DIEN Network: Detailed Information Extracting Network for Detecting Continuous Circular Capsulorhexis Boundaries of Cataracts

DA HAN1 AND LEI WANG2
1School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China
2College of Electrical Engineering, Hebei University of Architecture, Zhangjiakou 075000, China
Corresponding author: Lei Wang (wanglei3375@163.com)

ABSTRACT Robotic surgery is an arising area to satisfy the tremendous demand of modern clinical application, and it is becoming more and more acceptable by the normal. In this paper, we are dedicated to finding a more modern solution for continuous circular capsulorhexis of cataract surgery via deep learning method. We take inspiration from former works and propose a detailed information extracting network structure that is suitable applied in the area of clinical, where more side-output layers are used in the convolutional modules, rather than single side-output of specific layer. Moreover, to balance the positive samples and the negative samples and make the network model more stable convergence, we introduce focal loss as the loss function of the model. Instead of exploiting network structure deeper and deeper, the boundary of continuous circular capsulorhexis extracted by above approaches can satisfy the demand of the surgery robots need during a cataract surgery. We evaluate the results on the dataset provided by clinical ophthalmologists, and achieves $F = .808$.

INDEX TERMS Deep learning, edge detection, image processing, robotic surgery.

I. INTRODUCTION

Surgical robot has attracted tremendous attention from all fields of the medical industry since its inception. The application of MicroHand S surgical system [2] in China is one of the promising alternative treatment for sigmoid colon cancer. [3] offers instances of repairing a doubly committed juxta-arterial ventricular septal defect with Da Vinci robotic system via left thorax approach. The robotic technology in total knee arthroplasty is illustrated in [4]. Modern medicine combines advanced technology with the expertise of doctors to enable doctors to escape from the burden of burdens to guide more meaningful work.

Cataract, one of the most common diseases in ophthalmology, has received high attention from ophthalmologists and researchers. According to the survey conducted by [4], they find that age matters in the sufferings of cataract and blindness. Furthermore, recent studies have largely found higher rates of cataracts in women than in men. Meanwhile, due to unhealthy living habits and genetic factors, cataract infections are also becoming younger.

In order to reduce the ophthalmological problems caused by cataract, cataract surgery and other forms of surgery have emerged. [5] shows that conducting cataract surgery has a positive influence on one’s sight. [6] studies the incidence of endophthalmitis after cataract surgery can be 0.023%. As a conclusion, conducting cataract surgery is a practical means of dealing with cataracts.

We focus on the field of robot-assisted cataract surgery. Surgical robots need to gather information from different modules to determine the status of the operation and to ensure it smoothly conducted. The slave end sensing modules need to be able to collect tiny force armor at the ophthalmologist to adjust the surgery in time. Influencing devices such as endoscopes require high computational efficiency to ensure the real-time performance of image information. The server is able to translate the operation of the ophthalmologist at the console into an operational range within the eyeball. As shown in Fig. 1.

Its structure is similar to the Da Vinci surgical robot, and it is a master-slave composition. This structure ensures that the ophthalmologist can collect real-time information of the patient under the microscope while performing the operation, thereby adjusting the surgical method and handling accidents.
in a timely manner, ensuring that the entire state of the patient can be detected by the doctor and improving the safety performance of the operation. The slave end can detect the real-time condition of the patient’s body through a variety of sensors installed on the end robot arm during the operation of the ophthalmologist, at the same time, the server restrains the strength and operating range by algorithms while performing the operation to ensure the safety of the operation.

In order to ensure the safety of robotic surgery, we process the images collected by the microscope to identify the boundary of the continuous circular capsulorhexis (CCC) of cataract surgery and limit the movement of the robot within this area. In addition to the boundary of CCC obtained through image processing methods to provide clear and visible assistance to the ophthalmologist, the robot’s motion control module simultaneously uses algorithms to limit the force and movement of the multi-degree-of-freedom robot arm, making it able to withstand safety in the eyeball. The aim of the above methods is to complete the operation within the limit, and try to avoid the adverse consequences caused by prolonged surgical fatigue of the ophthalmologist, meanwhile, it can also eliminate the jitter caused by the novice due to tension.

With the breakthrough progress of deep learning in various fields in recent years [30]–[36], we have considered combining it with clinical medicine to provide powerful assistance to ophthalmologists in artificial cataract surgery. The problem of the extracting boundary of CCC can be regarded as an edge detection problem, which is one of the basic problems in the field of image processing. Our main goal is to extract the boundary of CCC, thus providing virtual constraints for the surgical robot, which makes it easier for the ophthalmologist to conduct surgery and better complete the operation. The method of edge detection in image processing can be divided into two categories: traditional detection approaches [7], [8], [26] and deep-learning based approaches [9]–[17].

II. RELATED WORKS

Our work is related to the field of edge detection and the field of robotic surgery. We introduce the related works in two areas respectively.

A. ROBOTIC SURGERY

[18] is a review of the literature is undertaken, conclusion that robotic surgery is still in its infancy and its niche has not yet been well defined, its current practical uses are mostly confined to smaller surgical procedures is obtained. [19] as part of robotic surgery, it focusses on binary instrument segmentation, where the objective is to label every pixel as background or instrument part. [20] summarizes the application of surgical robots in different fields of diseases, and also statistics the application status in various fields at the current stage, and made an objective evaluation of the development trend of surgical robots. [21] uses both transoral robotic surgery (TORS) for exposure and the Sonopet bone scalpel under navigational guidance to achieve en bloc resection of a cervical chordoma. Surgery results confirm the effectiveness of robotic surgery and provide a new therapy for this kind of disease. Postoperative pain feedback of gynecological cancer patients with traditional surgery and robotic surgery is counted in [22]. The results confirm that robotic surgery for treating gynecological cancer resulting in the least impact on short-term and long-term pain. Most patients (~90%) do not need opioids and are very satisfied with the surgery. In addition to robot-assisted surgical accidents for adults, [23] also targets this application to children, enabling them to better escort all human health.

B. EDGE DETECTION

Also known as boundary detection, is a fundamental problem of image processing. It can be classified into two categories according to the principle: traditional detection
method mainly based on the gradient and grayscale, some well-known approaches like Canny [7], Sobel [8], moreover, some similarly approaches such as Extended Difference-of-Gaussians (X-DoG) [26], Semi-supervised structured random forests (SRF) [1], Oriented Edge Forests (OEF) [15] are all receive brilliant results in their fields, respectively. These methods help people solve problems to some extent in the field of machine vision. However, with Lecun’s concept of deep learning [24], [25] being proposed, the traditional image processing method has received severe impact, and the principle of image processing has also undergone fundamental changes. [9] is the first deep learning method to make the detection task segmentation effect, which has created a groundbreaking effect for the following series of work.

III. PROPOSED METHOD

In this section, we introduce our proposed network structure in details. The size and depth of the network structure determine the amount of computation and the number of parameters, and the loss function determines the convergence speed of the training procedure. We will introduce the network structure and the loss function separately according to the order in this experiment.

A. NETWORK STRUCTURE

The HED network [9] pioneered the detection of segmentation results, and RCF [12] is based on it to make the experimental results more precise and accurate. We draw inspiration from the above two methods and apply these richer convolution features to different convolutional layers. VGG16 [27] pioneers an era of neural networks that continue to influence subsequent deep learning researches, it includes 5 stages and a fusion layer at the end of the network, and it is totally made up of 13 convolutional layers and 3 fully connected layers. We mainly modify the structure of network. Based on the foundation of VGG16 [27], we think that the side-output edge detection is meaningful for finer pixel detection, so we extend this change to the model’s convolution operation. In the network structure, in each convolution calculation of each stage, each path combines the idea of side-output, so that it can fully learn subtler edge features. Therefore, the structure of our proposed network is listed as below. By the way, we modify the structure without changing its kernel size, so that the network still fits arbitrary size of input images.

1) TOTAL NETWORK

The structure of our network is listed in Fig. 2, it is made up of 5 stages. According to the convolution operation and
feature learning in the 5 stages, we divide the five stages into two categories. The first two stages are the same, and so are the last three stages. They are named shallow convolution stages and deep convolution stages, respectively. The final fusion layer is used to fuse the output from different stages together. Different image features obtained by different convolution stages are merged in the end to obtain a prediction map of the input images, making full use of the convolution features of different depths to predict results of the input images via the sigmoid operation. In the feature learning procedure of each stage, the edge information of the picture is fully mined.

In the first two stages, the image is first input into a shallow block for processing, and then its output is sent to the deep block. The convolution operation is then performed on the output of the shallow block and the deep block to learn image features. At the end of the process, the outputs of the two convolution operations are fused and sent to the next stage after the maxpooling layer’s processing, and the fused results are also processed by a deconvolution layer to calculate each loss of the stage.

The results of the image after shallow stage’s processing is input into the deep stage for being processed. The difference between a deep stage and a shallow stage is that it processes the image deeper. Except that it contains more operations, the dimension of image features’ learning is also deeper than the shallow stages. This is because compared with the shallow stages, the deep ones have one more deep block. The convolution operation is also performed after the deep block. As shown in Fig. 2, the difference between deep stage and shallow stage is that deep stage owns one more deep block-convolutional structure, other structures remain the same.

In the above procedures, no matter whether it is a deep stage or a shallow stage, the labels of the images are input into each stage as a feature learning guidance into the model to strengthen the model’s ability to learn the target features. The pipeline is shown in Fig. 2.

The shallow and deep blocks and the convolution operations involved in the network are used to learn the features of the boundaries in the images. The maxpooling layers after each stage is used to ensure the consistency of the size of each image in each final connection layer. In our network structure, whether shallow convolution stages or deep convolution stages, the key arithmetic modules include both shallow and deep blocks. Both shallow and deep blocks are a very critical component of our network structure, and their role is to learn more detailed texture features. We’ll describe the two modules in details in the next stage.

In the structure diagram above, the input images are processed in the first shallow stage. In this stage, the input image first passes through the shallow block, and the output here is divided into two paths: the first goes through the convolution layer operation, and the second is sent into the deep block convolution layer. The two paths pass through the fusion of this stage, and then calculate the loss of this stage through the deconvolution layer, or pass the output as the input of the next stage through the maxpooling layer. Stage-2 and stage-1 are exactly the same. Compared with stage-1, stage-3 adds a deep block-convolution calculation path, ensuring that the model learns the edge features of the images at a deeper level. In the procedure of learning and processing images in each stage, the label of the data is input to each stage for reinforcement as supervised learning. At the same time, the results after each maxpooling are fused and calculated, and the Sigmoid-Convolution-Deconvolution is used to predict the probability images.

Next, we focus on the composition and principles of shallow convolution stages and deep convolution stages. The internal details of the network structure are shown below.

2) MODULES
Both the shallow block and the deep block make full use of the idea of side-output. The difference between the two is whether the block contains a convolution operation. The specific structure of the shallow block is shown in the Fig. 3. It undergoes two convolution-batch normalization (BN)-ReLU operations and fuses the output images with a single convolution-batch normalization(BN)-ReLU operation to obtain the output images.

Similarly, the specific structure of the deep block is shown in Fig. 4. Its calculation procedure is deeper than the shallow blocks. A total of three convolution operations are performed. However, in the side-output branch structure, it is not used. Any operation is just a fusion of the input image and the image after three convolution-BN-ReLU operations to obtain the output image. The main aim of deep block is to obtain the deeper feature of image, besides, the deeper the convolution structure, the more thoroughly tiny noises will be filtered. In order to utilize the side-output idea to the deep block, a fusion structure is also designed inside the deep block. Taking into account the relationship between the total number of model parameters and the speed of image processing, the subsequent experiments on the number of convolution-BN-ReLU operations of the main path and the side path verified the most suitable parameters for deep blocks is (3, 0). Therefore, the structure of the deep block is three times convolution-BN-ReLU operations on the main path, and the input image is directly merged in the end of side path.

B. LOSS FUNCTION
The loss function plays an important role in a network model, it effectively determines whether the network can converge and how fast it converges. Since the detection problem can be regarded as a binary classification problem, there is a huge difference between the positive and negative samples of the result. In the experiment, we need to avoid this defect in the result to ensure the good convergence of the model. Traditional binary classification problems often use cross-entropy or its different forms as the loss function. They also perform better and can objectively measure the performance of the network structure. We have found through experiments that, compared to cross-entropy, focal loss proposed by [28]
performs even better. It is formulated as follows:

$$L_{fl} = \begin{cases} 
-\alpha (1 - y')^\gamma \log y', & y = 1 \\
-(1 - \alpha)y'^\gamma \log(1 - y'), & y = 0 
\end{cases} \quad (1)$$

Meanwhile, we also tested the performance of different parameters in this experiment. The results show that when $\alpha = 0.25$ and $\gamma = 2$, the model performs better.

IV. EXPERIMENTS

A. DATASETS AND MODEL SETTING

In order to verify and test the adaptability and effectiveness of our network model for surgery, we seek professional ophthalmologists for surgical videos recorded by special equipment when they performed cataract surgery for different types of patients, and then part of continuous circular capsulorhexis is selected to make a dataset. The dataset contains children, adolescents, young people, middle-aged and old-aged groups, and also covers different samples of different stages of cataract. Testing with this data set can ensure that the proposed network model is suitable for ophthalmology during surgery, meanwhile, it can provide guidance and assistance to help the ophthalmologists perform cataract surgery better and safer.

In the dataset we produced, the images are processed from multiple surgical videos, including 1762 pictures containing the capsulorhexis boundary, and the ground truth is manually labeled by the ophthalmologist. These images are divided into two categories: test data including 984 images and validation data including 778 images to train and test the network model.

In the course of our experiments, we find that the number of deep and shallow convolutional layers has a significant effect on the experimental results. In order to explore the ideal results that are most suitable for this experiment, we test different parameters of deep and shallow blocks and compare the results conducted. We use the idea of comparative experiments, that is, making other variables fixed and apply a single variable to explore the results of different convolutional layer numbers, so as to obtain the optimal parameters. Here, when testing the parameters of shallow blocks, the parameters of deep blocks keep constant, and vice versa. We record and organize the experimental results into Table 1 and Table 2 to
TABLE 1. The number of different convolution operations included in the shallow block main path and the side-output path and their corresponding F-measures.

| Parameter setting and corresponding F-measure value | Number of convolution in the main path |
|-----------------------------------------------------|----------------------------------------|
|                                                     | 0           | 1           | 2           | 3           |
| Number of convolution in the side-output path       | ×           | 0.322       | 0.463       | 0.591       |
| 1                                                   | ×           | 0.677       | 0.808       | 0.779       |
| 2                                                   | ×           | 0.653       | 0.785       | 0.718       |
| 3                                                   | ×           | 0.557       | 0.692       | 0.665       |

TABLE 2. The number of different convolution operations included in the deep block main path and the side-output path and their corresponding F-measures.

| Parameter setting and corresponding F-measure value | Number of convolution in the main path |
|-----------------------------------------------------|----------------------------------------|
|                                                     | 0           | 1           | 2           | 3           |
| Number of convolution in the side-output path       | ×           | 0.348       | 0.629       | 0.808       |
| 1                                                   | ×           | 0.526       | 0.584       | 0.727       |
| 2                                                   | ×           | 0.504       | 0.495       | 0.658       |
| 3                                                   | ×           | 0.475       | 0.447       | 0.603       |

provide an intuitive comparison. Table 1 lists the corresponding parameters of the fixed deep block and the experimental results on the parameters of the shallow block. Table 2 shows the corresponding parameters of the fixed shallow block and the experimental results of the deep block parameters.

From the above two tables, we can conclude that the parameters corresponding to the best results of the shallow block main path and the side-output path should be (2,1), and the parameters corresponding to the deep block main path and the side-output path to obtain the optimal result. The parameter should be (3,0).

B. RESULTS

1) QUALITATIVE ASSESSMENT

We applied different boundary detection methods to the produced dataset to evaluate the applicability of different methods in this experiment. Specifically, the methods involved are deep learning detection methods and traditional detection methods. Among them, deep learning detection methods include HED [9], CED [13], and RCF [12]. Traditional detection methods include XDoG [26], Edge Boxes [29], and Canny [7].

During our experiments, three similar deep learning-based methods: HED, RCF, and the proposed DIEN network structure are saved and compared with some of their intermediate layer output. Since all three networks use VGG16 as the backbone network, we compared their outputs at stage-2 and stage-4 and listed the results in Fig. 5 to provide an intuitive comparison.

As can be seen from the output graph of the intermediate process, all three methods can learn the image features we expect during the training process. However, it can be concluded through horizontal comparison that in the output map of the three methods of stage-2, although all three methods can learn the image characteristics of boundary of CCC, HED and RCF retain a lot of interference information of the image. At the same time, it can be noted that the RCF has fully learned the boundary information of the pupil, and has not fully learned the expected CCC features. Observing the output maps of the three methods in stage-4, we can conclude that as of stage-4, only the method we mentioned has fully learned the boundary features of boundary of CCC, and HED and RCF can no longer fully learn the image features of boundary of CCC.

By analyzing of the output of the intermediate layers, it can be concluded that the reason for the difference lies in the
model structure. The proposed structure distinguishes and filters different features in the images through the use of deep and shallow blocks, while RCF only relies on side-output and cannot distinguish the various features contained in the images. HED also does not correctly filter different features contained in the images before multi-scale image fusion. The images contain a lot of interference, so the fused images are not the expected results.

After experimental tests, we show the representative results obtained by different methods in the graph to visually compare the effectiveness of each method.

From the comparison of the results listed in Figure 6, it can be seen that RCF can detect the boundary of CCC we need, but this method can also detect the tissues inside the boundary of CCC as the target pixel. This result will provide ophthalmologists with wrong auxiliary information. As a result, the failure of the operation may seriously threaten the life of the patient. The HED method can make predictions for boundary of CCC, but it is only a small part and does not fully predict the correct result. In combination with the previous intermediate layer output, it can be seen that during the model’s learning procedure of the image target features, the features of boundary of CCC are not fully learned. Therefore, this method cannot be used as a method for providing auxiliary information to an ophthalmologist. The next method is the CED. It can be seen from the figure that the model’s prediction of boundary of CCC in the image is even worse than HED. Therefore, CED cannot be used to provide guidance to ophthalmologists. The above three detection methods are deep learning-based detection and extraction methods, and the following comparison is compared with traditional detection methods. It can be seen from the comparison chart that compared with the deep learning detection methods, the detection results of the traditional detection methods are not so clear and accurate. Moreover, the results obtained by the deep learning detection methods contains less noise and interference. Therefore, this method can provide better auxiliary information for ophthalmologists.

2) QUANTITATIVE EVALUATION

In addition to the qualitative evaluation results, we conduct quantitative evaluations of the results obtained. We adopt the Precision(Pr) and Recall(Re) commonly used in the field of edge detection, and the final measurement according to the F-measure(Fm). The calculation methods of the above three indicators are also listed, as shown in the following equation:

\[
Pr = \frac{TP}{TP + FP} \quad (2)
\]

\[
Re = \frac{TP}{TP + FN} \quad (3)
\]

\[
Fm = 2 \times \frac{Pr \times Re}{Pr + Re} \quad (4)
\]

In the above three equations, Positive indicates that a positive judgment is made on the sample, T indicates that the judgment is correct, and F indicates that the judgment is wrong (Negative is similar). For example, TP indicates that the sample is positive, and the model judges it as positive, FP indicates that the model judges it as positive, but the judgment is wrong, and the sample is negative, while TN indicates that the sample is positive, but the model output’s judgement is negative, and FN indicates that it is originally negative, the data is judged as positive by the model.

The edge detection problem can be regarded as a binary classification problem, but in our application, the proportion of positive and negative samples is seriously unbalanced, so TPR and FPR are used to make receiver operating characteristic (ROC) images, and the area under curve (AUC) area is calculated to quantitatively compare different methods. The TPR and FPR calculations are defined as follows:

\[
TPR = \frac{TP}{TP + FN} \quad (5)
\]

\[
FPR = \frac{FP}{TN + FP} \quad (6)
\]
By using the above indicators, ROC images are drawn to quantify the performance of different methods, as shown in Fig. 7.

We can conclude from the comparison of the result graphs that the detection results of the proposed network structure achieve a result of $F = 0.808$ and are superior to those of other methods under the same conditions.

In order to illustrate persuasiveness of the verification experiments, we count the number of parameters of the four deep learning-based methods, combined with the processing speed listed later, to objectively evaluate the different methods. The number of parameters of the five stages of the proposed DIEN scheme has been given, followed by the number of parameters of the RCF and HED and the last CED network model. It can be seen that the CED of the Encoder-Decoder structure contains the most parameters. The proposed DIEN is about 1/6, and the HED and RCF are almost the same, which is about 1/8 of the number of CED parameters. As shown in Fig. 8.

At the same time, for the ophthalmologists who is undergoing surgery, the processing speed of the algorithm is also related to the accuracy of the surgery and to correctly handling and response to unexpected situations, that is, the processing speed of the algorithm for a single image or the unit time. The number of images is just as important as the ability to predict and locate the required information, therefore, we record the processing speed of the images by the different methods mentioned above. In order to better compare their performance, we summarize them into Table 3.

It can be concluded from the table that although the proposed DIEN network model has a processing speed lower than 30FPS of HED and RCF. Based on the statistics of the model parameters, it can be known that although the parameters of the proposed network are more than twice as high as those of RCF and HED, the processing speed of the images is not much lost, which can fully meet the real-time detection criteria and can provide real-time assistance for ophthalmologists performing surgery, making it more accurate and more error free. During the speed test, it can be concluded from the chart that proposed network owns almost the same speed as Canny. One possible reason is that the statistical results include the time of manually debugging of Canny, and at the same time, the rapid processing of the data set by the processor does not have such errors.

### V. CONCLUSION

In this paper, a DIEN network model for detecting CCC boundaries during cataract surgery is proposed. It uses VGG16 as the backbone network and combines the side-output detection idea. Meanwhile, it introduces deep blocks and shallow blocks for the purpose of descriptive learning of detailed features of images. Comparison experiments with other six methods including deep learning methods and traditional detection methods under the same conditions. The conclusion that the proposed scheme is superior to other schemes is obtained, and it can satisfy the application at the same time, the need for surgical robots provides intraoperative assistance to ophthalmologists. However, the result of $F$-measure $= 0.808$ can still get improved, because the
detection results contain a large number of false positive pixels. These pixels contain ground truth pixels, so they have no effect on the surgical procedure. However, as a detection task, it is a problem worth studying and solving.

REFERENCES

[1] Z. Zhang, F. Xing, X. Shi, and L. Yang, “SemiContour: A semi-supervised learning approach for contour detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 251–259.

[2] D. Luo, Y. Liu, H. Zhu, X. Li, W. Gao, X. Li, S. Zhu, and X. Yu, “The MicroHand S robotic-assisted versus Da Vinci robotic-assisted radical resection for patients with sigmoid colon cancer: A single-center retrospective study,” Surgical Endoscopy, vol. 34, pp. 3368–3374, 2020. [Online]. Available: https://doi.org/10.1007/s00464-019-07107-z

[3] Y. Hu, J. Deng, S. Zhao, and Q. Zhong, “Left thorax approach to repair doubly committed juxta-arterial ventricular septal defect with Da Vinci robotic system,” J. Cardiac Surg., vol. 34, no. 6, pp. 495–498, 2019.

[4] C. M. Lee and N. A. Afshari, “The global state of cataract blindness,” J. Cataract Refract. Surg., vol. 40, no. 1, pp. 98–103, Jan. 2017.

[5] D. S. Grewal, T. Schultz, S. Basti, and H. B. Dick, “Femtosecond laser-assisted cataract surgery—Current status and future directions,” Surv. Ophthalmol., vol. 61, no. 2, pp. 103–131, Mar. 2016.

[6] M. Jabbarvand, H. Hashemian, M. Khodaparast, M. Jouni, A. Tabatabaei, and S. Rezaei, “Endophthalmitis occurring after cataract surgery: Outcomes of more than 480 000 cataract surgeries, epidemiologic features, and risk factors,” Ophthalmology, vol. 123, no. 2, pp. 295–301, 2016.

[7] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[8] W. Gao, X. Zhang, L. Yang, and H. Liu, “An improved sobel edge detection,” in Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol., vol. 5, Jul. 2010, pp. 67–71.

[9] X. Xie and Z. Tu, “Holistically-nested edge detection,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 1395–1403.

[10] H. He, S. Zhang, M. Yang, Y. Shan, and T. Huang, “Bi-directional cascade network for perceptual edge detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 3828–3837.

[11] D. Xu, W. Ouyang, X. Alamada-Pineda, E. Ricci, X. Wang, and N. Sebe, “Learning deep structured multi-scale features using attention-gated CRFs for contour prediction,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 3961–3970.

[12] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, “Richer convolutional features for edge detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 3000–3009.

[13] Y. Wang, X. Zhao, and K. Huang, “Deep crisp boundaries,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 3892–3900.

[14] K. K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, “Convolutional oriented boundaries,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, Oct. 2016, pp. 580–596.

[15] S. Hallman and C. C. Fowlkes, “Oriented edge forests for boundary detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 1732–1740.

[16] Y. Wang, M. Fu, and L. Wang, “A robust farmland edge detection method combining anisotropic diffusion smoothing and a canny edge detector,” in Proc. IEEE Int. Conf. Prog. Informat. Comput. (PIC), Dec. 2015, pp. 296–300.

[17] P. Dollar and C. L. Zitnick, “Fast edge detection using structured forests,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 8, pp. 1558–1570, Aug. 2015.

[18] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, and W. C. Meyers, “Robotic surgery: A current perspective,” Ann. Surg., vol. 239, no. 1, p. 14, 2004.

[19] D. Pakhomov, V. Premachandran, M. Allain, M. Azizian, and N. Navab, “Deep residual learning for instrument segmentation in robotic surgery,” in Proc. Int. Workshop Mach. Learn. Med. Imag. Cham, Switzerland: Springer, Oct. 2019, pp. 566–573.

[20] G. W. Yin and Y. T. Kim, “Robotic surgery in gynecologic cancer,” Current Opinion Obstetrics Gynecol., vol. 24, no. 1, pp. 14–23, 2012.

[21] D. Petrov, M. Spadola, C. Berger, G. Glauser, A. F. Mahmoud, B. O’Malley, and N. R. Malhotra, “Novel approach using ultrasonic bone curettage and transoral robotic surgery for en bloc resection of cervical spine chordoma: Case report,” J. Neurosurg. Spine, vol. 30, no. 6, pp. 788–793, Jun. 2019.

[22] J. Abitbol, S. Lau, A. V. Ramanakumar, N. Drummond, Z. Rosberger, R. Gotlieb, J. How, R. Kessous, and W. H. Gotlieb, “Evaluating postoperative pain and satisfaction among women treated by robotic surgery for gynecologic cancer,” Gynecology Pelvic Med., vol. 2, p. 6, Mar. 2019.

[23] D. Ferrari-Light and R. J. Cerfolio, “Robotic thoracic surgery in children and adolescents—Technology and adaptability,” Pediatric Med., vol. 2, p. 11, Mar. 2019.

[24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[25] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, p. 436, 2015.

[26] H. Wimmermüller, “XDoG: Advanced image stylization with extended difference-of-Gaussians,” in Proc. ACM SIGGRAPH/Eurograph. Symp. Non-Photorealistic Animation Rendering, Aug. 2011, pp. 147–156.

[27] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2980–2988.

[29] C. L. Zitnick and P. Dollar, “Edge boxes: Locating object proposals from edges,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, Sep. 2014, pp. 391–405.

[30] H. Guo, S. Wang, J. Fan, and S. Li, “Learning automata based incremental learning method for deep neural networks,” IEEE Access, vol. 7, pp. 41164–41171, 2019.

[31] H. Rathore, A. K. Al-Ali, A. Mohamed, X. Du, and M. Guizani, “A novel deep learning strategy for classifying different attack patterns for deep brain implants,” IEEE Access, vol. 7, pp. 24154–24164, 2019.

[32] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, “Learning affective video features for facial expression recognition via hybrid deep learning,” IEEE Access, vol. 7, pp. 32297–32304, 2019.

[33] M. Yan, “Adaptive learning knowledge networks for few-shot learning,” IEEE Access, vol. 7, pp. 119041–119051, 2019.

[34] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, “Deep learning approach for intelligent intrusion detection system,” IEEE Access, vol. 7, pp. 41522–41550, 2019.

[35] C. Wu, B. Ju, Y. Wu, X. Lin, N. Xiong, G. Xu, H. Li, and X. Liang. “UAV autonomous target search based on deep reinforcement learning in complex disaster scene,” IEEE Access, vol. 7, pp. 117227–117245, 2019.

[36] L. Jiao and J. Zhao, “A survey on the new generation of deep learning in image processing,” IEEE Access, vol. 7, pp. 172231–172263, 2019.

DA HAN received the B.S. degree. He graduated from the Hebei University of Science and Technology, Shijiazhuang, China, in 2016, where he is currently pursuing the M.S. degree. His research interests include image processing and deep learning.

LEI WANG received the M.S. degree. He is currently working with the Hebei University of Science and Technology, Shijiazhuang, China, in 2016, where he is currently pursuing the M.S. degree. His research interests include image processing and deep learning. * * *