Summary of the apparel collocation recommendation system based on deep learning

Tian Jiaxin[1]
Information Engineering Institute
Beijing Institute of Fashion Technology
Beijing, China
617724382@qq.com

Guo Fei[3*]
Digital media and interactive Key Laboratory
Beijing Institute of Fashion Technology
Beijing, China
lunwen95@163.com

Du Yang[2]
Information Engineering Institute
Beijing Institute of Fashion Technology
Beijing, China
18310236998@163.com

Peng Shengqiong[3]
Digital media and interactive Key Laboratory
Beijing Institute of Fashion Technology
Beijing, China
ysylab@bift.edu.cn

***Abstract*—With the booming development of apparel e-commerce, there is a huge amount of clothing image data on the Internet. Because the semantic attribute cannot accurately describe and locate the garment image, the recommendation system based on deep learning has been developed rapidly. This paper describes a large amount of relevant literature, summarizes the research achievements of various universities in related fields in recent years, and summarizes the general steps of using deep learning to process clothing images, such as target positioning, feature processing, multi-attribute classification, etc., as well as the realization method of clothing retrieval and collocation system based on deep learning.**

Keywords—deep learning; image processing; cross-domain retrieval; collocation recommendation

I. INTRODUCTION

In recent years, the Internet and e-commerce have been booming, and online shopping has become the main form of living consumption for many people. People can buy all kinds of desirable commodities through mobile phones and computers without leaving their homes. Online shopping has attracted more and more consumers with its complete categories, high quality, low price and convenience, and gradually becomes the preferred shopping channel. At the same time, apparel e-commerce has also achieved unprecedented development.

With the advent of the era of big data, the amount of multimedia data has increased dramatically, and the huge amount of clothing information makes it impossible for consumers to quickly and accurately locate the goods they are interested in. This is also because the basic attributes of clothing on the e-commerce platform are mostly based on text labeling technology. The labeling process is very tedious and there is no uniform standard for text labels. The labeling and maintenance process mainly relies on the subjective understanding of merchants, which lacks certain professionalism and accuracy. In recent years, many scholars have turned to the study of features based on image contents, which can directly read the low-level features of images. With the rapid increase of computer processing speed and the greatly reduced cost of computer hardware, deep learning has become more and more popular. The depth image processing combined with deep learning and image processing can further extract the abstract features of images and describe the image information more comprehensively. When these technologies are applied to the personalized recommendation system of clothes, users will enjoy more high-quality recommendation of clothes matching, and the e-commerce platform of clothes will usher in better development.

II. IMAGE PROCESSING BASED ON DEEP LEARNING

Image processing using deep learning generally follows several steps:

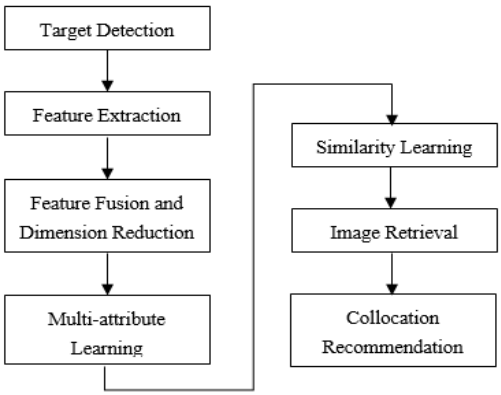

Fig. 1. *Image processing process based on Deep Learning*

A. Target Detection

Nanjing Information Engineering University Li Zhi [11] in the task of target detection for fashion and human body parts, is

proposed based on FAST-RCNN and FASTER-RCNN framework of garment parts. Both of them are basically similar in recognition accuracy, and FASTER-RCNN has reached the speed of real-time detection. It can also locate the clothing position and remove the background in the complex picture background, which improves the classification results of clothing styles. Ao Hongfei [13] of Beijing University of Posts and Telecommunications used convolutional neural network to optimize the performance of the system. In view of the fact that traditional face detection algorithms cannot detect face occlusion, two kinds of deep network models, Faster-RCNN and SSD, are trained using WIDER FACE data set to solve the problem of face occlusion in garment area location.

B. Feature Extraction, Fusion and Dimension Reduction

Zhao Guangming [2] of Northeastern University used an 8-layer AlexNet deep convolutional neural network model to extract clothing image features. On the one hand, the matching recommendation is made according to the image characteristics of clothing by looking for the nearest neighbor of clothing matching. On the other hand, the images of the upper and lower clothes were merged. The characteristics of clothes were extracted through the AlexNet model, and the matching was modeled according to the extracted features and the matching given by experts. Chen Qijin [4], a researcher from Zhejiang University, gave the feature extraction technology and feature description method of clothing image from the four aspects of color, texture, shape and local feature, and proposed the extraction of clothing skeleton and node optimization algorithm based on constrained triangulation, which expressed the basic shape structure and shape of clothing in a concise form. Zheng Senlie of Sun Yat-Sen University [7] used deep learning to automatically extract features, which were more representative of pictures than those obtained by previous artificial feature extraction methods, and could greatly extract the accuracy of picture classification.

Literature [4] combined IF-IDF lexical weight, SIFT feature quantization accuracy and local word frequency, proposed a method for the setting of Bundled feature weights, and proposed a method for the calculation of multi-bundled feature similarity by weighting the Bundled features in the garment image and combining the proportion of Bundled features in the image. Bao Qingping [1] of Zhejiang University used SIFT descriptor as a traditional feature extraction method. When using the convolutional neural network to extract the features, the fusion feature map of the convolution layer and the full connection layer is used as the feature representation. In order to reduce the sparsity of feature vectors, the dimensional reduction is carried out by using PCA. Fan Yuhang [5] of the University of Electronic Science and Technology expressed costume semantic information and global information well based on the fusion features of depth features and local features. The features extracted by convolutional neural network and the SIFT features encoded by BoF are fused, and then the fusion features are reduced and maintained by the denoising autocoder. In literature [13], Ao Hongfei improved the network structure of ResNet50, proposed the concept of multi-layer Feature fusion with multi-layer features, enhanced the network's Feature sensitivity to small-scale images, and improved the network performance. In literature [4], HSV-T feature space was constructed based on the color and texture fusion features of garment images. Dimensionality reduction technology and clustering technology were used to define the clothing style visual space.

C. Multi-attribute Learning and Classification

Traditional image classification methods generally require two steps: first, artificial feature extraction, and then classification learning based on artificial feature extraction. In the convolutional neural network, there is no need for artificial feature extraction. Feature extraction and classification learning are trained together, and parameter estimation is conducted through random gradient descent. At the same time, features extracted based on convolutional neural network are more advantageous than those extracted manually, and have stronger image recognition ability. Although convolutional neural network has achieved good results in the field of image, its interpretability is still a problem.

In literature [1], Bao Qingping adopted multi-task learning and designed a convolutional neural network model combined with multi-task learning in accordance with the requirement of recognizing multiple semantic attributes simultaneously in clothing images. In literature [5], Fan Yuhang proposed a model of clothing attribute multi-label classification based on convolutional neural network, which combines clothing attribute classification with convolutional neural network to express the detailed characteristics of clothing by training clothing attribute classifier. Zheng Senlie [7] designed and implemented a multi-label classifier for images based on deep learning, and realized a web-based automatic labeling system for clothing pictures. The training set and test set in LeveIDB or LMDB database format were converted from the original multi-label image, which accelerated the training of the classifier. Chang Chunhwan, Zhou Cailan and Liang Yuan [9] proposed an optimized convolutional neural network (convolutional neural network) garment classification algorithm based on residuals, which can improve classification accuracy and have a faster image processing speed. Three optimization methods are used in the network :(1) adjust the order of the batch normalization layer, activation function layer and convolution layer in the network; (2) parallel pooling structure of "pooling layer + convolution layer"; (3) replace the full connection layer with the global mean pooling layer. The research has obtained the best classification accuracy rate known at present on the DeepFashion data set. Liu Weiwei of Beijing University of Posts and Telecommunications [8] proposed a model that can learn the representation of continuous vector of clothing attributes, so as to change the clothing attributes from a binarized 0-1 value to a distributed continuous vector, and each segment of continuous vector represents one attribute of clothing image. The effect of attribute learning is further clarified in the visual analysis of the learning model.

D. Similarity Measure Learning

Similarity learning aims at learning an appropriate similarity measure, which can reduce the distance between similar samples and increase the distance between different

samples, so as to improve the classifier's classification ability, optimization matching and clustering performance. In literature [1], in order to improve the system's robustness on factors such as background environment and deformation in images, metric learning is introduced, specifically, Siamese or Triplet structures are introduced in the convolutional neural network. Similarity measurement learning method is also used in literature [5] to optimize clothing characteristics. Based on the Triplet similarity measurement learning method, a dual-convolution neural network structure based on Triplet Loss was proposed, and the triad training set was constructed to optimize the network parameters, so that the features of cross-scene retrieval of clothing were more matched, and the ranking of results was more reasonable.

E. Garment Image Retrieval Algorithm

Bao Qingping [1] used the Faster-RCNN algorithm to detect clothes for garment images, which improved the accuracy of retrieval. K-means is used to cluster feature vectors to speed up the retrieval speed. Literature [4] has constructed an inverted sequence index structure of clothing images and realized the clothing retrieval technology, which is effective for clothing images generated under the conditions of clothing deformation, occlusion and complex background. Literature [5] adopts the random Kd tree algorithm to realize the fast retrieval of clothing features. The random Kd tree algorithm is a kind of ANN, which is based on the realization of Kd tree. However, it is different from Kd tree that increases the search node to improve the search accuracy. The random Kd tree improves the search effect by improving the relative independence of the search node. Liu Weiwei [8] proposed a retrieval scheme of partial attribute replacement. After the first retrieval of the garment image, when the user proposed a new requirement for a certain attribute, the vector of the corresponding attribute in the image was replaced with the attribute vector of the user's requirement. While the rest was unchanged, the newly generated image features were used for retrieval. Li Zhi [11] constructed the hash index by using the method of iterative quantization, and generated the hash code. Experiments showed that the hash index of convolution features had the highest accuracy. Ao Hongfei [16] built a cross-domain image retrieval system based on the scene of film and TV drama, and realized the cross-domain matching function of clothing, with the accuracy rate of TOP 1 reaching 87.3% and TOP5 reaching 94.7%.

F. Clothes Matching Recommendation

Zhao Guangming [2] of Northeastern University constructed the recommendation model of clothing collocation from two aspects. On the one hand, he searched for the pattern of clothing collocation through unsupervised learning, and constructed it by mining association rules and describing the correlation degree of the content of clothes. On the other hand, it constructs the model of clothing collocation through the way of supervised learning, and makes full use of the data of the combination package of clothing collocation given by experts to construct the supervised learning model. Literature [4] constructed a clothing collocation data set using TPO collocation principle, scanned the clothing collocation set with Apriori algorithm, searched for frequent collocation item set, and mined association rules, and finally proposed an intelligent clothing collocation technology for ordinary users.

III. DYNAMIC SCENE RECOGNITION BASED ON DEEP LEARNING

A. Target Detection and Feature Analysis

Chen Yanjie of Beijing University of Posts and Telecommunications uses the structure of the most accurate Faster-Rcnn's RPN+fast for reference, and simultaneously detects clothes of the upper body, lower body and whole body. The model's upper body, lower body and full body clothing are precisely positioned. Then, the background and skin were removed with the Grabcut, and the clothing information was saved completely and the redundant information was removed. Feature extraction is carried out for the garment after positioning. By using the deep learning network to extract features as the basic framework, residual units and attention units are introduced to enable the algorithm to screen invalid image areas and concentrate the expression of features on the garment itself. Chen Donghao [18] used pedestrian detection and face detection in the study to locate the model, adjusted the size of the detection return area, established the Gaussian model of mixed channel to remove the background through frame video, and removed the model skin color through color space transformation.

Tu Qingqing [3] of the University of Electronic Science and Technology puts forward a method of fashion style recognition based on key points, which combines the geometric features of external contour and the internal topological structure. The edge information is extracted by Canny operator and the contour is corrected and connected. Then the skeleton of clothing is extracted through the key points of collar, waist, shoulder, sleeve, chest, trunk and so on. The clothing style features are described comprehensively. The problem of fashion style recognition is solved. The method of exact contour extraction of models in complex dynamic scene is used to solve the problem of feature analysis of network video clothing.

B. Dynamic Garment Matching and Recommendation

In literature [3], Tu Qingqing established a hierarchical evolutionary fashion multimedia information filtering method to reduce the input information by layers, gradually reduce the interference of irrelevant information in the network, and filter out network multimedia information directly related to apparel information and with high attention, so as to improve the efficiency of system analysis. Chen Yanjie [6], considering the real-time performance of online fashion show video, optimized the inter-frame voting and video segment, the model acceleration method based on mimick improved the real-time computation efficiency of the system. Chen Donghao [14] also used the interframe voting mechanism of video to stabilize the output of video retrieval results. In addition, the clothing matching system under the video walking scene was successfully constructed, and accurate matching was completed. The average TOP1 accuracy rate reached 95.2%, and the average TOP5 accuracy rate reached 99.7%. Moreover, the similarity measurement network was further trained, and the

cross-domain retrieval problem of street photography was migrated to the automatic clothing similarity recommendation under the scene of video catwalk.

IV. Other autonomous annotation systems based on deep learning

Tang Denglong [10] of Donghua University proposed an autonomous development method of clothing style based on DBN, which is well adapted to the characteristics of strong and rapid development of clothing style personalization and solves the problem of autonomous classification of clothing style personalization. DBN neural network algorithm is used to classify the clothing line drawing, and a model based on the clothing training DBN autonomous development is proposed. This model can train a large number of target line drawing, and the accuracy is improved to some extent. The deep belief network is decomposed into a series of restricted Boltzmann machines composed of two adjacent layers of mountains. The training parameters are layer by layer to ensure that feature vectors are mapped to different feature Spaces, so as to retain feature information as much as possible, initialize the deep belief network, and fine-tune the whole network with BP neural network for better identification.

Wang Shanna [12] of Zhejiang university of science and technology, by taking advantage of the basic characteristics of tie flower patterns and focusing on the aesthetic evaluation of tie flower patterns by convolutional neural network, proposed a high-level aesthetic classification method based on parallel convolutional neural network and an emotional labeling method based on parallel convolutional neural network. The subjective aesthetic evaluation of the necktie pattern was carried out, the manual aesthetic features were designed and extracted, and the model was obtained by inputting the parallel convolutional neural network training, so as to realize the classification of the high and low aesthetic feeling of the tie pattern image. The emotional labeling of the tie pattern image is carried out, and the manual emotional features are designed and extracted. The model is obtained by inputting the parallel convolutional neural network training, so as to realize the emotional labeling of the tie pattern image.

V. Development trend

The deep network is still changing the field of computer vision, but the amount of data needed to train the deep network is relatively large, and the construction of image annotation and data set is the bottleneck in the research. The training time of deep learning is too long. If the training speed can be improved, the practicability of deep machine learning will be greatly improved. In addition, the physical meaning of the knowledge representation learned in deep machine learning is very unclear. If the knowledge learned in each layer can be represented as the knowledge with physical significance, it will increase the comprehensibility of the knowledge learned.

At present, the existing clothing collocation method is implemented based on the given clothing collocation data set, and it is generally recommended for most people. However, each user has his own unique aesthetic and dressing style, and the user-oriented dressing matching technology can recommend different dressing matching cases according to the user's own needs and habits. In the following research, users' personal information can be added into the clothing matching technology, and quick 3D experience effect can be added into the interface design. The system presents clothes to users in the form of three-dimensional model, which can greatly improve the user experience and enrich the system.

Acknowledgment

This article is one of the research achievements of the 2017 Beijing Municipal Education Commission Science and Technology Program on the surface of the project (SQKM201710012006), Innovative design of garment ergonomics and apparel functions Beijing key Laboratory (KYTG02170202), and Digital media and interactive Key Laboratory of Beijing.

References

[1] Bao Qingping. Classification and Retrieval of Garment Image based on Deep Learning [D]. Zhejiang University, 2017.

[2] Zhao Guangming. Research and implementation of Garment matching recommendation algorithm based on Deep Learning [D]. Northeast University, 2016.

[3] Tu Qingqing. Research on Intelligent Fashion recommendation system based on key Point Garment style recognition [D]. University of Electronic Science and Technology, 2014.

[4] Chen Qijin. Research on Garment Retrieval and collocation based on Image content [D]. Zhejiang University, 2013.

[5] Fan Yuhang. Research on Garment Retrieval and collocation Technology based on Deep Learning [D]. University of Electronic Science and Technology, 2017.

[6] Chen Yanjie. Online clothing Retrieval and recommendation based on depth Network [D]. Beijing University of posts and Telecommunications, 2018.

[7] Zheng Senlie. Design and implementation of automatic Apparel Picture labeling system based on Deep Learning [D]. Sun Yat-sen University, 2015.

[8] Liu Weiwei. Research and Application of clothing attribute representation based on continuous Vector [D]. Beijing University of posts and Telecommunications, 2017.

[9] Zhang Zhenhuan, Zhou Cai-lan, Liang Yuan. Optimal Convolutional Neural Network Garment Classification algorithm based on residual error [J]. Computer Engineering and Science, 2018 40 (02): 354-360.

[10] Tang Denglong. Research and implementation of clothing style Self-development based on DBN [D]. Donghua University, 2014.

[11] Li Zhi. Clothing attribute Research based on Deep convolution Neural Network and its Application [D]. Nanjing University of Information Engineering, 2016.

[12] Wang Shanna. Research on fabric Aesthetic Classification and emotional labeling based on convolution Neural Network [D]. Zhejiang University of Technology, 2018.

[13] Ao Hongfei. Clothing matching system based on Deep Learning for Film and Television Pictures [D]. Beijing University of posts and Telecommunications, 2018.

[14] Chen Donghao. Clothing matching and recommendation in Video Walking scene [D]. Beijing University of posts and Telecommunications, 2017.