Research on Convolutional Neural Network Model for Sonar IMAGE Segmentation

Shengxi Jiao¹, Chunyu Zhao ¹and Ye Xin²

¹Northeast Electric Power University, Advanced Sensor Technology Institute, 132011 Jilin Jilin, China

²Northeast Electric Power University, Fault Diagnosis Institute, 132011 Jilin Jilin, China

Abstract. The speckle noise of sonar images affects the human interpretation and automatic recognition of images seriously. It is important and difficult to realize the precision segmentation of sonar image with speckle noise in the field of image processing. Full convolution neural network (FCN) has the advantage of accepting arbitrary size image and preserving spatial information of original input image. In this paper, the image features are obtained by autonomic learning of convolutional neural network, the original learning rules based on the mean square error loss function is improved. Taking the pixel as the processing unit, the segmentation method based on FCN model with relative loss function(FCN-RLF) for small submarine sonar image is proposed, sonar image pixel-level segmentation is achieved. Experimental results show that the improved algorithm can improve the segmentation accuracy and keep the edge and detail of sonar image better. The proposed model has better ability to reject sonar image speckle noise.

1 Introduction

Due to the presence of seabed reverberation and various environmental noises, the sonar image has severe speckle noises which will affect sonar image segmentation accuracy. More serious, the human interpretation and automatic recognition of small target sonar images will also be affected. In this paper, a kind of submarine small target sonar image segmentation algorithm is proposed to extract image features rapidly and achieve target segmentation precisely.

Due to the low resolution and severe speckle noise of sonar images, it is difficult to achieve accurate segmentation by ordinary image segmentation methods [1]. At present, there are many sonar image segmentation methods, such as clustering method [2], edge detection method [3], Markov Random Filed method [4] and combining with specific mathematical theory methods [5,6]. However, these methods have poor anti-speckle noise capabilities and cannot achieve automatic segmentation. With the advent of convolutional neural networks(CNN), many researchers apply CNN to image recognition and segmentation. CNN are not only improving the whole-image classification[7,8], but also making process on object detection [9,10], part and keypoint prediction [11,12]. In 1998, Lecun et al. [13] proposed the LeNet-5 network structure, the convolutional neural network was applied to handwritten character recognition firstly. In 2012, Krizhevsky et al. [14] designed AlexNet to further deepen the network structure and improve the accuracy of image recognition. Based on the classical network model, new convolutional

neural network models have been proposed successively, such as VGG [15] of Oxford University, GoogLeNet of GoogLe [16], ResNet of Microsoft [17] and so on. Paper [18] apply CNN to Medical Ultrasound Images Automatic Segmentation. However, using CNN for image segmentation, the original image needs to be divided into a large number of image blocks as the input of the network, resulting in low computational efficiency, and the network cannot utilize the connections between contexts in the image fully. To cope with this problem, a fully convolutional neural network was proposed for image semantic segmentation, which can accept arbitrary size image input while preserve the spatial information of the original input image [19]. All the above studies applied CNN and FCN to the image segmentation process greatly and achieved Satisfactory segmentation result, which played a good reference role to sonar image segmentation process.

In order to solve the low accuracy problem of sonar image segmentation, this paper applies FCN to sonar image segmentation process. By improving the loss function of FCN back propagation process, the improved network can take the target segmentation and detail preservation into account much better while improving the segmentation accuracy.

2 Sonar image segmentation of FCN model

2.1 Submarine small target sonar image

The submarine small target sonar image contains three areas: the target bright area, the dark shadow area and the seabed reverberation area, see Fig. 1. The shape of the shadow region also contains the target features and needs to be segmented separately. Sonar images often have severe speckle noise, poor image contrast and low resolution. In particular, the target area small target sonar image has less pixels and the edges is serious deterioration. Thus, an effective segmentation algorithm is necessary to improve the accuracy of segmentation.

Figure 1. Submarine small target sonar image

Figure 2. The values of a and b.

Figure 3. Data set building process.

2.2 Submarine small target sonar image data set

It is difficult to obtain a large amount of comprehensive sonar image data. Therefore, sonar image data set for training and testing need to be manually

established on the basis of limited sonar images. Through the analysis of a large number of sonar images, the noise statistical laws obey the Rayleigh distribution. Thus, speckle noise with different intensities and spot sizes are added on the basis of limited sonar images to expand the sonar image training data set. Rayleigh noise can be achieved by average noise, see equation (1):

$$z = a + \sqrt{-b \ln[(1 - U(0,1))]} \quad (1)$$

The values of a and b represent noise intensity and spot size respectively. Based on the original 200 sonar images, three kinds of speckle noise with different intensities were added to obtain 1800 images. The establishment process are shown in Fig. 2 and Fig. 3 .

The partial sonar images data set results are selected in Fig. 4 and Fig. 5.

Figure 4. The partial sonar images training data set results.

Figure 5. The partial sonar images testing data set results.

Figure 6. Full convolution neural network structure.

Table 1. The parameters of the model

Convolution section	Convolution layer	Number of convolution kernels	Convolution kernel Size/step size	Pooled window size/step size	Output image size/number per layer
1	conv1	64	3*3/1		222*222
	conv2	64	3*3/1		220*220/64
Max-pooling				2*2/2	110*110
2	conv3	128	3*3/1		108*108
	conv4	128	3*3/1		106*106/128
Max-pooling				2*2/2	53*53
3	conv5	256	3*3/1		51*51
	conv6	256	3*3/1		49*49
	conv7	256	3*3/1		47*47/256
Max-pooling				2*2/2	24*24
4	conv8	512	3*3/1		22*22
	conv9	512	3*3/1		20*20
	conv10	512	3*3/1		18*18/512
Max-pooling				2*2/2	9*9
5	conv11	512	3*3/1		7*7
	conv12	512	3*3/1		5*5
	conv13	512	3*3/1		3*3/512
Max-pooling				2*2/2	2*2
6	FC	4096	1*1/1		2*2
7	FC	4096	1*1/1		2*2/4096
8	FC	1000	1*1/1		2*2/1000

2.3 Model building

The FCN structure is shown in Fig. 6. The convolution process uses a 16-layer VGGNet network. The improved network can accept arbitrary size input image. The FCN deconvolution process needs to mirror the 16-layer VGGNet and restore the extracted feature map to the size and position of the original image. Both learning and inference are performed whole-image-at-a-time by dense feedforward computation and backpropagation.

2.4 FCN parameter setting and feature extraction

The FCN used in this paper has 8 convolution sections. Each section of the first 5 sections is connected with a maximum pooling layer, to reduce the image size, the following 3 sections are still use convolution method for feature extraction. The specific parameters settings of convolution kernel size, step size and number are shown in Table 1.

Taking submarine small target sonar image of 224 * 224 as an example shown in Fig. 1, after the first layer convolution operation, 64 sonar image feature maps with the size of 222 * 222 are output which is also used as the input for the second layer of convolution shown in Fig. 7.

Figure 7. Part of the image features obtained from the first convolutional layer

After the second layer of convolution operation, 128 sonar image feature maps with the size of 220 * 220 are output. After being processed by the first pooling layer, 128 feature maps with the size of 110 * 110 are generated correspondingly, and the image is reduced dimensionally, see Fig. 8.

Figure 8. The original image and part of the feature image obtained from the first sample layer.

The 10th convolutional output 512 features map with the size of 18 * 18, see Fig. 9. After the 16th convolution layer, the deconvolution sampling operation is performed to obtain 1000 original size images. Through the pixel-by-pixel operation, the maximum number of features in the deconvolution feature map is obtained. Further, predict the classification of each pixel, output image segmentation results finally.

Figure 9. The original image and part of the feature image obtained from the first sample layer: (a) Different areas of the same picture; (b) Different target sonar images; (c) The feature image of different areas of the same picture obtained from the first sample layer; (d) The feature image of different target sonar images obtained from the first sample layer;

3 Sonar image segmentation of FCN model

3.1 Loss function and its improvement

The convolutional neural network uses backpropagation chain derivation in the training process [20]. The weights are updated by minimizing the loss function to obtain the optimal weight parameters of the network. The optimal weight parameters are obtained from the mean square error loss function as shown in equation (2).

$$E(i) = \frac{1}{2} \sum_{k=1}^n \left(d_k^i - y_k^i \right)^2 \quad (2)$$

where $E(i)$ is the training error of a single sample, $d(i)$ is the expected output, and $y(i)$ is the predicted output of the network.

In convolution neural network learning process, the mean square error (MSE) loss function backpropagation algorithm has the shortcomings of small gradient change, slow update of weights and does not consider the non-target area pixels, resulting in larger pixel segmentation errors. In order to solve this problem, the paper presents a relative loss function algorithm, which target region and non-target region pixels are both taken into account. Because there is no obviously difference between the reverberation area and the target area, the algorithm calculates the probability of each pixel that belongs to the target area, so that the algorithm can speed up the convergence of the network and improve the segmentation accuracy. The relative loss function is as follows:

$$E = \sum_{i=1}^m E(i) = \sum_{i=1}^m \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp \left(- \left(c_k^i - b_k c_l^i \right)^2 \right) \quad (3)$$

$$b_k = \frac{F_k / n}{\max F / n} \quad (4)$$

where Y_i represents the target area pixels set of samples i ; $\bar{Y}_i$ represents the non-target pixels set of samples i ; c_i represents the set of target area network model output value; c_k^i represents network model output values of target area pixel set; c_l^i represents network model output values of non-target area pixel set; b_k represents the probability of the K th pixel with respect to the target area; F_k represents the number of the K th pixel corresponding to the target area; n represents the total number of samples;

3.2 Relative loss function algorithm flow

Convolution kernel weights W and bias b are mainly trained in FCN learning process, the operation of update and iteration according to the following steps:

- 1) Forward propagation process

Hidden layer input:

$$\mu_m = \sum_{h=1}^{n_m} W_{m,h} c_h + b_m = \sum_{h=1}^{n_m} W_{m,h} X_m + b_m \quad (5)$$

Hidden layer output:

$$c_m = \frac{1}{1 + e^{-\mu_m}} \quad (6)$$

Output layer input:

$$c_m = \frac{1}{1 + e^{-\mu_h}} \quad (7)$$

- 2) The establishment of a single sample i loss function:

$$E(i) = \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp \left(- \left(c_k^i - b_k c_l^i \right)^2 \right) \quad (8)$$

- 3) Reverse derivation process:

$$\frac{\partial E(i)}{\partial c_h^{ki}} = \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp \left(- \left(c_k^i - b_k c_l^i \right)^2 \right) \times 2 \left(c_k^i - b_k c_l^i \right) \quad (9)$$

$$\begin{aligned} \frac{\partial E(i)}{\partial W_{m,h}} &= \sum_{h=1}^{nhd} \frac{\partial E}{\partial c_{kh}^i} f' \left(\mu_h \right) W_{m,h} c_{km}^i \left(1 - c_{km}^i \right) x_m^i \\ &= \sum_{h=1}^{nhd} \sigma_{mh}^i W_{m,h} c_{km}^i \left(1 - c_{km}^i \right) x_m^i \\ &= \sigma_{h}^i x_m^i \end{aligned} \quad (10)$$

Figure 10. Image segmentation process.

4) W, b Update Process:

$$W_{m,h} = W_{m,h} - \delta \frac{\partial E(W,b)}{\partial W_{m,h}} \quad (11)$$

$$= W_{m,h} - \delta \sum_{i=1}^I \sigma_h^i x_m^i$$

$$b_m = b_m - \delta \frac{\partial E(W,b)}{\partial b_m} \quad (12)$$

$$= b_m - \delta \sum_{i=1}^I \sigma_h^i$$

5) Cycle Iteration:

Repeat steps 1) to 4) until the loss function $E(W,b)$ is within the tolerance of the error, and then the optimal weight parameter are obtained. The FCN network segmentation process is shown in Fig. 10.

4 Results and analysis

Submarine small target sonar images data set which contains 1800 training images and 25 test images are used in the experiment. Two groups comparative experiments are carried out, the first experiment compared FCN-MSE and FCN-RLF methods; the second experiment adds two classic image segmentation methods Otsu and FCM for comparison.

4.1 FCN network training results

In FCN network, the image misclassification rate decreases with the increase of the training iterations number during the training process, and tends to be stable finally.

The training results of the two methods are shown in Fig. 11 and Fig. 12.

Figure 11. The error rate of the FCN-MSE model over epoch number

Figure 12. The error rate of the FCN-RLF model over epoch number

In Fig. 11, the lowest training and testing error rate are 0.140 and 0.160 respectively. After algorithm improved, the lowest training and testing error rate are 0.086 and 0.122 respectively, which are lower than FCN-MSE method.

4.2 Sonar image segmentation results and analysis

4.2.1 The first group experiment

5 pieces of 25 test images are chosen to analyze the segmentation results, the sonar image segmentation results for the first group experiment are shown in Fig. 13.

Table 2. PixelAcc experiment results

FCN-MSE	9.16	53.08	58.48	67.71	66.44	65.58	64.73	65.12	64.8	64.62	64.11	64.58	64.36
	63.26	63.4	63.5	63.2	62.83	62.5	65.52	62.09	62.36	62.66	62.62	62.38	*
FCN-RLF	9.22	54.76	60.35	69.45	67.66	66.37	65.67	65.87	65.4	64.99	64.8	64.88	64.7
	63.53	63.6	63.26	63.27	62.97	62.64	62.71	62.38	62.62	62.99	62.94	62.69	*

Table 3. MeanAcc experiment results

FCN-MSE	99.49	92.47	91.56	92.08	92.45	92.47	92.55	92.1	91.83	91.25	90.96	91.02	90.88
	90.46	90.49	90.49	90.54	90.34	90.32	90.41	90.38	90.34	90.23	90.21	90.11	*
FCN-RLF	99.49	92.47	91.56	92.08	92.45	92.47	92.55	92.1	91.83	91.25	90.96	91.02	90.88
	90.46	90.49	90.49	90.54	90.34	90.32	90.41	90.38	90.34	90.23	90.21	90.11	*

Table 4. MeanIU experiment results

FCN-MSE	9.45	65.17	72.8	83.06	82.89	81.66	80.84	80.45	79.69	78.48	77.21	77.5	77.14
	76.52	76.38	76.19	76.04	75.7	75.52	75.46	75.65	75.65	76.03	75.87	75.68	*
FCN-RLF	9.44	66.34	74.07	84.22	83.63	82.06	81.35	81.13	80.23	79.05	77.58	77.82	77.48
	76.75	76.53	76.25	76.05	75.67	75.48	75.46	75.22	75.68	76.14	75.99	75.86	*

As can be seen from the segmentation results: The proposed method can segment the target outline and preserve the target details much better. Although the proposed method also misclassifies the reverberation area due to the existence of speckle noises, the improvement is obvious.

Figure 13. From left to right are: The original sonar image; Ground truth segmentation results; FCN-MSE method segmentation results; FCN-RLF method segmentation results;

Figure 14. PixelAcc detail results.

The first group experiment compares PixelAcc, MeanAcc, and MIU of FCN-RLF and FCN-MSE segmentation methods. The relationship between the three evaluation indexes are shown in Fig. 14-16, Table 2-5 are the evaluation index values of the twenty five images, and Table 3 is the mean value of the experimental data results.

Figure 15. MeanAcc detail results

Figure 16. MeanIU detail results

Table 5. The average of experimental results

	MeanIU	PixelAcc	MeanAcc
FCN-MSE	0.6235	0.9002	0.7574
FCN-RLF	0.6269	0.9133	0.7686

In table 5, the MeanAcc and PixelAcc of the FCN-RLF method are 0.7686 and 0.9133 respectively, which are 1.12% and 1.46% higher than the FCN-MSE method. The segmentation results are much better.

4.2.2 The second group experiment

In the second experiment, Otsu and FCM methods are added for comparison. The segmentation results are shown in Fig. 17. The precision and recall values are shown in Table 6,7.

Figure 17. From left to right are the segmentation results of OTSU, FCM, FCN-MSE and FCN-RLF.

Table 5. Precision

Otsu	FCM	FCN-MSE	FCN-RLF
0.7548	0.5548	0.9911	0.9970
0.8365	0.9133	0.9722	0.9743
0.7963	0.4255	0.9000	0.9068
0.8126	0.6339	0.9677	0.9727
0.8253	0.6485	0.9678	0.9694

Table 6. Rcall

Otsu	FCM	FCN-MSE	FCN-RLF
0.9923	0.9932	0.9832	0.9976
0.9873	0.987	0.997	0.9911
1	1	1	1
0.9979	0.9978	0.999	0.999
0.9983	0.9989	0.9972	0.9971

It can be seen that the recall of the segmentation methods is generally higher, Indicates that the methods used are all suitable for sonar images segmentation, however, the Otsu and FCM segmentation methods are affected by speckle noise seriously. The Otsu segmentation precision rate is up to 83.65%, but it is also lower than the FCN-MSE and FCN-RLF segmentation

methods where the convolutional neural network is added, the precision of the two method can reach 99.11% and 99.7% respectively. The improved method has a strong robustness to the sonar image speckle noise, and can improve the accuracy of sonar image segmentation effectively.

5 Conclusions

A submarine sonar image segmentation method based on FCN model with relative loss function is proposed in this paper. The experimental results show that the improved method has better robust to speckle noise than the traditional sonar image segmentation methods. Network training error reduced 31.14% compared with FCN-MSE method. The edges and details of the target image can be better preserved. The averaged segmentation accuracy and precision reached 91.33% and 96.4% respectively, which meets the sonar image processing accuracy requirements. The proposed method plays an important role in the special task of marine target recognition.

Since the network used in this paper belongs to the deep network, the introduction of deconvolution makes the network structure more complicated. Therefore, improving the segmentation accuracy and simplifying the network structure will be the major research direction in the future.

References

1. Mignotte, M.; Collet, C.; Perez, P.; Bouthem, P. Sonar image segmentation using an unsupervised hierarchical mrf model. *IEEE Transactions on Image Processing*, 9, 1216-31, Publisher Item Identifier S 1057-7149(00)05669-4, (2000).
2. Chabane, A. N.; Islam, N.; Zerr, B. Incremental clustering of sonar images using self-organizing maps combined with fuzzy adaptive resonance theory. *Ocean Engineering*, 2017, 142, 133-144. Available online: <http://dx.doi.org/10.1016/j.oceaneng/> (accessed on 05 July 2017).
3. Li, Q.; Ma, G.; Huo, G.; Yan, Z.; Sheng, H. New segmentation method of side-scan sonar image based on edge detection in NSCT domain. *Chinese Journal of Scientific Instrument*, 34, 1795-1801, DOI:10.19650/j.cnki.cjsi.2013.08.016, (2013).
4. Collet, C.; Thourel, P.; Perez, P.; Bouthem, P. Hierarchical MRF modeling for sonar picture segmentation. In *Processings of the 3rd IEEE International Conference on Image Processing*; Lausanne, Switzerland, September pp. 979 - 982,(1996).
5. Samiee, K.; Rad, G. A. R. Textural Segmentation of Sidescan Sonar Images Based on Gabor Filters Bank and Active Contours without Edges. In *Processings of the IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, pp.3-8,(2008).
6. Huo, G.; Yang, S. X.; Li, Q.; Zhou, Y. A Robust and Fast Method for Sidescan Sonar Image Segmentation

- Using Nonlocal Despeckling and Active Contour Model. *IEEE Transactions on Cybernetics*, 47, 855-872, DOI: 10.1109/TCYB,(2016)
7. Le, Q. V.. Building high-level features using large scale unsupervised learning. In *Processings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8595-8598,(2013)
8. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D. Going deeper with convolutions. In *Processings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9,(2015)
9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Computer Science*, 580-587, DOI: 10.1109/CVPR.2014,(2013)
10. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. Eprint Arxiv, arXiv:1312.6229v4,(2013)
11. Long, J.; Zhang, N.; Darrell, T. Do Convnets Learn Correspondence?. *Advances in Neural Information Processing Systems*, 2, 1601-1609, arXiv:1411.1091v1,(2014)
12. Zhang, N.; Donahue, J.; Girshick, R.; Darrell, T. Part-Based R-CNNs for Fine-Grained Category Detection. 8689, 834-849, arXiv:1407.3867v1,(2014)
13. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278-2324,(1998)
14. Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Processings of the International Conference on Neural Information Processing Systems*, pp. 1097-1105,(2012)
15. Wang, L.; Guo, S.; Huang, W.; Qiao, Y. Places205-VGGNet Models for Scene Recognition. *Computer Science*, arXiv:1508.01667v1,(2015)
16. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D. Going deeper with convolutions. In *Processings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9,(2014)
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition.,770-778, arXiv:1512.03385v1, (2015)
18. Shen, D.; Wu, G.; Suk, H. I. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19, 221,DOI: 10.1146/annurev-bioeng-071516-044442,(2017)
19. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In *Processings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.,(2015)
20. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Back propagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541-551, DOI: 10.1162/neco.1989.1.4.541, (1989).