Research Advance in Deep Learning Image Segmentation Algorithms

Junhua Shao* and Qiang Li

Research Institute, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China

*Corresponding author e-mail: shaojunhua@mail.lzjtu.cn

Abstract. Based on the problem of the decline of the accuracy of image segmentation, this paper proposes a image segmentation technology based on the deep learning. Thorough segmentation of images in the depth of network structure can be achieved by convolution and pooling, activation function, empty convolution, transpose convolution, loss function, and segmentation of network and other steps, and simulation and accuracy test were finished.

Keywords: Deep Learning, Image Segmentation, Convolution

1. Introduction
With the existence of massive image information, image processing technology has gradually become a topic with increasing interest to people [1]. Reasonable image segmentation and preservation of local images have become an urgent problem to be solved. Aiming at this problem, this paper proposes an image segmentation algorithm based on deep learning, aiming to ensure the accuracy of pixel points in the image to the maximum extent, so as to facilitate utilization for people and viewing.

2. Overview of Key Technology

2.1 Deep Learning Technology
A common structure of Convolutional Neural Networks (CNN) [2] is a feedforward neural network with back propagation and consists of basic artificial neurons, and learning and substantial parameters and bias constants were required. As an independent unit, each neuron makes an effective response to the input data within the perception range and outputs a series of values to represent the probability of classification [3-4]. The default input of the convolutional neural network is image, and the network structure is adjusted according to the input data of specific structure, so that the structural information of the data is fully utilized to improve the efficiency of the feedforward function, reduce parameters and shorten the time of model training. According to the spatial structure characteristics of images, the convolutional neural network neurons are designed as three-dimensional structures with three dimensions of width, height and depth, where depth is the depth of neurons. Before and after the convolution between the network layer in network is progressive, the relationship between the output of each layer is the next layer of input [5], which means each convolution layer extracts the current characteristics of the most effective transmission to the next layer, with the increase of layer depth...
(network), the extracted features more and more high-level and abstract, eventually forming the abstract description of the whole image.

2.2 Image Segmentation Technology

Fully Convolutional Networks (FCN) is an image segmentation network model proposed by Jonathan Long et al from UC Berkeley, which creatively realizes end-to-end image segmentation for the first time, restoring image semantics from abstract image features and outputting the categories of each pixel [6]. After a convolution and pooling operation, the size of the output result is reduced by one point. In order to restore the output result to the size of the input picture and upsample the output result, deconvolution is used here to realize the upsampling effect. Deconvolution operation is carried out on the output of the last layer of FCN network to restore the size of the segmentation result to the size of the input image. However, due to the loss of precision, some image details cannot be restored, resulting in the lumpiness of the output classification result. In order to repair details as much as possible, and improve the segmentation effect, the output of the convolution layer was respectively made with the same multiples of deconvolution based on the the same times of contraction and according to result accuracy of different multiples on sampling image segmentation, it is known that the smaller the better segmentation result will be received.

FCN does not show limit to the size of input images, and the size of training set and test set images is not required to be uniform. FCN is also more efficient, without the problems caused by the use of pixel blocks, which avoids repeated operation and repeated storage of convolution [7-8]. At the same time, FCN also has an obvious disadvantage: although the up-sampling technique improves the segmentation effect, the segmentation accuracy is still far more enough.

3. Image Segmentation Algorithm Based On Deep Learning

3.1 Image Preprocessing

In order to enrich the training set, better extract features, and prevent the overfitting of the generalization model, it is often necessary to preprocess the data set [9] to complete the data enhancement.

For image data, geometric transformation method is usually adopted to enhance data. The main methods are as follows:

1. Rotation transformation: To rotate the image randomly by a certain angle, so that the orientation of the target object in the image changes;
2. Flip transformation: To flip the image along the horizontal or vertical direction;
3. Zoom transformation: To enlarge or shrink the image according to a certain proportion;
4. Translation transformation: The image is translated in a certain way to achieve the purpose of changing the target position. The direction and distance of translation can be defined manually or generated randomly;
5. Random cutting: To select a position on the image at random, and select a size at random for cutting. Generally, the size of the cutting is greater than half of the original image, and try to ensure that the cutting results include the target object;
6. Scale transformation: To enlarge or shrink the length and width of the image respectively according to the specified proportion, and change the size or blur degree of the image content;
7. Contrast transformation: To change the contrast of the image, in normal conditions, to change the value of brightness and saturation in the HSV mode image without changing the hue value. The change of illumination can be simulated by the exponential calculation of maximum brightness and the change of brightness.
8. Noise disturbance: To randomly select the position in the image, and randomly change the value of the pixel in the position by using Gaussian noise and salt and pepper noise;

3.2 Image Segmentation
For image segmentation technology, the main process includes steps such as convolution and pooling, activation function, empty convolution, transpose convolution, loss function, and etc. The specific flow chart is shown in Figure 1.

![Flowchart](image)

**Fig. 1** Image segmentation flow chart

First of all, convolution and pooling operations are carried out to extract image features through convolution verification [10]. The parameters of convolution kernel are obtained through training and learning, and the value of $F$ is determined by the model depth. Namely, the larger the size of the convolution kernel is, the less the image eigenvalue is. The calculation formula of eigenvalue is as follows:

$$N = \frac{W - F + 2 \times P}{S} + 1$$

Besides, $W \times W$ refers to the size of the image, $F \times F$ is the size of the convolution kernel, $S$ denotes the moving pixel value, $P$ is the image filling width, that is, the patch width.

The main functions of pooling in deep learning technology are as follows:

1. Reducing the input characteristic dimension value, and then reducing the number of parameters and their calculations to avoid overfitting problems and ensure that the model is easy to be trained.

2. To ensure that the image transformation and redundancy has no deformation.

3. To ensure the rapid detection of pixels as much as possible.

The pooling operation in this paper adopts maximum pooling, that is, the maximum value of all the eigenvalues in the image neighborhood is obtained and taken as the eigenvalue of the pixel. The operation flow chart for maximum pooling is shown in Figure 2.
Fig. 2. Maximum pooling operation diagram

As can be seen from the figure, maximum pooling can adapt to small changes in values, so it has translational invariance. At the same time, it is most sensitive to the eigenvalue obtained by the convolution kernel, and can quickly obtain the eigenvalue information required by users. Secondly, activation function is mainly used to solve nonlinear problems. The activation function used in this paper is shown in Figure 3.

Fig. 3 Activation function diagram

As can be seen from the figure, as the input signal is 0, the output results are all 0; When the input signal is greater than 0, the output is equal to the input signal. The random gradient descending bracelet is going to be zero. As , the function is in a hard saturated state, while as , the function is in an unsaturated state, thus speeding up the fast training of deep learning.

However, in order to make the output result close to 0, when is input, the output result is set as a function , where is the smallest value as far as possible. It is obtained based on prior knowledge, so as to ensure the rationality of the output result.

In order to increase the number of pixel points and ensure the accuracy of eigenvalue extraction, this paper uses void convolution to increase the range of receptive field that the eyes can see. Assuming that the convolution kernels are , and the convolution step size is 1, then the size of the pixel block used to activate the unit is , that is, the receptive field has a linear relationship with the number of image layers. In order to facilitate the processing of multi-dimensional images, the definition of two-dimensional space void convolution is as follows:

\[ (F \ast k)(P) = \sum F(s)k(t) \]

\( F \) represents the input two-dimensional signal, and \( S \) is the domain of definition; \( K \) is the kernel function, \( T \) is the domain; \( L \) signifies the coefficient of void convolution, and \( P \) is the domain. So, \( s + lt = P \) denotes the convolution condition.

In the calculation process of void convolution, the relationship between void convolution and
receptive field is shown in Fig. 4.

![Fig. 4 Relationship between void convolution and receptive field](image)

According to the figure, the sizes of all convolution kernels are $3 \times 3$. (a) represents the receptive field when the void convolution coefficient is 1, and the eigenvalues obtained are the same as those obtained in ordinary images. (b) represents the receptive field when the void convolution coefficient is 2. The size of image area of the receptive field is $7 \times 7$, and 9 key pixel points are obtained. (c) represents the receptive field when the void convolution coefficient is 4, the image area of the receptive field is $15 \times 15$, and the range of the receptive field is increased.

Through convolution and pooling, the image size is reduced. For the end-to-end output of images, the transpose convolution operation is needed. The forward propagation process of transpose convolution is the back propagation process of convolution.

Assuming that the convolution check image is used for the transpose convolution processing with step size of 3. The steps are as follows:

1. Step 1. Convolution is used to check each pixel for convolution processing, and $4 \times 4$ feature maps can be obtained.
2. Step 2. Fusing the feature images according to the specified step size. The relative positions of the feature images are relative to the positions of the original pixels, and the pixel differences between different feature images are directly summed up.
3. Step 3. Assuming that the input size is $L_i$, the convolution kernel size is $K$, and the convolution Step size is $S$, then the output result is $L_o$. The calculation formula is as follows:

   $$L_o = (L_i - 1) \times S + k$$

Considering the problem of image distortion, this paper uses the calculation of variance loss function to reduce this problem. The calculation formula of the loss function is as follows:

$$C = \frac{(y-a)^2}{2}$$

The calculation formula of variance loss function is as follows:

$$\frac{\partial C}{\partial w} = (a - y)\sigma'(z)x = a\sigma'(z)$$

$$\frac{\partial C}{\partial b} = (a - y)\sigma'(z) = a\sigma'(z)$$

The updated calculation of $w$ and $b$ values is as follows:

$$w \leftarrow w - \eta \times \frac{\partial C}{\partial w} = w - \eta \times a \times \sigma'(z)$$

$$b \leftarrow b - \eta \times \frac{\partial C}{\partial b} = b - \eta \times a \times \sigma'(z)$$

As can be seen from the above formula, $\eta \times a \times \sigma'(z)$ is usually close to 0.
Finally, there is the problem of network layer segmentation, which can also be understood as the classification of each pixel. Through convolution and pooling, it can be obtained that the dimension of the convolution kernel determines the dimension of the output result, while in the network layer, each pixel should be classified and the output pixel-level classification processing should be completed. First, to execute the Sigmoid function for the classification task, and the calculation formula is as follows:

\[ \sigma(z)_j = \frac{e^{z_j}}{\sum e^{z_i}} \]

The calculation results are mapped to the space (0,1) to complete the dichotomy processing, that is, the k-dimensional vector-value \((a_1, a_2, ..., a_k)\) is mapped to a vector, and \(b_i \in (0,1)\) and \(\sum b_i = 1\) \(i \in \{1,2, ..., k\}\).

4. Performance Analysis
In the training process, the variation trend curve of the model's accuracy rate on the verification set is shown in Fig. 5.

![Fig. 5 Trend of accuracy](image)

As can be seen from the figure, the accuracy and loss values of each round of training model on the verification set are recorded in the picture, and 50 iterations are one round. As can be seen from the figure, as the number of iterations increases, the accuracy of the model on the verification set keeps increasing, while the loss keeps decreasing. The rate of change is fast at first and then slow down. After the number of iterations reaches the 600th round, there is almost no changes. We set the total number of iterations to be 40,000, but as can be seen from the figure, it stops when the number of iterations reaches 30,000. It shows that the training stop strategy plays a role, that is, the training model stops training when the validation machine fails to achieve better accuracy for five consecutive times.

5. Conclusion
Based on the deep learning technology, this paper studies the image segmentation technology, completes all the image segmentation processing operations through steps such as convolution and pooling, activation function, void convolution, transpose convolution, loss function and segmentation network layer, and finally to verifies the accuracy of image segmentation technology.

References
[1] Li Mengyao. Image enhancement based on MATLAB. Information Recording Materials, 2020, 21(03): 237-238. (in Chinese)
[2] Shi Yali, Tang Liang. Target detection system based on fusion of radar and video. Electronic
Technology and Software Engineering, 2020(05): 76-77.

[3] Ke Pengfei, Cai Maoguo, Wu Tao. Face recognition algorithm based on improved convolutional neural network and ensemble learning. Computer Engineering, 2020, 46(02): 262-267+273.

[4] Zhang Jiaming, Wang Xiaoman, Jing Wenbo. Speech emotion recognition based on deep convolutional network and spectral image. Journal of Changchun University of Science and Technology (Natural Science Edition), 2020, 43(01): 76-81.

[5] KONG Mingxin. Research on Radar Emitter Identification Algorithm Based on Deep Convolutional Neural Network. Hangzhou Dianzi University, 2020.

[6] Li Xuan, Sun Xinnan. Image segmentation algorithm based on convolutional neural network. Journal of Shenyang Aerospace University, 2020, 37(01): 50-57.

[7] Yang Zhen, Wang Jun, Xin Chunhua. Automatic segmentation and classification of MR images using DCE-MRI combined with improved convolutional neural network. Journal of Chongqing University of Technology (Natural Science), 2020, 34(02): 147-157.

[8] Zhang Yongpeng, Liu Yunpeng, Wang Renfang, Sun Dechao, Jin Ran, Dong Chen. Overview of medical image segmentation. Electronic World, 2020(03): 85-86.

[9] Hongyuan Yao, Haipeng Wang, Xueyuan Lin, Xinlong Pan. Research on improved image preprocessing algorithm in remote sensing image Mosaic. Computer & Digital Engineering, 2020, 48(02): 428-432.

[10] Xie Yuan, Miao Yubin, Zhang Shu. Blade segmentation algorithm based on Cavity Full Convolutional Network. Modern Computer, 2020(06): 88-92.