Contour extraction of medical images using an attention-based network.

LV, J.J., CHEN, H.Y., LI, J.W., LIN, K.H., CHEN, R.J., WANG, L.J., ZENG, X.X., REN, J.C. and ZHAO, H.M.

2023

Contour extraction of medical images using an attention-based network

Ju Jian Lv[1], Hao Yuan Chen[1], Jia Wen Li[1,2,3*], Kai Han Lin[4*], Rong Jun Chen[1], Lei Jun Wang[1], Xian Xian Zeng[1], Jin Chang Ren[1,5], Hui Min Zhao[1]

[1]School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China

[2]Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin 541004, China

[3]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430065, China

[4]Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China

[5]National Subsea Centre, Robert Gordon University, Aberdeen AB21 0BH, UK

[*]Corresponding authors: Jia Wen Li, lijiawen@gpnu.edu.cn; Kai Han Lin, ninohan@foxmail.com

Abstract:

A comprehensive analysis of medical images is important, as it assists in early screening and clinical treatment as well as subsequent rehabilitation. In general, the contour information can elaborately describe the shape and size of lesions in a medical image, which accurately reflects specific and valuable properties that facilitate the identification of abnormalities, so contour extraction is meaningful. However, the traditional method usually depends on the output of image segmentation, which causes blurred edges and loss of details. To address these issues, an effective attention-based network for contour extraction is proposed, where a model mixed with U-Net and an attention network is utilized to extract image features, and a multilayer perceptron (MLP) is employed to classify those features to obtain a clear contour. Compared with the existing methods, the experimental results on three datasets (Herlev, Drosophila, and ISIC-2017) show that the accuracy reaches approximately 93–98% by using the proposed network, and the number of parameters is 46.4% less than the deep active contour network (DACN). Such performances are impressive when considering accuracy and the number of parameters as

the key concerns. Therefore, this study reduces the model computation with almost no loss of accuracy, which can satisfy clinical requirements for medical image analysis.

Keywords:

Contour extraction, Medical image, Attention-based network, Multilayer perceptron (MLP), Deep learning

1. Introduction

With the development of imaging technology, large numbers of medical images are produced from various mechanisms, such as computed tomography (CT), X-rays, magnetic resonance imaging (MRI), and microscopy imaging [1]. Undoubtedly, a comprehensive analysis of medical images is important, as it assists in early screening and clinical treatment as well as subsequent rehabilitation [2]. When physicians and specialists observe an image with abnormalities, for instance, an MRI of the brain in patients with multiple sclerosis, they will search a large collection of MRIs and retrieve those cases with tomographic levels that contain abnormalities showing similar size, shape, and location [3]. This is a visual manner used to understand the diseases, so traditionally, physicians and specialists determine the conditions of patients through their own experiences in observing the size, shape, and location of abnormalities on the medical image [4]. This subjective solution is manual, which costs considerable time and effort. Hence, to reduce the heavy workload for visual inspection, an automatic algorithm is desired. Considering that the contour information can elaborately describe the shape and size of lesions in a medical image, which accurately reflects specific and valuable properties that facilitate the determination of abnormalities, the contour extraction of medical images is meaningful [5].

Typically, two methods have been applied to achieve contour extraction: the traditional image segmentation algorithm and the deep learning method. Traditional algorithms are usually highly interpretable because they are developed by modeling the mathematical processes underlying specific problems or conditions [6]. Therefore, the whole calculation procedure is

interpretable. For example, Li et al. [7] utilized a multithreshold optimization method to solve the problems of fuzzy features and nonconvergent thresholds in the image denoising of details during fine segmentation. In another work, Mathur et al. [8] offered an approach to improve the Sobel edge detector and evaluated it on the MRI image, as the Sobel operator has been extensively employed in the edge detection algorithm, where it creates an image emphasizing edges. This property helps in determining the location of a brain tumor in an MRI image.

However, medical images have the characteristics of unclear lesion edges, uneven gray levels, and interfering objects, which can influence the performances of traditional algorithms [9]. In this regard, the deep learning method is good at extracting contours more precisely [10]. For instance, He et al. [11] proposed a mask region-convolutional neural network (R-CNN) based on a faster R-CNN, in which the region of interest (ROI)-align replaces the ROI-pooling to improve the segmentation accuracy. Zunair and Hamza [12] introduced a sharp U-Net architecture by designing connections between the encoder and decoder subnetworks through a depthwise convolution of the encoder feature maps with a sharpening spatial filter. Therefore, the gap issue between the encoder and decoder features is addressed, and the segmentation accuracy of edges is improved. To overcome the cluttered lines in traditional contour detection, Rampun et al. [13] presented a multilevel edge detection network named holistically nested edge detection (HED). But, it has a limitation in that it only focuses on the main component of the image, while the edge is not clear. Ding et al. [14] applied a generative adversarial network (GAN) for contour extraction, in which the contour of the target object is produced by the generator, and the authenticity is determined by a discriminator that obtains the contour information. Nonetheless, this approach lacks spatial consistency, resulting in the contour becoming incoherent and broken. To this end, Deng et al. [15] proposed a contour enhancement module to make the network efficiently learn the shape information.

In addition, selecting a suitable contour representation method is vital, as different contour representations have an influence on computational complexity and the results [16]. In this regard, Huang et al. [17] employed a snake model for contour extraction. It is an active contour representation method that deploys a deformable parametric curve to define an

energy function that weighs the internal and external forces. When the energy is minimal, the curve will fit the contour of objects and be smooth, i.e., the active contour converges to the edge of the target. Nevertheless, it requires an initial contour line to start the iteration, meaning that it is sensitive to the initial profile setting. Then, Soora et al. [18] modified the active contour model to be differentiable and combined it with a deep neural network. Meanwhile, a new loss function incorporating external forces and regional information was used, which enhanced the segmentation robustness.

Inspired by previous studies, it is necessary to solve blurred lesion edges and neglected lesion details such that the accuracy of contour extraction can be further enhanced. To this aim, an attention-based network for contour extraction of medical images is proposed in this paper. In particular, the main contributions are summarized below:

(1) An attention-based network that combines U-Net and a convolutional block attention module (CBAM) is applied to address edge blurring and interference. This network combines the feature extraction capability of U-Net with the attention capability of CBAM, so it facilitates removing interferences and improving accuracy. An experiment on the Drosophila dataset shows that the sensitivity is enhanced by 6.3% compared to U-Net only.

(2) A multilayer perceptron (MLP) is adopted to define an implicit function, which implies the probability that the current position is far from the target. It assists in refining the edge of the upstream output feature map to enhance accuracy. The experiment on the Drosophila dataset reveals that the accuracy is improved by 0.3% compared to U-Net only.

(3) During the process of extracting the feature map as a contour, the manual setting of the threshold has an influence on the result. To avoid manual setting, a one-hot coding method is considered, which can be integrated into the neural network and compared to the adaptive setting of the threshold.

(4) The marching squares algorithm is good at extracting the contour, but the output points are not related. Hence, the union-find method and single-linked list to connect all points into a curve are exploited to recover these chain relations.

The remainder of this paper is organized as follows: Section 2 reviews the related works concerning contour extraction. Section 3 presents the proposed network through three subsections: feature extraction, MLP, and contour extraction. Section

4 analyzes the performance evaluation and assesses the results from the experiments using three datasets (Herlev, Drosophila, and ISIC-2017). Finally, Section 5 describes the conclusion of this paper.

2. Related Works

Typically, a convolutional neural network (CNN) becomes difficult to train when it is deepened, as the deeper framework usually causes the gradient to disappear. To this end, He et al. [19] used ResNet to solve this issue. ResNet is based on residual blocks that deploy connections to skip several layers in a network. When network degradation appears, it will manifest as transformation rather than direct gradient disappearance. This framework is flexible and can be further extended by different numbers of layers [20]. Moreover, the encoder and decoder of ResNet can be embedded in other networks, such as U-Net, so it is appropriate for improving the object recognition performance in medical images [21].

U-Net is a pixel-based approach that has been extensively employed in medical image segmentation [22]. It has a U-shaped symmetric structure that consists of a feature extraction module and a feature fusion module. The feature extraction module adopts eight convolutions and downsampling on the input image to obtain the features. On the other hand, the feature fusion module is a reverse operation that aims to recover the image from the features. To achieve this, two convolutions and upsampling are needed, and then this procedure is repeated four times such that the output feature map obtained by the U-Net is the same as the input resolution. In addition, to reduce the loss of features due to downsampling, U-Net utilizes skip connections in each upsampling layer to fuse the corresponding downsampling feature maps, so the segmentation accuracy can be improved as the network gradually converges to the target region [23]. Although U-Net is available for image segmentation with a clear boundary, it exhibits the same resolutions on the input and output feature maps. Hence, the traditional U-Net is unsuitable for solving the interference problem in medical images [24]. In this regard, the combination of U-Net and attention modules is more appropriate, such as the channel spatial sequence attention module (CSSAM) proposed by Song et al. [25]. It is based on U-Net and an embedded CBAM in skip connections to achieve a

higher-precision segmentation of welding engineering drawing contours. CBAM is a block attention-based model that combines spatial and channel domains [26]. In the spatial domain, two feature maps are acquired by compressing the channels with maximum pooling and average pooling and then performing a convolution to obtain the result. This procedure is similar to a spatial transformer network (STN) [27]. Simultaneously, in the channel domain, the attention function is implemented by squeezing and excitation, such as the squeeze-and-excitation network (SENET) [28]. As found, CBAM considers both spatial and channel attention and possesses a symmetry-like structure. Thus, the input and output dimensions remain constant. Such effects are beneficial to embed into any module to solve the interference problem in a medical image. However, the attention boundary using CBAM is usually blurred, so directly applying it in contour extraction is improper [29]. To this end, Wang et al. [30] proposed the VGG-style base network (VSBN) as the backbone and embedded the CBAM module after each convolution block. Its novelties include the utilization of salt-and-pepper noise in data argumentation and good performance in the diagnosis of coronavirus disease 2019 (COVID-19). Similarly, Zhang et al. [31] designed an attention network based on CBAM and fine-tuned it to find COVID-19 relevant regions accurately. This framework can be interpreted by gradient-weighted class activation mapping (Grad-CAM), and 18 approaches for data augmentation were employed to avoid overfitting.

In addition, the marching squares algorithm is a graphics approach that extracts contours from the image [32]. If the input is a binary image, this algorithm will include 16 cases. Then, these cases are stored in a preestablished look-up table to enhance the efficiency of producing the edges from the point array. Based on this step, by performing the operations for each pixel, all point and edge sets can be acquired and used to extract contours from the image. If the input is a gray image, the algorithm interpolates between adjacent points to calculate the exact contour position. Similar to the marching cubes, the marching squares algorithm generates contour lines based on values and processes each pixel independently. Hence, when the input volume is enormous, it can be implemented in parallel to accelerate the processing [33].

Furthermore, the MLP simulates the operation of neurons by setting multiple layers and employing a full-connection

framework. Generally, it has fewer parameters, and the output layer is one or more neurons that can classify the features. Consequently, the MLP can be regarded as a fine-tuning method in the network [34]. Saito et al. [35] designed a PIFu network for a three-dimensional image reconstruction of the human body, where the CNN obtains a heatmap of the main subject, the MLP represents the isometric surface, and the marching cubes algorithm finds the boundary lines and constructs the model. Recently, several studies combined U-Net and the MLP to develop a low-complexity and low-parameter framework. For instance, Valanarasu and Patel [36] adopted UNeXt, a U-Net-like framework, and embedded a tokenized MLP block for medical image segmentation. In another work, Tu et al. [37] proposed a MAXIM architecture for the image processing task, in which the encoder and decoder are designed according to the U-Net, and a multiaxis gated MLP serves as an efficient and flexible general-purpose vision backbone. In [38], TransClaw U-Net was developed by Yao et al., in which the encoder adopted a similar approach to the transformers and embedded an MLP for medical image segmentation. In addition, Peng et al. [39] utilized the combination of U-Net and an MLP to solve the temperature field reconstruction, which demonstrated the generalization performance of the model.

This paper develops a segmentation network for two-dimensional medical images, which improves the efficiency of training and prediction by transforming the task of contour extraction into a binary classification issue. Specifically, the novelties include an improved U-Net framework that reduces the feature loss caused by the downsampling step. Then, combined with the attention module, the lack of global and local information in the U-Net network is solved, which enhances the segmentation accuracy and enables faster convergence accordingly. In addition, to avoid the hyperparameter setting in the contour extraction and reduce human interference, the foreground and background prediction method is applied, and the segmentation results are extracted as the contours through the level-set algorithm, which mitigates the iterative process of the snake model. Finally, to demonstrate the performance effectiveness of the proposed network, evaluations through three datasets are performed based on various metrics, and comparisons with different approaches are also provided.

3. Methodology

The proposed attention-based network is depicted in Fig. 1, which includes three parts: feature extraction, the MLP, and contour extraction. The input is a red–green–blue (RGB) medical image. Then, feature extraction is achieved by an improved U-Net, where one is the encoding network with downsampling and the other is the decoding network with upsampling. Compared to the traditional U-Net, this variant deploys the attention module before each upsampling layer. The subsequent ablation experiments demonstrate that the attention module effectively removes the interference information from the images. Here, the output feature map has a C-dimensional representation for each pixel, which describes the global context information [25]. In addition, a combination of U-Net and CBAM is available to fuse local and global context information, which helps to produce precise segmentation accuracy [40]. However, the downsampling process of U-Net loses considerable feature information, and the edge details are ignored, so it is necessary to involve the MLP on each pixel of the feature map. The MLP performs a binary classification task and acquires two-dimensional information representing the probabilities of the target and nontarget. Therefore, a binary image is obtained by comparing these probabilities, and this mechanism can simplify the manual setting of thresholds. Finally, the marching squares algorithm is applied to extract the contour of the binary image and chain all points into a curve. Consequently, contour extraction of the image is accomplished. The following subsections describe the above stages in detail.

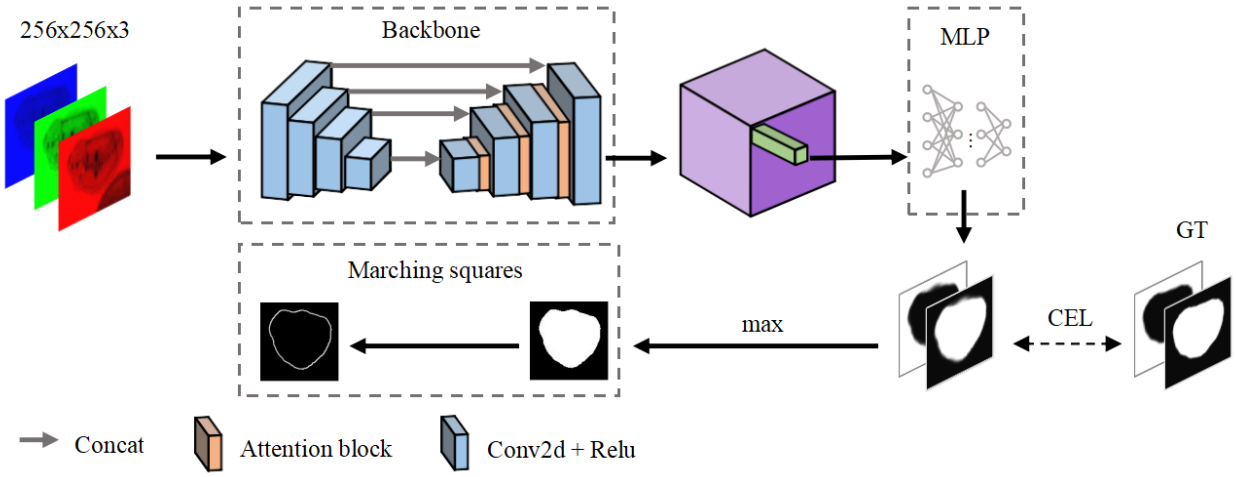

Fig. 1. Contour extraction of medical images using the proposed attention-based network.

3.1 Feature extraction

The feature extraction contains encoding and decoding. The feature encoding utilizes the pretrained ResNet18 model [19], which reduces the losses from the U-Net downsampling process through skip connections, so the network converges. The feature decoding is similar to the feature decoding of U-Net, while the novelty is that CBAM is embedded before each upsampling layer. It is helpful to solve the grayscale unevenness presented in the medical image and reduce unwanted holes in the output. As mentioned, CBAM is a network embedded with an attention mechanism that includes channel attention and spatial attention [26]. Mathematically, the channel attention is expressed as (1) and (2):

$$f_a(x) = MLP(AvgPool(x)) + MLP(MaxPool(x)) \tag{1}$$

$$M_c(F) = F * Sigmoid(f_a(F)) \tag{2}$$

where $M_c(F)$ denotes the channel attention generated by the average pooling and maximum pooling. The variable F refers to the output of the upsampling module. The two MLPs in (1) are shared MLPs, and the same weights are employed in these MLPs.

The spatial attention $M_s(F)$ is denoted as (3) and (4), where A $\oplus$ B refers to the concatenations of A and B by channels:

$$f_c(x) = Conv(AvgPool(x) \oplus MaxPool(x)) \tag{3}$$

$$M_s(F) = F * Sigmoid(f_c(F)) \tag{4}$$

Now, the CBAM is obtained by combining channel attention and spatial attention, as presented in (5):

$$M(F) = M_s(M_c(G(F))) + F \tag{5}$$

where

$$G(F) = Conv_2(Conv_1(F)) \tag{6}$$

CBAM first employs two convolutions to the input to obtain *G(F)*, and then performs channel attention and spatial attention, which aim to compensate for the lack of position perception using U-Net. In this regard, the fusion of CBAM and U-Net effectively senses the contour location and accurately extracts the contour. It also assists in removing the interference near the lesion to improve the segmentation performance. After multiple convolutions and downsampling, the low-level feature details that are valuable for recovering the original image can be achieved. Nonetheless, the feature details include both

target lesions and noise. Hence, the attention module is employed to reduce the impact of noise such that the contour extraction accuracy can be enhanced. Undoubtedly, a more accurate feature map facilitates the decision, such as the MLP, which can determine the lesion location and shape through this information. Consequently, after downsampling and upsampling, the feature descriptions fuse global and local information to provide the model with a larger field of perception.

3.2 The multilayer perceptron (MLP)

The proposed network adopts the MLP to classify the extracted features. A preliminary test reveals that to correctly find the distinctions between lesions and interferences, the all features method is more suitable than the selected features method, as the selected features method is not good at converging correctly. Based on that, the MLP is embedded in each feature description, where each feature contains a C-dimensional vector that describes the global information of the pixel. Assume that the feature map obtained is U^{W*H*C}, where C denotes the number of feature descriptions (from a preliminary test, to provide satisfying results, $C = 512$). Then, the algorithm for generating a binary map is as follows:

Algorithm 1. MLP fine-tuning process

Input: U^{W*H*C}

Output: M^{W*H}

initialize: Set M^{W*H}
for x = 1, 2, …, W **do**
    for y = 1, 2, …, H **do**
        block^{1*1*2}←$MLP(U(x, y))$
            if block(1, 1, 0) > block(1, 1, 1) **then**
            $M(x, y)$←1
        else
            $M(x, y)$←0
        end if
    end for
end for

As illustrated in Fig. 2, the MLP includes three hidden layers with 512, 256, and 128 elements. It can refine the edge and clearly distinguish the target lesion from the background. Then, the task is a binary classification issue, so the output is two values that represent the probabilities of foreground and background, indicating how far the current location away from the target is. Finally, with the help of the softmax activation function, the two heatmaps are obtained simultaneously. Thus, comparing the probabilities of foreground and background, a binary map is acquired. If the foreground probability is larger than the background probability, this pixel will be considered an internal region of the lesion.

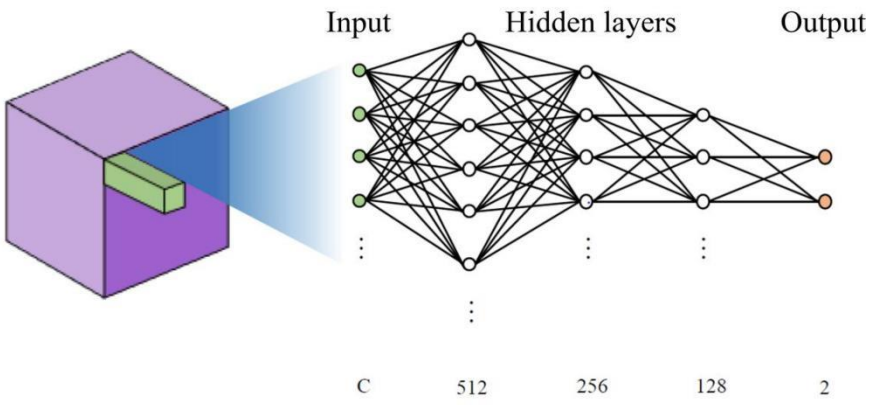

Fig. 2. The MLP in the proposed network.

Overall, there are two reasons for the above operations. First, compared with the previous image segmentation algorithm, the probability map usually ignores much information. Thus, the probability calculation is performed on the foreground and background, which helps to enhance the accuracy. Second, compared with the previous contour extraction algorithm, the probabilistic map has not yet accurately distinguished the border and generally requires a manual threshold setting to obtain a binary image. Nevertheless, this method is subjective, and the exact threshold depends on the empirical value. Hence, the operations applied can avoid setting the hyperparameter and solve the need for a manual threshold setting.

3.3 Contour extraction

After acquiring a binary map, the marching squares algorithm is employed, as it is appropriate for two-dimensional contour extraction. This algorithm constructs edges based on the adjacent values around the pixels, and linear interpolation is performed to yield a smooth curve from an input binary map. When determining a threshold value, the algorithm iterates all pixels, and by comparing the current pixel to its three adjacent pixels, 16 cases are provided. Then, these cases can be simplified by rotating and mirroring into six, as depicted in Fig. 3. As a result, the number of cases is reduced. In addition, a template that describes the edges constructed is used, and a 4-bit hash code is obtained when comparing them in a certain direction. Finally, the edges are acquired based on the template, and this operation can be parallel, which decreases time consumption and reduces the computational cost accordingly.

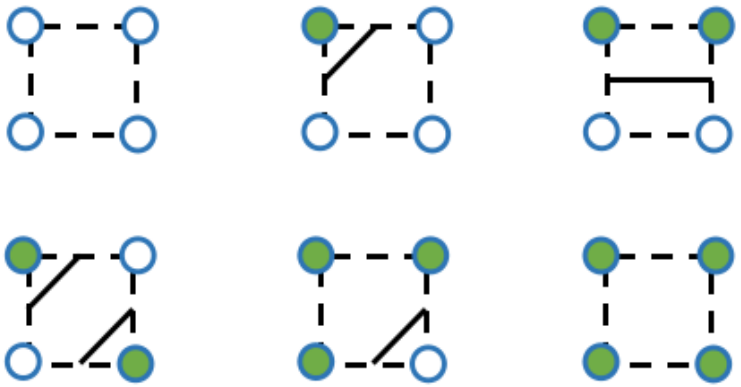

Fig. 3. The six basic cases from the marching squares algorithm.

At this point, each line is separate, as the output points are not related, but certain relations should exist in the two-dimensional space. Hence, it is meaningful to recover the relations of each line. To this aim, the union-find algorithm and a single-linked list are adopted to decide the chain relations. First, for fetching the vertices in the chain, a single-linked list is exploited to store the vertices index of the chain in reverse order. Subsequently, the path is output by iterating these chains in reverse order, so the contour having the directed lines can be extracted. Finally, to set a direction and chain root point for merging the points along this direction, a directional acyclic curve is needed. Therefore, the extractions of simple lines containing the rings are achieved by such extensions, recovering the chain relations correspondingly.

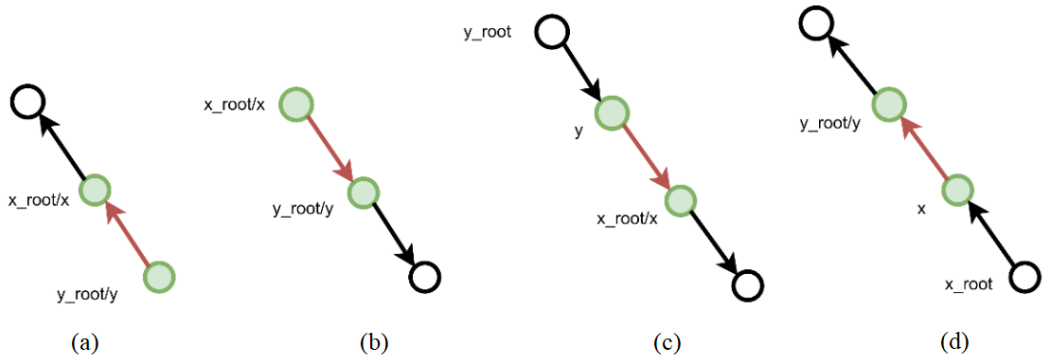

Fig. 4. The four basic cases when connecting two points in a map.

As an illustration, the green dots shown in Fig. 4 represent two vertices (*x* and *y*) that are assumed to be connected, and seven cases of connections exist, where six of them can be simplified into four. When both *x* and *y* are chain roots and one of the chains has only one vertex, as drawn in (a) and (b) of Fig. 4, a point that is in a single point chain to another that is in a long chain (in red arrow) can be connected. If *x* and *y* are both chain roots and there is only one vertex, the two points can be connected directly. In addition, when one of *x* and *y* is the root of the chain and both points are in a long chain with the same direction, the two points in this direction can be connected, as displayed in (c) and (d) of Fig. 4. When both *x* and *y* are not the chain roots and their directions are opposite, as presented in (a) of Fig. 5, it is necessary to invert the short chain to produce the case in (d) of Fig. 4 and then solve the connection problem in the same way. Finally, when *x* and *y* are both chain roots and the directions are opposite, as depicted in (b) of Fig. 5, the short chain can be reversed to acquire the case in (c) of Fig. 4. Therefore, all cases of chain connections are considered, and a connected curve is generated after this step.

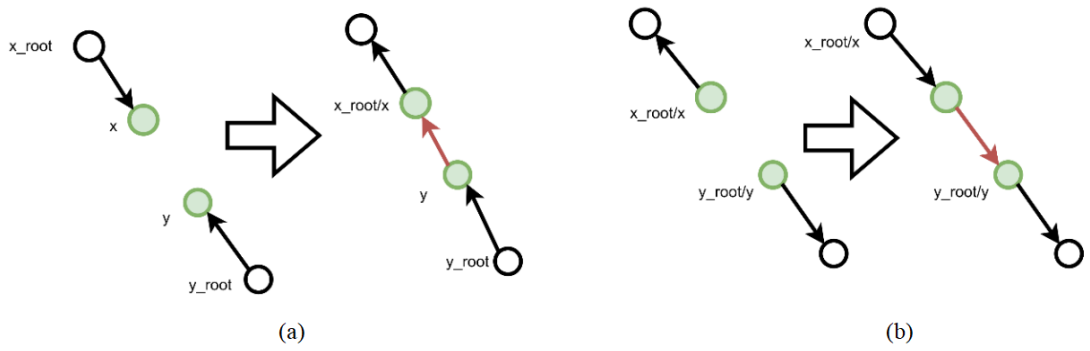

Fig. 5. Two extra cases can be converted into one of the basic cases.

4. Experiments and Results

4.1 Datasets

To thoroughly evaluate the proposed network, three public datasets are assessed in this paper: Drosophila [41], Herlev [42], and ISIC-2017 [43]. The Drosophila dataset contains 399 training images and 100 testing images, each with the corresponding mask image. In addition, the ROIs of Drosophila embryos are annotated, and other similar cells are located within the field of view, which can be regarded as interferences. Hence, the aim is to identify the ROIs and extract the contours using the proposed network. The Herlev dataset includes 917 cervical cell images that are categorized into seven classes. The ROIs are labeled, but due to the complex and diverse shapes of the cervical cells, the image edge contrast is mostly weak. Thus, this challenge helps to assess the segmentation accuracy using the proposed network. Finally, the ISIC-2017 dataset comprises a large number of dermoscopic images (2000 training images and 600 test images), and each has a corresponding label. In particular, this dataset contains various types of melanomas and moles, along with interferences such as hairs and lights. Thus, the generalization performance of the proposed network can be validated. Furthermore, the three datasets vary in image size, so they are scaled to 256 × 256 in the preprocessing stage. Meanwhile, to ensure the robustness of the model, random horizontal and vertical flips with brightness and contrast adjustments (set to ±0.4 in the experiments) are performed.

Regarding the experimental conditions, Python is employed, the optimizer is Adam, the initial learning rate is 0.001, the learning rate is adjusted by StepLR, and the encoding of the network is based on a pretrained ResNet18 model. Due to the different scales of the datasets, the training strategy is varied for each one to avoid overfitting. In addition, the following setting values are applied, as they helpfully enhance the accuracy: for the Drosophila dataset, the learning rate is 0.9 times the original rate per 6 epochs; for the Herlev dataset, the learning rate is modified to 0.9 times the original rate per 4 epochs; and for the ISIC2017 dataset, the learning rate changes to 0.8 times the original rate per 3 epochs. The network is trained on a single NVIDIA GeForce RTX 2080 and Ubuntu, and the batch size is 6.

4.2 Evaluation Metrics

In the experiments, several typical metrics, such as the dice similarity coefficient (DSC), precision, sensitivity, accuracy, and Hausdorff distance, are obtained. Note that the predicted mask generated by the proposed network is *Pred*, and the ground truth mask from the dataset is *GT*. The calculation of each metric is explained below.

DSC reveals the similarity between the two sets of data. Assume that $|X|$ represents the area of Subject X, and $|X \cap Y|$ is the area of the intersection of subjects X and Y. Then, the definition of DSC is (7):

$$DSC = \frac{|Pred \cap GT|}{|Pred| + |GT|} \times 2 \tag{7}$$

Precision indicates the correct percentage of the content predicted by the algorithm. Its calculation is (8):

$$Precision = \frac{|Pred \cap GT|}{|Pred|} \tag{8}$$

Sensitivity, also known as recall, refers to the percentage of correct regions determined by the algorithm. It demonstrates the ability of the classifier to discriminate between positive samples, as calculated by (9):

$$Sensitivity = \frac{|Pred \cap GT|}{|GT|} \tag{9}$$

Accuracy describes the percentage of all samples that are correctly classified. It is flawed in the case of uneven samples, so it is mostly applied in combination with other metrics. The calculation is expressed by (10):

$$Accuracy = \frac{|U - Pred \cup GT + Pred \cap GT|}{|U|} \tag{10}$$

where U is the area of the whole image obtained by multiplying the width and height.

Moreover, denote the contour extracted from the prediction map as *PredC* and the contour of the ground truth as *GTC*. Both are point sets acquired from the marching squares, without containing the duplicated points. The Hausdorff distance is adopted to measure the distance between the two point sets. For each point, the distance between it and the closest point set is acquired, and then the maximum distance is utilized as Hausdorff distance h by (11):

$$h(PredC, GTC) = \max_{a \in PredC} \{ \min_{b \in GTC} \{ d(a, b) \} \} \tag{11}$$

where $d(a, b)$ refers to the distance between points a and b.

4.3 Results and Discussion

First, the experimental results are compared with the U-Net [22], deep active contour network (DACN) [44], and several state-of-the-art approaches [45-47]. On the one hand, U-Net has a low number of parameters and computation costs, so it is suitable for a lightweight framework. However, it lacks an attention mechanism and is inappropriate for processing images with interference. Hence, the traditional U-Net model is listed as one of the comparison methods to validate the improvement employing the proposed network. On the other hand, to perform a fair comparison, methods using approximate frameworks and with similar experimental conditions are considered. In this regard, DACN is a framework that has achieved good results in medical image contour extraction in recent years, as it utilizes a differential active contour model to enhance accuracy. Thus, it is chosen in the comparative study. Furthermore, ASCNet [45], Deeplab V3 [46], and FocusNetAlpha [47] are selected, as they are appropriate for comparisons.

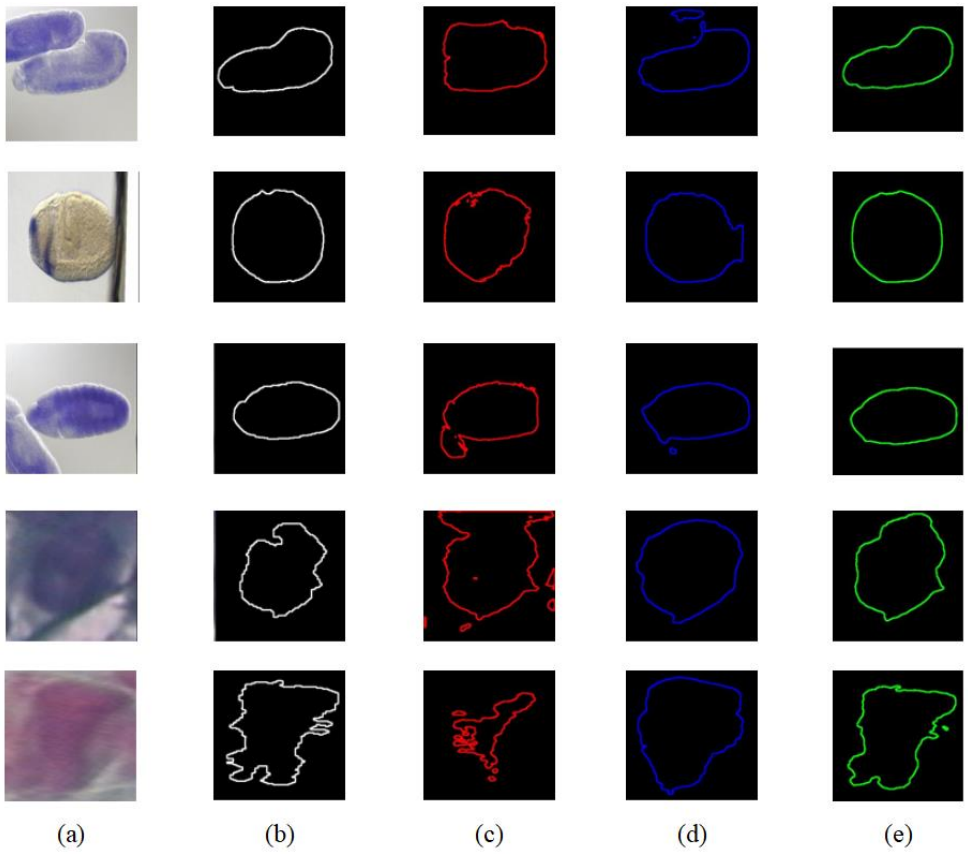

Fig. 6. The contour extraction results from different cases (Drosophila and Herlev): (a) source; (b) ground truth; (c) U-Net; (d) DACN; (e) proposed network.

For the results from Drosophila and Herlev, the proposed network is trained on Drosophila with 103 epochs and on Herlev with 120 epochs. Then, the contour extraction results from different cases are depicted in Fig. 6, and the details of the evaluation metrics are summarized in Tables 1 and 2, where the best metric is shown in bold. On the one hand, the results of the metrics reveal that the network is more impressive than U-Net and DACN because it yields a higher DSC, precision, sensitivity, and accuracy and a lower Hausdorff distance, meaning that the performances are improved. On the other hand, in Fig. 6 (e), more clear contours are observed visually than in Fig. 6 (c) and (d), and their shapes are more similar to the source and ground truth in Fig. 6 (a) and (b). As a result, from both a quantitative analysis and visual check, the proposed network demonstrates that its contour extraction can be more precise than U-Net and DACN.

Table 1. Quantitative analysis of different methods on the Drosophila dataset.

Model	Dice	Precision	Sensitive	Accuracy	Hausdorff distance
U-Net [22]	92.60%	94.31%	91.43%	94.56%	28.19
DACN [44]	96.77%	97.79%	95.83%	97.53%	17.84
Proposed network	**97.76%↑**	**98.39%↑**	**97.20%↑**	**98.24%↑**	**7.97↓**

Table 2. Quantitative analysis of different methods on the Herlev dataset.

Model	Dice	Precision	Sensitive	Accuracy	Hausdorff distance
U-Net [22]	90.42%	90.53%	93.08%	95.88%	36.49
DACN [44]	94.54%	94.74%	95.08%	97.63%	15.58
ASCNet [45]	91.50%	91.00%	93.80%	--	--
Deeplab V3 [46]	91.30%	91.70%	92.60%	--	--
Proposed network	**95.15%↑**	**95.78%↑**	**95.14%↑**	**97.96%↑**	**13.08↓**

In addition, the ISIC-2017 dataset provides the imaging of damaged skin under a microscope, and the skin tissues in various locations show larger variations. It includes numerous interferences, such as hairs and lights, and is more complex, so the

results exhibit lower performances than the above two datasets, as presented in Table 3. Here, several metrics are satisfying, except for the sensitivity and Hausdorff distance. The reason may be due to interferences that are not easy to identify, so the lesion areas are not obvious, which could lead to inference failure or the edges inaccurately surrounding the lesion areas.

Table 3. Quantitative analysis of different methods on the ISIC-2017 dataset.

Model	DSC	Precision	Sensitivity	Accuracy	Hausdorff distance
U-Net [22]	78.26%	87.66%	76.83%	91.35%	39.17
DACN [44]	84.62%	90.76%	**84.32%**	93.19%	**24.46**
FocusNetAlpha [47]	84.04%	80.02%	82.22%	**93.49%**	--
Proposed network	**84.71%↑**	**93.79%↑**	81.60%	93.48%	26.89

Next, Table 4 shows the complexity comparisons between DACN and the proposed network by considering the number of parameters (Params) and floating point operations (FLOPs). The results are calculated automatically using the ptflops library (a python module), and the input images are 256 × 256 and C is 256. As seen, the proposed network is simpler to implement, as it involves 46.4% fewer Params than DACN. In addition, it is easily trained and benefits acceleration during the convergence process. Therefore, it can be said that the network achieves higher efficiency with fewer computational resources and less complexity.

Table 4. Comparisons of complexity between DACN and the proposed network.

Model	Params (M)	FLOPs (GMac)
DACN [44]	48.90	704.90
Proposed network	**26.20↓**	**66.10↓**

Subsequently, to validate the contributions of different components used in the backbone of the network, ablation experiments regarding the attention module and the MLP are conducted, as displayed in Fig. 7 and Table 5. In Table 5, the

accuracy of the attention module improves by 0.17–0.23%, and the accuracy of the fine-tuning with the MLP is enhanced by 0.02–0.29%. Furthermore, as seen in Fig. 7, the absence of either attention or MLP could cause several pieces to be improperly determined, so the accuracies are decreased correspondingly. For instance, as highlighted in the red square, without the attention module, the bottom subfigure in Fig. 7 (c) yields an error, while the same case in Fig. 7 (e) that uses the proposed network can avoid it. Thus, the attention module and the MLP play vital roles in making the size and shape of the results similar to the ground truth.

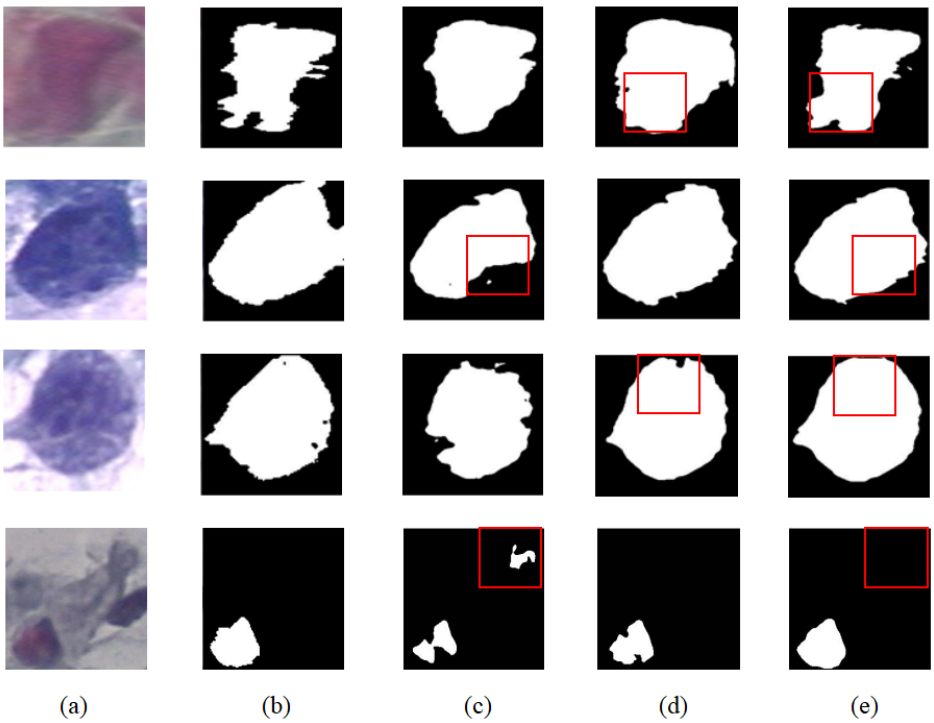

Fig. 7. Ablation experiment results (Drosophila and Herlev): (a) source; (b) ground truth; (c) w/o attention; (d) w/o MLP; (e) proposed network.

Table 5. The accuracies of ablation experiments based on the Herlev and Drosophila datasets.

Dataset	w/o Attention	w/o MLP	**Proposed network**
Herlev	97.73%	97.94%	**97.96%**↑
Drosophila	98.07%	97.95%	**98.24%**↑

Finally, as a flexible solution, the experimental results imply that the network with only U-Net and an MLP can satisfy

most medical cases, as the MLP fixes the errors of decisions and conducts a similar behavior as attention. However, the MLP is not perfect enough to overcome the interferences in several specific medical images with hairs and lights. Therefore, when considering accuracy as a vital concern in the segmentation task, it is preferred to apply a fusion of more appropriate modules in the network backbone, such as the proposed network that combines the advantages of U-Net, MLP, and CBAM. In addition, the proposed network only adds the attention module before the upsampling process, as adding it to the downsampling suppresses much valuable feature information. Based on the above considerations, as an overview, a comparison chart to spot the differences between the proposed network and the other methods is illustrated in Fig. 8, where the variations are highlighted in red squares.

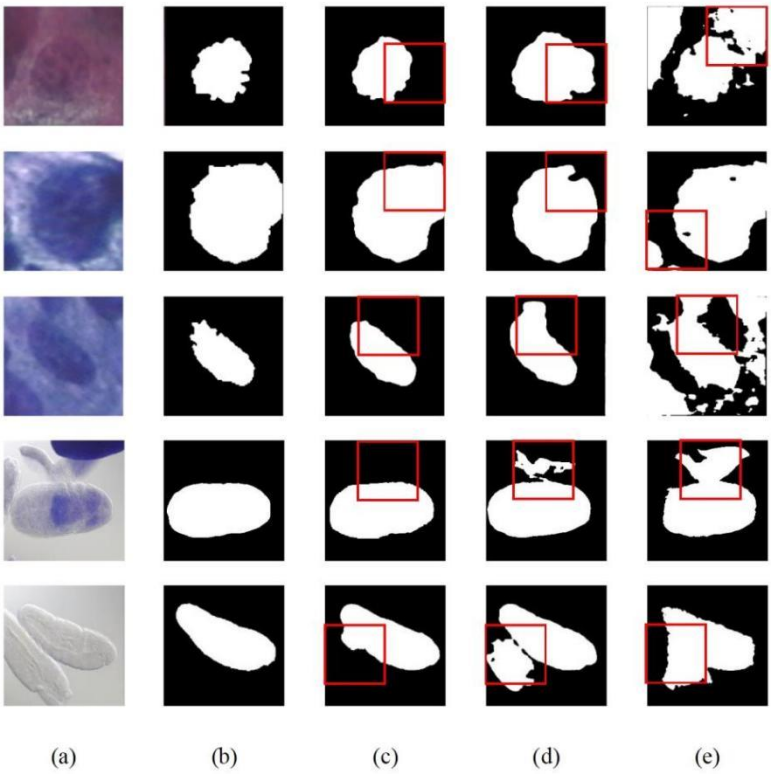

Fig. 8. A comparison chart of different cases (Drosophila and Herlev): (a) source; (b) ground truth; (c) proposed network; (d) DACN; (e) U-Net.

5. Conclusion

This paper proposes an attention-based network for medical image contour extraction, and the main novelties are that this

network combines the advantages of U-Net and CBAM, which are helpful to find the lesion position and shape information in a medical image. In addition, an MLP is applied to refine the edges and enhance the boundaries of internal and external regions of the lesion, so a high contour extraction accuracy is acquired. Furthermore, five commonly used evaluation metrics are detailed and analyzed. Compared with the existing approaches, on the one hand, the inference time of the network is reduced, and the model complexity is decreased, while the accuracy is almost unchanged. On the other hand, the experiments on the Drosophila and Herlev datasets demonstrate that the network contributes impressive performances in contour extraction. Consequently, the proposed network reduces the number of parameters and computation costs with almost no loss of accuracy, which satisfies the clinical requirement for medical image analysis. In the future, this framework will be improved by a transformer with a superior attention module that further enhances the capabilities to capture medical image details.

Conflicts of Interest:

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgments:

This work was supported in part by the National Natural Science Foundation of China under Grant 62072122, in part by the Scientific and Technological Planning Projects of Guangdong Province under Grant 2021A0505030074, in part by the Special Projects in Key Fields of Ordinary Universities of Guangdong Province under Grant 2021ZDZX1087, in part by the Guangzhou Science and Technology Plan Project under Grant 202102020857, in part by the Research Fund of Guangdong Polytechnic Normal University under Grant 2022SDKYA015, in part by the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security under Grant MIMS22-02, and in part by the Fund of Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System (Wuhan University of Science and

Technology) under Grant ZNXX2022005.

Author Contributions:

Conceptualization: Ju Jian Lv and Hao Yuan Chen; Methodology: Ju Jian Lv and Hao Yuan Chen; Formal analysis and investigation: Ju Jian Lv, Hao Yuan Chen, Jia Wen Li, and Kai Han Lin; Funding acquisition: Jia Wen Li, Rong Jun Chen, Lei Jun Wang, Xian Xian Zeng, Jin Chang Ren, and Hui Min Zhao; Resources: Jia Wen Li, Rong Jun Chen, Jin Chang Ren, and Hui Min Zhao; Writing - original draft preparation: Ju Jian Lv, Hao Yuan Chen, Jia Wen Li, and Kai Han Lin; Writing - review and editing: Ju Jian Lv, Hao Yuan Chen, and Jia Wen Li; Supervision: Ju Jian Lv, Jia Wen Li, Jin Chang Ren, and Hui Min Zhao. All authors have read and agreed to the final manuscript.

References:

[1] Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, Muhammad Khurram Khan, Medical image analysis using convolutional neural networks: A review, Journal of Medical Systems 42 (2018) 226.

[2] Kun Wang, Xiaohong Zhang, Yuting Lu, Xiangbo Zhang, Wei Zhang, CGRNet: Contour-guided graph reasoning network for ambiguous biomedical image segmentation, Biomedical Signal Processing and Control 75 (2022) 103621.

[3] Hemant D. Tagare, C. Carl Jaffe, James Duncan, Medical image databases: A content-based retrieval approach, Journal of the American Medical Informatics Association 4 (1997) 184-198.

[4] Daniel García-Lorenzo, Simon Francis, Sridar Narayanan, Douglas L.Arnold, D. Louis Collins, Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging, Medical Image Analysis 17 (2013) 1-18.

[5] Yashwant Kurmi, Vijayshri Chaurasia, Multifeature-based medical image segmentation, IET Image Processing 12

(2018) 1491-1498.

[6] Vishal Monga, Yuelong Li, Yonina C. Eldar, Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing, IEEE Signal Processing Magazine 38 (2021) 18-44.

[7] Wei Li, Qian Huang, Gautam Srivastava, Contour feature extraction of medical image based on multi-threshold optimization, Mobile Networks and Applications 26 (2021) 381-389.

[8] Neha Mathur, Shruti Mathur, Divya Mathur, A novel approach to improve Sobel edge detector, Procedia Computer Science 93 (2016) 431-438.

[9] Xinjian Chen, Lingjiao Pan, A survey of graph cuts/graph search based medical image segmentation, IEEE Reviews in Biomedical Engineering 11 (2018) 112-124.

[10] Xiangbin Liu, Liping Song, Shuai Liu, Yudong Zhang, A review of deep-learning-based medical image segmentation methods, Sustainability 13 (2021) 1224.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, Mask R-CNN, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2020) 386-397.

[12] Hasib Zunair, A. Ben Hamza, Sharp U-Net: Depthwise convolutional network for biomedical image segmentation, Computers in Biology and Medicine 136 (2021) 104699.

[13] Andrik Rampun, Karen López-Linares, Philip J. Morrow, Bryan W. Scotney, Hui Wang, Inmaculada Garcia Ocaña, Grégory Maclair, Reyer Zwiggelaar, Miguel A. González Ballester, Iván Macía, Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network, Medical Image Analysis 57 (2019) 1-17.

[14] Yi Ding, Chao Zhang, Mingsheng Cao, Yilei Wang, Dajiang Chen, Ning Zhang, Zhiguang Qin, ToStaGAN: An end-to-end two-stage generative adversarial network for brain tumor segmentation, Neurocomputing 462 (2021) 141-153.

[15] Qiao Deng, Rongli Zhang, Siyue Li, Jin Hong, Yu-Dong Zhang, Winnie Chiu Wing Chu, Lin Shi, Voting-based contour-aware framework for medical image segmentation, Applied Sciences 13 (2022) 84.

[16] Mingrui Zhuang, Zhonghua Chen, Hongkai Wang, Hong Tang, Jiang He, Bobo Qin, Yuxin Yang, Xiaoxian Jin, Mengzhu Yu, Baitao Jin, Taijing Li, Lauri Kettunen, Efficient contour-based annotation by iterative deep learning for organ segmentation from volumetric medical images, International Journal of Computer Assisted Radiology and Surgery (2022) 1-16, DOI: 10.1007/s11548-022-02730-z.

[17] Xing Huang, Haozhi Zhu, Jiexin Wang, Adoption of snake variable model-based method in segmentation and quantitative calculation of cardiac ultrasound medical images, Journal of Healthcare Engineering (2021) 2425482.

[18] Narasimha Reddy Soora, Ehsan Ur Rahman Mohammed, Sharfuddin Waseem Mohammed, N. C. Santosh Kumar, Deep active contour-based capsule network for medical image segmentation, IETE Journal of Research (2022) 2098184, DOI: 10.1080/03772063.2022.2098184.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2016, 770-778.

[20] Padmavathi Kora, Chui Ping Ooi, Oliver Faust, U. Raghavendra, Anjan Gudigar, Wai Yee Chan, K. Meenakshi, K. Swaraja, Pawel Plawiak, U. Rajendra Acharya, Transfer learning techniques for medical image analysis: A review, Biocybernetics and Biomedical Engineering 42 (2022) 79-107.

[21] Ayat Abedalla, Malak Abdullah, Mahmoud Al-Ayyoub, Elhadj Benkhelifa, Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures, PeerJ Computer Science (2021) e607, DOI: 10.7717/peerj-cs.607.

[22] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2015, 234-241.

[23] Nahian Siddique, Sidike Paheding, Colin P. Elkin, Vijay Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, IEEE Access 9 (2021) 82031-82057.

[24] Jian Liu, Jian Wang, Weiwei Ruan, Chengshan Lin, Daguo Chen, Diagnostic and gradation model of osteoporosis

based on improved deep U-Net network, Journal of Medical Systems 44 (2020) 15.

[25] Zhiwei Song, Hui Yao, Dan Tian, Gaohui Zhan, CSSAM: U-net network for application and segmentation of welding engineering drawings, arXiv preprint arXiv:2209.14102, 2022.

[26] Murat Canayaz, C+EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet, Chaos, Solitons & Fractals 151 (2021) 111310.

[27] Inwan Yoo, David G. C. Hildebrand, Willie F. Tobin, Wei-Chung Allen Lee, Won-Ki Jeong, ssEMnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features, in: Proceedings of the International Workshop on Deep Learning in Medical Image Analysis 2017, 249-257.

[28] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018, 7132-7141.

[29] Hongchun Lu, Shengwei Tian, Long Yu, Lu Liu, Junlong Cheng, Weidong Wu, Xiaojing Kang, Dezhi Zhang, DCACNet: Dual context aggregation and attention-guided cross deconvolution network for medical image segmentation, Computer Methods and Programs in Biomedicine 214 (2022) 106566.

[30] Shui-Hua Wang, Steven Lawrence Fernandes, Ziquan Zhu, Yu-Dong Zhang, AVNC: Attention-based VGG-style network for COVID-19 diagnosis by CBAM, IEEE Sensors Journal 22 (2022) 17431-17438.

[31] Yudong Zhang, Xin Zhang, Weiguo Zhu, ANC: Attention network for COVID-19 explainable diagnosis based on convolutional block attention module, Computer Modeling in Engineering & Sciences 127 (2021) 1037-1058.

[32] Sui Gong, Timothy S. Newman, Fine feature sensitive marching squares, IET Image Processing 11 (2017) 796-802.

[33] Adam Huang, Hon-Man Liu, Chung-Wei Lee, Chung-Yi Yang, Yuk-Ming Tsang, On concise 3-D simple point characterizations: A marching cubes paradigm, IEEE Transactions on Medical Imaging 28 (2009) 43-51.

[34] Ahmed Ghoneim, Ghulam Muhammad, M. Shamim Hossain, Cervical cancer classification using convolutional neural networks and extreme learning machines, Future Generation Computer Systems 102 (2020) 643-649.

[35] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, Hao Li, PIFu: Pixel-aligned

implicit function for high-resolution clothed human digitization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision 2019, 2304-2314.

[36] Jeya Maria Jose Valanarasu, Vishal M. Patel, UNeXt: MLP-based rapid medical image segmentation network, arXiv preprint arXiv:2203.04967, 2022.

[37] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, Yinxiao Li, MAXIM: Multi-axis MLP for image processing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, 5769-5780.

[38] Chang Yao, Menghan Hu, Guangtao Zhai, Xiao-Ping Zhang, TransClaw U-Net: Claw U-Net with transformers for medical image segmentation, arXiv preprint (2021) arXiv:2107.05188.

[39] Xingwen Peng, Xingchen Li, Zhiqiang Gong, Xiaoyu Zhao, Wen Yao, A deep learning method based on partition modeling for reconstructing temperature field, International Journal of Thermal Sciences 182 (2022) 107802.

[40] Yutong Cai, Yong Wang MA-Unet: An improved version of Unet based on multi-scale and attention mechanism for medical image segmentation, in: Proceedings of the International Conference on Electronics and Communication, Network and Computer Technology, 2022, 205-211.

[41] Qingzhen Xu, Zhoutao Wang, Fengyun Wang, Yongyi Gong, Multi-feature fusion CNNs for Drosophila embryo of interest detection, Physica A: Statistical Mechanics and its Applications 531 (2019) 121808.

[42] Jan Jantzen, Jonas Norup, George Dounias, Beth Bjerregaard, Pap-smear benchmark data for pattern classification, in: Proceedings of the Nature inspired Smart Information Systems, 2005, 1-9.

[43] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern, Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC), in: Proceedings of the IEEE International Symposium on Biomedical Imaging 2018, 168-

172.

[44] Mo Zhang, Bin Dong, Quanzheng Li, Deep active contour network for medical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2020, 321-331.

[45] Mo Zhang, Jie Zhao, Xiang Li, Li Zhang, Quanzheng Li, ASCNET: Adaptive-scale convolutional neural networks for multi-scale feature learning, in: Proceedings of the IEEE International Symposium on Biomedical Imaging 2020, 144-148.

[46] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2018) 834-848.

[47] Chaitanya Kaul, Nick Pears, Suresh Manandhar, Divided we stand: A novel residual group attention mechanism for medical image segmentation, arXiv preprint (2019) arXiv:1912.02079.