Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 2, No. 3, 2022

Semantic Segmentation Based on Improved Robinson Operator

Xiaoshuo Jia*, Zhihui Li

School of Computer Science, Guangdong University of Science and Technology, Dongguan 523079, Guangdong, China

Abstract: Classical pixel-level semantic segmentation methods are easily affected by factors such as environment and pixel values, resulting in low accuracy of their methods. Therefore, in order to improve the accuracy of pixel-level semantic segmentation under practical conditions, this paper designs a lightweight network based on the nonlinear characteristics of residual structure and the characteristics of optimized edge operators. The experimental results show that the designed lightweight residual network can effectively improve the accuracy by about 2% compared with the traditional residual network, and the model size is reduced from the original 40.2MB to 18.9MB.

Keywords: Semantic segmentation, Residual structure, Pooling layer, Lightweigth model, Robinson Operator.

1. Introduction

In image processing, pixel level semantic segmentation [1-3] is an extremely important and complex task. CNN algorithm [4-7] has made excellent achievements in image classification, segmentation and tracking of kaggle and AI Challenger competitions by taking advantage of the characteristics of multi parameters. FCN uses deconvolution for up sampling to make the extracted features more detailed. U-net uses network symmetric structure to fuse high-dimensional features and low-dimensional features to weight edge features. In cpfnet, the concept of dilated convolution is proposed, which can expand the field of view of convolution layer to extract more feature information. Then, the context based feature fusion is realized in combination with the inception module, which has achieved excellent results in medical data sets. STDC is based on FPN for multi-scale fusion, so its performance is better than cpfnet algorithm. Although the algorithm based on CNN has high accuracy, the extracted features usually lose a lot of spatial information due to the pool layer. Finally, problems such as redundancy of semantic segmentation network structure, large amount of calculation and segmentation errors appear.

Here, we analyze the statistical properties and distribution properties of the data on the data distribution of the image edge features, and design a correlation pooling layer rel layer. Then the NR layer is designed on the basis of Robinson operator to extract the edge contour features. Finally, renrnet is designed by combining NR layer and rel layer in residual structure. Renrnet and SOTA algorithms were compared under the SBD dataset. The experimental results show that ReNRnet has certain advantages in accuracy and speed.

2. Method

2.1. Optimized Edge Operator NR

On the basis of the Robinson operator, an optimized edge operator NR is designed, at the orientation of the polar coordinate system of the image, and 8 polar coordinate convolution kernels are given. The picture pic and the 8 azimuth convolution kernels R_1...R_8 do the convolution operation and then accumulate the result cov, such as formula (1):

$$cov = \sum_{i=1}^{8} R_i * pic \tag{1}$$

Obtain the minimum value min and maximum value max in cov, and then perform the operation, and finally obtain the pixel value $P_{m,\,i,\,j}$ of the edge feature image P. m, i, and j represent the width, length, and pass numbers of the image, respectively.

2.2. ReNRnet

In feature extraction, NR can effectively extract the spatial position of edge features and reduce the interference of unnecessary features. Here, to ensure that the correlation between features is not lost during forward propagation, we propose a Rel layer, as formula (2) and (3).

$$S^{m,n} = (\sum_{i,j}(p_{i,j}^{m} * p_{i,j}^{n}) - \sum_{i,j} p_{i,j}^{m} * \sum_{i,j} p_{i,j}^{n}) / h * w \tag{2}$$

$$r_{m,n} = S^{m,n} \Big/ \sqrt{S^{m,m} S^{n,n}} \tag{3}$$

The feature map P obtains n special detection maps P_1,..., P_N (only five feature maps are illustrated here) under the sliding window movement with the step size s of 1. P_2,..., P_N all intersect with P_1. P_1, P_2,..., P_5 get the corresponding correlation features $R_{1,2}$,..., $R_{1,5}$ under formula 2 and formula 3, and then regression the correlation features $R_{1,2}$,..., $R_{1,5}$ to the positions where P_1, P_2,..., P_5 interact with each other, and finally get the correlation feature map $rel_{1,2}$,..., $rel_{1,5}$. P_1 and the surrounding characteristic map P_2,..., P_5 are calculated by algorithm2 to obtain the correlation coefficient characteristic map $relp^1$ of P_1. The eigenvalue of $relp^1$ indicates the degree of correlation between P_1 and the surrounding feature maps P_2,..., P_5. Finally, the result of point multiplication is regressed to the spatial position corresponding to P_1.

Table 1. The network structure parameters are shown in the following table

Layer	Kernel/stride
Conv1	128*11*11/4
NR	5*3/15
Conv3/5/7/9/11/13	16*1*3/2 16*3*1/2
Rel pool	3*3/2
Conv2/4/6/8/10/12	8*3*3/2

It can be seen from the principle that the Rel layer can not only effectively filter the interference of non-related features and improve the accurate removal rate, but also strengthen the features according to a certain proportion, thus helping NR improve the efficient positioning of edge features. Finally, ReNRnet achieves efficient semantic segmentation effect.

3. Experiments

3.1. Dataset

The Pacal VOC (07 + 12) dataset is divided into 4 categories and 20 subcategories. There are 33K photos in total, including 16K train-val data set and 16K test data set.

SBD belongs to the enhanced dataset of VOC dataset. It contains 11355 marker images in the VOC. We used segmentation of VOC datasets as annotations to train the network. SBD divided the whole data set into two parts, 8498 training images and 2857 test images.

3.2. Comparison with SOTA

The training platforms are RTX 3060. The optimization function is Adam optimizer. The loss function uses formula 4 to calculate the error between the true value and the predicted value, and uses IoU to evaluate the test results. yp and yt represent the predicted values and actual values respectively.

$$\mathrm{dice}(\mathrm{p},\mathrm{t}) = 2 * |y_p \cap y_t| / (|y_p| + |y_t|) \quad (4)$$

Here we compare Resnet with some SOTA algorithms, such as Deep snake[8], Unet[9], PANet[10], FCIS[11], ESE[12-13], etc. The results for the SBD datasets are shown in Table 2 below.

Table 2. Comparison results of SBD datasets.

Network	DisCnet	Deep	UNet	PANet	FCIS	ESE	SegNet
AUC(%)	42.6	40.7	34.4	40.1	36.4	28.2	33.6
Fps	28.6	20.1	19.6	23.1	17.4	15.9	20.3

The SBD dataset is an enhanced dataset for Pascal VOC. We used the model trained under Pascal VOC as the pre training model and tested it under the SBD dataset. The results are shown in Table 2, and the segmentation effect is shown in Figure 1. Under the SBD dataset, the model of ESE, UNet and SegNet are relatively small. ReNRnet adopts a residual structure based on correlation features, which can effectively reduce the weight of the model and effectively improve the calculation speed. It can be seen from the segmentation effect in Figure 1 that whether it is a complex background or a small target, ReNRnet can extract the related features through the NR and the Rel layer under the residual structure, thereby realizing the lightweight model and pixel level segmentation.

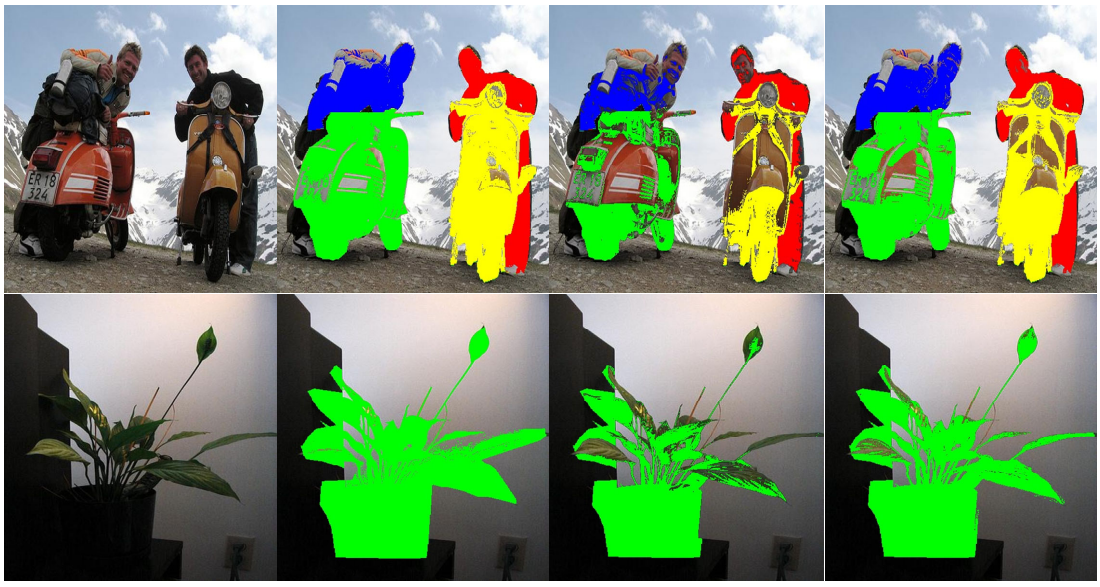

Figure 1. Image represent the original image. And ReNRnet, ESE, Deep Sanke correspond to the segmentation effect image of these algorithms respectively.

4. Conclusion

In this paper, the statistical features of discrete data and Robinson operator are analyzed to extract the correlation features in the image, and the corresponding correlation pooling layer and NR layer are designed. Then combined with residual structure, ReNRnet is designed. Then we make a comprehensive comparison of some SOTA algorithms under the SBD dataset, and prove that ReNRnet performs well in accuracy and speed.

Acknowledgment

This work is supported by Natural Science Program of Guangdong University of Science and Technology under the Grant No.GKY-2021KYQNK-2.

References

[1] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[2] Li Y, Qi H, Dai J, et al. Fully convolutional instance-aware semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2359-2367.

[3] Xu W, Wang H, Qi F, et al. Explicit shape encoding for real-time instance segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5168-5177.

[4] Yuan Z W, Zhang J. Feature extraction and image retrieval based on AlexNet[C]//Eighth International Conference on Digital Image Processing (ICDIP 2016). SPIE, 2016, 10033: 65-69.

[5] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[6] Szegedy C , Vanhoucke V , Ioffe S , et al. Rethinking the Inception Architecture for Computer Vision[J]. IEEE, 2016:2818-2826.

[7] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[8] Peng S, Jiang W, Pi H, et al. Deep snake for real-time instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8533-8542.

[9] Zhao X, Vemulapalli R, Mansfield P A, et al. Contrastive learning for label efficient semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10623-10633.

[10] Armato S G , Roberts R Y , Mcnitt-Gray M F , et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans.[J]. Academic Radiology, 2007, 14(12):1455-1463.

[11] Jetley S, Sapienza M, Golodetz S, et al. Straight to shapes: real-time detection of encoded shapes[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 6550-6559.

[12] Ze Yang, Yinghao Xu, Han Xue, Zheng Zhang, Raquel Urtasun, Liwei Wang, Stephen Lin, and Han Hu. Dense reppoints: Representing visual objects with dense point sets. arXiv preprint arXiv:1912.11473, 2019.

[13] Xingyi Zhou, Dequan Wang, and Philipp Kr¨ahenb¨uhl. Objects as points. Arxiv preprint arxiv:1904.07850, 2019.