Shirt Semantic Classification Based On Convolutional Neural Network

Bo Qiu¹, Xiang Liu¹, Yunyu Shi¹ and Binjie Xin ²

¹School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, China
²Fashion College, Shanghai University of Engineering Science, China
xliu@sues.edu.cn

Abstract. With the rapid development of e-commerce, a large number of clothing images have flooded into the Internet replacing the keyword search with a graph search has become a new trend. However, the inaccurate semantic classification of clothing leads to the low accuracy of graph searching and the poor retrieval effect. In order to solve this problem, this paper proposes a parallel shirt-style classification algorithm which greatly improves the accuracy. This paper mainly completed the following work: Firstly, the shirt sample library was established by means of questionnaire survey and expert voting. Then, the mapping mechanism was used to extract the ROI area and construct the shirt-attention branch network. Last, the backbone is combined with the branch to make up the ShirtNet. The experimental results show that the introduction of parallel network makes the classification accuracy rate up to 91% which is much improved compared with traditional machine learning and convolutional neural networks.

1. Introduction

In recent years, the development of e-commerce industry and the transformation of cultural trends have prompted people to pay more attention to the style, emotion, semantics and other aspects of clothing. As one of the most popular upper garments, shirt occupies a large part of online shopping. However, some merchants lack professional knowledge and misuse modifiers that do not match clothing, resulting in consumers doing improper keyword search, low accuracy in graph search method and finding goods more difficultly. Therefore, extracting valuable style semantic features from shirts has become the top priority of current shirt recognition.

Experts previously conducted research on clothing classification based on machine learning, factor analysis and other methods. Gao et al.¹ improved the shortcomings of using only shallow features by extracting and combining multiple color features of clothing for classification. Lu et al.² firstly used linear interpolation to remove the image background, and then used LBP operator to extract color and texture features of clothing in HSV space for classification and recognition. The relationship between the collar angle and the angle of the collar surface was studied by Qiao et al.³ to calculate and subdivide the collar angle change according to the average value. Admittedly, these traditional machine learning and manual extraction methods effectively implement clothing classification, but there exists shortcomings such as low recognition accuracy and low generalization ability. The development of deep learning has prompted to a qualitative leap in clothing classification. Wang et al.⁴ proposed a method of merging handmade emotion features by transferring the AlexNet structure to the
emotional analysis of tie patterns. Liu et al.\(^5\) based on the region's fully connected convolutional network analyzed the efficiency and effectiveness of clothing image classification, and developed an algorithm to identify suit targets of various shapes and sizes.

This paper designs ShirtNet that can effectively classify shirt styles. Compared with traditional CNN models, it has the advantages of high accuracy. Firstly, combining the three aspects of current product semantics, cognitive psychology, and perceptual engineering\(^6\), the modifiers of the currently suitable shirt style were determined. Through expert screening and determination of style categories, a shirt style image library was established. Then, according to the principle of the mapping mechanism, the region of interest features of the sample tie was extracted. Finally, a parallel network was established to identify the shirt style accurately.

2. Algorithm Flow

The experiment has three main steps: building a shirt style library, constructing a Shirt-attention network, and a shirt backbone network, as shown in Figure 1.

![Figure 1. Algorithm flow](image)

1) Build a sample library of shirt styles. This paper collects existing shirt samples in the market, judging them by the experts of fashion Institute and evaluating by the experts from the School of Fashion, combined with questionnaire to distinguish shirt styles.

2) Establish parallel branches. Aiming at the region of interest in the shirt collar, a mapping mechanism was used to build a shallow shirt-attention network.

3) Construct the ShirtNet backbone network. In this paper, the shirt-attention network and backbone are combined to build a parallel network. The Adam algorithm was used to train until the network successfully converges.

3. Define shirt modifiers

Based on the previous researcher's approach, this paper collected mainstream shirt styles of e-commerce websites and uses a questionnaire survey to preliminarily screen out 10 different styles. The experts of the School of Clothing subdivided the items based on function, occasion, style, and collar type\(^7\), and they are shown in Table 1. Colors, styles, collars, patterns and other factors are blended to divide the shirt into the following 6 categories: "formal", "college British", "leisure", "fashion", "fresh", "personality".

| Categories       | Colours       | simple | plain | multicoloured | Regular graphics | Sparse stripes | Elegant colours | complicated patterns |
|------------------|---------------|--------|-------|---------------|------------------|----------------|------------------|----------------------|
| formal           |               | ✓      |       |               |                  |                |                  |                      |
| college British  |               | ✓      |       |               | ✓                |                |                  |                      |
| leisure          |               |       | ✓     |               |                  |                |                  |                      |
| fashion          |               |       |       |               | ✓                | ✓              |                  |                      |
| fresh            |               |       |       |               |                  |                | ✓                |                      |
| personal         |               |       |       |               |                  |                |✓                |                      |
4. Network Frame

4.1 Shirt_Attention Network
Shirt-Attention is a multi-layer antecedent network including convolutional layers, pooling layers, fully connected layers and so on. Among them, the receptive field of each neuron is the upper local receptive field. The calculation formula is as follows:

\[ Z^l = W^l * A^{l-1} + b^l \]  
\[ A^l = ReLU(Z^l) \]  

Where, \( W^l \) represents the weight matrix of the lth layer, \( A^{l-1} \) represents the output matrix of the (l-1)th layer, and \( b^l \) represents the offset matrix of the lth layer and outputs \( A^l \) through the lth layer activation function. In this paper, the nonlinear activation function ReLU is used, as in equation (3) is as follows:

\[ f(x) = \max(0, x) \]  

The additional maximum pooling layer is the feature mapping layer, which not only preserves the effective features in the receptive field, but also plays a role in dimensionality reduction.

4.2 Parallel Network
The parallel ShirtNet classification network constructed in this paper is shown in Figure 2. The entire model consists of two parallel branches. The ShirtNet backbone network contains four convolutional layers, two maximum pooling layers. The Shirt-attention branch includes two convolutional layers and a pooling layer. After the parallel branch features are flattened, network fusion is performed and the recognition results are output.

![Figure 2. Structure of parallel ShirtNet](image)

5. Experiment

5.1 Sample Library
Experts from the Institute of Clothing Technology vote on existing samples. More than half of the experts made the same evaluation, and then put the style into one category to divided into six styles
finally: "formal", "academic England", "casual", "fashion", "fresh", "personality", which constitutes the shirt style library, as shown in Figure 3.

5.2 Training Results

In this paper, data enhancement methods such as up-and-down inversion and mean deviation are used to expand the sample set. This paper uses 1000 pictures as training samples and 200 pictures as validation sets. The experimental parameters are shown in Table 2. The loss and accuracy of the training set and validation set are shown in Figure 4.

Table 2. Model parameters

| Batch-size | Learning-rate | GPU       | Epochs |
|------------|---------------|-----------|--------|
| 16         | 0.0001        | Nvidia-1080Ti | 80     |

5.3 Ablation experiments

In order to further verify the superiority of ShirtNet, three main networks were selected for ablation experiments. The experimental results are shown in Table 3. The recognition accuracy of VGG-16 is very poor. The fundamental reason is that the model has too many layers and is not friendly enough for small sample data sets. The accuracy of LeNet-5 has reached 68%. The model contains only two convolutional layers and one pooling layer, which is not enough to learn semantic features leading to underfitting such as high precision of training set and low precision of verification set. AlexNet reduces the number of convolution layers compared with vgg-16, and uses a more stable activation function ReLU compared with lenet-5, further improving the accuracy. Compared with other models, the precision of the proposed network improves greatly, reaching 91%. The modification of the backbone network coupled with the dynamic capsule network can well cope with small sample tasks.

Figure 3. Shirt style sample library
Figure 4. Loss and Accuracy in different epochs

Table 3. The style recognition accuracy of mainstream network structure.

| CNN model | precision |
|-----------|-----------|
| LeNet-5   | 68%       |
| AlexNet   | 76%       |
| VGG-16    | 59%       |
| ShirtNet  | 91%       |

6. Conclusions
In this paper, a parallel ShirtNet based on shirt style recognition is constructed. Experimental results prove that the network has higher accuracy than other CNN structures under the same data set. The parallel mapping mechanism can be used to extract the region of interest for more complex style recognition, and this paper proposes a new thinking for recognizing clothing style. ShirtNet can effectively improve the accuracy of graphical search with large data sets and reduce the incidence of misuse of labels by merchants who lack professionalism and misdescribed products by buyers.

Acknowledgments
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project is sponsored by the National Natural Science Foundation (61876106).

References
[1] Gao M, Du Y, Ai H, et al. A hybrid approach to pedestrian clothing color attribute extraction[C]// Iapr Int. Conf. on Machine Vision Applications. IEEE, 2015. (Tokyo National Museum of Emerging Science and Innovation)
[2] LI Yue ling, WU Guoping, GENG Xiuxiu, DU Xiaojun. Clothing Image Retrieval Based on LBP-GLCM Texture Feature Extraction[J] Video Engineering | Video Eng, 2015, 39(12):99-103.
[3] QIAO Tiantian, JI Yanbo, ZHOU Jie. Sensory evaluation on men’s shirt collars with different angle design[J]. Journal of Xi’an Polytechnic University, 2019, 32(02):123-128.
[4] WANG Shanna, ZHANG Huaxiong, KANG Feng. Emotion classification of necktie pattern
based on convolution neural network [J]. *Journal of Textile Research*, 2018, 39(08):117-123.

5. LIU Zhengdong, LIU Yihan, WANG Shouren. Depth learning method for suit detection in images [J]. *Journal of Textile Research*, 2019, 40(04):158-164.

6. LV Jia, CHEN Dongsheng. Evaluation of clothing emotion based on customer's psychological cognition [J]. *Journal of Textile Research*, 2015, 36(09):100-107.

7. Sabour S, Frosst N E, Hinton G. *Dynamic routing between capsules*[EB/OL]. (2017) [2019]. https://arxiv.org/abs/1710.09829.