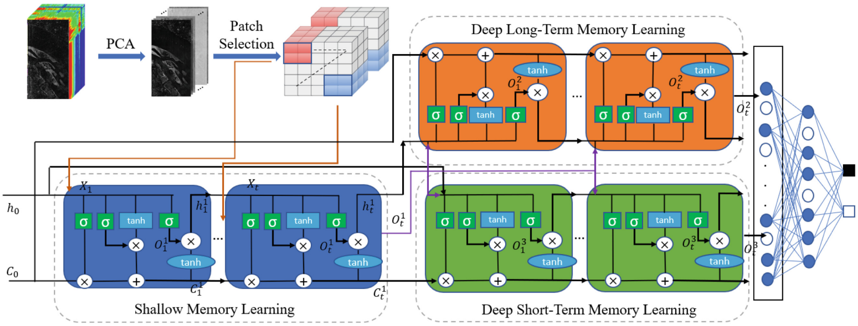


Fig. 1. The architecture of the proposed MLM-LSTM model

2 The Proposed Approach

Figure 1 shows the architecture of the proposed MLM-LSTM method, which is composed of two main steps, i.e., 1) spectral feature extraction and dimension reduction, 2) shallow memory learning and hierarchical deep memory learning.

2.1 Spectral Feature Extraction and Dimension Reduction

Given two HSIs $T^1, T^2 \in \Re^{W*H*B}$ acquired on the same geographical area at times t_1 and t_2, where W, H and B denote the numbers of rows, columns and spectral bands, respectively. To analyze the behaviors of spectral differences between the two images, let us compute the HS difference image T^D by subtracting bitemporal images from each other pixel by pixel, i.e.,

$$T^D = \left|T^2 - T^1\right| \tag{1}$$

Then principal component analysis (PCA) is introduced to reduce the high-dimensional original input while retaining more spatial and spectral features. The spectral feature of T^D can be represented as $T^D(PCA) \in \Re^{H*W*q_{PCA}}$, where q_{PCA} is the number of bands after PCA dimensionality reduction. T^D will be feed into LSTM modules in the next stage.

2.2 Shallow and Deep Memory Learning

2.2.1 LSTM Unit

An LSTM unit comprises a forget gate f_t, an input gate i_t, and an output gate O_t. The LSTM cell memories values at arbitrary time intervals and these three gates control the flow of information at each time step t which can be calculated as follows:

$$f_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \tag{2}$$

$$i_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \tag{3}$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{ci}C_t + b_o) \tag{4}$$

where W, b, σ represent the coefficient matrix, bias vector and sigmoid function, respectively. These three gates are crucial parts of the LSTM unit, which is used to update the current memory state of this unit, obtain the short-term memory C_t and long-term memory h_t, that can be represented as:

$$\tilde{C}^t = \tanh(W_{\tilde{C}}h_{t-1} + W_{\tilde{C}}h_{t-1}X_t + b_{\tilde{C}}) \tag{5}$$

$$C_t = f_tC_{t-1} + i_t * \tilde{C}^t \tag{6}$$

$$h_t = O_t * \tanh(C_t) \tag{7}$$

2.2.2 MLM-LSTM: Multi-layer Memory Feature Extraction

T^D Will be divided into a group of overlapped 3-D neighboring patches denoted as $Z_{(\alpha,\beta)} \in \Re^{S*S*q_{PCA}}$, where S is the patch size of Z, (α, β) denote the coordinates of the patch centre in the spatial domain where $\alpha\epsilon[1, W]$, $\beta\epsilon[1, H]$ (we set S = 3 in this study). The total number of 3-D patches from T^D will be $(W - S + 1) \times (H - S + 1)$. If we split the patched across the spectral channels, then Z can be considered as an q_{PCA}-length sequence $\{(Z^1_{(\alpha,\beta)}, Z^2_{(\alpha,\beta)}, \ldots, Z^{q_{PCA}}_{(\alpha,\beta)})|\ Z^q_{(\alpha,\beta)} \in \Re^{S*S*1}, 1 \leq q \leq q_{PCA}\}$ The image patches in the sequence are fed into the memory feature extraction module one by one to extract the spectral feature via a recurrent operator. The proposed memory feature extraction module is composed of three LSTM layers with t hidden size and m LSTM layers. In order to fully extract all the features of the input, we set $t = 256$ and m = 2 in this study. The first LSTM is used to extract long-term memory h^1, short-term memory C^1_t and output O^1_t from the input patches, h_o and C_o are initialized to zero. After initial feature extraction by using the shallow memory learning, the extracted outputs O^1_t, h^1_t, C^1_t have the same size as Z.

Next, we used two LSTM modules for deep extraction of the long-term memory and short-term memory, respectively. The outputs O^2_t, O^3_t can be obtained by Eqs. (8–9).

$$O^2_t = \sigma(W_{xo}O^1_t + W_{ho}h^1_t + W_{ci}C_0 + b_o) \tag{8}$$

$$O_t^3 = \sigma(W_{xo}O_t^1 + W_{ho}h_0 + W_{ci}C_t^1 + b_o) \tag{9}$$

O_t^2, O_t^3 are used to extract the deep hierarchical feature from the long-term and short-term spatial-spectral memories, respectively. Then these two outputs are further concatenated together and fed into to a linear layer for feature integration.

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.

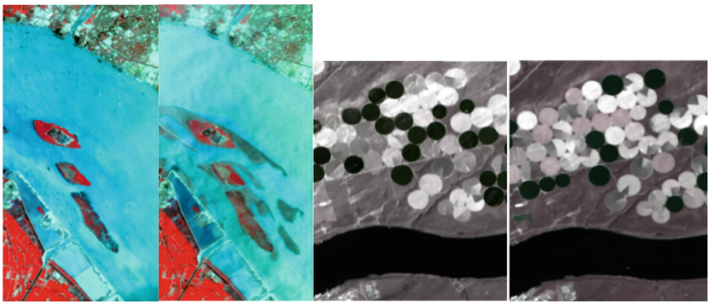
$$Loss_{(pred,label)} = -\frac{1}{u}\sum_{i=1}^{n}(l * \log(p) + (1-l) * \log(1-p)) \tag{10}$$

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) with the initial learning rate of 0.0001.

3 Experiments and Results

3.1 Experimental Settings

Change detection task can be treated as a binary classification problem, therefore, three commonly used evaluation metrics for classification, including the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (KP), were adopted in our experiments for quantitative performance assessment [9]. Two datasets (i.e., River and Hermiston) shown in Fig. 2 [10] are adopted in this study for performance evaluation.



(a) (b) (c) (d)

Fig. 2. Pseudo-colored images of the three datasets, including the River dataset captured on May 3, 2013 (a) and December 31, 2013 (b) and the Hermiston dataset captured on May 1, 2004 (c) and May 8, 2007 (d), respectively.

3.2 Results and Analysis

In this session, we evaluate the effectiveness of the proposed method by comparing it with a few start-of-the-art unsupervised methods, which include the change vector analysis (CVA) [11], principal component analysis (PCA-KM) [12] and absolute distance (AD) [13] as well as several deep-learning based methods such as 2-D-CNN [14], 3-D-CNN [15], HybridSN [16], and Traditional Long-short-term-memory (LSTM) [17]. The proposed MLM-LSTM and all other DL-based methods are trained based on the PyTorch on an NVIDIA RTX A2000, with the batch size set to 128 and the number of training epoch as 200. 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. The quantitative assessment results of OA and KP on River and Hermiston datasets are shown in Table 1.

Table 1. Quantitative assessment of different methods on River and Hermiston datasets

	River		Hermiston	
	OA	KP	OA	KP
AD	0.9431	0.7137	0.9342	0.7904
CVA	0.9253	0.6528	0.9287	0.7705
PCA-KM	0.9517	0.7476	0.9224	0.7472
LSTM	0.9569 ± 0.0011	0.7216 ± 0.0070	0.9537 ± 0.0024	0.8580 ± 0.0009
HybridSN	0.9671 ± 0.0019	0.7826 ± 0.0087	0.9579 ± 0.0007	0.8789 ± 0.0005
3-D-CNN	0.9700 ± 0.0008	0.8045 ± 0.0053	0.9639 ± 0.0002	0.8966 ± 0.0047
2-D-CNN	0.9682 ± 0.0007	0.7946 ± 0.0033	0.9585 ± 0.0013	0.8794 ± 0.0050
MLM-LSTM	**0.9723 ± 0.0005**	**0.8248 ± 0.0023**	**0.9708 ± 0.0014**	**0.9194 ± 0.0013**

3.2.1 Results on River Dataset

Table 1 presents the quantitative assessment results and Fig. 3 provides visual comparison maps for the River dataset. The ground truth map in Fig. 3(i) reveals noticeable changes such as sediment accumulation and alterations in building cover along the riverbank. Figures 3(a–c) demonstrate that the unsupervised algorithms yield numerous false alarms, particularly in the lower left and upper left corners of the maps where non-changing pixels are incorrectly classified as changed. As a result, the KP values for all unsupervised methods remain below 80%. In contrast, DL-based algorithms effectively classify most of the false alarms. Among the benchmarked DL methods, traditional LSTM performs the worst with an average KP of 0.7261 and OA of 95.69%, indicating the highest number of missing pixels. The 2-D CNN and 3-D CNN produce similar detection results, with an OA of around 97% and KP of approximately 0.80, slightly outperforming the unsupervised methods. Our proposed MLM-LSTM method achieves the highest OA and KP among all compared methods, with an OA value of 0.9723 and KP value of 0.8248, surpassing the second-place method by 3%. Additionally, MLM-LSTM exhibits the smallest variance, further confirming its effectiveness.

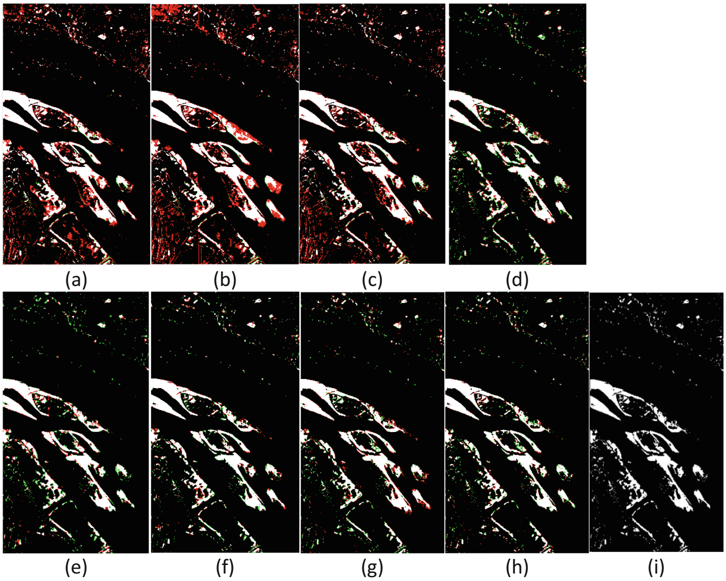


Fig. 3. Extracted change maps on the River Dataset from different methods of AD (a), CVA (b), PCA-KM (c), LSTM (d), HybridSN (e), 3-DCNN (f), 2-DCNN (g), MLM-LSTM (h) in comparison to the Ground-truth map (i), where the false alarms and missing pixels are labelled in red and green. (Color figure online)

3.2.2 Results on Hermiston Dataset

The Hermiston dataset's quantitative assessment results and extracted change maps are presented in Table 1 and Fig. 4, respectively. The changing areas mainly consist of crop regions characterized by simple round shapes. Among the unsupervised methods, there is a significant number of undetected pixels, resulting in OA values below 94% and KP values lower than 0.8. The CNN-based methods, including 2-D CNN, 3-D CNN, and HybridSN, exhibit similar detection results, with OA around 97% and KP values below 0.9. However, both 2-D CNN and 3-D CNN show significant variance, leading to unstable detection performance. Traditional LSTM consistently demonstrates the lowest detection results among all DL-based benchmarks. In contrast, our MLM-LSTM method shows a significant improvement in performance, achieving the highest detection accuracy among all methods. It achieves an OA value of 0.9708, which is 0.23% higher than the second-place method, and a KP value of 0.9194, which is 6.14% higher than the

KP of the traditional LSTM method. This demonstrates the effectiveness and robustness of our proposed MLM-LSTM in handling variations in different sizes of changes.

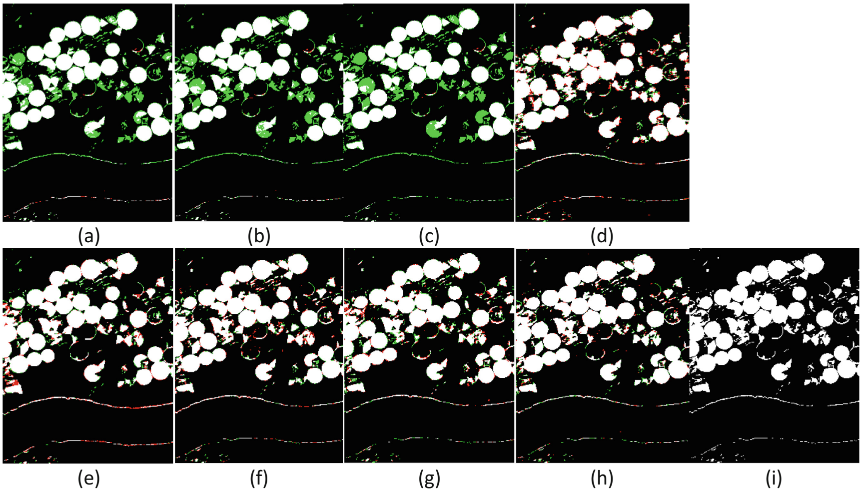


Fig. 4. Extracted change maps on the Hermiston Dataset from different methods of AD (a), CVA (b), PCA-KM (c), LSTM (d), HybridSN (e), 3-DCNN (f), 2-DCNN (g), MLM-LSTM (h) in comparison to the Ground-truth map (i), where the false alarms and missing pixels are labelled in red and green. (Color figure online)

3.3 Ablation Experiments

In our experiments, we investigated the impact of different factors on the performance of the MLM-LSTM model. Firstly, we tested patch sizes of 3 × 3, 5 × 5, 7 × 7, and 9 × 9, finding that increasing the patch size had minimal impact on KP and OA (Fig. 5(a)). A patch size of 5 × 5 showed improved performance on the River dataset but lacked robustness on the Hermiston dataset. Considering computational efficiency, a patch size of 3 × 3 struck a suitable balance. We also explored the optimal number of layers and found that 3 layers maximized KP on the River dataset, while 2 layers performed best on the Hermiston dataset (Fig. 5(b)). In terms of hidden size (Fig. 5(c)), increasing the size in the LSTM module resulted in higher KP values, and a hidden size of 256 was selected for optimal accuracy. Finally, we varied the training ratios from 10% to 50% and observed that increasing the number of training pixels improved detection accuracy (Fig. 5(d)). Our MLM-LSTM consistently outperformed other methods, achieving exceptional performance with a highest KP of 0.8656 on the River dataset. These results highlight the robustness and effectiveness of MLM-LSTM in hyperspectral change detection tasks.

3.4 Hyperparameter Analysis

To assess the efficiency of our proposed MLM-LSTM, we compared the hyperparameters and floating-point operations (FLOPs) of different methods, as shown in Table 2. It is

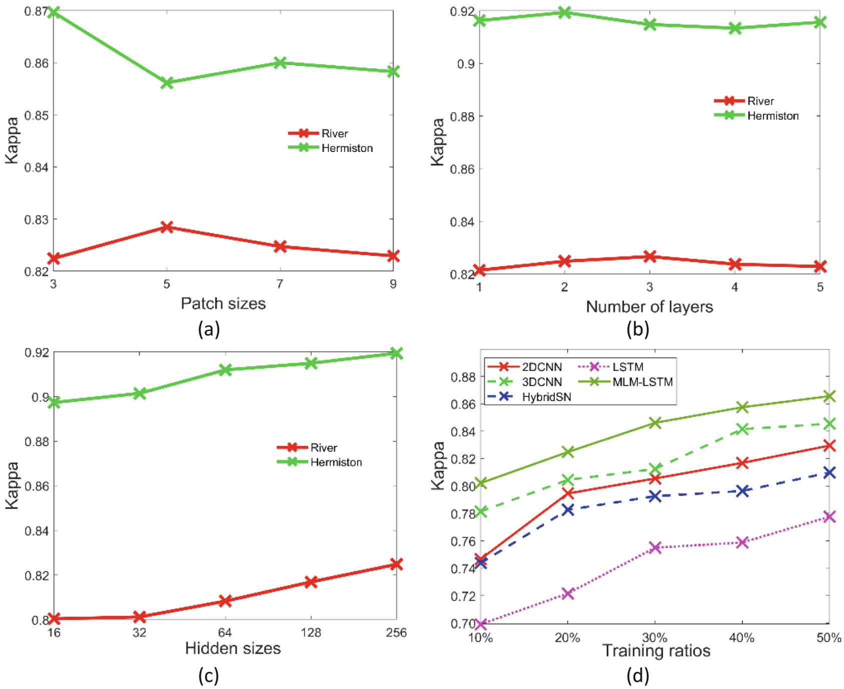


Fig. 5. Ablation experiments and results od the MLM-LSTM in different setting, including the KP values of different patch sizes (a), different number of layers (b), different hidden sizes (c) and different training ratios on the River dataset (d)

evident that the 3-D CNN and HybridSN methods have significantly more hyperparameters compared to the LSTM-based methods. Furthermore, the FLOPs associated with CNN-based methods are several dozen or even hundreds of times higher than those of LSTM-based methods. This discrepancy arises from the nature of convolutional layers used in CNN, which involve convolution operations, pooling operations, and non-linear activations. Image analysis tasks often require a large number of filters and larger input sizes, resulting in higher FLOPs for CNN. In contrast, LSTM primarily involves matrix multiplication and element-wise operations, resulting in lower FLOPs.

Table 2. Comparing the parameters of different DL-based methods on River dataset

	LSTM	HybridSN	3-D CNN	2-D CNN	Proposed
Hyperparameters (k)	213.79	5128.74	1613.03	607.43	1411.46
FLOPs	3.51	1579.24	215.35	368.21	12.77

4 Conclusion

This paper introduces a novel end-to-end DL-based network called MLM-LSTM for HCD. The proposed MLM-LSTM leverages shallow memory learning and hierarchical long-term and short-term memory learning modules to effectively capture the spectral-spatial features. This leads to more precise binary classification. Experimental results on two publicly available HCD datasets demonstrate that the proposed MLM-LSTM surpasses other benchmark models in terms of performance. It exhibits better stability compared to benchmark methods. These results provide comprehensive validation of the effectiveness and efficiency of the proposed model for HCD tasks.

There are still some limitations of our proposed method. For example, in the deep memory extraction modules of long-term memory and short-term memory, the model parameters are relatively large, and the outputs after long-term memory extraction and short-term memory extraction are not fully utilized. To address this limitation, we plan to incorporate a Siamese network [18] in our future work. This approach will allow us to share parameters during the extraction of long-term memory and short-term memory, enabling the extraction of distinctive change features in a more efficient manner. Additionally, we intend to explore the inclusion of more LSTM layers to further improve the feature representation ability of the network.

References

1. Ma, P., et al.: Multiscale superpixelwise prophet model for noise-robust feature extraction in hyperspectral images. IEEE Trans. Geosci. Remote Sens. **61**, 1–12 (2023)
2. Luo, F., Zhou, T., Liu, J., Guo, T., Gong, X., Ren, J.: Multiscale diff-changed feature fusion network for hyperspectral image change detection. IEEE Trans. Geosci. Remote Sens. **6**, 1–13 (2023)
3. Yan, Y., Ren, J., Liu, Q., Zhao, H., Sun, H., Zabalza, J.: PCA-domain fused singular spectral analysis for fast and noise-robust spectral-spatial feature mining in hyperspectral classification. IEEE Geosci. Remote Sens. Lett. (2021)
4. Hasanlou, M., Seydi, S.T.: Hyperspectral change detection: an experimental comparative study. Int. J. Remote Sens. **39**, 7029–7083 (2018)
5. Song, R., Ni, W., Cheng, W., Wang, X.: CSANet: cross-temporal interaction symmetric attention network for hyperspectral image change detection. IEEE Geosci. Remote Sens. Lett. (2022)
6. 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. **57**, 3–13 (2018)
7. Song, A., Choi, J., Han, Y., Kim, Y.: Change detection in hyperspectral images using recurrent 3D fully convolutional networks. Remote Sens. **10**, 1827 (2018)
8. Qu, J., Hou, S., Dong, W., Li, Y., Xie, W.: A multilevel encoder–decoder attention network for change detection in hyperspectral images. IEEE Trans. Geosci. Remote Sens. **60**, 1–13 (2021)
9. Yan, Y., et al.: Non-destructive testing of composite fiber materials with hyperspectral imaging—evaluative studies in the EU H2020 FibreEUse project. IEEE Trans. Instrum. Meas. **71**, 1–13 (2022)
10. Li, Y., Ren, J., Yan, Y., Petrovski, A.: CBANet: an end-to-end cross band 2-D attention network for hyperspectral change detection in remote sensing. IEEE Trans. Geosci. Remote Sens. (2023)

11. Malila, W.A.: Change vector analysis: an approach for detecting forest changes with Landsat. Presented at the LARS Symposia LARS Symposia, p. 385 (1980)
12. Li, Y., Ren, J., Yan, Y., Liu, Q., Petrovski, A., McCall, J.: Unsupervised change detection in hyperspectral images using principal components space data clustering. J. Phys.: Conf. Ser. (2022)
13. Hotelling, H.: Relations between two sets of variates. Methodol. Distrib. 162–190 (1992)
14. He, N., et al.: Feature extraction with multiscale covariance maps for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **57**, 755–769 (2018)
15. Hamida, A.B., Benoit, A., Lambert, P., Amar, C.B.: 3-D deep learning approach for remote sensing image classification. IEEE Trans. Geosci. Remote Sens. **56**, 4420–4434 (2018)
16. 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 Geosci. Remote Sens. Lett.Geosci. Remote Sens. Lett. **17**, 277–281 (2019)
17. Liu, Q., Zhou, F., Hang, R., Yuan, X.: Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. Remote Sens. **9**(12), 1330 (2017)
18. Li, X., et al.: Siamese residual neural network for musical shape evaluation in piano performance assessment. Presented at the 31st European Signal Processing Conference (2023)