Advances in Intelligent Systems Research, volume 153
International Conference on Applied Mathematics, Modeling and Simulation (AMMS 2017)

Automatic Segmentation Method of Garment Figure Based on Convolutional Neural Network

Chong Chen
Fashion Institute, Shanghai University of Engineering Science, Shanghai 201620, China

***Abstract*—Aiming to the problem that the rapid acquisition of clothing style map requires manual participation and takes a long time and heavy task, an automatic image segmentation method based on convolution neural network is proposed. Firstly, we denoised normalized and semantically annotated the sequence image in order to produce the data set. Then, we trained the convolution neural networks, which were improved with fusion multi-scale feature and residual connection, and obtained the optimized convolution neural network segmentation model. Finally, it loaded the pre-segmentation image into the optimized model to get the normalized mask pattern, and used the cubic spline interpolation methods to restore the resolution and result of the HD segmentation with the original mask. In this paper, the results of Photoshop segmentation are as the reference standard. The experimental results show that the accuracy of the method is close to the reference standard, and the batch segmentation can be realized automatically, which can solve the heavy problem of the target segmentation task in 3D reconstruction.**

Keywords-computer vision; image segmentation; convolutional neural networks; multi-scale feature fusion; residual connection

I. INTRODUCTION

In 3D reconstruction based on image sequences, the accuracy and speed of target segmentation play a decisive role in the quality and efficiency of reconstruction. Image segmentation is widely used in military, remote sensing, meteorology, medicine and other fields, and its main difficulty is in image processing. Domestic and foreign experts put forward four kinds of methods for the specific application of image processing: threshold segmentation, edge detection segmentation, region extraction and segmentation combined with the specific theory[1]. At present, there is still no suitable method for automatic segmentation. At present, Photoshop is used to segment the image sequence and reduce the efficiency of 3D reconstruction through manual participation. This paper presents a method of automatic segmentation based on improved convolutional neural network image sequences. The data set is used to train a convolutional neural network with improved multi-scale features and integrated residual connectivity to obtain an optimized convolutional neural network segmentation model. The pre-segmentation image is normalized, compressed and loaded into the optimized segmentation model. Through the convolutional network forward transfer to obtain the target segmentation mask, and through the cubic spline interpolation method to improve the resolution of the split mask to the size of the original image with the original map custom masking operation, and access to high-resolution segmentation result. In this paper, the segmentation result of Photoshop is taken as the reference standard. The experimental results show that this method is close to the reference standard in terms of the accuracy, realize automatic batch segmentation and better solve the problem for the target segmentation 3D reconstruction.

II. RELATED THEORY

A. Convolution Neural Network

Convolution neural network is a deep neural network model with sparse connection and weight sharing. In recent years, it has been widely concerned by various scientific fields. Figure 1 is a schematic diagram of a convolutional neural network for image segmentation tasks: mainly by the convolution layer, the pool layer and deconvolution layer[2]. In convolutional layers, the convolution kernel has the same effect as the filter. Since the RGB color mode image is equivalent to a two-dimensional matrix per channel, the convolutional layer convolutes the convolution kernel and the input image through a sliding window channel to extract different types of features. They are called feature maps and convolution kernels. They have the same number of features and the number of features. The sub-sampling layer, also known as the pooling layer, reduces the dimensionality by pooling the operation data, reduces the size of the input data and reduces the calculation amount. Usually, the calculation method includes the largest pool, the average pool, and the random pool. The convolutional network is optimized by the error backpropagation algorithm, and the convolution kernel weights are updated regularly. The degree of network optimization is determined by the error and accuracy of the convolutional network model in the data set.

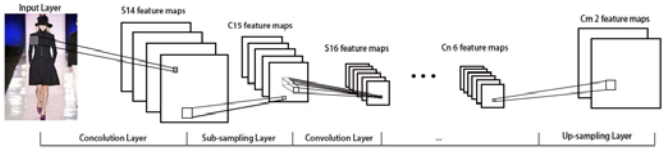

FIGURE I. CONVOLUTION NEURAL NETWORK STRUCTURE

B. Residual Connection

In depth learning, the most important factor to obtain excellent performance is the depth of convolutional networks. Deep networks can extract higher-level characteristic information. However, the resulting gradient dispersion caused the network to fail to converge, even network degradation. Network-level Increasing will lead to greater error. The performance is not degenerated when the depth increases by

superposing the xy's identity mapping on a shallow network, theoretically allowing the training of any depth network, which is essentially independent of the depth network[3]. Figure 2 shows the residual connection.

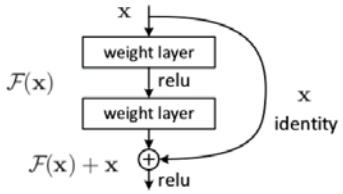

FIGURE II. RESIDUAL LEARNING DIAGRAM

The output equation with residual connection to the network is as follows:

$$y = F(x, \{W_i\}) + x$$

Where x and y are the inputs and outputs, respectively, $F(x, \{W_i\})$ represents the learned residual function[4]. The above formula holds when the input and output dimensions of the residual connection are the same, and requires a linear transformation to achieve the same dimension when the dimensions are different:

$$y = F(x, \{W_i\}) + W_s x$$

The use of residual connections is to ensure that the neuron's overall output is closer to the original input, and the main features of the maximum retention, so that the convolution of the network is approximately equal to the map, and to minimize the error.

C. *Cubic Spline Interpolation*

The cubic spline interpolation method is the most effective interpolation method for calculating pixel values at present. We fit 4 × 4 adjacent pixels in the vicinity of the source image, and then take the cubic spline value corresponding to the target pixel as the value of the target image. Through pixel interpolation can enhance the resolution of the image. The actual calculation requires the introduction of boundary conditions to complete the calculation. The boundary usually has a natural boundary (the second order of the boundary points is 0), the boundary of the boundary (giving the derivative of the boundary point), the non-kink boundary (the third-order guide of the two ends are equal to the third-order guide of the neighboring point).

III. IMAGE SEGMENTATION BASED ON OPTIMIZED CONVOLUTIONAL NEURAL NETWORK

Based on the optimization of convolutional neural network image segmentation framework is divided into two phases:

- Neural network training phase. After the preprocessing of the original image, the dataset is made and then input to the optimized convolutional neural network (multi-scale feature fusion and integrated residual connection) for optimization to obtain an optimized image segmentation neural network model.
- Automatic segmentation batch picture stage. The normalized compressed pre-segmentation image is input to the optimized convolutional neural network segmentation model. After propagating forward, the target segmentation mask-S is output and the resolution is restored to the original image size to obtain a mask-B, which is used to obtain a high-resolution segmentation result by a custom mask operation.

A. *Data Set Creation*

This article takes the model fashion figure as the research example, and the clothing surface has rich and meticulous patterns. We choose 40 sets of clothing. On average, each garment has 30 to 60 different angles of the image. The specific steps to make a data set are as follows:

- Denoising. The V channel in the HSV color space model indicates the brightness of the color. Therefore, Gaussian smoothing is performed on the V channel to reduce the noise. The pictures before and after processing are shown in Figure 3 (a) and 3 (b), respectively.
- Normalize operation. After denoising, the less relevant background portion of the image is removed and normalized to the same size as the normalized image as shown in Figure 3 (c).
- Data annotation. The normalized image was semantically annotated to form a label map by using Photoshop software, where in the target area is labeled as 1. The background area is labeled as 0, and the labeled area is labeled as Figure 3 (d).

FIGURE III. DATA SET IMAGE

- Amplify the data set. By rotating the normalized image and its label image, the dataset is expanded by 8 times the original number.

Because the clothing has a variety of complex details of the composition of the target detection interference caused by the use of the current image segmentation is more difficult to solve, this paper uses deep convolution neural network[5] to solve the problem.

B. *Optimized Convolution Neural Network*

1) *Multiscale feature fusion*

Since the convolution kernel size is fixed in the existing convolutional neural network structure, smaller objects tend to be ignored when features are extracted through larger

convolutional operations. When the convolution kernel size is small, the result is not continuous. In this paper, the method of multi-scale features is integrated, and the global and local information images are integrated to enhance the robustness of the target scale. Since different scale feature maps contain different scales of the detail, low-frequency (low-resolution) images provide target contour position information and high-frequency (high-resolution) images containing more detail information. Figure 4 shows the convolutional network structure of multi-feature fusion. Assuming that the size of the input image is m, the concrete steps for implementing multi-scale feature fusion are as follows:

- The input image is averaged and is input to the neural network. After pooling, a three-dimensional characteristic map of m, m / 2 and m / 4 is obtained .
- The three-dimensional feature maps are respectively input to the three Base Net branches in the (b) stage to extract the feature maps of the corresponding features.
- The feature map of scale m / 4 obtained by Base Net-3 was up sampled to m / 2 on the deconvolution and pixel-by-pixel merged by the feature map of Base Net-2 to obtain the m / 2 feature map.
- The characteristic map of m / 2 obtained by (3) was sampled to m by deconvolution and merged with the feature pattern output by BaseNet-1 by pixel-wise pixel addition to obtain a mask map with a scale of m.

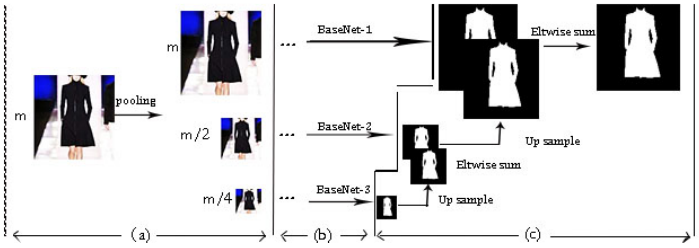

FIGURE IV. CONVOLUTIONAL NETWORK STRUCTURE

2) Integration residual connection

In order to solve the problem of information loss and loss caused by the increase of the convolutional network depth, we use the residual connection to transfer the input information directly to the output through the branch. Even if some neurons are not activated and transferred to the next new layer, to ensure the integrity of the information, to a certain extent, it can solve the problem. The following is the concrete realization of the residual connection:

- The output of PRE is used as input for the residual connection.
- The output of the residual connection is combined with the POST-Conv output by adding the pixels for the POST input.
- Since the series of convolutions connected by each residual does not change the size of the graph before and after convoluting, the size of the convolution kernel in the residual connection is 3, and the padding space and the step size are 1.
- Since the number of POST-Conv features in the network structure is twice the PRE output, the number of convolution kernels in the residual connection is set to 2N.

C. Automatic Sequence Image Segmentation

The automatic segmentation of sequence images is based on the optimized segmentation model, which realizes the automatic segmentation of targets through the characteristics of convolutional networks and realizes the automatic process through continuous input and output. Due to limited hardware computational power, inputting directly into the neural network directly raw images of higher resolution often leads to network crashes because of variables that go beyond the data type. Therefore, according to the method of Figure 4, the high-resolution map is compressed and normalized, then input into the optimized convolutional neural network segmentation model. Then, the original image resolution is restored to the original image resolution of the convolutional neural network by cubic spline interpolation. Finally, the masking map of the high-resolution target segmentation for the result B is obtained through the mask operation. Figure 5(a) is the original image; (b) is a mask view of enhanced resolution-B, where the background area is 0; (c) operates as a mask for high-resolution target segmentation results.

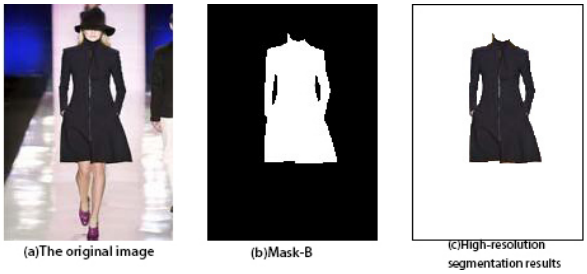

FIGURE V. SPLIT RESULTS

The neural network divides a target image when every time it propagates forwardly. Therefore, after several consecutive images enter the neural network and propagate multiple times in succession, multiple segmentation mask images are continuously output. All image sequences in a group of objects are segmented. Then the high resolution segmentation results are obtained by resuming the resolution and mask operation, so as to realize the automatic segmentation of the target image sequence.

IV. EXPERIMENT AND ANALYSIS

Experimental hardware environment for the CPU E3-1245 v5, GPU Nvidia quadro m2000, memory 32G. Programming language Python 3.5. Use Tensorflow for neural network training and learning[5]. Stochastic gradient descent method to solve optimization, VGGNet-D network initialization, neural network initial learning rate: 10^{-9}[6].

In this paper, the convolutional neural network is designed to minimize the loss function, which is a measure of the overall

gap between experimental results and a reference standard. In the network training optimization process, as the number of iterations increases, the loss value gradually decreases, and the accuracy rate gradually increases. When both ends are stable, the neural network is fully trained and trained in a convergent state. Network training status is determined by comparing the loss curve with the accuracy curve. When the experiment iteration is about 50,000 times to achieve the best condition.

In this paper, we use the accuracy P and recall R, F1-score three indicators to assess the accuracy of image segmentation. The artificial result of Photo Shop is taken as a reference standard and compared with the more classic semantic segmentation method in image segmentation. Table 1 is the result of comparison of segmentation results. As can be seen from Table 1, the coefficients of the proposed method in all three indexes are higher than those of semantic segmentation.

TABLE I. SEGMENTATION EFFECT CONTRAST

	P	***R***	***F1-score***
Semantic segmentation	0.8999	0.8097	0.8546
This Peper	0.9565	0.8877	0.9205

In order to more intuitive comparison of segmentation results, Figure 6 is for the Photoshop segmentation, semantic segmentation method and the results of this comparison.

Figure 6 (a) is the artificial segmentation result of the experimental reference standard, and (b) is the segmentation result of the semantic segmentation method. This method can remove most of the background outside of the target area, but can not accurately segment the background within the target area, and the segmentation edges produce a significant discontinuity. (c) Segmentation results for this article, in contrast to the segmentation results in (b), avoid large target loss and discontinuity margins. For a better distinction between goals and backgrounds, the separation of goals is relatively complete, and to ensure that the segmentation effect and details are close to artificial segmentation standards.

From the experimental analysis, we can see that this method can simultaneously improve the segmentation speed of the image sequence, and meet the target segmentation accuracy of 3D reconstruction at the same time, to achieve batch automatic segmentation, and solve the problem of large target segmentation workload in 3D reconstruction. Because this method has a good ability to migrate, as long as the mission data set training, you can get better results.

V. CONCLUSIONS

Aiming at the problem that the rapid acquisition of clothing style map requires manual participation and takes a long time and heavy task, an automatic image segmentation method based on convolution neural network is proposed. On the one hand, this paper integrated use of local and global information through the integration of multi-scale functions; on the other hand, in the process of extracting residual connection features, it can add missing information under the supplementary neural network. Finally, the cubic spline interpolation is used to improve the resolution and mask operation, and the high-resolution segmentation result is obtained. The paper takes Photo Shop's segmentation results as a reference standard for comparison. Experimental results show that the accuracy of segmentation using this method is close to the reference standard, which can quickly and accurately auto-segment the auxiliary functions such as style drawing and style drawing recognition. However, due to limited hardware computing power, resulting in reduced precision after image scaling, edge segmentation is not enough precise. Therefore, how to ensure the accuracy of the image in the case of automatic segmentation is the focus of future research.

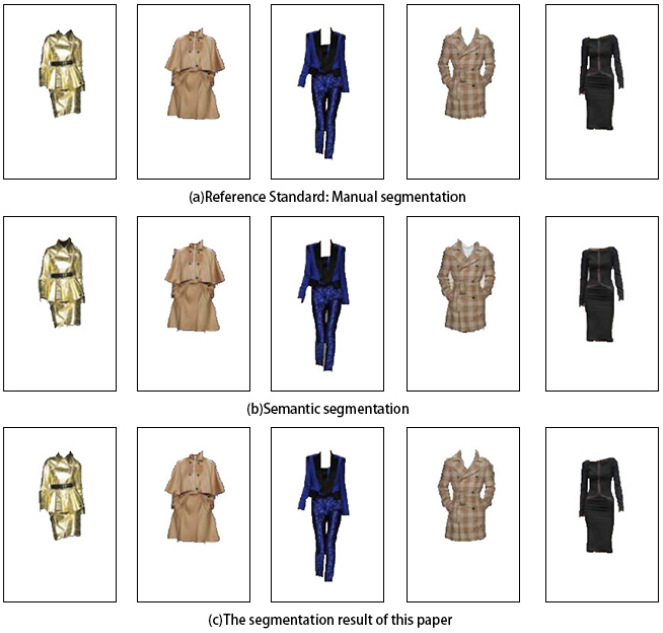

FIGURE VI. COMPARISON OF THE THREE METHODS OF SEGMENTATION

REFERENCES

[1] Gonzalez R C, Woods R E. Digital image processing [M] .3rd ed. Nguyenqiu Qi, Ruan Yuzhi, translated. Beijing: Publishing House of Electronics Industry 2015.

[2] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [C]// Proc of European Conference on Computer Vision. [S. l.] : Springer International Publishing, 2014: 818-833.

[3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. ar Xiv preprint ar Xiv: 1409. 1556, 2014.

[6] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. ar Xiv preprint ar Xiv: 1207. 0580, 2012.