Neurocomputing 351 (2019) 77–86

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deeply supervised full convolution network for HEp-2 specimen image segmentation

Hai Xie[a], Haijun Lei[b], Yejun He[a], Baiying Lei[c,d,*]

[a] College of Information and Engineering, Shenzhen University, Guangdong Engineering Research Center of Base Station Antennas and Propagation, Shenzhen Key Lab of Antennas and Propagation, Shenzhen, China
[b] School of Computer and Software Engineering, Shenzhen University, Guangdong Province Key Laboratory of Popular High-performance Computers, Shenzhen, China
[c] School of Biomedical Engineering, Health Science Center, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China
[d] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, China

ARTICLE INFO

Article history:
Received 2 November 2018
Revised 28 February 2019
Accepted 10 March 2019
Available online 18 April 2019

Communicated by Dr. Shen Jianbing Shen

Keywords:
HEp-2 specimen images
Deeply supervised full convolutional network
Hierarchical supervision
Dense deconvolution layer

ABSTRACT

Human Epithelial-2 (HEp-2) cell images play an important role for the detection of antinuclear autoantibodies (ANA) in autoimmune diseases. Segmentation is the primary step for classification, further treatment and diagnosis. However, the staining patterns and scales of HEp-2 specimen images have different variances, which still make segmentation quite a challenging task. To solve it, we propose a novel deeply supervised full convolutional network (DSFCN) for robust segmentation of different HEp-2 cell images dataset. DSFCN is based on a very deep network, which integrates the dense deconvolution layer (DDL) and hierarchical supervision structure (HS). Specifically, The DDL uses the up-sampling to restore the high resolution of the original input images to replace the traditional deconvolution layer, and the hierarchical supervision is added to capture feature information of the shallow layers. The high-resolution predictive output is obtained by capturing local and global information between layers. Without relying on the prior knowledge and complex post-processing, DSFCN can be effectively trained in an end-to-end manner. The proposed method is trained and tested on the I3A-2014 public dataset, and the segmentation result demonstrates that the performance of our model outperforms other state-of-the-art methods.

1. Introduction

HEp-2 cells with indirect immunofluorescence (IIF) is a commonly-used technique for detecting anti-nuclear anti-bodies (ANA), which can be visualized via a fluorescence microscope. Segmenting HEp-2 specimen images is indispensable due to its importance in daily clinical practice to improve the efficiency of computer-aided diagnosis and detection. However, manual analysis from a large number of IIF images still has the limitations (e.g., high clinical experiences, time-consuming and inter-variability among doctors' knowledge). As a result, the subjective results and inter-laboratories diversity restrict the true expression of the reading results [1]. To address these limitations, a number of automatic and robust HEp-2 cell classification models have been proposed in recent years [2–4]. In these methods, segmentation is the first step for HEp-2 cell images classification since the accurate segmentation results are beneficial for the subsequent classification processing [3,4].

Intensity thresholding is one of the most popular and preliminary approaches for cell segmentation. Petra et al. proposed a HEp-2 cell image segmentation method using Otsu by utilizing the first thresholding [5]. Jiang et al. proposed a novel approach based on the framework of verification-based multi-threshold probing for HEp-2 cell image segmentation [6]. Many studies aimed at the segmentation of HEp-2 cells [7,8] due to the large variances of appearances among different HEp-2 cell categories. However, most of the previous works achieved accurate segmentation for images containing a certain pattern of cells while failing to achieve good results when different staining patterns were provided. Examples of the staining patterns of HEp-2 specimen images are illustrated in Fig. 1. However, there is still room for robustness improvement of the HEp-2 specimen images segmentation method.

To improve the segmentation performance of HEp-2 specimen images, a method with an impressive feature is highly desirable. In recent years, the deep convolutional neural networks (CNNs) have attracted wide attention due to their impressive performance

* Corresponding author.
E-mail address: leiby@szu.edu.cn (B. Lei).

https://doi.org/10.1016/j.neucom.2019.03.067
0925-2312/

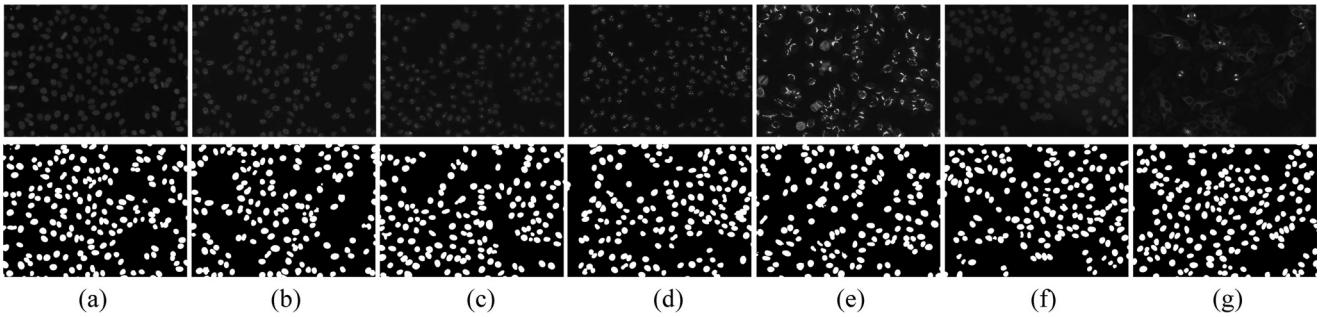

Fig. 1. An example of different HEp-2 specimen images staining patterns, the first row represents the raw images and the second row indicates the corresponding segmentation masks. (a)–(g) represents Homogeneous, Speckled, Nucleolar, Centromere, Golgi, Nuclear membrane, and Miotic spindle, respectively.

in various image processing tasks [9–11]. The fully convolutional network (FCN) extends the traditional CNN, which is one of the most representative models [12,13] for segmentation. The main idea under FCN model is to apply the classification networks (AlexNet [14], VGG net [15], GoogLeNet [16], and ResNet [17]) to the segmentation task by transforming the last classifier layers to the deconvolutional layers. In fact, the deeper level of network layer information and fusion can further improve the segmentation performance. Hence, the full convolution residual network (FCRN) with a deeper residual network (ResNet) was proposed [18]. However, when the fully convolutional connection is applied to the fully convolutional layer in the deep network, the resolution of the feature map of the output layer will be reduced. This results in loss of information which is highly undesirable for the segmentation of medical images.

In order to tackle this problem, a lightweight neural network called U-Net was proposed [19]. The U-Net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. However, the network is inefficient to capture the edge information for some HEp-2 specimen patterns. To better express the image information, Isola et al. explored generative adversarial networks (GANs) in the conditional setting and proposed pix2pix network framework based on U-Net [20]. This architecture makes conditional GANs suitable for image-to-image translation tasks, where an input image is fed into the network and a corresponding output image is generated. With the adversarial learning, the network can learn rich edge information. Nevertheless, the feature information is easily decreased in the processing of skip connection. To overcome this limitation, the Dense Deconvolutional Layer (DDL) structure is fetched in this paper. This idea has been proposed in recent years and achieved considerable segmentation performance [21–25].

The DDL consists of a series of skip connection layers between the previous and later layer instead of performing a summation operation. This architecture also resolves the gradient vanishing problem effectively. Inspired by the previous works, we propose a novel end-to-end Deep Supervised Fully Convolutional Network (DSFCN), which utilizes DDLs without requiring prior knowledge and post-processing. The improved loss functions are introduced in the two lateral output layers to optimize the output feature maps so that the hierarchical supervision (HS) depth is fully exploited. Our proposed DSFCN framework is able to learn rich hierarchical features and captures the local and global contextual information effectively. Due to the perfect performance, the proposed approach can be regarded as a general technology for image segmentation. In summary, the main contributions of this paper are three-fold:

- We propose a novel end-to-end deep learning framework based on FCN. Due to the DDL structure, the network can learn rich boundary information for HEp-2 specimen images.
- An improved HS mechanism is added to the network, which can optimize the output feature maps.
- Experimental results demonstrate that the proposed method achieves the state-of-the-art segmentation performance on the I3A-2014 dataset.

The rest of this paper is organized as follows. The related work is presented in Section 2. Section 3 introduces the proposed network framework in detail. The experiment settings and comparison results are illustrated in Sections 4 and 5. Sections 6 and 7 are dedicated to discussions and conclusions, respectively.

2. Related work

2.1. Image segmentation

We all know that image segmentation is usually the basic research for other visual tasks, such as visual tracking [26,27], classification [28–34], detection [35–37] and cropping [38]. Recently, there are many outstanding image segmentation methods [39–41]. For example, Shen et al. proposed a novel method to optimize the higher-order energy with appearance entropy by transforming a higher-order energy function to a lower-order one at a local region, which is used to solve the image segmentation problem [42]. Based on the sub-Markov random walk, Dong et al. proposed a novel framework for interactive seeded image segmentation [43]. In addition, Shen et al. presented a new image superpixel segmentation approach by using the density-based spatial clustering of applications with noise [44]. In terms of medical image segmentation, Jia et al. proposed an automated coarse-to-fine segmentation method by utilizing a probabilistic atlas constructed for each scan and a cohort of trained CNNs for prostate MR studies [45].

2.2. Full convolutional networks

Usually, it is necessary for image segmentation [12] and image generation [46] algorithms to make the prediction of the size and space for the original pictures. However, the stride of convolution decreases the size of the input image. As a result, the deconvolution acts as an up-sampling role. In 2015, Long et al. proposed an FCN for semantic segmentation [12]. This framework based on VGG architecture leverages deconvolution layers instead of softmax layer, which outputs the predicted images the same size as the original images. Based on ResNet, Wu et al. proposed a new network architecture called FCRN for semantic segmentation [18], which achieved state-of-the-art segmentation performance. Liu et al. proposed a collaborative deconvolutional neural network (C-DCNN) to exploit the semantic and geometric properties of images for image segmentation [47]. In addition, the full convolution networks (FCN) have also been frequently used in the medical image segmentation tasks. Ronneberger et al. proposed a symmetric network structure called U-Net for medical images segmentation [19]. Based on U-Net, the pix2pix with adversarial learning mechanism

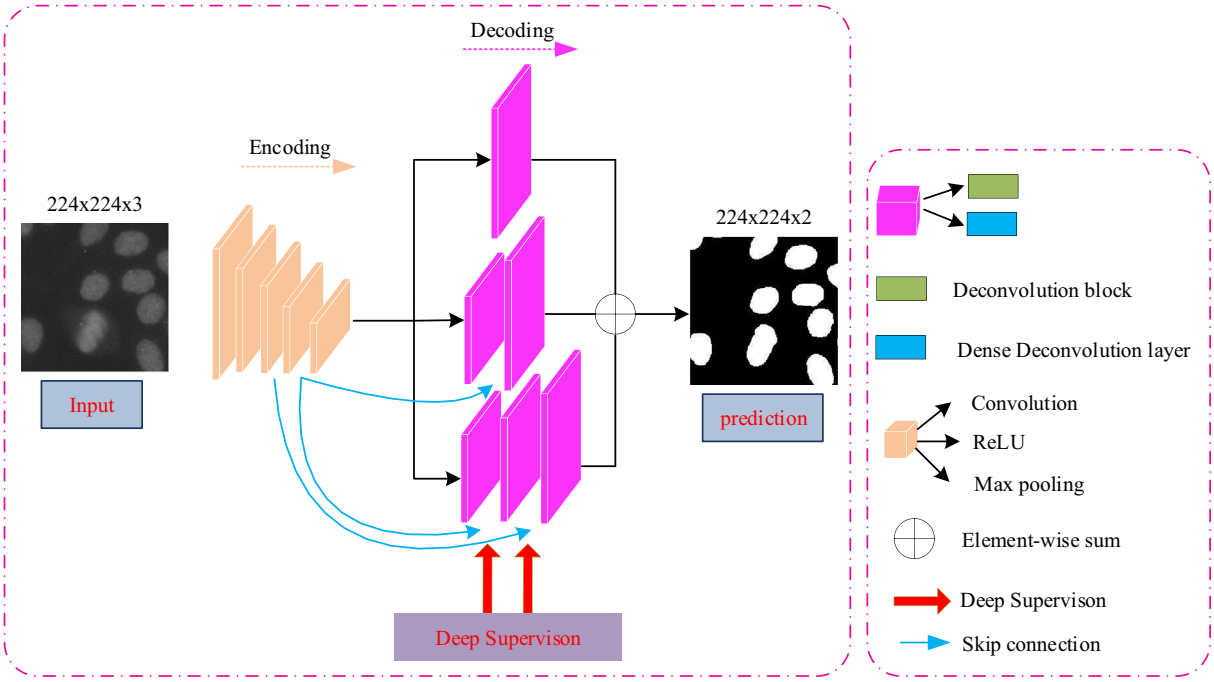

Fig. 2. The architecture of our proposed method. The long-range skip connections are used to incorporate multi-level features between encoding and decoding module to void gradient vanishing. We use DDL to refine and fuse different layer feature graphs to obtain a high resolution. The HS scheme is added into the last two output layers of a three-layer deconvolution operation to refine the output results.

was proposed. Bi et al. proposed a new semi-automated skin lesion segmentation method that incorporates FCNs with multi-scale integration [48]. He et al. proposed a novel skin lesion segmentation network via a very deep dense deconvolution network using dermoscopic images [49].

2.3. Hierarchical supervision mechanism

Deep supervision mechanism is often used to improve the performance of learning tasks especially segmentation task. For example, Wang and Shen proposed a deep model that is trained in a deep supervision way, where the supervision mechanism is fed into the mutil-level layers to provide mutil-level saliency information [50]. Fan et al. proposed a deep learning approach for image registration by utilizing the difference between images as additional information to supervise the training [51]. In addition, the architecture employs hierarchical loss layers in the up-sampling path of U-Net so that the proposed network can be more constraint and convergent. To alleviate the destruction of some correlations within image regions, Wang et al. connected a classification layer in each deconvolutional layer that is ahead of up-pooling layer and supervised it with pixel-wise ground-truth [52]. With the hierarchical predictions, the network can use cooperatively unpooling and bilinear interpolation for resolution recovery.

3. Methodology

Our proposed deep supervised full convolution network (DSFCN) consists of the adaptive convolution unit, DDL, skip connection unit and HS. The architecture of our proposed model is illustrated in Fig. 2. Similar to FCN, the adaptive convolution unit in DSFCN is used to adjust the weight parameters. We use DDL to optimize and fuse the feature maps to generate a higher resolution image. Thus, DSFCN can sharpen object boundaries in an end-to-end way. Here, we adopt DDL instead of the original multi-layer fusion since DDL not only restores the size of the original input pixel, but also can effectively obtain the global and contextual information. The captured global features can effectively identify the whole image, which helps to correctly classify the pixels in the region of interest. Since there is no direct relationship between the adjacent pixels of the generated output feature maps, DSFCN solves this issue.

In the DSFCN framework, we integrate different resolution feature maps extracted in the down-sampling process through the depth of VGG-16. The low-level boundary information is generated using the pre-trained VGG-16 model, while the advanced semantic information is obtained by the segmentation network. The segmentation network restores the size of the feature mappings, then reconstructs the spatial dimension information, and finally obtains the fine structure of objects. The thinning network combines the characteristics of the low layer boundary with the advanced semantic information.

3.1. Dense deconvolution layer

In dense CNN, each layer is directly connected to all the other layers in a feed-forward fashion. In the original FCN-8s [12], the multi-resolution features are fused by the summation function, which may lead to the loss of the boundary information resulting in an unsatisfactory result. In the present study, we tackle this problem by adding a skip connection layer between the previous and later layer instead of performing a summation. In addition, the direct link between the intermediate feature mappings increases the dependencies, which leads to faster convergence for the training process and also improves the information flow [24].

The proposed DDL architecture is presented in Fig. 3 and explained in details as follows. Let $G_1, G_2, \ldots, G_{l-1}$ indicate the connection of the generated feature graphs in the middle layers, G_l is defined as follows

$$G_l = C([G_1, G_2, \ldots, G_{l-1}]). \quad (1)$$

where $C(\cdot)$ indicates the convolution operation. The dense connection reduces the degree of gradient vanishing, enhances the

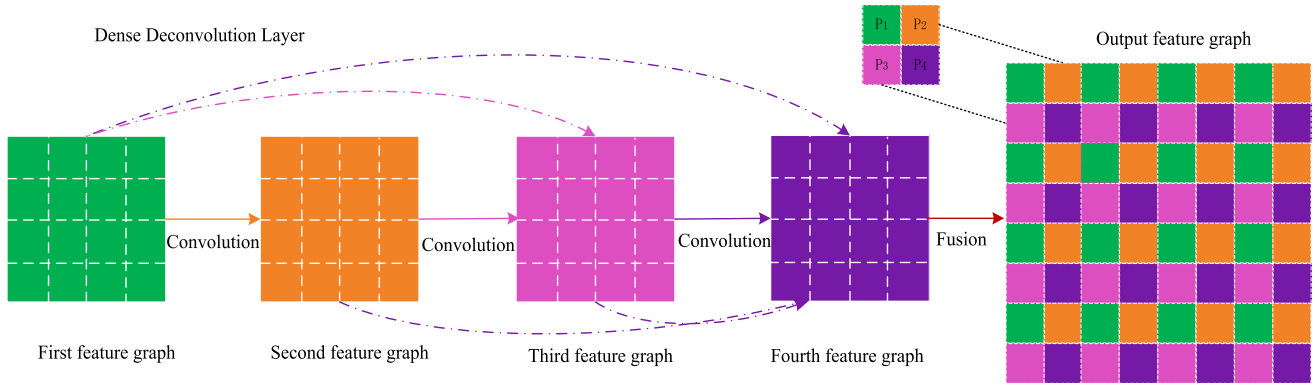

Fig. 3. The architecture of the dense deconvolution layer of the proposed method. The four intermediate feature maps are generated by different convolution operations and are fused in the stage of decoding to enrich the information of the final feature map so that a higher resolution graph is obtained.

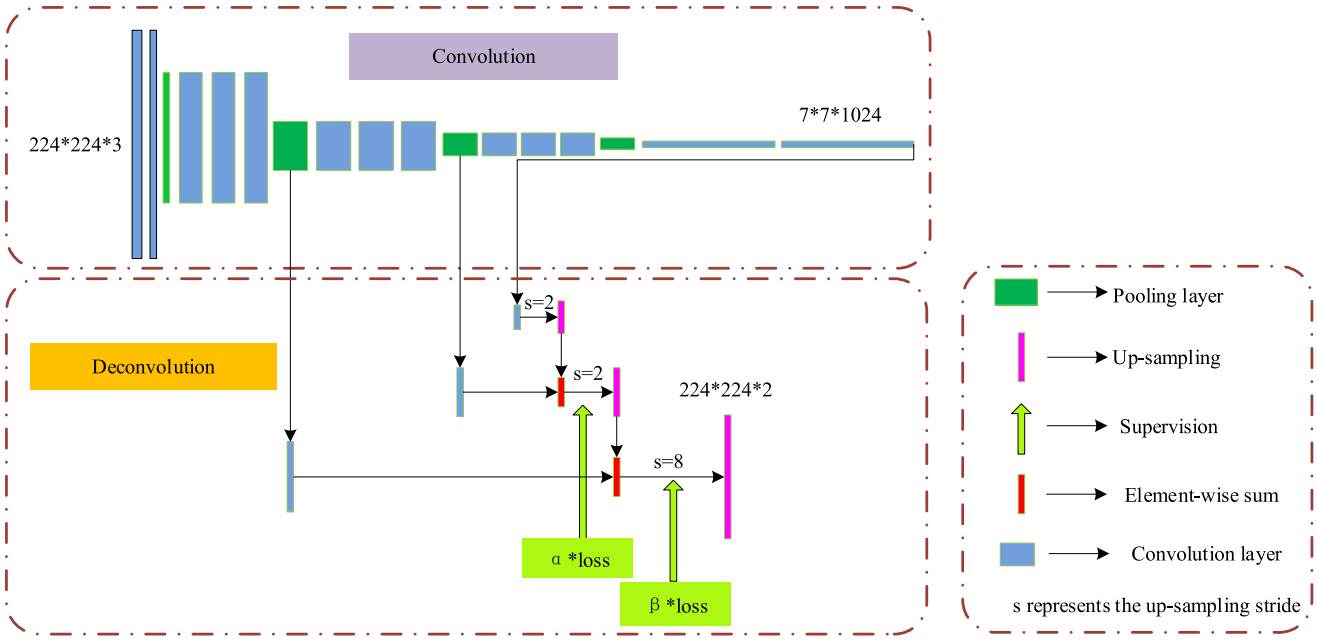

Fig. 4. Detail illustration of the proposed hierarchical supervision mechanism. The HS mechanism is added to the last two output.

gradient propagation, and reuses features more effectively. To extract the final feature map, we merge the four intermediate feature maps together. Supposing that (x', y') is the location of the corresponding pixel, s is stride, which is set as 2 in our study, the pixel value of the final feature graph is calculated as follows

$$P\left(x', y'\right) = P_{m,n}(x * s + m,\ y * s + n), \tag{2}$$

$$m = x' \bmod s,\ n = y' \bmod s. \tag{3}$$

DDL is vital in the up-sampling operation as it helps to restore the high resolution of the original input images. In the decoding phase, it is useful to restore the detailed low-layer features generated by the encoding module. Hence, we use DDL to refine and fuse different layer feature graphs to obtain a higher resolution.

3.2. Hierarchical supervision layer

We propose a deep-level supervision mechanism to restore the characteristics of the shallow layer. Inspired by Lee et al. [53], the supervision mechanism is added to the output of the two convolution layers, as shown in Fig. 4. Furthermore, we try to add deep-level supervision mechanism to the output of a three-layer deconvolution operation to refine edge information. Since there is no direct relationship between the predictive output and the real label, the performance of the output layer is relatively low. To tackle this problem, we conduct deep supervision in the last two output layers. The loss functions of the branch networks can be expressed as

$$L\mathcal{B},\ \mathcal{W} = \sum_{s} w_s L_s(\mathcal{B},\ \mathcal{W}) + w_m L_m(\mathcal{B},\ \mathcal{W}), \tag{4}$$

$$L_s(\mathcal{B},\ \mathcal{W}) = -log\left(p_k\left(x_{i,j},\ t_{i,j}\right)\right), \tag{5}$$

$$L_m(\mathcal{B},\ \mathcal{W}) = \frac{2\sum_i^N \sum_j^M x_{i,j} t_{i,j}}{\sum_i^N \sum_j^M x_{i,j}^2 + \sum_i^N \sum_j^M t_{i,j}^2} \tag{6}$$

where $L(\mathcal{B},\ \mathcal{W})$ is the loss function of the output of the branch networks; $L_s(\mathcal{B},\ \mathcal{W})$ is the loss function of the output value and the real value of the main network; $L_m(\mathcal{B},\ \mathcal{W})$ represents the loss layer

of every pixel categorization. The w_s and w_m are the weights of the network, and $p_k(x_{i,j}, t_{i,j})$ indicates the probability value of the final output layers of the proposed network. In addition, the loss functions in different layers are multiplied by different weights.

As shown in Fig. 4, the deep supervision scheme is added into the last two output layers of the proposed network. In addition, the resolution of the feature maps generated by these two layers is different, which inspires us to set different weights for the loss function. In our experiments, the weight α and β are set to 0.5 and 0.3, respectively.

4. Experiment settings

4.1. Database

Our experiments are evaluated on the public dataset I3A-2014. The I3A dataset was first released in the fluorescent image based cell classification contest organized by ICIP 2013 [54], and later used in the contest organized by ICPR 2014. The dataset records 252 specimens from seven categories: Homogeneous (53), Speckled (52), Nucleolar (50), Centromere (51), Golgi (10), Nuclear membrane (21), and Mitotic spindle (15). The number in brackets indicates the number of specimen samples for the corresponding type of cells. For each specimen, four images were captured in different locations with a size of 1388×1040. In total, 1008 grayscale specimen images of the I3A dataset were used in the present study. Our experiments are conducted on MatConvNet toolbox written in MATLAB R2017a using a computer with CPU Intel Xeon E5-2680 @ 2.70 GHz, GPU NVIDIA Quadro K4000, and 128 G of RAM. The stochastic gradient descend (mini-batch size=20, weight decay=0.0001, momentum=0.9) is used to optimize the target function. We observe that the training process starts to converge after 10 epochs.

4.2. Data augmentation

Although the I3A-2014 dataset comprises of 1008 specimen images, which are still insufficient to train a deep learning network model as smaller dataset can easily lead to overfitting issue. Therefore, we conduct data augmentation to enrich our dataset by using mirroring (M), cropping (C), and rotation (R) operations. We use a combination of data enhancement techniques in our experiment namely: (1) Cropping (C): each specimen image from the original dataset is randomly cropped into 30 pieces of 224×224, resulting in a total of 30,240 generated images. (2) Crop + Mirror (CM): Based on the dataset generated by (1), each image is generated by mirroring operation, resulting in a dataset of 60,480 images. (3) Crop + Mirror + Rotate (CMR): Each image in (2) was rotated at four different angles i.e.0°, 90°, 180° and 270°. As a result, we obtain a total of 241,920 augmented images.

4.3. Evaluation metrics

In this paper, we utilize the most commonly used evaluation criteria to assess our segmentation model, which consists of segmentation accuracy (SEG), sensitivity (SE), Jaccard index (JA) and accuracy (AC). The SEG is a similarity metric obtained by comparing the prediction results of our model and the ground truth. The JA measures the overlap between the predicted results and ground truth and is expressed as their intersection over union. The metrics for evaluating segmentation results are denoted as

$$SEG = \frac{2 \times precision \times Recall}{precision + Recall} \tag{7}$$

$$precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \tag{8}$$

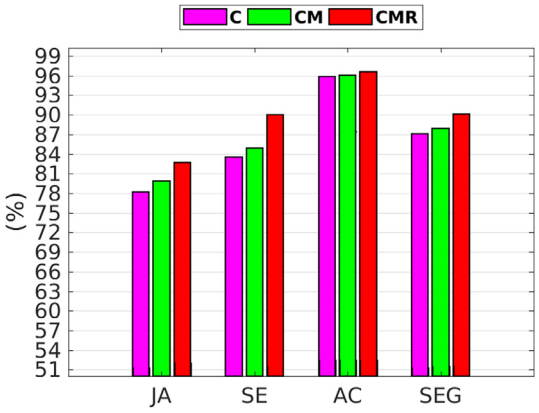

Fig. 5. The comparison results of different data augmented methods.

Table 1
Segmentation results of different data augmentation methods with HS and without HS.

No.	Method	Number of instances	SE (%)	JA (%)	AC (%)	SEG (%)
1	C	30,240	83.55	78.09	95.72	87.06
2	CM	60,480	85.73	79.07	95.93	87.69
3	CMR	**241,920**	89.06	81.64	96.35	89.35
4	C+HS	30,240	83.55	78.20	95.84	87.13
5	CM+HS	60,480	84.85	79.84	96.02	87.93
6	CMR+HS	**241,920**	**89.96**	**82.68**	**96.56**	**90.10**

$$JA = \frac{TP}{TP + TN + FP} \tag{9}$$

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$SE = \frac{TN}{FP + TN} \tag{11}$$

where TP, TN, FP, and FN present the number of true positive, true negative, false positive, and false negative, respectively.

5. Experiment results

5.1. Comparison results of different augmented datasets

Due to the limitation of the number of instances in I3A-2014 dataset for a deep network, different data augmentation strategies are made to avoid over-fitting in the training process. In addition, the augmented dataset is beneficial for improving the segmentation performance. The proposed method is evaluated on different augmented datasets that are described in Section 3.1, respectively. Fig. 5 shows that the segmentation performance on I3A-2014 dataset for different data augmented methods. In addition, Table 1 summarizes the segmentation results of different augmentation strategies. It can be observed that the larger dataset can result in a better performance. The reason is that the dataset with mass images is fed into the deep network, which can provide rich information for the proposed network, even some edge information can also be learned.

5.2. Comparison results with and without HS

To demonstrate the effectiveness of the proposed HS mechanism in improving segmentation accuracy, we also conduct the experiments with and without HS. The segmentation results are also listed in Table 1. From Nos. 3 and 6, we can see that the SEG

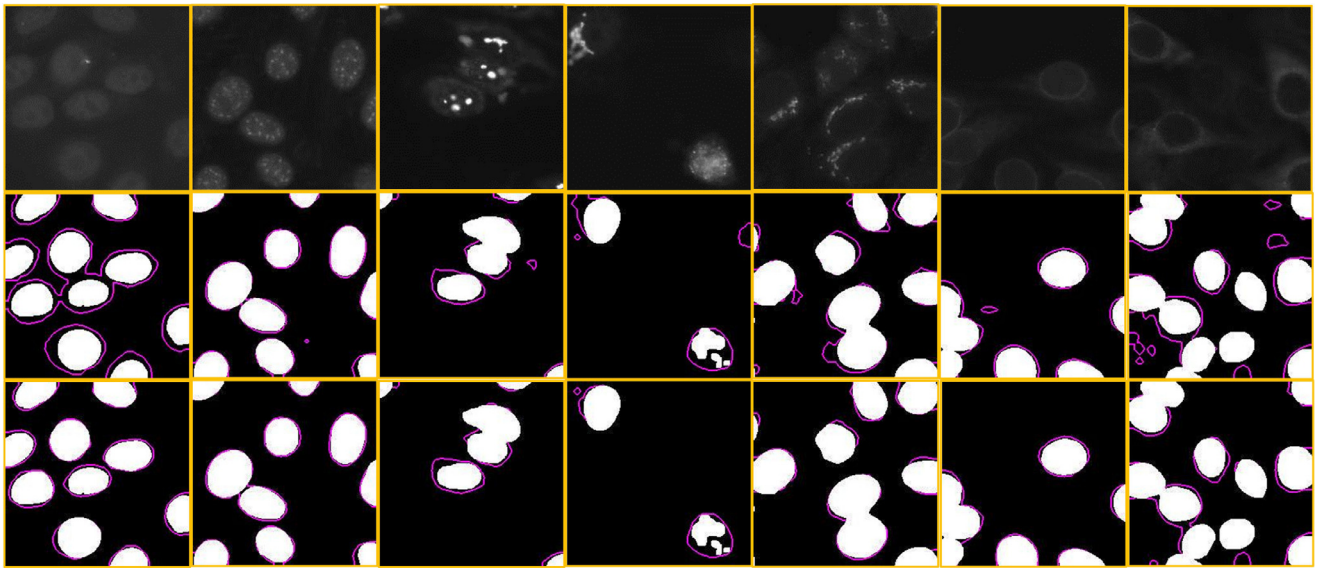

Fig. 6. The comparisons of our method with and without HS. From the top line to the bottom line: the original images, the results without HS and the results with HS. From left to right: Homogeneous, Speckled, Nucleolar, Centromere, Golgi, Nuclear membrane, and Mitotic spindle.

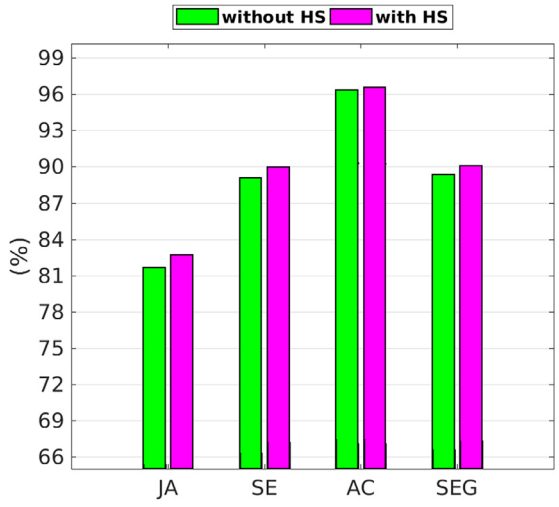

Fig. 7. Segmentation results with and without HS on the augmented I3A-2014 dataset.

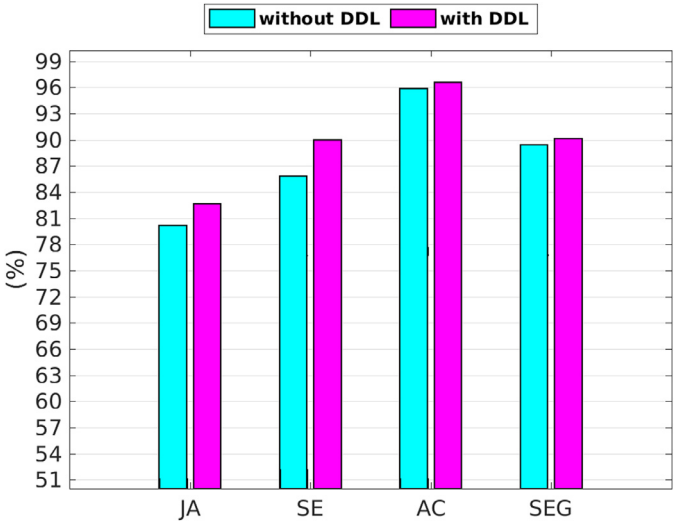

Fig. 8. Segmentation results with and without DDL on the augmented I3A-2014 dataset.

with HS is 0.75% higher than SEG without HS and JA with HS is 1.04% higher than JA without HS. The results demonstrate the effectiveness of our model with HS. In addition, Fig. 6 also provides more intuitive results to demonstrate the effectiveness of our method with HS, in which the first row indicates the original images, the second row and the third row represent results without HS and results with HS respectively (the purple lines indicate the boundary of segmentation results and the white areas represent ground-truth.).

From the results in Fig. 6, we can see that the segmentation results with HS are closer to the ground-truth regions. For those relative complex staining patterns (e.g., Golgi, Nuclear membrane, and Mitotic spindle), the incorrect segmentation regions with HS are smaller than that without HS. The main reason is that the HS mechanism can learn rich hierarchical features and refine coarse output prediction. With HS, the multi-level and multi-scale features can be learned. As a result, the network can integrate the detailed information of the edge corner of each image. Therefore, HS has more discriminative features than the network without HS. In fact, when the layers become deeper, the size and resolution of the feature maps become smaller, and the receptive field becomes larger. As a result, the global contextual information can be captured as well. HS not only optimizes the detailed edges, but also boosts the segmentation performance globally.

We also conduct experiments with and without HS, and the segmentation results are shown in Fig. 7. It can be seen that our proposed method improves the performance in terms of JA, SE, AC, and SEG, respectively. This experiment shows that the network with HS is better than the approach without HS on the augmented I3A-2014 dataset, which improves the effectiveness of our proposed HS structure.

5.3. Comparison results with and without DDL

To prove the effectiveness of the DDL structure, we carry out some experiments with and without DDL. The experimental results are shown in Fig. 8. It can be seen that the SEG with DDL is higher than that without DDL in terms of JA, SE, AC, and SEG,

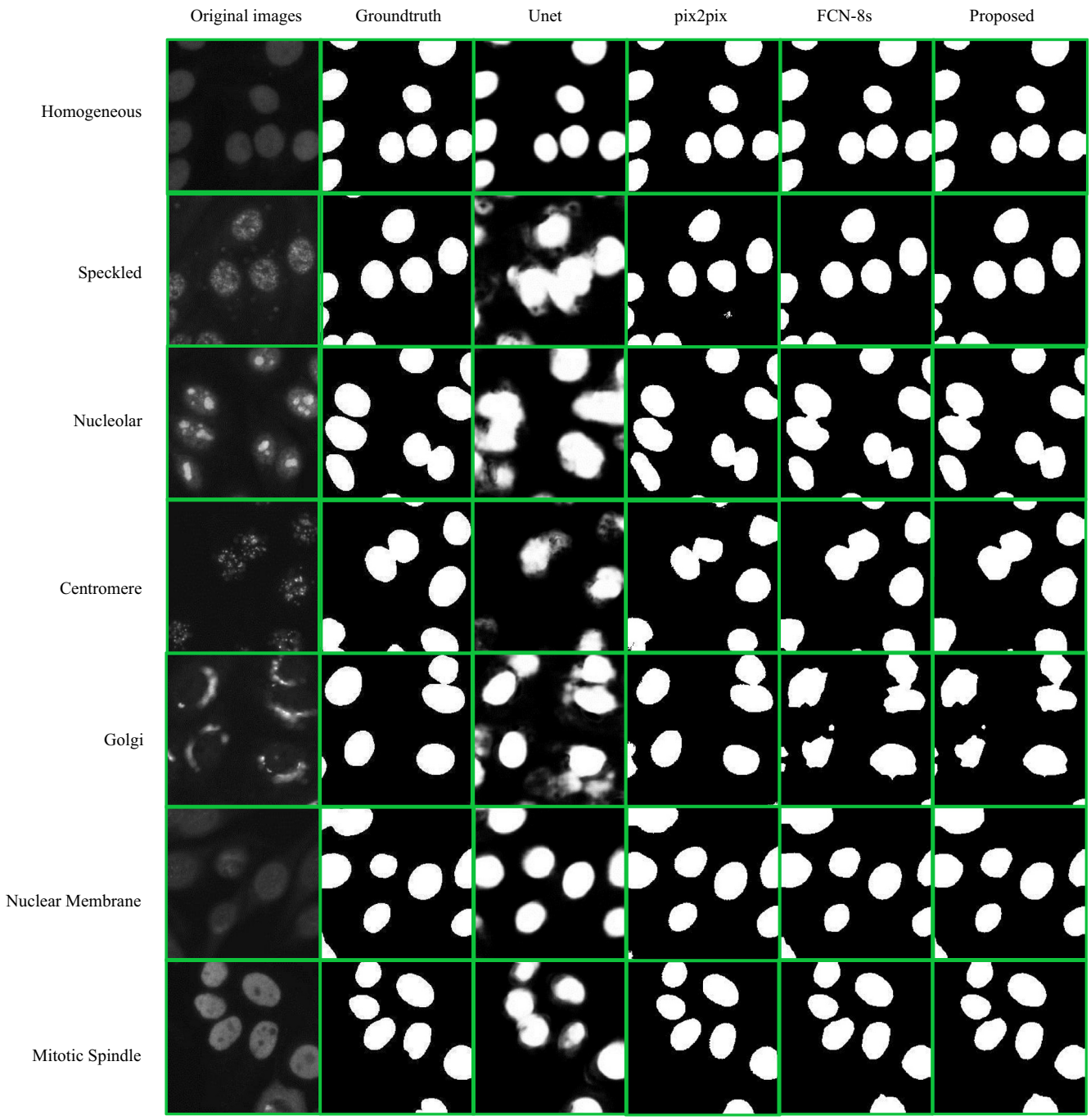

Fig. 9. Segmentation results on augmented I3A-2014 dataset. The first column shows examples from each cell category. The corresponding ground truths are presented in the second column. The following columns present the results from different frameworks.

respectively. The reason is that the dense connection in dense deconvolutional layers can help the model to acquire rich information of shallow layers. As a result, the boundary information can be better extracted by our proposed model, which shows that the DDL architecture can improve the semantic segmentation performance.

5.4. Comparison results with different staining patterns

It is known from the clinical practice that the images in the IIF dataset refer to seven different classes of specimen level staining pattern: Homogeneous, Speckled, Nucleolar, Centromere, Golgi, Nuclear Membrane, and Mitotic Spindle. The different staining patterns of different HEp-2 specimen images result in great interclass differences. As shown in Fig. 1, we can see that cells of the first class appear as elliptical compact regions with a bright core and a sharp contour, while those belonging to the last class present a dark core enclosed by a very hazy contour. This affects the performance of the foreground detection approach significantly. Table 2 shows the average performance of JA, SE, AC, and SEG over all the images of different specimens, respectively.

As shown in Table 2, we can observe that the best performance of the proposed architecture is obtained in the first class. This is close to our expectation, since the homogeneous class is the staining pattern, which shows the lowest intensity variations in the cell body.

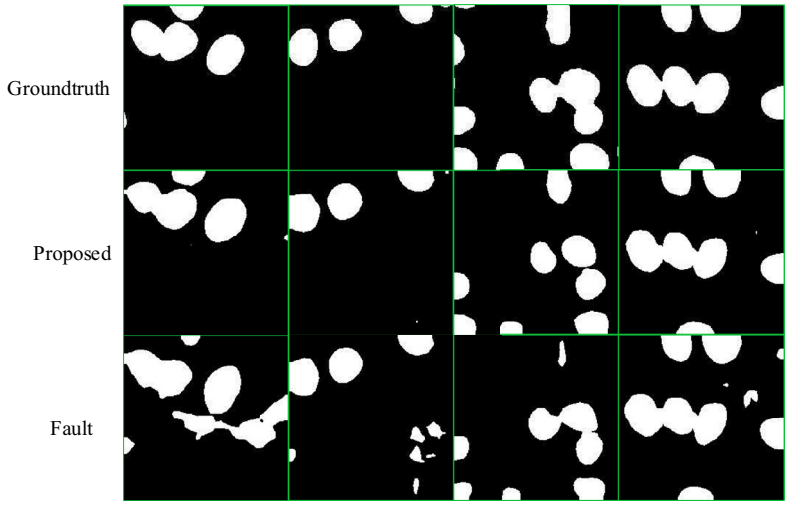

Fig. 10. Comparison results from the proposed and fault network architecture.

Table 2
The performance of the proposed method for different staining patterns of the I3A-2014 dataset.

Staining pattern	JA (%)	SE (%)	AC (%)	SEG (%)
Homogeneous	**89.10**	**95.04**	**97.93**	**94.15**
Speckled	86.46	93.33	97.57	92.68
Nucleolar	79.92	88.27	96.12	88.63
Centromere	84.46	91.16	97.17	91.51
Golgi	59.15	69.36	90.72	73.67
Nuclear membrane	81.40	90.16	95.82	89.37
Mitotic spindle	67.46	75.30	92.47	78.70

Table 3
Results of different segmentation algorithms on the augmented I3A-2014 dataset.

Method	JA (%)	SE (%)	AC (%)	SEG (%)
U-Net [19]	62.10	68.33	92.30	74.80
Pix2pix [20]	75.94	81.96	95.38	85.85
FCN-8s [12]	81.63	88.97	96.35	89.38
DSFCN (ours)	**82.68**	**89.96**	**96.56**	**90.10**

5.5. Comparison with other methods

In order to further demonstrate the effectiveness of our DSFCN method, we compare the proposed method with different segmentation methods based on the I3A dataset. Table 3 summarizes the segmentation results from different approaches. It can be observed that the accuracy achieved from our method is 0.72% higher than the best method among comparative methods (e.g. FCN-8s [12]) in terms of SEG for the augmented I3A-2014 dataset. The reason is that hierarchical supervised neural network accelerates the optimization speed which means the gradient is easy to propagate back to the previous layers from the later layer.

Fig. 9 presents segmentation results of different deep learning frameworks for each cell category of the augmented I3A-2014 dataset. It can also be seen that the segmentation result from the first class HEp-2 specimen images (e.g., Homogeneous) is better than other classes. In addition, the segmentation results of each method in this staining pattern are almost the same (e.g., the homogeneous class has the lowest intensity variations in the cell body), which verifies our expectation. It can also be observed that lots of false segmentation regions were generated in U-Net for the segmentation result of the Speckled class. By comparing the segmentation result of Centromere class, we can observe that the segmentation performance of our proposed method is slightly better than FCN-8s.

6. Discussions

As described in the above sections, we present an automated deep supervised full convolution network for HEp-2 specimen images segmentation in an end-to-end way. In the following, we will discuss the effect of the dataset size and network architecture.

6.1. The effect of the dataset size

The I3A-2014 dataset has only 1008 HEp-2 specimen images and the data sample sizes are small. Hence, a deep framework may be unsuitable for this task. For this reason, we conduct several data augmentation experiments. Thus, the number of different staining patterns is balanced and the augmented dataset also results in a better segmentation performance. From Table 1, we can see that only crop, crop & mirror, and crop & mirror & rotate achieve the segmentation accuracy of 87.13%, 87.93%, 90.10%, respectively. We can draw a conclusion that the larger dataset can obtain better segmentation accuracy. The main reason is that a large amount of image data increases the readability of the input image in terms of position and direction for the proposed network framework.

6.2. The effect of network depth

Our proposed DSFCN architecture is based on VGG-16 model, which is pre-trained on the PASCAL VOC 2012 dataset. On the basis of this framework, we leverage DDL that consists of a series of skip connection, which tackles the problem of gradient vanishing and makes the convergence easier. We observed that the network converges almost at the tenth epoch in the training stage. In addition, the layer of HS is also added to the proposed network architecture, which is useful for improving the segmentation performance on I3A-2014 dataset.

However, the deep network does not always perform well. For example, in order to enhance the representations and utilization of features, we attempt to add the residual connection in our framework. Nevertheless, the deeper structure causes the network overfitting so that the object cannot be segmented from the background exactly. As shown in Fig. 10, many segmentation results of the modified deeper network are false. By contrast, our proposed framework is quite closer to the ground-truth, which demonstrates the effectiveness of our proposed method as well.

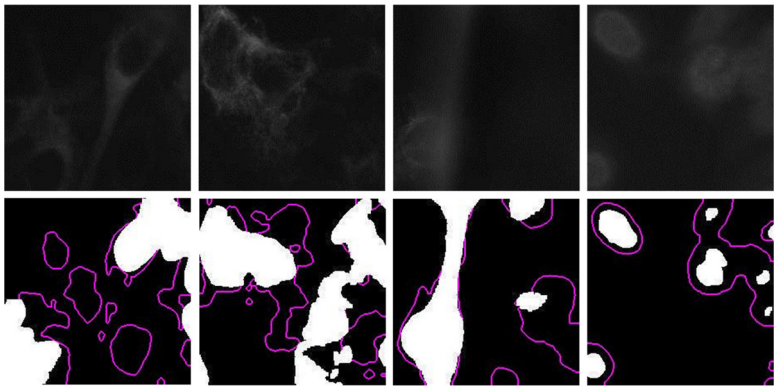

Fig. 11. Failure example of our proposed model. The first row represents the original images and the second row indicates the segmentation results of our method, in which the white areas represent ground-truth and the purple lines indicate the boundary of segmentation results.

6.3. The failure cases

Although our proposed approach achieves the state-of-the-art performance, there are some limitations in our proposed method. The main limitation of our approach is that the distribution and illumination of the dataset are uneven and insufficient. As a result, our proposed model cannot extract discriminative features for object regions, which leads to some failing examples occur, as shown in Fig. 11. It can be seen that these failure instances mainly happen in these images with uneven illumination and blurry boundaries.

7. Conclusion

In this paper, we propose an automated DSFCN framework for HEp-2 specimen images segmentation, which is able to tackle the problem of localization for classification. The proposed model includes DDL and HS mechanism. DSFCN is able to learn discriminative feature representation and effective integration of multi-level contextual information. Obviously, our method can automatically and accurately segment the region of interest. DSFCN can build a feature connection on DDL to learn the characteristics of the shallow network and reuse it. By adding the hierarchical supervision to solve the gradient vanishing problem and enhance the propagation of multi-level features in the whole network, we improve the segmentation performance of HEp-2 specimen images.

Acknowledgments

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, and 61801305), Guangdong Province Key Laboratory of Popular High Performance Computers (No. 2017B030314073), Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (No. KQTD2016053112051497), Shenzhen Key Basic Research Project (Nos. JCYJ20170302153337765 and JCYJ20170818142347251), NTUT-SZU Joint Research Program (No. 2018006), and Open Fund Project of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (No. MJUKF201711).

References

[1] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking HEp-2 cells classification methods, IEEE Trans. Med. Imaging 32 (10) (2013) 1878–1889.
[2] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, B. Lei, A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning, Pattern Recognit. 79 (2018) 290–302.
[3] Y. Li, L. Shen, S. Yu, HEp-2 specimen image segmentation and classification using very deep fully convolutional network, IEEE Trans. Med. Imaging 36 (2017) 1561–1572.
[4] Y. Li, L. Shen, cC-GAN: a robust transfer-learning framework for HEp-2 specimen image segmentation, IEEE Access 6 (2018) 14048–14058.
[5] P. Petra, P. Horst, M. Bernd, Mining knowledge for HEp-2 cell image classification, Artif. Intell. Med. 26 (1) (2002) 161–173.
[6] X. Jiang, G. Percannella, M. Vento, A verification-based multithreshold probing approach to HEp-2 cell segmentation, in: Proceedings of the Computer Analysis of Images and Patterns, 2015, pp. 266–276.
[7] S. Di Cataldo, S. Tonti, A. Bottino, E. Ficarra, ANAlyte: a modular image analysis tool for ANA testing with indirect immunofluorescence, Comput. Methods Prog. Biomed. 128 (2016) 86–99.
[8] M. Merone, P. Soda, On using active contour to segment HEp-2 cells, in: Proceedings of the International Symposium on Computer-Based Medical Systems, 2016, pp. 118–123.
[9] P. Tobias, H. Alexander, M. Markus, L. Bastian, Full-resolution residual networks for semantic segmentation in street scenes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3309–3318.
[10] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, A. Agrawal, Context encoding for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7151–7160.
[11] Z. Yu, E.-L. Tan, D. Ni, J. Qin, S. Chen, S. Li, B. Lei, T. Wang, A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition, IEEE J. Biomed. Health Inform 22 (2018) 874–885.
[12] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[13] D. Nie, L. Wang, Y. Gao, D. Shen, Fully convolutional networks for multi-modality isointense infant brain image segmentation, in: Proceedings of the IEEE International Symposium on Biomedical Imaging, 2017, pp. 1342–1345.
[14] I.S.A. Krizhevsky, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.
[15] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556v6 (2015).
[16] W.L.C. Szegedy, Y. Jia, P. Sermanet, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[18] Z. Wu, C. Shen, A.v.d. Hengel, High-performance Semantic Segmentation Using Very Deep Fully Convolutional Networks, arXiv:1604.04339 (2016).
[19] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, 2015, pp. 234–241.
[20] P. Isola, J.Y. Zhu, T. Zhou, A. Alexei, Image-to-Image Translation with Conditional Adversarial Networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[21] H. Chen, Q. Dou, L. Yu, J. Qin, P.-A. Heng, VoxResNet: Deep Voxelwise Residual Networks for Brain Segmentation from 3D MR Images, NeuroImage 170 (2018) 446–455.
[22] H. Gao, H. Yuan, Z. Wang, S. Ji, Pixel Deconvolutional Networks, arXiv:1705.06820 (2017).
[23] G. Huang, Z. Liu, V.D.M. Laurens, K.Q. Weinberger, Densely Connected Convolutional Networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[24] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1175–1183.
[25] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, B. Lei, Dense deconvolutional network for skin lesion segmentation, IEEE J. Biomed. Health Inform. 23 (2) (2018) 527–537.

[26] X. Dong, J. Shen, Triplet loss in siamese network for object tracking, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 459–474.
[27] X. Dong, J. Shen, W. Wang, Y. Liu, L. Shao, F. Porikli, Hyperparameter optimization for tracking with continuous deep q-learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 518–527.
[28] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, IEEE Trans. Med. Imaging. 38 (4) (2018) 991–1004.
[29] J. Shi, Z. Xue, Y. Dai, B. Peng, Y. Dong, Q. Zhang, Y. Zhang, Cascaded multi-column RVFL+ classifier for single-modal neuroimaging-based diagnosis of Parkinson's disease, IEEE Trans. Biomed. Eng. (2018), doi:10.1109/TBME.2018.2889398.
[30] J. Zhang, Y. Xia, Y. Xie, M. Fulham, D.D. Feng, Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features, IEEE J. Biomed. Health Inform. 22 (5) (2018) 1521–1530.
[31] Y. Xie, J. Zhang, Y. Xia, M. Fulham, Y. Zhang, Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT, Inf. Fusion 42 (2018) 102–110.
[32] J. Shi, Q. Jiang, R. Mao, M. Lu, T. Wang, FR-KECA: fuzzy robust kernel entropy component analysis, Neurocomputing 149 (2015) 1415–1423.
[33] J. Shi, Q. Jiang, Q. Zhang, Q. Huang, X. Li, Sparse kernel entropy component analysis for dimensionality reduction of biomedical data, Neurocomputing 168 (2015) 930–940.
[34] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification, IEEE Trans. Med. Imaging (2019), doi:10.1109/TMI.2019.2893944.
[35] J. Zhang, Y. Xia, H. Cui, Y. Zhang, Pulmonary nodule detection in medical images: a survey, Biomed. Signal Process. Control 43 (2018) 138–147.
[36] W. Wang, J. Shen, L. Shao, Video salient object detection via fully convolutional networks, IEEE Trans. Image Process. 27 (1) (2018) 38–49.
[37] J. Zhang, Y. Xia, H. Zeng, Y. Zhang, NODULe: combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection, Neurocomputing 317 (2018) 159–167.
[38] W. Wang, J. Shen, H. Ling, A deep network solution for attention and aesthetics aware photo cropping, IEEE Trans. Pattern Anal. Mach. Intell. (2018), doi:10.1109/TPAMI.2018.2840724.
[39] J. Peng, J. Shen, X. Li, High-order energies for stereo segmentation, IEEE Trans. Cybern. 46 (7) (2016) 1616–1627.
[40] X. Dong, J. Shen, L. Shao, M.-H. Yang, Interactive cosegmentation using global and local energy optimization, IEEE Trans. Image Process. 24 (11) (2015) 3966–3977.
[41] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, IEEE Trans. Image Process. 23 (4) (2014) 1451–1462.
[42] J. Shen, J. Peng, X. Dong, L. Shao, F. Porikli, Higher order energies for image segmentation, IEEE Trans. Image Process. 26 (10) (2017) 4911–4922.
[43] X. Dong, J. Shen, L. Shao, L. Van Gool, Sub-Markov random walk for image segmentation, IEEE Trans. Image Process. 25 (2) (2016) 516–527.
[44] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, L. Shao, Real-time superpixel segmentation by DBSCAN clustering algorithm, IEEE Trans. Image Process. 25 (12) (2016) 5933–5942.
[45] H. Jia, Y. Xia, Y. Song, W. Cai, M. Fulham, D.D. Feng, Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging, Neurocomputing 275 (2018) 1358–1369.
[46] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434 (2015).
[47] J. Liu, Y. Wang, Y. Li, J. Fu, J. Li, H. Lu, Collaborative deconvolutional neural networks for joint depth estimation and semantic segmentation, IEEE Trans. Neural Netw. 29 (11) (2018) 5655–5666.
[48] L. Bi, J. Kim, E. Ahn, D. Feng, M. Fulham, Semi-automatic skin lesion segmentation via fully convolutional networks, in: Proceedings of the IEEE International Symposium on Biomedical Imaging, 2017, pp. 561–564.
[49] X. He, Z. Yu, T. Wang, B. Lei, Y. Shi, Dense deconvolution net: multi path fusion and dense deconvolution for high resolution skin lesion segmentation, Technol. Health Care 26 (S1) (2018) 307–316.
[50] W. Wang, J. Shen, Deep visual attention prediction, IEEE Trans. Image Process. 27 (5) (2018) 2368–2378.
[51] J. Fan, X. Cao, P.T. Yap, D. Shen, BIRNet: Brain Image Registration Using Dual-Supervised Fully Convolutional Networks, Med. Image Anal. 54 (2019) 193–206.
[52] Y. Wang, J. Liu, Y. Li, J. Fu, M. Xu, H. Lu, Hierarchically supervised deconvolutional network for semantic video segmentation, Pattern Recognit. Lett. 64 (2016) 437–445.
[53] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, in: Proceedings of the Artificial Intelligence and Statistics, 2015, pp. 562–570.
[54] P. Hobson, B.C. Lovell, G. Percannella, M. Vento, A. Wiliem, Benchmarking human epithelial type 2 interphase cells classification methods on a very large dataset, Artif. Intell. Med. 65 (3) (2015) 239–250.

Hai Xie received the B.E. degree in department of information science and Engineering, Wanfang College of Science and Technology HPU, Zhengzhou, Henan Province, China, in 2013 and M.E. degree in College of Computer Science and Software Engineering, Shenzhen University, China, in 2015. He is currently a Ph.D. candidate in the College of Information and Engineering at Shenzhen University, Shenzhen, China. His search interest is medical image analysis and deep learning.

Haijun Lei received the M.E. degree in department of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China, in 1997 and the Ph.D. degree in Institute for Image Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan in 2001. Since 2006, he has been with the faculty of the College of Computer Science and Software Engineering, Shenzhen University, China. His current research interests include image processing, and pattern recognition.

Yejun He (SM'09) received the Ph.D. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2005. Since 2011, he has been a Full Professor with the College of Information Engineering, Shenzhen University, Shenzhen, China, where he is currently the Director of the Guangdong Engineering Research Center of Base Station Antennas and Propagation and the Shenzhen Key Laboratory of Antennas and Propagation, Shenzhen, China, and the Vice Director of Shenzhen Engineering Research Center of Base Station Antennas and Radio Frequency, Shenzhen, China. He has authored or co-authored over 100 research papers and books (chapters) and holds about 20 patents. His current research interests include wireless mobile communication, antennas, and RF. Dr. He is a Fellow of the IET.

Baiying Lei received her M. Eng. degree in electronics science and technology from Zhejiang University, China in 2007, and Ph.D. degree from Nanyang Technological University (NTU), Singapore in 2013. She is currently with School of Biomedical Engineering, Health Science Center, Shenzhen University, China. Her current research interests include medical image analysis, machine learning, and pattern recognition. Dr. Lei has coauthored more than 100 scientific articles, e.g., IEEE TCYB, IEEE TMI, IEEE TBME, IEEE JBHI. Pattern Recognition and Information Sciences. She is an IEEE senior member and serves as the editorial board member of Scientific Reports, Frontiers in Neuroinformatics, Frontiers in Aging Neuroscience, and Academic Editor of Plos One.