J. Meas. Sci. Instrum., 2024, 15(2): 195-203.
DOI: 10.62756/jmsi.1674-8042.2024020
http://jmsi.publish.founderss.cn/
jmsi@nuc.edu.cn

Refined edge detection model based on RCF

ZHAO Weidong*, ZHANG Yao, ZHANG Dandan, LING Qiang

School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan 243000, China

***Corresponding author:** ZHAO Weidong (zwd720819@163.com)

Received: May 14, 2023 **Revised:** July 4, 2023 **Accepted:** August 5, 2023

Abstract: Edge detection is a fundamental method in image processing and computer vision. Aiming to address the issues of roughness and blurriness in edges generated by deep learning-based edge detection technology, a refined edge detection (RED) model based on richer convolutional features (RCF) for edge detection was proposed. In this model, RCF was used as the baseline network. Some downsampling operations in the backbone network were removed, and the coordinate attention (CA) module and hybrid dilated convolution were added to the backbone network. The number and parameters of the compression layers were changed in the deep supervision module, and smooth compression for reducing feature dimensionality was adopted. In the final fusion module, a cross-layer cross-fusion module was used to fuse the information from high and low layers. The RED model was trained and tested on the extended BSDS500 dataset. The optimal dataset scale (ODS) and the optimal image scale (OIS) of the dataset were 0.809 and 0.832, respectively, as evaluated on the BSDS500 benchmark. The experimental results showed that RED model extracted clearer and more detailed edge contours, and the extracted edge information was more comprehensive and abundant.

Key words: deep learning; edge detection; dilated convolution; coordinate attention; cross-layer fusion

0 Introduction

Edge detection[1] is a fundamental method in image processing and computer vision, the essence of edge detection is to extract the information of mutation in the image and highlight the outline of the target object. According to the available research work, edge detection is of great significance to many areas such as high-level feature extraction, feature description, target recognition, and image segmentation[2-4].

Edge detection algorithms can be broadly classified into traditional edge detection algorithms and deep learning-based edge detection algorithms[5] based on deep learning[6]. Traditional detection methods not only generate a lot of noise in image processing[7-12], but also the edge localization is not very accurate, and the detected edges are not smooth enough to effectively suppress the background texture. Traditional detection methods rely on low-level features. They are easily affected by the environment and difficult to characterize high-level information. This limitation hinders the industrial application of traditional edge detection techniques. In recent years, deep learning technology has significantly improved the performance of edge detection tasks[13], and the use of convolutional neural networks for edge detection has become a new trend. Xie et al.[14] proposed holistically-nested edge detection (HED). The training and prediction of HED is not only based on the whole image, but also combines multi-scale and multi-level feature learning. However, HED fails to fully exploit the abundant hierarchical properties of convolutional neural networks. To address this problem, Liu et al.[15] proposed richer convolutional features (RCF) for edge detection. By using the feature pyramid network (FPN)[16] and combining the feature maps of the upper and lower levels through RCF, more accurate features can be extracted at different scales.

Although the RCF algorithm has been greatly improved in the HED, it uses excessive max-pooling layers to reduce dimensionality, causing lines to appear blurred and excessively thick. The backbone network's ability to extract detailed features from the image is insufficient, and the method of information fusion across different scales is too simplistic, resulting in incomplete edge information.

A refined edge detection (RED) model based on RCF was proposed. It was built on RCF by removing certain pooling layers from the backbone network to prevent excessive loss of detailed information. Additionally, an

attention mechanism was introduced by adding CA attention modulesat each stage to enhance the ability of the backbone network to extract features[17]. Furthermore, a hybrid dilated convolution technique was used to expand the receptive field of the backbone network[18]. Finally, the multi-scale features of each layer were fully fused using cross-layer cross-fusion techniques. This approach helped to enhance the sharpness of the edges by focusing on detailed information while also effectively extracting global features.

1 Related work

RCF uses the VGG16 (Visual geometry group) as the backbone network for feature extraction, and removes the fully connected layer and the last pooling layer, resulting in a fully convolutional network (FCN) architecture. The convolutional features are encapsulated in stages 1 to 5 to leverage the rich feature layer structure[19]. The structure of RCF is shown in Fig. 1. Each stage is connected by a pooling layer. The convolutional layer of the backbone network is connected to a convolutional layer with a kernel size 1×1 and channel depth 21. The resulting feature maps at each stage are accumulated by using an eltwise layer to attain hybrid features, and a 1×1 convolutional layer is followed each eltwise layer, and then a deconvolutional layer is used to upsample this feature map. A cross-entropy loss/sigmoid layer is connected to the upsampling layer at each stage. Finally, a 1×1 convolutional layer is used to fuse the feature maps from each stage and obtain the fusion output.

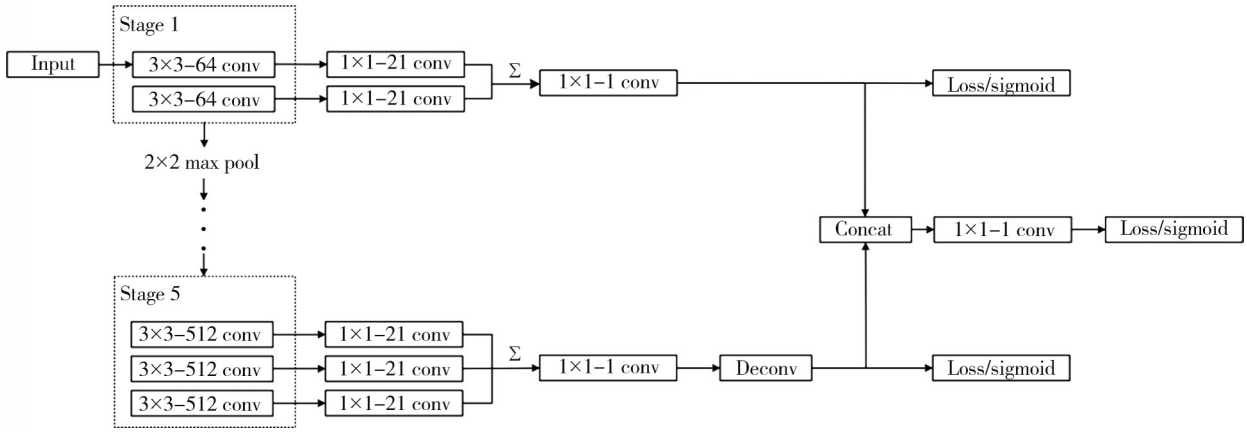


Fig. 1 Structure of RCF

2 Proposed method

The RCF network was used as the baseline network by RED, and the model architecture is shown in Fig.2. It mainly consisted of the backbone network, the deep supervision module, and the feature fusion module. Among them, "CA" is referred to the CA attention module, "Dil" is represented the hybrid dilated convolution, "+" is denoted the summation operation, and "F" is stood for the cross-layer cross-fusion module.

2.1 Backbone network

The pruned VGG16 network was also used by the backbone network of RED. In RCF network, a downsampling operation was performed after each stage except for the last stage. With the downsampling process, the size and resolution of the images were constantly reduced, resulting in the loss of a significant amount of detailed information. However, experiments have shown that if multiple pooling layers were directly removed, it would clutter the extracted edge information and also make the model less capable of generalization. So a balance was struck between the two by using pooling layers but reducing their number. The pooled layer with stage 3 and stage 4 removed was finally selected through experimentation. This not only avoided the loss of detailed information and edge blur caused by frequent downsampling, but also improved the accuracy of the image and suppressed the complex background texture in the image.

RED also incorporated CA modules in the backbone network to extract additional semantic and global features. The squeeze-and-excitation (SE) attention module primarily addressed the issue of loss caused by varying importance of different feature map channels during the convolutional pooling process[20]. However, it only considered the encoding of the information between channels while ignoring the spatial information. A spatial attention mechanism was incorporated into the SE attention module by the convolutional block attention module (CBAM)[21]. This approach reduced the number of channels and used large size convolutions to leverage

location information.

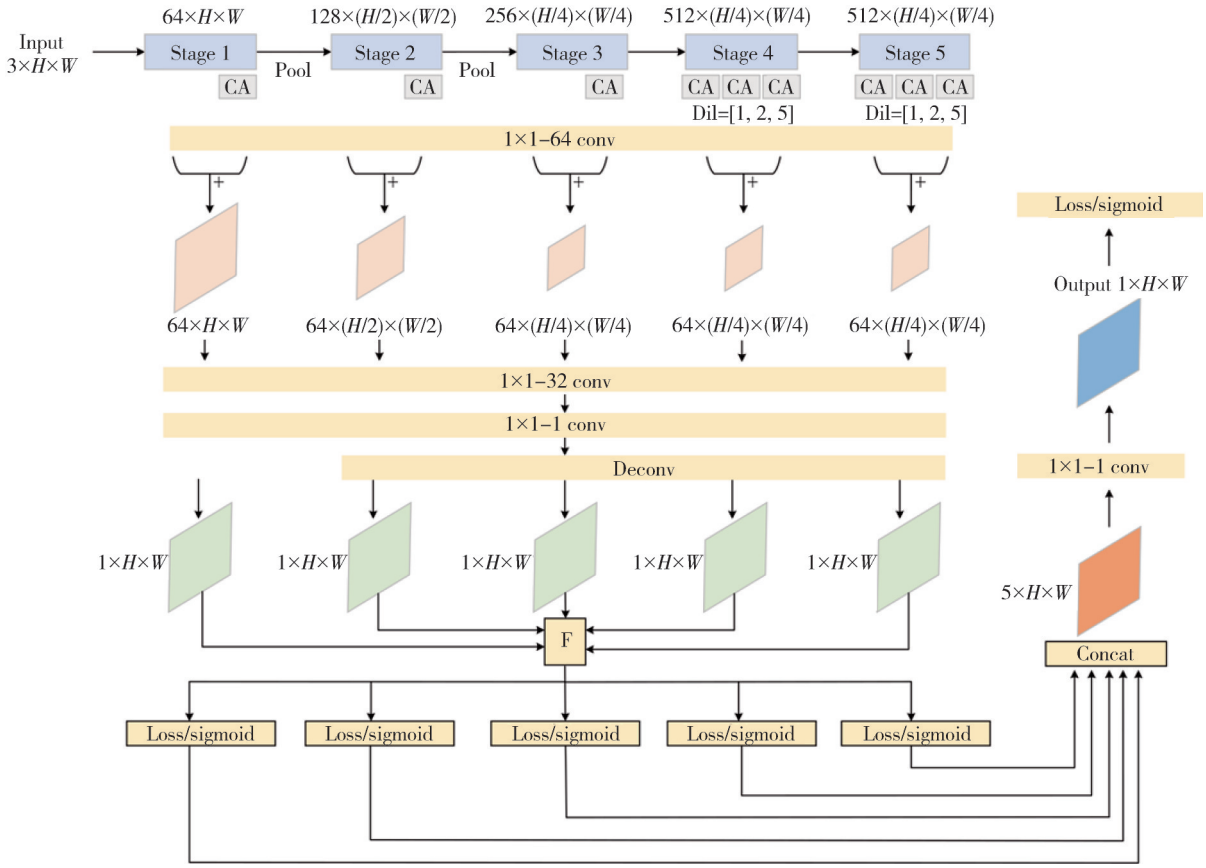


Fig. 2 Structure of RED

In this way, although the importance of different positions in various feature maps could be learned, the issue of long-range dependence was ignored. The CA attention module considered both channel information and directional location information, and incorporated the location information into the channel attention. The architecture of the CA attention module is shown in Fig.3.

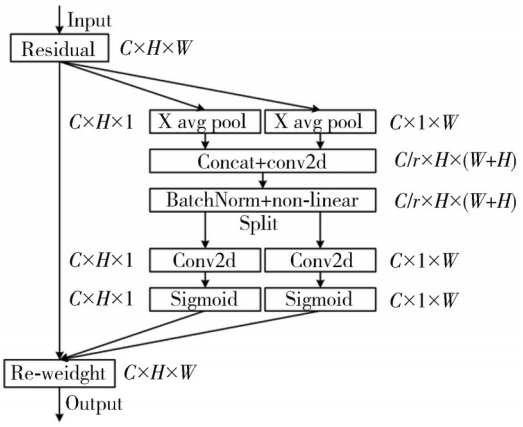


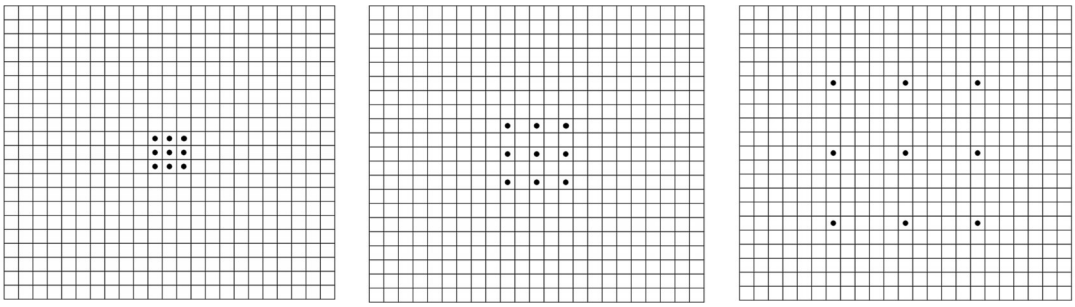
Fig. 3 Structure of CA attention module

The specific operation could be divided into two steps: coordinate information embedding and coordinate attention generation. First, in order to enable the attention module to capture remote spatial interactions with precise location information, the global discretization was decomposed and transformed into a pair of one-dimensional feature encoding operations. Specifically, given an input x, and the pooling kernel with size of $(H, 1)$ or $(1, W)$ is used to encode each channel along the horizontal and vertical coordinates. These two transformations enabled the attention module to capture long-term dependencies along one spatial direction and preserve precise location information along another spatial direction. This helped the network to accurately locate the target of interest.

In order to leverage the resulting representations, a second transformation was proposed. After undergoing the transformations in the information embedding, a stitching operation was performed on the aforementioned transformations. The convolution transformation function was then applied to them, called coordinate attention generation.

To address the issues of low resolution of high-level features and limited perception of detailed information resulting from multiple convolution and pooling operations, the convolution in stage 4 and stage 5 were replaced with a dilated convolution. A hyper-parameter called the dilation rate was introduced by dilated convolution, which was

applied to ordinary convolution. The dilation rate refers to the number of intervals between each point of the kernel. Dilated convolution can increase the receptive field while maintaining the same computational conditions. The contrast plots of the dilated convolution kernel for dilation rates of 1, 2, and 5 are shown as Fig.4.



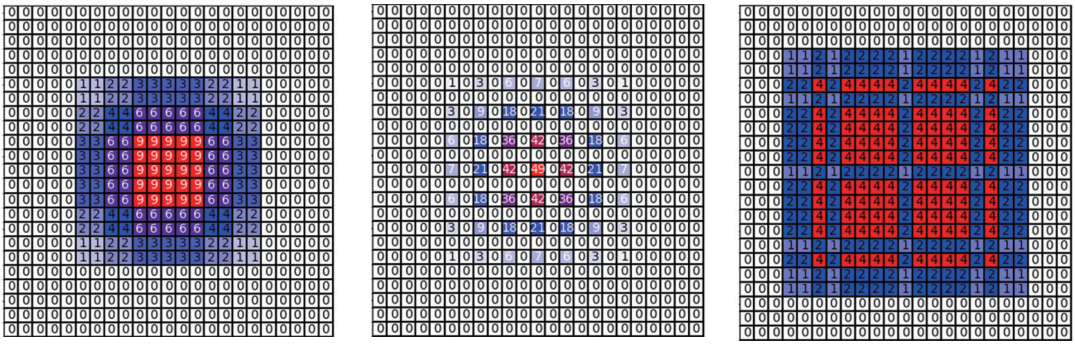
(a) Dilation rate=1 (b) Dilation rate=2 (c) Dilation rate=5

Fig. 4 Comparison of receptive fields of dilated convolution with different dilation rates

However, when using successive dilated convolutions with the same dilation rate, a gridding effect could occur. This effect resulted in an uneven distribution of the number of times each element participates in the convolution kernel operation. As a result, each pixel became independent of one another, lacking interdependence. This lack of interdependence could lead to a loss of local information and a loss of relevance between distant information. Therefore, the hybrid dilated convolution (HDC) principle was followed when using extended convolution. Each layer used a different dilation rate, and the dilation rate could not have a convention greater than 1, turn dilation rate was transformed into a sawtooth form. For example, dilation rate=[1, 2, 5], and it also needs to satisfy the condition $M_i \leqslant K_i$, in which

$$M_i = \max\left[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i\right], \quad (1)$$

where M_i is the maximum distance between two non-zero elements in the layer i; r_i is the dilation rate of the layer i; and K_i is the size of the convolution kernel in the layer i. The number of times that each element participates in the operation across various combinations of dilation rates is shown in Fig.5. When the dilation rate = [2, 2, 2], the receptive field was 13×13. But the number of times each element participated in the operation was extremely uneven. When the dilation rate = [1, 2, 3], the receptive field remained 13×13. However, the number of times each element participated in the operation was not uniform enough. When the dilation rate=[1, 2, 5], not only the receptive field was increased to 17×17, but also the application times of each element were evenly distributed, mostly ranging between 1 and 4. Therefore, the dilation rate of the three successive dilated convolutions in stage 4 and stage 5 was set to [1, 2, 5].



(a) Dilation rate=[2,2,2] (b) Dilation rate=[1,2,3] (c) Dilation rate=[1,2,5]

Fig. 5 Frequency with which each element participates in operation across various combinations.

2.2 Deep supervision module

The deep supervision module was primarily used to enhance the efficiency of feature extraction in the backbone network and improve the training effectiveness of the model. To prevent the number of features from changing too quickly, the number and parameter values of the convolution layers used for dimensionality reduction were adjusted. First, each convolution layer in VGG16 was connected to a 1×1 convolution layer with a channel depth of 64 for feature compression. Each stage was then accumulated with an eltwise layer to obtain hybrid features. 1×1−32 convolution layers were used for transition, and finally 1×1−1 convolutions were used for

compression. The feature map was then scaled up using a transposed convolution initialized by bilinear interpolation[22] and aligned by cropping. The smooth variation of the features can make the entire learning process more stable. The addition of small convolution kernels can not only extend the learning capacity but also mitigate overfitting in image datasets.

2.3 Fusion module

The original paper on RCF network constructed a simple network to generate the side output of the middle layer. The output edge images of RCF at each stage are shown in Fig. 6. It could be observed that the output image became progressively more coarse as the depth increased. For the final fusion output, the RCF network initially combined the five side outputs by splicing them together. However, this method of fusion, which only involved a 1×1 convolution, was inadequate for effectively processing multi-scale information.

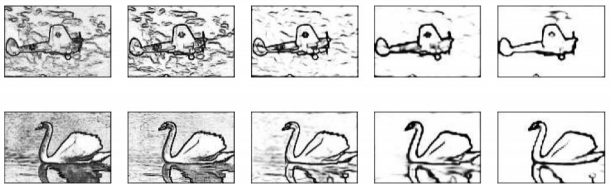


Fig. 6 Output edge images of RCF at each stage

In convolutional neural networks, deep features are characterized by high semantic information and a large receptive field, while the shallow network is exposed to the original information of the image and can capture the most detailed information. As the depth increases, more semantic features of the overall edge contour will be obtained. Therefore, the depth of the features can be increased by merging the features from different layers, resulting in deeper features that contain more comprehensive target edge information. Visual cross-image fusion (VCF) using deep neural networks for image edge detection[23] extracts features from a multi-layer hierarchical structure through cross-fusion. RED fuses the output information from different stages using a cross-layer cross-fusion module.

Through numerous experiments, researchers have discovered that the side output of the third stage was the most suitable foundation for cross-fusion. The specific connection mode is shown in Fig. 7. The transposed convolution and clipped aligned feature maps of stage 3 were added to the feature maps of stages 1, 2, 4, and 5, and new side outputs were created. These outputs were then combined with the original stage 3. Since all levels of features have been fully integrated beforehand, only one 1×1 convolution was required for the overall integration. The final edge image was outputted.

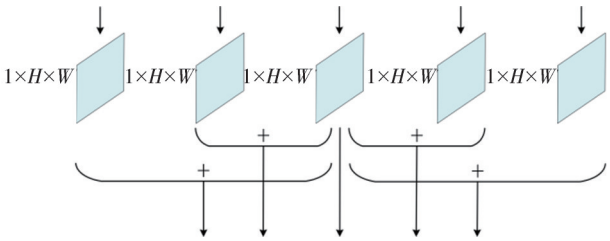


Fig. 7 Structure of cross-layer cross-fusion module

3 Experiment

3.1 Dataset and parameter setting

To evaluate the performance of RED, the extended BSDS500 dataset[24] was used for training and validation. The original BSDS500 dataset consisted of 200 training images, 100 validation images, and 200 test images. The HED paper extends the original dataset to 28 800 images. NYUD dataset consisted of 1 449 densely labeled pairs of aligned RGB and depth images. The NYUD dataset was split into 381 training images, 414 validation images, and 654 test images. Depth information was obtained by using the horizontal-vertical-angle (HHA) method. The HHA features can be represented as a color image through normalization.

All experiments were conducted on a Windows 10 operating system using Python 3.7, which was based on PyTorch 1.11.0 framework. The experiments were performed on a single NVIDIA GeForce RTX 3080 Ti GPU with 12 GB of video memory. The stochastic gradient descent (SGD) method was used to train the model, with the momentum set to 0.9, the batch size set to 8, the initial learning rate set to e^{-6}, and the weight decay set to 0.000 2. The pre-trained model was not used during training, and the network parameters were initialized using a Gaussian distribution.

3.2 Evaluation method

The main evaluation indexes of the edge detection models are optimal dataset scale (ODS), optimal image scale (OIS), and average precision (AP). ODS represents the detection result obtained by applying a fixed threshold to all images in the dataset, while OIS represents the detection result obtained by applying an optimal threshold to each individual image. In addition, the PR curve was used to evaluate the overall performance. The precision rate represents the probability that the boundary pixels

generated by the model are real boundary pixels, while the recall rate represents the probability of detecting all real boundary pixels. The recall rate is abscissa and the precision rate is ordinate.

One evaluation method is standard evaluation, where post-processing operations, such as non-maximum suppression, are performed and then compared with the ground truth labels. The other method is clarity evaluation, which involves direct evaluation without any post-processing operations. The two methods were used to verify the performance of RED.

4 Result and discussion

4.1 BSDS500 dataset

RED was compared with various algorithms, including non-deep learning algorithms[25-27] such as Canny, as well as recent deep learning methods such as CED[28] and LPCB[29]. During the test, the image pyramid technique was also employed to enhance the quality of the edges and fuse the output to obtain a multi-scale edge image. The evaluation indices of each model under the standard evaluation on the BSDS500 dataset are shown in Table 1, and "ms" represents the multi-scale results.

Table 1 Standard evaluation of each algorithm on BSDS500 dataset

Algorithm	ODS	OIS	AP
Canny[8]	0.611	0.676	0.520
Pb[25]	0.726	0.757	0.652
SE[12]	0.743	0.763	0.800
DeepEdge[26]	0.753	0.772	0.815
DeepContour[27]	0.757	0.780	0.800
HED[14]	0.788	0.804	0.815
RCF[15]	0.806	0.823	0.787
RCF-ms[15]	0.811	0.830	0.830
CED[28]	0.794	0.811	0.847
CED-ms[28]	0.815	0.833	0.889
LPCB[29]	0.808	0.824	0.849
LPCB-ms[29]	0.815	0.834	0.869
RED	**0.809**	**0.832**	**0.834**
RED-ms	**0.815**	**0.840**	**0.840**

It can be seen that the performance of RED has been improved based on RCF. Its single-scale ODS value was 0.809, which was 0.3% higher than RCF. Its single-scale OIS value was 0.832, which was 0.9% higher than RCF. Furthermore, its single-scale AP value was 0.834, which was 4.7% higher than RCF. In terms of multi-scale metrics, RED's ODS value was 0.815, which was 0.4% higher than RCF's. Similarly, its multi-scale OIS value was 0.840, which was 1% higher than RCF. Lastly, its multi-scale AP value was 0.840, which was 1% higher than RCF. Compared to other algorithms, the ODS value or OIS value has been improved to some extent.

The evaluation indices of RED for the clarity evaluation of the BSDS500 dataset are shown in Table 2. In the absence of non-maximum suppression, the ODS value was 0.601, which was 1.6% higher than RCF. Additionally, the OIS value was 0.621, which was 1.7% higher than RCF. It can be seen that RED can not only effectively detect edges, but also make them finer and improve the detection performance.

Table 2 Clarity evaluation of each algorithm on BSDS500 dataset

Algorithm	ODS	OIS
Huamn	0.803	0.803
HED[14]	0.588	0.608
RCF[15]	0.585	0.604
RED	**0.601**	**0.621**

The PR curve of RED with other algorithms on the BSDS500 dataset is shown in Fig.8.

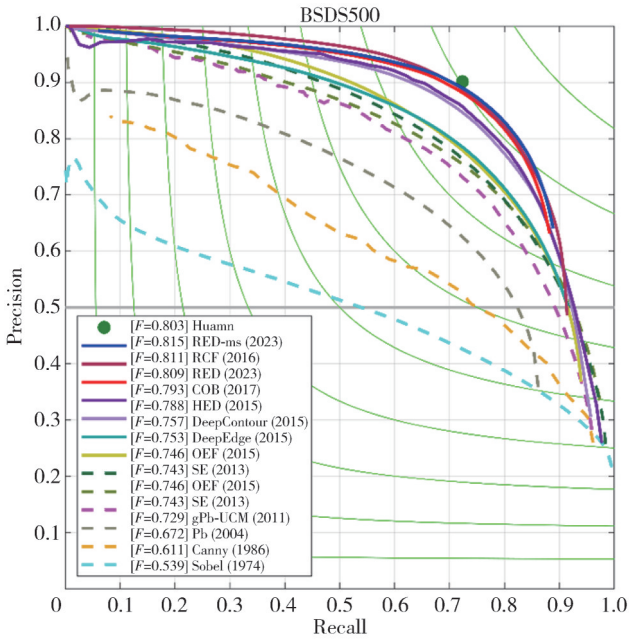


Fig. 8 PR curves of RED with other algorithms on BSDS500 dataset

The human eye has a performance in edge detection with an ODS F-measure value of 0.803. The ODS F-measure of the RCF network was 0.811, which surpassed that of the human eye. Additionally, RED has been improved based on the RCF network, resulting in an increased ODS F-measure of 0.815.

The comparison of the effects of RED with RCF is shown in Fig. 9. Through comparison, it could be observed that the RCF network generated rough lines in the edge images, with a poor ability to handle details and incomplete extraction of edge information.

On the other hand, RED can not only extract edge

lines more clearly, but also effectively handle details such as the outline of the window well. The resulting edge image is more complete, and the edge outline is more detailed.

(a) Original image (b) Ground truth (c) RCF (d) RCF-NMS (e) RED (f) RED-NMS

Fig. 9 Effects of comparison of RED with RCF. Among them, -NMS represents that edge map has been subjected to non-maximum suppression operation

4.2 NYUD dataset

To validate the performance of RED, it was trained and validated on the NYUD dataset and also compared with other algorithms. The ODS and OIS values for each algorithm are shown in Table 3.

Table 3 Standard evaluation of each algorithm on NYUD dataset

Algorithm	ODS	OIS
HED-RGB[14]	0.705	0.728
HED-HHA[14]	0.681	0.695
HED-RGB-HHA[14]	0.739	0.755
RCF-RGB[15]	0.720	0.738
RCF-HHA[15]	0.694	0.711
RCF-RGB-HHA[15]	0.746	0.761
RED-RGB	**0.729**	**0.742**
RED-HHA	**0.707**	**0.719**
RED-RGB-HHA	**0.758**	**0.778**

In Table 3, "-RGB" indicates that RGB images are used as training data, "-HHA" indicates that HHA images are used as training data, and "-RGB-HHA" indicates that RGB and HHA images are used together as training data. When training on RGB images, RED had an ODS value that was 0.9% higher than RCF and an OIS value that was 0.4% higher than RCF. When training on HHA images, RED had an ODS value that was 1.3% higher than RCF and an OIS value that was 0.8% higher than RCF. Lastly, when training on mixed data, RED had an ODS value that was 1.2% higher than RCF and an OIS value that was 1.7% higher than RCF.

4.3 Ablation experiment

In order to verify the impact of each module in the RED model, relevant experiments were conducted, and the experimental results are shown in Table 4. "F" indicates the cross-layer cross-fusion module.

Table 4 Comparison of improvement effect of each module

Program	HDC+CA	Smooth compression	F	ODS	OIS
1	√			0.807	0.825
2		√		0.806	0.829
3			√	0.808	0.827
4	√	√		0.808	0.830
5		√	√	0.808	0.832
6	√		√	0.809	0.828

After removing part of the pooling layer in the backbone network and incorporating hybrid dilated convolution and CA attention modules, the model's ODS value was increased by 0.1% and the OIS value was increased by 0.2%. It has been proven that removing part of the pooling layer can increase the level of edge detail. Additionally, incorporating hybrid dilated convolution and CA attention modules can expand the receptive field of the network and enhance the feature extraction capability of the backbone network. After applying smooth compression in the deep supervision module, the ODS value of the model remained unchanged, but the OIS value was increased by 0.6%. Additionally, the loss converges faster during training, which demonstrates that smooth compression can enhance the stability of the network. After introducing the cross-layer cross-fusion technology in the feature fusion module, the model's ODS value was increased by 0.2% and the OIS value was increased by 0.4%. It demonstrated that the cross-layer cross-fusion module effectively integrated features from different layers, enhancing the model's performance. When any two modules were used, the ODS value or OIS value of the model were improved to varying degrees.

5 Conclusions

A RED model was proposed based on RCF to enhance the quality of rough edge lines and incomplete contours produced by the RCF network. It removed part of the downsampling layer in the backbone network and replaced the convolution layers in the fourth and fifth stages with hybrid dilated convolution. Additionally, it incorporated the CA attention module into the backbone network. The number and parameters of the feature compression layers in the deep supervision module have been modified to enhance the smoothness of the compressed features. The cross-layer cross-fusion technology was integrated into the feature fusion module. With the aforementioned improvements, the model achieved an ODS value of 0.809 and an OIS value of 0.832 on the BSDS500 dataset. These values were 0.3% and 0.9% higher than RCF, respectively. The experiment showed that the improved network effectively utilized edge information and multi-scale information. The network can simultaneously focus on the details and global information of the image, resulting in more detailed and rich extracted edges. It improved the completion of edge detection.

Acknowledgement

This work was supported by Anhui Provincial Natural Science Foundation Project (No. 2108085MF225).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] DUAN R L, LI Q X, LI Y H. A review of image edge detection methods. Optical Technology, 2005, 31(3): 415-419.

[2] SHAO S, GE H W. MAAUNet: Exploration of U-shaped encoding and decoding structure for semantic segmentation of medical image. Journal of Measurement Science and Instrumentation, 2022, 13(4): 418-429.

[3] LIU B Y, YUAN W H, DONG X S, et al. Research on hydrophobic image segmentation of insulators based on improved edge connection Canny algorithm. High Voltage Engineering, 2022, 58(1): 162-169.

[4] LI C Y, BAI J, ZHEN L. A U-Net based contour enhanced attention for medical image segmentation. Journal of Graphology, 2022, 43(2): 273-278.

[5] LI C J, QU Z. Review of image edge detection algorithms based on deep learning. Journal of Computer Applications, 2020, 40(11): 3280-3288.

[6] ZHENG L, SHEN L, LU T, et al. Scalable person re-identification: A benchmark//International Conference on Computer Vision (ICCV), December 13-16, 2015, Santiago, Chile. New York: IEEE, 2015: 1116-1124.

[7] KITTLER J. On the accuracy of the Sobel edge detector. Image and Vision Computing, 1983, 1(1): 37-42.

[8] WANG L, SUN Y. Improved canny edge detection algorithm//The 2nd International Conference on Computer Science and Management Technology (ICCSMT), November 2-14, 2021, Shanghai, China. New York: IEEE, 2021: 414-417.

[9] YU X S, MENG X Y, JIN T F, et al. Object edge detection method based on improved canny algorithm. Lasers and Optoelectronics Progress, 2023, 60(22): 2212002.

[10] GUO Y C, LI M J, LIU M G, et al. Improved algorithm for edge detection of building cracks based on Canny operator. Computer Simulation, 2022, 39(11): 360-365.

[11] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.

[12] DOLLÁR P, ZITNICK C L. Fast edge detection using structured forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(8): 1558-1570.

[13] HUANG S, RAN H S. Refined edge detection method based on semantic information. Computer Engineering, 2022, 48(3): 204-210.

[14] XIE S, TU Z. Holistically-nested edge detection. International Journal of Computer Vision, 2017, 125(1-3): 3-18.

[15] LIU Y, CHENG M M, HU X, et al. Richer convolutional features for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1939-1946. .

[16] LIN T Y, P DOLLÁR, GIRSHICK R, et al. Feature pyramid networks for object detection//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 936-944.

[17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 21-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 13708-13717.

[18] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Nevada, USA. New York: IEEE, 2018: 1451-1460.

[19] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[20] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-

excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
[21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module. Switzerland: Springer, Cham, 2018.
[22] DUMOULIN V, VISIN F. A guide to convolution arithmetic for deep learning. Machine Learning, arXiv: 1603.07285.
[23] QU Z, WANG S Y, LIU L, et al. Visual cross-image fusion using deep neural networks for image edge detection. IEEE Access, 2019, 7: 57604-57615.
[24] MARTIN D R, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics//IEEE International Conference on Computer Vision, July 7-14, 2001, Vancouver, BC, Canada. New York: IEEE, 2001: 416-423.
[25] MARTIN D R, FOWLKES C C, MALIK J. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(5): 530-549.
[26] BERTASIUS G, SHI J, TORRESANI L. DeepEdge: A multi-scale bifurcated deep network for top-down contour detection. Computer Vision & Pattern Recognition, arXiv: 1412.1123.
[27] CHEN H, QI X, YU L, et al. DCAN: deep contour-aware networks for accurate gland segmentation//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2487-2496.
[28] WANG Y, ZHAO X, HUANG K. Deep crisp boundaries//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 1724-1732.
[29] QU Z, WANG S, LIU L, et al. Visual cross-image fusion using deep neural networks for image edge detection. IEEE Access, 2019, 7: 57604-57615

基于RCF的细化边缘检测模型

赵卫东*, 张　瑶, 张丹丹, 凌　强

安徽工业大学 电气与信息工程学院, 安徽 马鞍山 243000

摘　要: 边缘检测是图像处理和计算机视觉中的基本问题。针对基于深度学习的边缘检测技术(如RCF网络)存在生成的边缘线模糊粗糙及边缘信息不全等问题,本文提出一种基于RCF网络的细化边缘检测模型RED。该模型在RCF模型的基础上,去除主干网络中部分下采样,并在主干网络中引入CA注意力模块和混合扩张卷积;在深监督模块改变压缩层的数量和参数,采用平滑压缩的方式进行特征降维;在最后的融合模块,采用跨层交叉融合的方式来融合高低层间的信息。改进后的模型在扩充后的BSDS500数据集上进行了训练和测试,通过在BSDS500基准上进行评估得到数据集最优尺度(ODS)和单图最优尺度(OIS),分别为0.809和0.832。实验结果表明,该模型提取的边缘轮廓更加清晰细致,提取到的边缘信息也更加全面丰富。

关键词: 深度学习;边缘检测;扩张卷积;坐标注意力;跨层融合

引用格式: ZHAO Weidong, ZHANG Yao, ZHANG Dandan, et al. Refined edge detection model based on RCF. Journal of Measurement Science and Instrumentation, 2024, 15(2): 195-203.