Universe Scientific Publishing

ORIGINAL ARTICLE

Saliency Contour Extraction Based on Multi-level Feature Channel Optimization Coding

Linling Fang, Yingle Fan*

Hangzhou Dianzi University, Hangzhou 310018, China. E-mail: 182060252@hdu.edu.cn

Abstract: A biomimetic vision computing model based on multi-level feature channel optimization coding is proposed and applied to image contour detection, combining the end-to-end detection method of full convolutional neural network and the traditional contour detection method based on biological vision mechanism. Considering the effectiveness of the Gabor filter in perceiving the scale and direction of the image target, the Gabor filter is introduced to simulate the multi-level feature response on the visual path. The optimal scale and direction of the Gabor filter are obtained based on the similarity index, and they are used as the frequency separation parameter of the NSCT transform. The contour sub-image obtained by the NSCT transform is combined with the original image for feature enhancement and fusion to realize the primary contour response. The low-dimensional and low-redundancy primary contour response is used as the input sample of the network model to relieve network pressure and reduce computational complexity. A fully improved convolutional neural network model is constructed for multi-scale training, through feature encoder to feature decoder, to achieve end-to-end pixel prediction, and obtain a complete and continuous detection image of the subject contour. Using the BSDS500 atlas as the experimental sample, the average accuracy index is 0.85, which runs on the device CPU at a detection rate of 20+ FPS to achieve a good balance between training efficiency and detection effect.

Keywords: Contour Detection; Convolutional Neural Network; Multi-level Features; Biological Vision Mechanism

1. Introduction

Contour information is of great significance to the segmentation and recognition of image data. It will realize the rapid outline of the target area of the image, which is helpful to the analysis and understanding of the image in the limited feature dimension[1]. Therefore, the automatic detection of image contours is one of the important research contents in the field of machine learning and image processing. Due to the rapid development of deep learning, the current deep-learning-based methods have received extensive attention in traditional contour detection. The multi-layer network structure of deep convolutional networks is used to simulate the analysis of human visual perception systems in the processing of visual information. Layer features, which can actively perform feature learning and extraction, effectively simplify the process of extraction and data reconstruction of complex features that have been manual or semi-automatic. The fully convolutional neural network proposed by Long promotes the development of visual images and realizes the end-to-end classification and detection of images at the pixel level[2]. The input images of any size are trained through a series of network training such as pooling and upsampling to obtain the segmented images with the same size as the original images.

doi: 10.18686/esta.v7i4.168

Based on this network, Deeplab v1[3] was proposed in 2016, which combines a fully convolutional neural network with a fully connected conditional random field and greatly improves the network segmentation performance.

However, the above methods generally have the following problems: (1) the direct use of fully convolutional neural networks for image segmentation and fusion will lead to imprecise segmentation results and generalization of feature information; (2) it fails to combine deep learning with traditional feature-based methods. The detection performance depends heavily on the number and quality of training samples, and the ability to filter redundant information including texture background is weak; (3) although some methods consider the extraction of multi-source features, they essentially lack the learning process represented by convolutional neural network training, so they cannot truly reflect the effectiveness of multi-source features in expressing contours.

2. Basic principles

This paper proposes a bionic vision computing model based on multi-level feature channel optimization coding. First, this article calculates the optimal scale and direction corresponding to the Gabor filter, and uses the obtained optimal scale and direction as the frequency separation parameters of the non-subsampled contourlet transform (NSCT). Next, the contour sub-image is obtained by NSCT and the original images are feature-enhanced and merged to obtain the primary contour response which is used as the input of the neural network. Then the primary contour response is passed to the fully convolutional neural network composed of FSC-32S, FSC-16S, and FSC-8S network units, and the convolution and pooling modules of the feature encoder are used to realize the active learning of network parameters. The deconvolution and up-sampling module of the decoder obtains the image contour mask image corresponding to the original image, and performs dot multiplication with the original image, and finally realizes the accurate detection of the image contour. The algorithm flow chart of this chapter is shown in **Figure 1.**

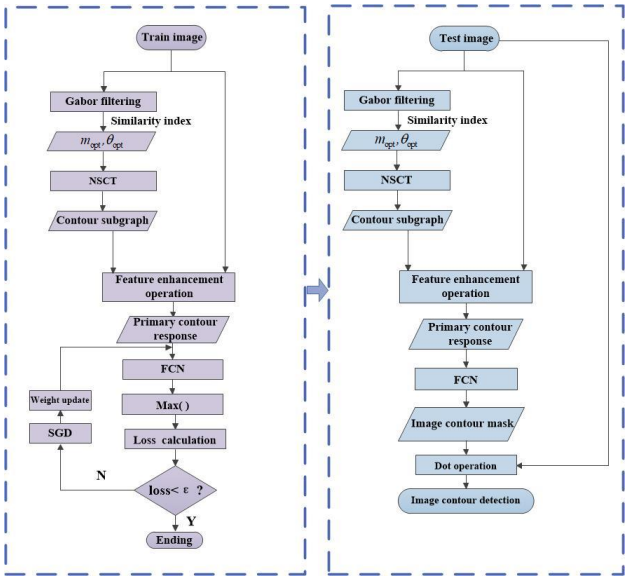

Figure 1. Algorithm flow chart.

2.1 Multi-level feature response on the visual path

The dynamic response characteristics of the Gabor function are very consistent with the physiological characteristics of the human visual system, and have a certain correlation[4]. In this paper, Gabor function is used to simulate the response characteristics of biological vision to the multi-scale and different orientation characteristics of the image on the visual path. The details are shown in formulas (1) ~ (4).

$$I_{m,n}^{\text{Gabor}}(x,y) = \Sigma_u \Sigma_v I(x-u, y-v)\psi_{m,n}(u,v) \tag{1}$$

$$\psi(x,y)=(\frac{1}{2\pi\sigma_x\ \sigma_y})exp\left[\frac{-1}{2}(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2})+2\pi j\omega x\right] \tag{2}$$

$$\psi_{m,n}(x,y)=\alpha^{-m}\psi(x_0,y_0) \tag{3}$$

$$(x_0,y_0)=\alpha^{-m}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} \tag{4}$$

In the formula, $I_{m,n}^{\text{Gabor}}(x,y)$ represents the Gabor feature information obtained on the scale m and direction $\theta=n\pi/K$ of the original image through the Gabor function; σ_x,σ_y respectively represent the standard deviation of the Gabor wavelet basis function along the x-axis and y-axis; ω is the compound modulation frequency of Gabor function. Taking $\psi(x,y)$ as the mother wavelet, through the scale and rotation transformation, the Gabor filter $\psi_{m,n}(x,y)$ can be obtained. Among them, u,v is the template size of $\psi_{m,n}(x,y)$; $m=0,...,S-1$, $n=0,...,K-1$, S and K represent the number of scales and the number of directions respec-

$$SSIM\left(I_{m,n}^{\text{Gabor}},I^{\text{mark}}\right)=L\left(I_{m,n}^{\text{Gabor}},I^{\text{mark}}\right)C\left(I_{m,n}^{\text{Gabor}},I^{\text{mark}}\right)S\left(I_{m,n}^{\text{Gabor}},I^{\text{mark}}\right) \tag{5}$$

Among them, $SSIM(\cdot)$ represents the ty between the characteristic response and the original image. When $SSIM(\cdot)$ takes the maximum value, the optimal scale m_{opt} and direction θ_{opt} are obtained; $L(\cdot)$ represents a quantitative similarity measure on brightness; $C(\cdot)$ represents a quantitative similarity measure on contrast; $S(\cdot)$ represents a quantitative similarity measure on structure.

The optimal scale and direction are used as the di-

$$E(x,y)=\begin{cases}I(x,y), C_{m_{\text{opt}}}^{\theta_{\text{opt}}}(x,y)<t\\ max\left(C_{m_{\text{opt}}}^{\theta_{\text{opt}}}\bullet I(x,y),255\right), C_{m_{\text{opt}}}^{\theta_{\text{opt}}}(x,y)\geq t\end{cases} \tag{6}$$

In the formulas, $N_{m_{\text{opt}}}^{\theta_{\text{opt}}}(x,y)$ represents the non-downsampled contourlet transform under the optimal scale and direction parameters; $C_{m_{\text{opt}}}^{\theta_{\text{opt}}}(x,y)$ represents the corresponding NSCT contour sub-image; t represents the brightness average of the contour sub-image $C_{m_{\text{opt}}}^{\theta_{\text{opt}}}(x,y)$; $max(\cdot)$ represents the maximum value function.

2.3 Fully convolutional neural network

This paper divides the network into two parts: feature encoder and feature decoder. From end to end, there is no need to select the target image area. In the first part of the feature encoder, in the convolution block (3×3, 1×1, 3×3) structure, the 1×1 convolution kernel is added to every two 3×3 convolution kernels. At the same time, in order to strengthen the non-linear and translation invariance of learning image features, a maximum pooling layer is added to each layer of convolution module. In the second part of the feature decoder, the primary contour response is continuously reduced to $1/8$, $1/16$ and $1/32$ times of the original after feature encoding, and the obtained feature map has a low

tively; α is the scale factor of $\psi(x,y)$, where: $\alpha>1$.

2.2 Frequency domain separation mechanism of visual information

Studies have found that the connecting pathway between the lateral geniculate body and the primary visual cortex carries the function of signal transmission, and can also effectively achieve signal separation[5]. Taking the optimal coding process of visual information into account, the introduction of NSCT transform is to achieve the frequency domain separation effect of the outer knee body[6], but the artificial setting of the weighting parameters in the image decomposition process makes the detection results have greater uncertainty. This paper proposes a method based on similarity index[7] to obtain the best direction and scale of the Gabor filter, which is used as the frequency separation parameter of NSCT. Such as formula (5).

rect basis for NSCT to set the frequency separation parameters, and then the contour sub-image obtained by NSCT and the original image are feature-enhanced fusion to obtain a low-dimensional and low-redundancy primary contour response, which is as shown in formulas (6) and (7).

resolution. Therefore, a feature decoder with bilinear up-sampling operation is added to achieve optimized coding of low-resolution contour feature maps.

3. Experiment and analysis

The experiment uses the BSDS500 data[8] set as the performance test and evaluation of the method in this paper. The data set is composed of 200 training sets, 200 test sets, and 100 verification sets, as well as hand-labeled images corresponding to the data set[9]. As shown in figure 2, through quantitative and qualitative comparison with the corresponding contour detection method, it is found that this method makes full use of the multi-source feature signal fusion coding ability under multi-scale, so that the main contour of the detected image is complete and continuous, and the irrelevant texture around the contour is effectively suppressed , which is consistent with the corresponding manual marking diagram.

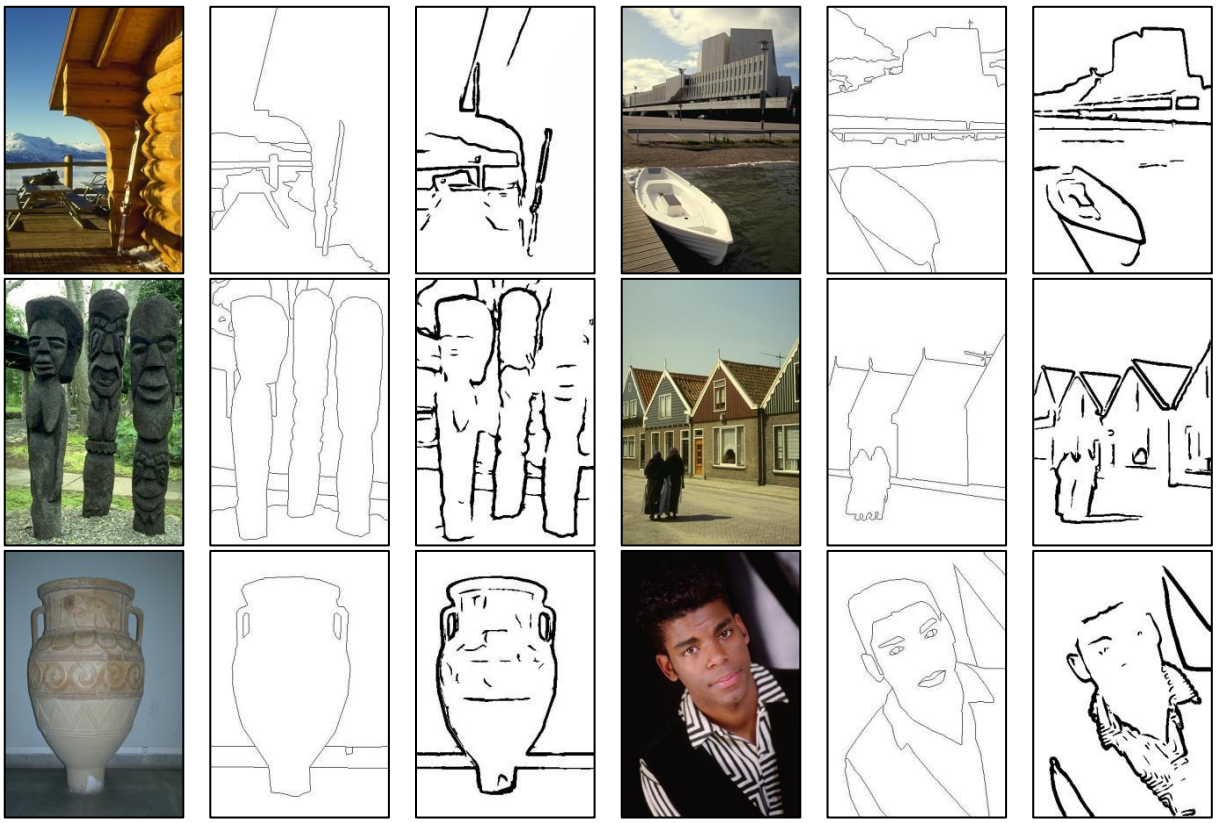

Figure 2. Example of test results on the BSD500 image library.

The detected contour image is acquired through non-maximum value suppression processing, with 1 representing the contour and 0 representing the background. Using 4 standard quantitative evaluation methods, (a) Optimal Dataset Scale, ODS; (b) Optimal Image Scale, OIS; (c) mean Average Precision, mAP; (d) Frames Per Second, FPS. F evaluation index is shown in formula (8).

$$F = \frac{2PR}{P+R} \tag{8}$$

In the formula, P represents the accuracy of pixel classification; R represents the recall rate of pixels.

This method selects RCF[11], COB[12], HED[13], HFL[14], DeepContour[15], DeepEdge[16], OEF[17] and other deep learning methods in contour detection applications, And MCG, EGB, Canny, MShift and other traditional methods based on biological vision mechanism[18,19] to compare the experimental results. As shown in **Figure 3** and **Table 1**, the evaluation indicators of the detection performance of each method are displayed. The evaluation index of the human manual map is 0.80. Only the detection effect of RCF and the method in this chapter surpasses the human A multimedia image label map, and the method in this chapter is 1% higher than the RCF on the AP index. RCF is a precise edge detector that re-

organizes the features obtained by convolution between all layers in the network channel into a new overall feature as output. It is not only effective in edge detection, but also has the possibility of being applied to medical images. The detection speed of the method in this chapter is 20+ frames/s, and even the test speed on GPU reaches 42 frames/s, which is significantly faster than other methods. It shows that this method can achieve the task requirements of rapid detection under the condition of high average accuracy. In addition, from the perspective of the PR curve, the recall rate of this method is higher than that of RCF, which means that more contour information can be more completely retained in the contour detection process, and the training efficiency and detection effect can be balanced.

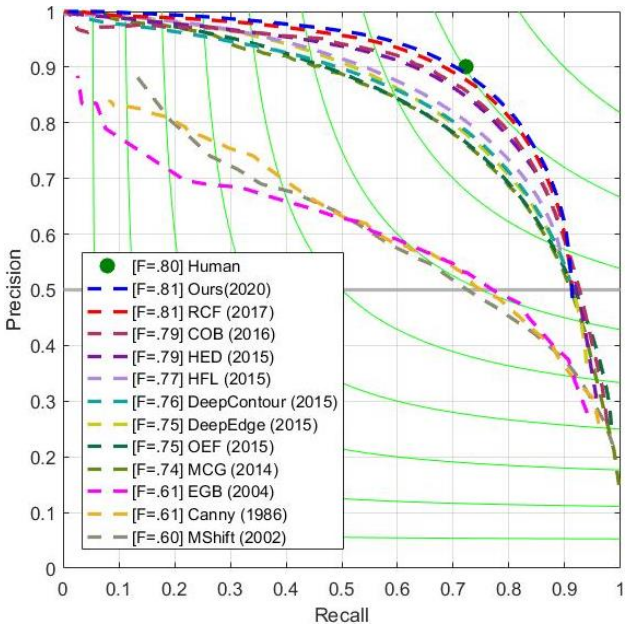

Figure 3. PR curve.

Table 1. Comparison of performance results with other methods

Method	ODS	OIS	AP	FPS
Human	0.80	0.80	-	-
RCF	0.81	0.82	0.84	30+
COB	0.79	0.82	0.85	-
HED	0.79	0.81	0.84	30+
HFL	0.77	0.79	0.80	5/6+
DeepContour	0.76	0.78	0.79	1/30+
DeepEdge	0.75	0.77	0.81	1/1000+
OEF	0.75	0.77	0.82	2/3
MCG	0.74	0.78	0.76	1/18
EGB	0.61	0.66	0.56	10
Canny	0.61	0.68	0.52	15
Mshift	0.60	0.65	0.50	1/5
Ours	0.81	0.82	0.85	20+

4. Conclusion

In this paper, a computational model of multi-level feature channel optimization coding is designed and applied to the rapid detection of the significant contour of the image. The innovation is mainly focused on combining deep learning with traditional feature-based methods, and adding auxiliary image features to extract image

feature information in depth.

References

1. Mouelhi A, Sayadi M, Fnaiech F, *et al*. Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. Biomedical Signal Processing and Control 2013; 8(5): 421–436.
2. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7–12; Boston. IEEE; 2015.
3. Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017; 40(4): 834–848.
4. Wu T, Bartlett MS, Movellan JR. Facial expression recognition using Gabor motion energy filters. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops; 2010 Jun 13–18; San Francisco. IEEE; 2010.
5. Chen CY, Sonnenberg L, Weller S, *et al*. Spatial frequency sensitivity in macaque midbrain. Nature Communications 2018; 9(1): 2852.
6. Da Cunha AL, Zhou J, Do MN. The nonsubsampled contourlet transform: Theory, design, and applications. IEEE Transactions on Image Processing 2006; 15(10): 3089–3101.
7. Dagher I, Mikhael S, Al-Khalil O. Gabor face clustering using affinity propagation and structural similarity index. Multimedia Tools and Applications 2020.
8. Dollár P, Zitnick CL. Structured forests for fast edge detection. 2013 IEEE International Conference on Computer Vision; 2013 Dec 1–8; Sydney. IEEE; 2014.
9. Martin D, Fowlkes C, Tal D, *et al*. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proceedings Eighth IEEE International Conference on Computer Vision; 2001 Jul 7–14; Vancouver. IEEE 2002.
10. Liu Y, Cheng MM, Hu X, *et al*. Richer convolutional features for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 2019; 41(8): 1939–1946.
11. Liu Y, Cheng M M, Hu X, *et al*. Richer convolutional features for edge detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017; 3000–3009.
12. Maninis K K, Pont-Tuset J, Arbeláez P, *et al*. Convolutional oriented boundaries. European Conference on Computer Vision. Springer, Cham 2016; 580–596.
13. Xie S, Tu Z. Holistically-nested edge detection. Proceedings of the IEEE International Conference on Computer Vision 2015; 1395–1403.
14. Bertasius G, Shi J, Torresani L. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. Proceedings of the IEEE International Conference on Computer Vision 2015; 504–512.
15. Shen W, Wang X, Wang Y, *et al*. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015; 3982–3991.
16. Bertasius G, Shi J, Torresani L. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4380–4389.
17. Hallman, Sam, Fowlkes, Charless C. Oriented edge forests for boundary detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015: 1732–1740.
18. Gong XY, Su H, Xu D, *et al*. An overview of contour detection approaches. International Journal of Automation and Computing 2018; 15(6): 656–672.
19. Lu Z, Wang X, Shang J, *et al*. A multimedia image edge extraction algorithm based on flexible representation of quantum. Multimedia Tools and Applications 2019; 78(17): 24067–24082.