Pattern segmentation based on PoolNet and boundary connectivity

Xiaohong Gu, and *Shuyuan* Shang*

Beijing Institute of Fashion Technology, Beijing, 100029, China

Keywords: Significance test, Boundary connectivity, Pattern segmentation, PoolNet.

Abstract. Traditional image segmentation algorithms need to design features manually. Therefore, based on the advantage of PoolNet in object detection, this paper proposes a clothing pattern segmentation algorithm combining PoolNet detection and boundary connectivity. This algorithm has more fine edges than traditional segmentation. Especially in the case of similar background before and after, we can get more complete pattern elements, and have better results in segmentation accuracy.

1 Introduction

Traditional image segmentation algorithms have reached a bottleneck in image processing. Because of its dependence on the characteristics of manual design and higher mathematical ability requirements, the research[1-3] process has many limitations. In order to break through the original image data in precision, accuracy, recall and other aspects of the calculation results, researchers developed a series of model structures combined with the depth learning theory. The strong representation ability of the depth model structure makes the extracted image features have stronger generalization performance than the manually set ones[4], and have been applied in many fields such as image, voice and text. It is mainly to use depth learning to accurately locate patterns on the acquisition of deep semantics, as well as to accurately extract the edges of local significance, so as to make the edges of acquired patterns complete. The algorithm flow is as follows: 1) Input the pattern image of any size into the pre trained PoolNet model to obtain the global saliency map. There are some background information and virtualization problems in the saliency map, which affect the accurate extraction of patterns; 2) The background probability is obtained by using the super pixel background prior algorithm, the background information is suppressed and the target area is refined to obtain the local saliency image; 3) The obtained saliency map is weighted and fused to obtain highlighted saliency map; 4) Adaptive threshold segmentation transforms saliency image into binary image; 5) Morphological processing eliminates some background information in black and white images; 6) The

* Corresponding author:shang@bift.edu.cn

optimized image is used as a mask to obtain patterns by multiplying the image to be segmented.

2 Significance test

Traditional image segmentation algorithms capture local information and global context information separately through manually designed features and strict algorithm design. For example, threshold segmentation, super pixel, etc. mostly focus on the underlying information of the image, unable to capture high-level semantic information, and perform poorly for some images with complex background textures. CNNs can automatically learn features from large-scale datasets. Because of the smaller receptive field and overlapping area, shallow convolution can extract more fine-grained information and retain image details. Deep semantic knowledge can obtain more accurate location information of image object. However, when high-level semantic information is transmitted to a shallow level, the location information captured at a deeper level will be diluted. PoolNet[5] improved on the basis of CNN, applied simple pooling, and introduced GGM (global guidance module) and FAM (feature aggregation module) on the architecture of FPN, so as to generate detailed saliency mapping and further sharpen the details of the target object. Since the U-shaped structure of the model can still achieve better results when there are few training sets, this paper uses the pre trained PoolNet to generate the global saliency map. See Fig. 1 for the model structure.

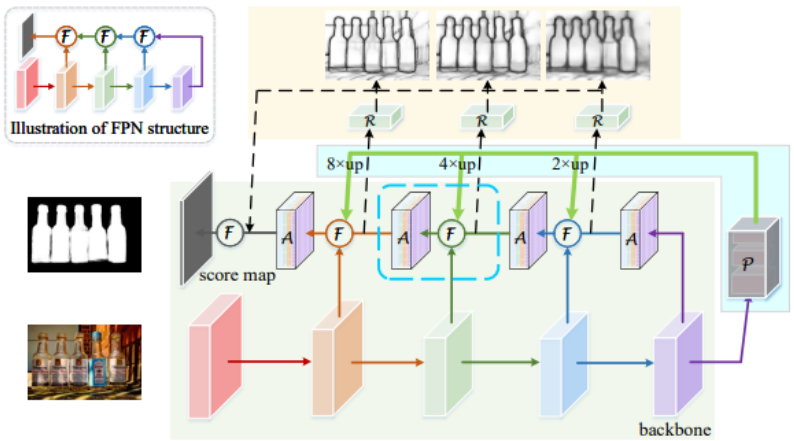

Fig. 1. PoolNetnet work structure.

The backbone network of the model is VGGNet16[6]. The last full connection layer is removed and all convolutional blocks are retained. For images with H×W , downsampling is performed according to decreasing spatial resolution, and the feature map of layer i is

$$\frac{w}{2^i}\times\frac{w}{2^i}\times C_i \tag{1}$$

Because the first layer convolution block can not extract useful information, it is easy to cause information redundancy, which greatly increases the computing cost. Therefore, only the features of conv2, conv3, conv4, and conv5 are used, and these features are marked as {Ci | i=2,3,4,5}. In order to improve the prediction of the object boundary, combined with edge detection training, the number of channels is compressed by 3×3 and 1×1 convolution layers in the top-down transmission process. By applying saliency detection and edge detection respectively, saliency features {Si|i=2,3,4,5} and edge features {Di|i=1,2,3} are

obtained. Compress them to 48 channels and send them to the decoder network to obtain the saliency map.

In the network training process, the batch size is fixed to 1, so that images of different scales can be input. Therefore, the iter size parameter is equivalent to the batch size, and the parameter is set to 64. The amount of training data is increased by means of data enhancement.The model is micro tuned on the basis of the ResNet50[7] pre-training model.Through the control variable analysis and comparison with the optimizer SGD, the training under Adam has a better weight result. Adjust the input size of epoch to obtain the optimal solution in the whole model training process and obtain the pattern saliency map.

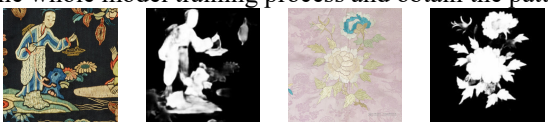

Fig. 2. Significance test results.

The deep learning saliency detection algorithm adopts an end-to-end approach. Because only features are extracted from RGB images, the information contained is limited. When the foreground and background are similar in color and texture, and the background is complex, there are problems such as difficult background suppression, incomplete saliency map, and blurry edges, as shown in Fig.2. Therefore, this paper uses a background prior method to fill and refine the target area.

3 Boundary connectivity

Because local features contain rich edge details of the image, it is unreasonable to completely discard this part of information in image segmentation, which will also lead to rough segmentation results. At present, most of them judge the background prior information of the region according to the correlation between the image region and the edge. Simply viewing all the edges as the background is easy to introduce foreground noise. The edge connectivity in the image is used to improve the robustness of the background prior. Boundary connectivity refers to the ratio of the perimeter of an area to the length of the entire image boundary, that is, the spatial layout of the image area is characterized by the image boundary[8]. Usually, there are few areas connected with the image edge, and the correlation between the area boundaries is small, then the significance value of the area is high, and the corresponding background probability is low. The formula for the connection tightness between region R and image boundary is as follows:

$$\mathrm{Bnd}Con(R)=\frac{Len_{\mathrm{bnd}}(P)}{\sqrt{Area(P)}} \tag{2}$$

where, p represents any super pixel in the image, Bnd is the collection of image boundary blocks, Len is the length associated with the area and image boundary, and $\sqrt{\mathrm{Area}(p)}$ is the area of the entire super pixel area. The formula is:

$$\mathrm{Len}_{bnd}(p)=\sum_{i=1}^{N}S(\mathrm{p},p_i)\bullet\delta(p_i\in Bnd) \tag{3}$$

$$\mathrm{Area}\ (\mathrm{p})=\sum_{i=1}^{N}\exp(-\frac{d_{geo}^2(p,p_i)}{2\delta_{clr}^2})=\sum_{i=1}^{N}S(p,p_i) \tag{4}$$

$D_{geo}(p,p_i)$ represents the geodesic distance between any two hyperpixels.The experimental setting δ_{clr}=10, when $p_i \in$ Bnd, $\delta(.)$=1,otherwise, it is 0. According to the background contrast weight of the super pixel, the background probability w_i^{bg} is introduced as the connectivity mapping of the super pixel p boundary, which hovers between (0,1). The formula is as follows:

$$w_i^{bg} = 1 - \exp(-\frac{BndCon^2(pi)}{2\delta_{bndCon}^2}) \tag{5}$$

Generally, it is set δ_{bndCon}=1 according to experience. When $\delta_{bndCon} \in$ [0.5,2.5], the background weighted contrast is defined as:

$$wCtr(p) = \sum_{i=1}^{N} d_{app}(p, p_i) w_{spa}(p, p_i) w_i^{bg} \tag{6}$$

$d_{app}(p,pi)$ is the super pixel center distance, $w_{spa}(p,p_i)$ is the spatial distance weight. When the background probability of the target area is high, the contrast is enhanced, then vice versa.

4 Significant fusion

The PoolNet model produces a salient map that more fully highlights the target area in the image, while the background probability map obtained by boundary connectivity tends to be local to the image. Based on the global and local characteristics of the obtained salient map, the optimized salient map S_f is obtained by using the weighted fusion method. The formula is as follows:

$$S_f = L_1(S_1, S_2) \tag{7}$$

Highlighted salient maps are obtained by linear fusion and binarized. Binary maps to an overall black and white image, visually highlighting the relationship between the part and the whole. The binarization of the saliency map not only highlights the edge structure of the target area, but also results in a small amount of data in subsequent image processing. At the same time, holes appear in the area of interest in the generated binary image. In subsequent processing, the image details of the internal area are not well presented. The binary image is filled with a hole filling algorithm in morphological processing. The basic idea of morphology is to use structural elements to measure or extract target shapes or features. Holes are background areas surrounded by foreground pixel borders and can be filled with a closed operation.Expansion increases the white highlight area before corrosion removes small non-critical areas, but a kernelsize needs to be set.Therefore, by constructing an array X_0 with an element of 0, the pixel value of the corresponding hole is 1, and the iterative process is used to fill all holes:

$$X_k = (X_{k-1} \oplus B) \cap I^c, k = 1,2,3... \tag{8}$$

X_k is the Marker image, X_{k-1} is the Mask image, which is used to constrain the expansion result, that is, Mask $\geqslant$ Masker, and B is the structural unit SE, which is used to define connectivity.

5 Experiment and analysis

To verify the accuracy of this algorithm, we compared it with other significance detection algorithms using data sets including Berkeley, DUTS, MSRA-B and the collected images related to garment patterns. The effectiveness and robustness of the algorithm are evaluated through analysis from both subjective and objective aspects. This section from the Recall rate, Precision, F-measure and aveMAE evaluation indicators to analyze and compare segmentation performance. The experimental environment is configured for 64-bit Windows with CPU: Intel Core i5-6500 and RAM: 8G.

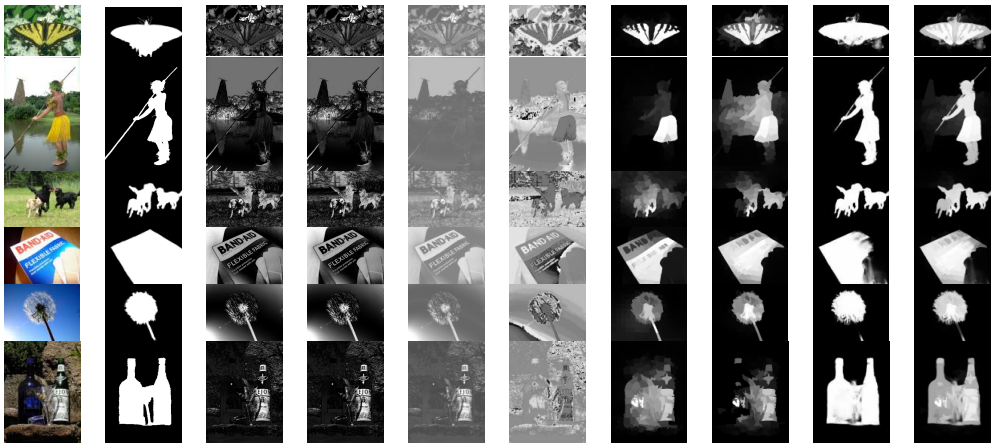

Fig. 3. Comparison of saliency maps of different algorithms.

The significant image effects listed in Fig.3. include FT[9], LC[10], HC[11], RC[11] all with background pixels, so it is not possible to get a more focused target. MR[12] has a better signature when it is a single target and it is inside the image. Background contrast weighted probability maps have the problems of incomplete salient maps and background virtualization for multiple targets with similar background before and after. The images obtained by PoolNet were the most significant in all experiments, but some edge information was ignored, such as dandelions in the second-to-last row of the graph. Obtaining the optimized salient map by using the linear fusion strategy effectively suppresses the corresponding background information and improves at the edge of the target..

Table 1. Algorithmic segmentation criteria.

Algorithm	Berkeley			DUTS			MSRA-B		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
FT	0.275	0.224	0.141	0.599	0.397	0.373	0.718	0.224	0.330
HC	0.284	0.590	0.204	0.902	0.719	0.495	0.692	0.494	0.321
LC	0.265	0.157	0.100	0.592	0.396	0.382	0.716	0.227	0.322
RC	0.319	0.651	0.290	0.536	0.490	0.297	0.588	0.529	0.239
MR	0.917	0.413	0.712	0.896	0.499	0.667	0.905	0.473	0.566
wCtr	0.983	0.572	0.657	0.972	0.552	0.744	0.980	0.455	0.645
Poolnet	0.958	0.897	0.936	0.894	0.832	0.942	0.912	0.944	0.917
Ours	0.972	0.777	0.846	0.917	0.723	0.875	0.991	0.754	0.808

It can be seen from Table 1 that background probability significance images have higher accuracy because they calculate the regional distribution according to the boundary, while PoolNet network model has better global positioning of images according to high-level semantic information, so the recall rate is higher. On the other hand, due to the difference in the correlation between model training data sets and prediction samples, general data sets with similar types will have higher target positioning accuracy; The optimized saliency map combines the advantages of the two aspects to obtain a better result. It is applied to the clothing data set for training analysis to verify the effectiveness of this method.

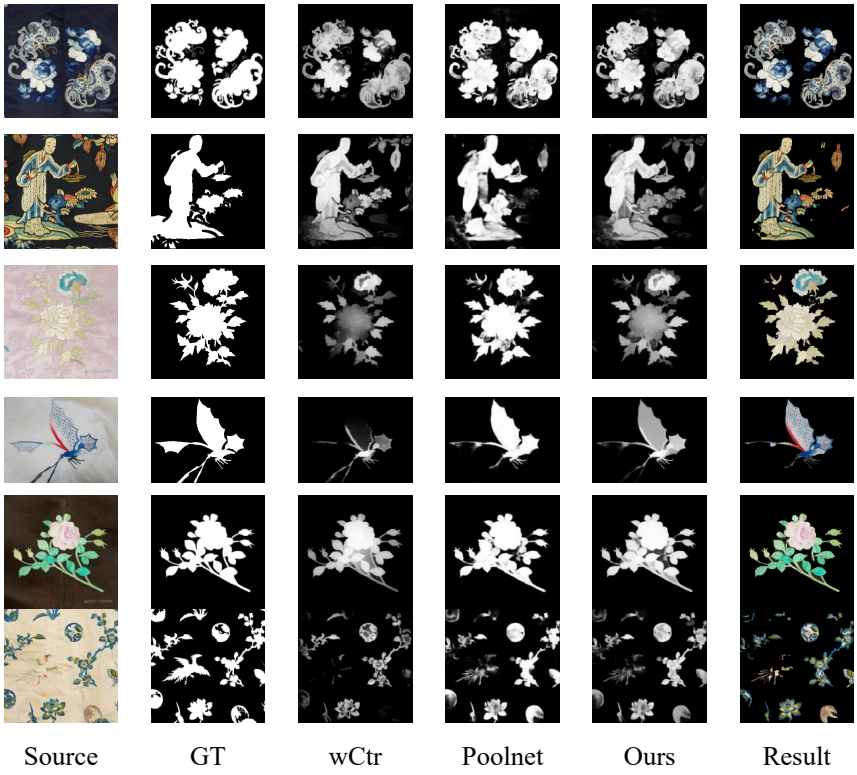

Fig. 4. Comparison of segmentation results on pattern dataset.

Fig. 4 shows the comparison of experimental results of different algorithms on the pattern dataset. It can be seen from the figure that the background robustness algorithm is relatively ineffective in preserving the integrity of significant targets, and the segmentation results in the fourth and sixth rows of images have more defects. Because PoolNet is an end-to-end training, the segmentation area obtained is relatively complete, but there is also a phenomenon of missing parts, such as the missing of flowers in the second picture. Compared with the first two algorithms,the algorithm proposed in this paper can maintain high segmentation integrity without manual labeling, while maintaining good edges. This shows that the method in this paper has certain competitiveness for a large number of multi-scale images with complex structures, and improves the performance of the model over other types when the background and target areas are similar.It can be seen from the data in Table 2 that the positioning accuracy of the algorithm for patterns has been improved, and the model performance has been greatly improved compared with other algorithms.

Table 2. Algorithm segmentation indicators.

Algorithm	Precision	Recall	F-measure	aveMAE
wCtr	0.948	0.734	0.885	0.095
PoolNet	0.870	0.784	0.845	0.105
Ours	0.934	0.847	0.910	0.066

6 Conclusion

Based on the edge connectivity of the image and the advantages of PoolNet in significant segmentation, the effective extraction of garment patterns is achieved. Because both global and local significance of the image are considered, the algorithm presented in this paper is more universal. The experiment is carried out from various angles, and the accuracy, recall rate, F value and MAE value are compared. The algorithm improves to some extent in obtaining the target image and improves the robustness.

Fund project:Research Innovation Project for Postgraduates of Beijing Institute of Fashion Technology (X2022-114), Beijing Philosophy and Social Science Planning Research Project (No. 13JDWYA005)

References

1. E. Shelhamer, J. Long, & T. Darrell (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 39(4):640-651.
2. H. Zhao, J. Shi, X. Qi, X. Wang, & J. Jia (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6230-6239.
3. Y. Wei, H. Hu, Z. Xie, Z. Zhang, Y. Cao, J. Bao, D. Chen, & B. Guo (2022). Contrastive Learning Rivals Masked Image Modeling in Fine-tuning via Feature Distillation. *ArXiv abs/2205.14141*, 1-17.
4. PAN Chenting (2020). Application of Deep Learning in Video Action Recognition. *Computing Technology and Automation*, 39(4), 123-127.(in chinese).
5. J. Liu, Q. Hou, M. Cheng, J. Feng, & J. Jiang (2019). A Simple Pooling-Based Design for Real-Time Salient Object Detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3912-3921.
6. K. Simonyan, & A. Zisserman (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR abs/1409.1556* ,1-14.
7. K. He, X. Zhang, S. Ren, & J. Sun (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
8. W. Zhu, S. Liang,Y. Wei, & J. Sun (2014). Saliency Optimization from Robust Background Detection. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2814-2821.
9. R. Achanta, S. Hemami, F. Estrada, & S. Süsstrunk (2009). Frequency-tuned salient region detection. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 1597-1604.
10. Y. Zhai, & M. Shah (2006). Visual attention detection in video sequences using spatiotemporal cues. *14th Annual ACM International Conference on Multimedia*, 815-824.

11. M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, & S. Hu (2015). Global Contrast Based Salient Region Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), 569-582.
12. C. Yang, L. Zhang, H. Lu, X. Ruan, & M. Yang (2013). Saliency Detection via Graph-Based Manifold Ranking. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3166-3173.