*ICIC Express Letters*
*Part B: Applications*
Volume **7**, Number **6**, June **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1371–1376

# DRESS IMAGE RETRIEVAL BASED ON IMPROVEMENT OF DEEP LEARNING

Yang Zhang[1] and Jinhua Fu[2,3,*]

[1]College of Mathematics and Information Science
[3]School of Computer and Communication Engineering
Zhengzhou University of Light Industry
No. 5, Dongfeng Road, Zhengzhou 450002, P. R. China
*Corresponding author: fujinhuazz@qq.com

[2]State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou Information Science and Technology Institute
Zhengzhou 450000, P. R. China

Received February 2016; accepted April 2016

Abstract. *Online dress shopping is becoming an increasingly popular way of shopping and nowadays image-based dress retrieval plays an important role in identifying their favorite items for customers. In this paper, we present an efficient retrieval of dress image via improvement of deep learning. The deep learning is usually realized by convolutional neural networks (CNNs). We design the CNN by four sublayers followed by an output, which gives a retrieval of dress image. Further, we improve CNN by activation function to induce the sparsity in the hidden units. The deep network can be efficiently trained using the improvement. The results show that the matching results can be efficiently acquired, which proves the effectiveness of our method such as accuracy and computation complexity.*

**Keywords:** Dress shopping, Dress retrieval, Convolutional neural networks, Computation complexity

## 1. Introduction.

Recently, online dress shopping is becoming an increasingly popular shopping model and the online shopping market has been expanded greatly. Image-based dress retrieval plays an important role in building an online recommendation system and offers a more convenient way for customers to identify their favorite items.

There is a large body of researches on clothing segmentations, recognition and retrieval in the past. Two efficient applications to clothing image [1,2] are proposed to give different processes of clothing image, but they only focus on segment clothing for grouping images. Ferrari and Zisserman [3] present a probabilistic generative model of visual attributes, together with an efficient learning algorithm. Using semi-automatic system, they represent clothes by combinations of attributes that describe various characteristics of clothing images. Then, an automated system [4] is proposed that is capable of generating a list of nameable attributes for clothes on human body in unconstrained images. Liu et al. [5] address a practical problem of cross-scenario clothing retrieval – given a daily human photo captured in general environment. However, they can only swap a region from an image for the corresponding region from another image. Similarly, Yamaguchi et al. [6] find similar styles from a large database of tagged fashion images and use these examples to parse the query. They develop the similar image retrieval to increase the accuracy of clothing segmentation. On the other hand, Fu et al. [7] address the problem of large scale cross-scenario clothing retrieval with semantic-preserving visual phrases (SPVP) and the SPVP significantly enhances the discriminative power of local features with a slight increase of memory usage. However, these works require indexing of patches and region segmentation as a preprocessing. Therefore, they are incapable of dealing with arbitrary user input which strides across multiple patches. More recently, Mizuochi et al. [8] develop a novel clothing retrieval system considering local similarity, where users can retrieve their desired clothes which are globally similar to an image and partially similar to another image. Wang et al. [9] compare a variety of color feature extraction methods and similarity measure methods. They present an image querying and retrieval system based on color feature. However, the retrieval rationality should be still improved via more features of clothing image. Nowadays, deep learning is essentially the neural network 2.0 revolution, which has shown its more powerfulness in feature representation from a large corpus, which is being applied to visual feature recognition, extraction and matching.

Recently, there has been a rise in GPU-accelerated (Graphics Processing Unit) algorithms in machine learning thanks to the rising popularity of deep learning algorithms. Deep learning via CNNs is a collection of algorithms for various problems in machine learning. It involves computationally intensive methods, such as convolutions, Fourier transforms, and other matrix-based operations in which GPUs are well-suited for computing. In this paper, we improve the CNN by an efficient activation function, and design deep learning for discriminating feature representations capable of identifying different clothes images and implement the effective retrieval of clothing image based on GPU.

The rest of this paper is organized as follows. The convolutional neural network is designed in Section 2. Then the improved convolutional neural network is related via activation function in Section 3. Finally, experimental results are discussed in Section 4, and conclusions are drawn in Section 5.

## 2. Design of Convolutional Neural Network.

The convolutional neural network is designed as shown in Figure 1. From Figure 1, our CNN consists of a series of sublayers, namely a convolution (filter bank) sublayer, a non-linearity sublayer, a local response normalization sublayer, and a feature pooling sublayer followed by an output. Each type of layer contains several feature maps, or groups of neurons, in a rectangular configuration.

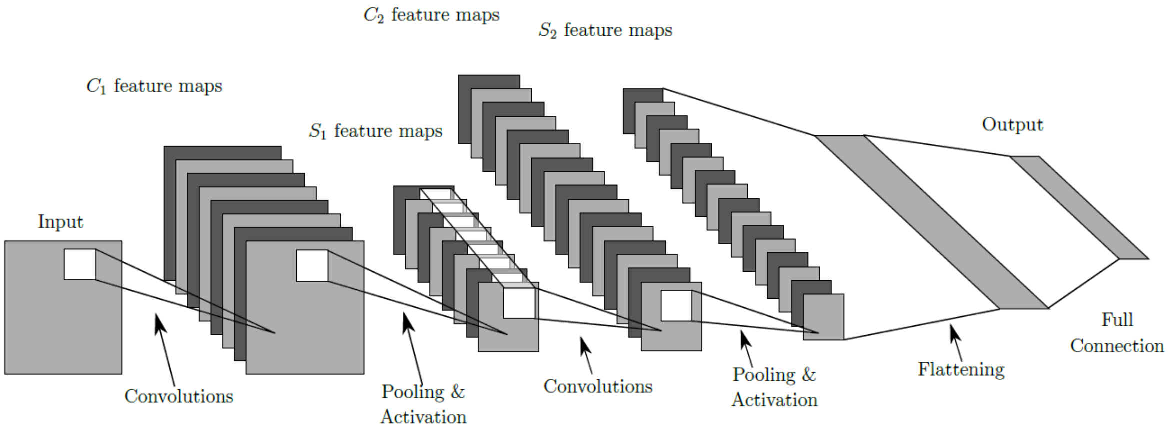

Figure 1. Model of convolutional neural network

### 2.1. Convolution sublayer.

The relationship between convolution and correlation is given in Equation (1), where $*$, $\star$ represent convolution and correlation, respectively.

$$A * \widetilde{B} = A \star B \tag{1}$$

where $A$ and $B$ must be two inputting vectors.

Equations (2) and (3) show the relationship of the output $a$, input $h$, weight $W$ and bias $b$. Each filter will identify, hopefully different and important, features of the image.

$$a = h \star W + b \tag{2}$$

$$a(x,y,z) = \sum_{p,r,s} h(x+p, y+r, s) W(p,r,s,z) + b_z \tag{3}$$

where $h$ is the input image having dimensions $n_{iy} \times n_{ix} \times n_c$. $W$ and $b$ are trainable parameters. $p$ and $r$ are offsets in the rows and columns of the image. $s$ is the offset in the channel of the image. $z$ indicates the layer of the image.

2.2. **Sub-sampling layer.** The sub-sampling layer produces down-sampled versions of the input maps. If there are $N$ input maps, then there will be exactly $N$ output maps, although the output maps will be smaller. More formally,

$$x_j^\ell = f\left(\beta_j^\ell down\left(x_j^{\ell-1}\right) + b_j^\ell\right) \tag{4}$$

where $\ell$ is the number of layers. $down()$ represents a sub-sampling function. Typically this function will sum over each distinct $n$-by-$n$ block in the input image so that the output image is $n$-times smaller along both spatial dimensions. Each output map is given its own multiplicative bias $\beta$ and an additive bias $b$.

2.3. **Activation sublayer.** This is a layer of neurons that use the non-saturating activation function such as *sigmoid* function.

$$f(x) = (1 + e^{-x})^{-1} \tag{5}$$

It increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer.

2.4. **Pooling sublayer.** In order to reduce variance, pooling layers compute the max or average value of a particular feature over a region of the image. This will ensure that the same result will be obtained, even when image features have small translations. This is an important operation for object classification and detection. The pooling sublayer can be expressed as

$$x_j^\ell = \max_{i \in N} f_i^{\ell-1} u(s,s) \tag{6}$$

where $N$ is the total number of down-sampling layer. $u(s,s)$ is the function of inputting window. $s$ is the window size.

3. **Improving CNN by Activation Function ReLu.** Any stage in the convolution system can be viewed as applying a certain function $g(y)$ to the input function $f(x)$, so analysis in the frequency domain comes to find the Fourier transform $F(g(f(x)))$ with respect to $F(f(x))$.

In general, problem does not have an analytical solution, but we can find one solution in case $g(y) = ReLu(y)$, which is the abbreviation of Rectified Linear Units (ReLu). For acting as data clipping in time domain, it creates sharp corners in the signal, so in the frequency domain this would add higher frequency harmonics to the spectrum.

Mathematically we can express $ReLu(f(x))$ function through $f(x)$ as a multiplication with the $sign(f(x))$, which is equal to 1 if $f(x) > 0$ and 0 otherwise:

$$ReLu(f(x)) = \max f(x), 0 = sign(f(x)) * f(x) \tag{7}$$

Because we are working with limited intervals (number of samples) of function $f(x)$, we can express ReLu through the multiplication with sum of delta functions:

$$sign(f(x)) * f(x) = f(x) * \sum_i \delta(x - x_i),\ f(x_i) > 0 \tag{8}$$

The Fourier transform of a delta function is given by:

$$F(\delta(x - x_0))(k) = e^{2\pi jkx_0} \tag{9}$$

Using linearity of FFTs (Fast Fourier Transforms) and convolution theorem, we can express the Fourier transform of $ReLu(f(x))$ through the Fourier transform of $f(x)$:

$$F(ReLu(f(x)))(k) = \left(\sum_i e^{2\pi jkx_i}\right) \otimes F(f(x)) \tag{10}$$

This shows that in frequency domain, $ReLu()$ acts as a convolution with the function of known form. *ReLu* does not face gradient vanishing problem as with *sigmoid* and *tanh* function. Also it can be used in Restricted Boltzmann machine to model real/integer valued inputs.

4. **Feature Analysis and Retrieval Results.** We employ a dataset from Yahoo Shopping that contains more than 160,000 cloth images to validate our approach. The dataset has several categories, which can show a great diversity of the images. All experiments are run on a PC with AMD Athlon II X4 Four Cores, and NVIDIA GeForce GT430. The software platform is based on MS Visual Studio 2012. Our final retrieval results clearly show the effective implementation of our technique.

4.1. **Texture feature and extraction.** To extract texture descriptors by convolutional neural network, we use a group of Gabor filters. A Gabor filter is basically a Gaussian multiplied by a cosine that detects edges at a certain frequency and angle. The steps to extracting texture features are as follows:

- Take center section of image;
- Convolve with conjugate of each Gabor filter;
- Calculate means and standard deviations of each image to obtain Gabor coefficients.

4.2. **Color feature and matching.** The clothing images with different numbers of color of each category are considered differently in the scoring process. If there is a primary color in the target image, the scoring process is to calculate the chroma difference between this primary color to the palette in the database. For different categories, the weighting is different and the matched primary color will have a higher weighting than those secondary or decorative colors. Those target images without a primary color but with secondary colors are considered as objects with mixed colors. The scoring process is to calculate the chroma difference between each of the secondary color of the target image to each of the primary colors and those palettes without primary color favored with higher weighting.

4.3. **Retrieval results.** Since most of textures in clothing images we are trying to detect consist of stripes in the vertical and horizontal direction, we do not bother including a large family of angles, and also only include low frequencies that might appear on the average article of clothing. We ended up using the family of filters with angles 0, $\frac{\pi}{4}$, $\frac{2\pi}{4}$, $\frac{3\pi}{4}$ and frequencies 0.03, 0.0725, 0.115, 0.1575, 0.2, shown in Figure 2.

Since extracting these descriptors involves about 20 2D CNN convolutions for each image, this is by far the most time-consuming part of CNN. After convolution on inputting images, we found that the descriptor vectors are too similar in all shirts. To simplify things, we decided to convolve on color images and this can improve our results, in which shirts with stripes will mostly return shirts with stripes in their top matches as shown in Figure 3.

Figure 4 gives another matching result of our application in Women's Tops. Since color is the most important attribute in this category, and that texture is secondary, we select two weights $W_{color}$ and $W_{texture}$, where $W_{color}+W_{texture} = 1$ and $0 < W_{texture} < W_{color} < 1$. So we use the following weights $W_{color} = 0.7$ and $W_{texture} = 0.3$.

As shown in Figure 3 and Figure 4, the matching results given by our method are quite similar to the comparison image. In the computational time, for an image with a size of $1024 * 1024$ pixels, using our method extracting features takes about 0.06 seconds

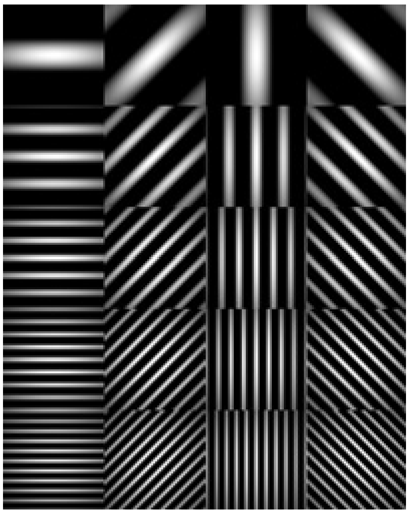

Figure 2. Set of Gabor filters used in this paper

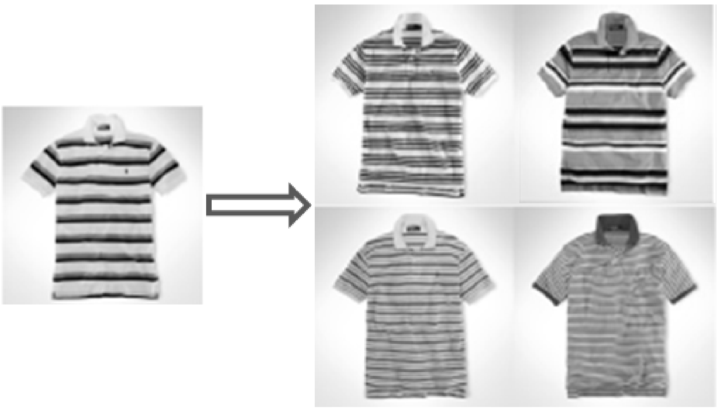

Figure 3. Matching result by texture feature

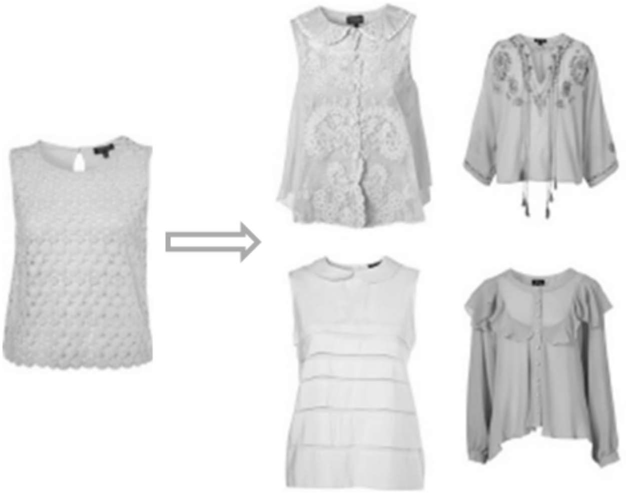

Figure 4. Matching result by both color and texture features

and performing a retrieval takes about 3 seconds, so our method can implement a fast retrieval speed. Overall, our method can increase the retrieval accuracy and rationality via both texture and color features compared with the method [9], which only considers color feature. In the computation complexity, our method seems to be $O(n^2 \log n)$ compared with $O(n^2k^2)$ of the method [10] which divides the image to equal sized blocks and is computation-intensive ($k$ is the number of image pixels).

## 5. Conclusion.

In this paper, an efficient retrieval of clothing image has been proposed and implemented via optimization of convolutional neural networks. Our method not only can implement a fast retrieval speed but also has a good accuracy and rationality. Based on dataset from Yahoo Shopping, experimental results show that we can acquire a better retrieval result by considering texture and color features. Hence users can easily upload a clothing image and query, and then the system processes the image, extracts features, measures similarity, and returns similar clothing images with the corresponding information.

In the future, we plan to embrace more attributes to provide the collected clothes images with better comprehensive description and efficient large-scale computing is also our future focus based on GPU platform.

**Acknowledgment.** This work was supported by the Scientific and Technological Project of Henan Province Science and Technology Department of China (No. 102102210544), and Key Project of Henan Province Higher Education Institution of China (No. 16A520096).

## REFERENCES

[1] B. Hasan and D. Hogg, Segmentation using deformable spatial priors with application to clothing, *Proc. of BMVC*, pp.1-11, 2010.

[2] N. Wang and H. Ai, Who blocks who: Simultaneous clothing segmentation for grouping images, *Proc. of ICCV*, pp.1535-1542, 2011.

[3] V. Ferrari and A. Zisserman, Learning visual attributes, *Advances in Neural Information Processing Systems*, pp.433-440, 2007.

[4] H. Chen, A. Gallagher and B. Girod, Describing clothing by semantic attributes, *Proc. of ECCV*, pp.609-623, 2012.

[5] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu and S. Yan, Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set, *Proc. of the 20th ACM International Conference on Multimedia*, pp.1335-1336, 2012.

[6] K. Yamaguchi, M. H. Kiapour and T. L. Berg, Paper doll parsing: Retrieving similar styles to parse clothing items, *Proc. of ICCV*, pp.3519-3526, 2013.

[7] J. Fu, J. Wang, Z. Li, M. Xu and H. Lu, Efficient clothing retrieval with semantic-preserving visual phrases, *Proc. of the 11th Asian Conference on Computer Vision-Volume Part II*, pp.420-431, 2012.

[8] M. Mizuochi, A. Kanezaki and T. Harada, Clothing retrieval based on local similarity with multiple images, *Proc. of the ACM International Conference on Multimedia*, pp.1165-1168, 2014.

[9] Y. Wang, W. Fu, Y. Wang and X. Huang, Retrieval of clothing images based on color feature, *International Conference on Automation, Mechanical Control and Computational Engineering*, pp.143-149, 2015.

[10] C. Vasanthanayaki and R. Malini, Image retrieval based on block color averaging, *Journal of Theoretical and Applied Information Technology*, vol.56, no.2, pp.209-216, 2013.