Journal of Fiber Bioengineering and Informatics 11:4 (2018) 209–216
doi:10.3993/jfbim00319

Classification of Female Apparel using Convolutional Neural Network

Qiao-Qi Li [a], Yue-Qi Zhong [a,b,*], Xin Wang [a]

[a] *Donghua University, Shanghai 201620, China*

[b] *Key Lab of Textile Science and Technology, Ministry of Education, China*

Abstract

With the vigorous development of clothing e-commerce, the amount of clothing image data on the internet has increased dramatically. A tedious effort was required to manually label and classify the semantic attributes of clothing images. Manual marking is time-consuming and laborious, so a method of automatic classification using convolutional neural networks was studied. In this paper, a female cloth dataset consisting of 10 types of female clothing was built. Convolutional Neural Network (CNN) was employed to learn the feature vectors for each type. Five different types of architectures, including ResNet50, Inception-v3, and VGG-19, AlexNet, and FashionNet were used for performance comparison. Experimental results have shown that Inception-v3 possesses the highest accuracy (98.07% for training and 96.91% for testing) in clothing classification compared with other methods.

Keywords: Female Clothing Image; Image Classification; Convolutional Neural Network; Deep Learning

1 Introduction

There is a rapid growth of e-commerce. Apparel online retail is a large part of that which account for about 25% of overall Internet transactions [1]. This phenomenon has led to a huge number of online fashion products data. Correctly labeling these data can facilitate applications such as similar clothing retrieval and relevant products recommendation. The rapid development of the e-commerce industry has brought convenience to our lives, but at the same time, it has put forward new and higher requirements for traditional search: identifying and classifying objects in pictures, and searching, one of the more extensive problems is the identification of clothing types. The traditional recognition and classification technology rely on manual or digital image processing, pattern recognition, through image detection and segmentation, feature extraction and classification recognition, etc. The disadvantage is that the extracted features are mostly characterized by the specified features, and can not describe all the images well. The properties of objects in the middle, and often require large calculations, etc., are time-consuming and labor-intensive, and are not suitable for calculation on portable devices.

*Corresponding author.
Email address: zhyq@dhu.edu.cn (Yue-Qi Zhong).

Dec. 2018

Currently, in major e-commerce shopping platform, the method to label fashion products mostly requires prior fine classification of clothing images and affix the corresponding label. At present, most of the methods of labeling are manual labeling. In the environment where the number of garment images is exploding, the method of manual labeling has many drawbacks, and there are several serious disadvantages as follows: 1) Keywords can only describe abstract features that are easy to extract and abstract. It can not fully reflect the visual characteristics of clothing images; 2), People need accurate keyword text annotation for each image. However, the number of images is huge and requires a great deal of manpower and resources for manual annotation; 3) Due to the subjectivity of human cognition, different people may have different understandings about the same image. When image keyword text is annotated, subjectivity and uncertainty, thus affecting the classification results. By contrast, clothing classification based on image content can save lots of human labor and also give more objective results. The final goal of this project is to be able to auto label or score each product based on one or more unconstrained clothing images. In this work, the experimental target was focused on female apparel images.

Various algorithms have been proposed to solve image-based clothing classification problem. The current research progress is as follows: Alem [2] classified clothing into 23 categories based on attitude estimation and color, SIFT and other features. Bourdev [3] developed a system to describe the appearance of people. In addition, for garment segmentation, Hu [4] proposed the use of foreground and background estimates based on the Constrained Delaunay Triangulation (CDT), which does not require any pre-defined garment models. Weber [5] and others proposed the use of attitude detectors to solve the problem of clothing occlusion. Manfredi [6] proposed a clothing segmentation algorithm based on popular store dataset. There were also some researchers dedicated to the classification of clothing attributes, such as color, collar, sleeves or other attributes. Chen [7] introduced a fully automatic system that produces garments with a list of named properties. Lorenzo-Navarro J [8] evaluated the LBP and HOG descriptors' ability to classify clothing attributes. The aforementioned work used handcraft features to address the classification problem.

Deep Learning is an evolving technology and also is able to solve apparel image classification efficiently. In recent years, the feature extraction method based on deep learning has received great attention and has a good application effect in natural image classification. It can effectively solve the problem of clothing image classification, and minimize the abstraction process on data features. Reduce human intervention. In 2012, Krizhevsky Alex et al. [9] used CNN to win the ImageNet Image Recognition Contest and took the second place. At present, CNN has achieved remarkable results in various image processing tasks. For example, the Deep ID model proposed by Sun et al. [10] had achieved more than 99% accuracy in face recognition. Simonyan et al. [11] proposed two-stream CNN that combines two factors of time and space for motion recognition and got a good result. VGG-NET [12] achieved first and second place respectively in 2014 ILSVRC positioning and classification. GoogLeNet [13] won the ILSVRC Classification Challenge in 2014 with a top 5 error rate of 6.67%. GoogLeNet proposed a new model component named "Inception Module", which is a NIN (Network In Network) structure. Simply by increasing the number of layers did not further improve network performance, and sometimes will be counterproductive. To solve this problem, Microsoft Research Institute proposed Deep Residual Network (Residual Network) in 2015 [14]. ResNet increased the network layers to 152. At this time, the training process of the whole network could be normalized and achieved a better performance.

Lao et al [15] trained Convolutional Neural Network and obtained 50.2% and 74.5% accuracy respectively in the types of clothing and attribute classification. Bao [1] adopted a Convolu-