ISSN: 2641-192X
Journal of
Textile Science & Fashion Technology
DOI: 10.33552/JTSFT.2020.07.000659
Iris Publishers

Research Article

Copyright © All rights are reserved by Zha Tao

Application Comparison of Textile Fabric Image Retrieval Algorithms Based on Content

Zha Tao*

Soochow university, China

***Corresponding author:** Zha Tao, Graduate Student, College of Textile and Clothing Engineering, Soochow university, China.

Received Date: September 10, 2020

Published Date: November 18, 2020

Abstract

In order to realize the process of textile fabric sales, textile fabric sales companies accurately retrieve the fabrics that meet the requirements from the massive textile fabric image database in order to meet the buyer's requirements for fabrics. At the same time, in order to overcome the shortcomings of traditional image retrieval methods, which need to label and save a large number of text tags, and if the data set is too large, it is impossible to mark the images one by one. In this paper, five content- based image retrieval algorithms are used in textile fabric image retrieval. Three of them are based on traditional feature engineering and two are retrieval methods based on neural network feature extraction. In order to get algorithms with high accuracy and meet the requirements of textile fabric image retrieval, the experimental results of these algorithms are compared. The results show that the neural network model based on deep learning has good stability and accuracy in the process of textile fabric retrieval and has certain practical value and research significance for textile fabric image retrieval.

Keywords: Textile fabric; Image retrieval; Feature extraction

Introduction

Background

Textile fabrics are the main resource for the production and development of the textile industry, and the image is the main presentation form of textile fabrics. In the process of transactions between buyers and sellers, due to the huge number of sample images in the fabric image library, how to accurately extract the fabric information that users really need from the massive inventory has become a problem faced by all manufacturers and sellers. Whether it is in pattern design, inventory management, or material procurement, judging whether the textile fabric pattern already exists, and the specific inventory location, are inseparable from image retrieval technology. Therefore, how to accurately retrieve textile fabric images from the massive image database has become a research hotspot. At present, there are two general methods for textile fabric image retrieval. One is Text-Based Image Retrieval (TBIR), which carries out semantic matching through text description of textile fabric image. The other is Content-Based Image Retrieval (CBIR), which extracts features from the color and texture of the image and achieve "search for images with images".

The work of this paper

Because in the retrieval by text search for images, the expression ability of keywords is limited, and it is powerless for some details and features that are difficult to describe. The traditional retrieval by text search method requires tags and saves a large number of text tags, so it is only suitable for Small data set, if the data set is too large, the images cannot be labeled one by one. Therefore, this paper applies five kinds of content-based image retrieval technology in textile fabric image retrieval. Three of them belong to the traditional feature engineering methods, which take color feature, shape feature and texture feature as entry points, and two are based on neural networks.

The feature extraction methods are vgg16 model and convolutional autoencoder. Use these five methods as the feature extraction

This work is licensed under Creative Commons Attribution 4.0 License

module to build an image feature database for the textile fabric image database. Extract features from the fabric image to be retrieved and construct the same feature vector. Finally, the retrieval module is designed, including similarity measurement criteria, sorting, searching, and returning results with higher similarity. The experimental results of these five methods are compared to obtain a retrieval algorithm with higher accuracy.

Purpose and significance

Textile fabrics are the main resource for the production and development of the textile industry. There are thousands of data flows every day. Therefore, it is extremely important for textile fabric sales companies to accurately extract the fabrics users really need from the massive inventory. This paper takes the textile fabric which is more important in the textile sales industry as the research object and introduces the content- based image retrieval algorithm to realize the textile fabric image retrieval, and to find an effective search method for the large-scale textile fabric image data. Because the color and texture of textile fabrics are more complex, the visual dimension is also relatively high, plus a huge amount of data. Therefore, the in-depth study of image retrieval technology can not only promote the development of textile fabric sales field, but also strengthen the research progress in the field of computer vision and information retrieval, which has certain research value.

Related Work

Mode of feature engineering

Traditional image retrieval methods need to use BOW, VLAD, FV and other encoding methods to encode image features. The extracted image features are generally hand-designed features, including global features such as color, texture, and shape, and features that represent local information.

Color features: There are statistical methods, segmentation methods and cluster analysis methods to extract color features [1]. The classic color features include color moments, color histograms, color aggregation vectors, etc., and many researchers have improved on these color features.

Texture features: Texture characteristics can distinguish differences in fabrics, including patterns, fabric roughness, and smoothness. There are three main methods of texture feature analysis: statistical method, model analysis method and spectrum modeling method [2]. The classical statistical methods are gray-level co-occurrence matrix (GCM) and local binary patterns (LBP). The gray-level co-occurrence matrix method mainly describes texture features through the correlation of image gray-level space; LBP method mainly obtains texture features by comparing the size of center LBP Operator and neighborhood LBP Operator. The statistical method is simple to implement and has a relatively strong generalization ability, but it does not consider the global information of the image. The classic model analysis methods include Markov random field model and autoregressive model. The model analysis method believes that the image texture is subject to a certain model distribution, so the method is mainly to learn the model distribution parameters. Spectrum modeling method is mainly used to extract texture features by analyzing the local area of the image in time domain and frequency domain. The classical methods include Gabor transform and wavelet transform [3].

Shape features: The extraction of image shape features is mainly divided into two categories: one is to extract the outline boundary information of the overall image; the methods are based on the Laplacian method and the canny operator method; the other is the region-based shape feature extraction method.

Local features: Generally, global features are used for retrieval, and the accuracy rate is not high, and local features can be used for better retrieval. Commonly used local features include HOG features, SIFT features and so on [4]. Sivic and Zisserman used BOW to encode the extracted features. Bossard, et al. proposed to use HOG, LBP and other features for fusion, and use machine learning methods such as SVM, random forest, etc. to classify clothing. Yi Yang et al. proposed to use the joint position feature of human body pose estimation for retrieval.

The way of neural network

Convolutional Neural Network (CNN) is a very important deep network in the current deep learning algorithm. Because it can directly operate on the input image, the application range is becoming wider. CNN can be seen as a multi-layer perceptron structure with hidden layers. Compared with other network structures, CNN introduces convolutional layer and pooling layer. Among them, the convolutional layer is used to extract features of the input image, and the pooling layer is mainly to aggregate the features extracted by the convolutional layer. Convolutional neural networks have the characteristics of local perception, that is, different convolution kernels perform feature convolution on different parts of the image, and different local perceptions share parameters. Another feature of CNN is parameter sharing. CNN generally includes input layer, convolutional layer, pooling layer, fully connected layer, and output layer, and the convolutional layer is followed by pooling layer.

Auto-Encoder (AE) is a neural network structure with multiple hidden layers. Unlike traditional neural networks, auto-encoders can automatically learn the inherent dependencies of data and extract feature data in an unsupervised way. The principle of autoencoder is not complicated, and it can be understood as a system that tries to restore the original input.

The autoencoder is mainly composed of two parts, encoding part and decoding part. The encoding part of the autoencoder maps the input data to the feature space, and the decoding part maps the encoded data in the feature space back to the original sample space. The training process of the autoencoder is to find the optimal network parameters to make the decoded data approximate the input data to the greatest extent. Since the acquisition of decoded data depends on the encoded data, to achieve the effect of the decoded data approaching the input data, it is necessary to grasp the internal distribution mode of the input data during encoding. This process does not care about the type of sample data at all. Therefore, for the autoencoder, we do not care about what the output is. The encoded data output by the encoding part is what we really need, which is the feature extracted by the autoencoder. It can be seen that the essence of autoencoder training is an unsupervised feature learning.

Since the autoencoder-based deep learning method was proposed, it has received extensive attention from scholars due to its excellent feature learning ability and simple model foundation. A variety of improvement methods have appeared on the original basis, such as Sparse Auto-Encoder (SAE) proposed by LeCun, et al. [5]; Denoise Auto- Encoder (DAE) proposed by Bengio, et al. [6,7]; Contractive Auto-Encoder (CAE) proposed by Salah Rifai et al. [8] (Figure 1).

Experiment

Experimental environment

The experimental platform is Collaboratory, based on the Keras deep learning framework, with Tensor flow as the backend. Tensor flow is a deep learning platform launched by Google. Its bottom layer is implemented in C++ language and the upper layer is packaged in Python. Its flexible and easy-to-use front-end language is very convenient for building complex algorithm models, and its distributed architecture can make the training and execution of deep learning models more efficient [9,10].

Pre-training data: For the Vgg16 model, the Fashion-MNIST data set is used for training. Fashion-MNIST is a clothing image data set released by the German research organization Zalando Research, which contains 60,000 training samples and 10,000 test samples, including 10 category labels: T-shirts, pants, pullovers, skirts, coats, sandals, shirts, sneakers, bags, ankle boots [11].

Experimental scheme design

Image retrieval of textile fabrics based on feature engineering:

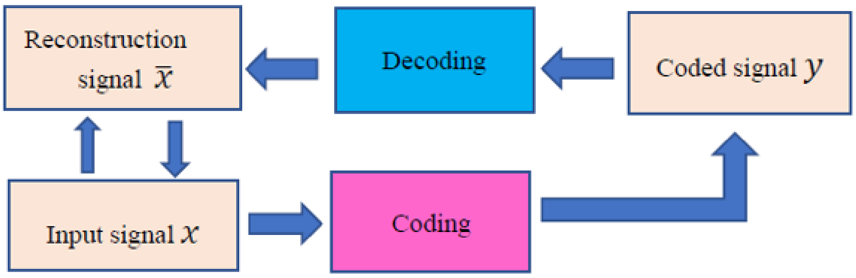

Figure 1: Schematic diagram of autoencoder structure.

Figure 2: Original textile fabric image.

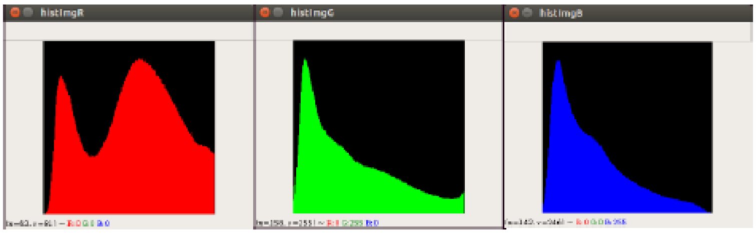

Figure 3: Component histogram.

Figure 4: Component graph.

Image retrieval of textile fabrics based on color histogram. For textile fabric image retrieval based on color histogram, first extract the color histogram of textile fabric from the textile fabric image database, and save the extracted features, that is, the establishment of the feature database. Extract the histograms on the R, G, and B components of the fabric, and store the color histogram as the retrieved feature index in the corresponding database. After the query fabric image is input into the system, the color histogram of the image is extracted, and then the extracted histogram is compared with the histogram in the feature database to find several similar color histograms, and finally return the corresponding similar image (Figures 2-4).

1. Image retrieval of textile fabrics based on SIFT features. The textile fabric image retrieval based on SIFT features is similar to the textile fabric image retrieval based on color histograms. It is necessary to extract the SIFT features of textile fabric images in the database and save them as a feature database. The SIFT feature mainly performs Gaussian filtering and down sampling on the part of the image to achieve feature extraction. Judgment of Image similarity based on SIFT features is related to the number of matched feature points, Therefore, the image feature library should contain the number of SIFT features of each textile fabric image and the location of each feature point. When the user inputs the image to be retrieved into the system, the textile fabric image with high matching degree can be found by matching the SIFT feature points of the image to be retrieved and the SIFT feature points in the feature database [12-16].

The algorithm uses the DOG operator to establish the scale space, and the algorithm definition is shown in the formula.

$$D(x,y,\sigma)=(\mathrm{G}(\mathrm{x},\mathrm{y},\sigma_1)-G(x,y,\sigma_2))*\mathrm{f}(\mathrm{x},\mathrm{y}) \qquad (1)$$

2. Image retrieval of textile fabrics based on Gabor wavelet texture features. In the retrieval process, it is first necessary to use Gabor wavelet to perform filtering operation on each image in the textile fabric image library to obtain Gabor images, and then these Gabor images are treated as texture features and stored in the feature database. When users retrieve images, the same operations are performed on the input images, the same operation is performed on the input image, and the obtained Gabor image is matched with the Gabor image in the feature database, and the mean and variance are used to determine whether the two images are similar [17,18].

The Gabor function g(x, y) can be defined by the following formula, sx and sy are the standard deviations on the two coordinate axes respectively.

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x{}^2}+\frac{y^2}{\sigma_y{}^2}\right)+2\pi jW_x\right] \qquad (2)$$

Image retrieval of textile fabrics based on neural network

1. Image retrieval of textile fabrics based on vgg16. The feature extraction network VGG-16 model used in the experiment is a kind of deep convolutional neural network VGGNet. It is a classic model composed of 16 layers of neural networks and uses weights pre-trained by ImageNet. Using the idea of small core stacking, 3 × 3 small convolution cores and 2 × 2 maximum pooling layers are stacked repeatedly, including 13 convolution layers and 3 full connection layers. The input data of 224 *224* 3 are extracted by multi-layer convolution and pooling. The overall structure is as follows (Figure 5):

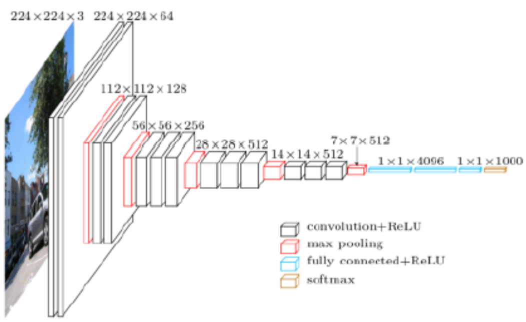

Figure 5: Vgg16 structure diagram.

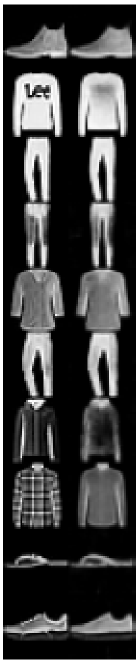

Figure 6: Reconstruction graph.

The training of convolutional neural network needs a lot of label data, and the data samples of textile fabric images is limited. Therefore, this section uses the vgg16 model trained on ImageNet, and trains the model based on fashion MNIST data set, so that it has a better recognition effect on textile fabric images which are similar to clothing images, so as to achieve good retrieval results [19,20].

Each textile fabric image in the image database is input into the trained vgg16, and the features of the following layers (Pool5, FC6 or FC7) are extracted and used as CNN feature database of textile fabric images. Input the image to be retrieved into vgg16, extract the features of the same layer as the image library, use cosine similarity to calculate the similarity between the retrieved image and the textile fabric image in the image feature library, and return the image with the smallest distance according to the demand, that is, the most similar textile fabric image.

2. Image retrieval of textile fabrics based on autoencoder. The convolutional autoencoder used in the experiment introduces convolution and pooling operations on the basis of traditional autoencoders to realize local receptive fields and weight sharing. The architecture of convolutional autoencoder is similar to the autoencoder, and the mathematical calculation process is as follows:

For input single-channel input x, the formula for extracting feature maps through the number of k convolution kernels is as follows:

$$h^k = \sigma(x * W^k + b^k) \qquad (3)$$

Where σ represents the activation function, and ∗ represents the convolution operation.

The reconstruction process uses the following formula:

$$y = \sigma(\sum_{k \in H} h^k * \tilde{W} + c) \qquad (4)$$

Where σ represents the activation function, h represents the number of feature map groups, $\tilde{W}$ represents the transposition of the weight of the coding part, and C represents the bias. The loss function generally uses the mean square error, the formula is as follows:

$$E(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (5)$$

Like other neural networks, the back-propagation algorithm is used to calculate the gradient of the error function with respect to the parameters. This can be easily obtained using convolution. The formula is as follows:

$$\frac{\partial E(\theta)}{\partial w^k} = x * \delta h^k + \tilde{h}^k * \delta y \qquad (6)$$

Among them, δh and δy represents the features extracted by the hidden layer and the result of reconstruction respectively.

First, the encoder accepts an input of 28×28×1, and then applies convolution, activation, and normalization twice. Each step is 3×3. After the last normalization process, the obtained dimension7×7×64 is flattened to a vector of 3136, and the fully connected layer is used as a potential space representation (Table 1).

The decoder accepts the potential space representation output from the encoder, constructs a new 3136 full connection layer, Change the dimension back to 7×7×64 format, and perform convolution, activation, and normalization again in the same way. The convolution used here is to restore the original channel depth of the image, and to change the dimensions of the image back to the original size (Table 2).

Table 1: Encoder structure.

Model: encoder		
Layer(type)	Output shape	Parm
Input_1	(None,28,28,1)	0
Conv2d	(None,14,14,32)	320
LeakyRelu	(None,14,14,32)	0
Batch_normalization	(None,14,14,32)	128
Conv2d_1	(None,7,7,64)	18496
LeakyRelu_1	(None,7,7,64)	0
Batch_normalization_1	(None,7,7,64)	256
Flatten	(None,3136)	0
Dense	(None,16)	50192

Table 2: Decoder structure.

Model: encoder		
Layer(type)	Output shape	Parm
Input_2	(None,16)	0
Dense_1	(None,3136)	53312
Reshape	(None,7,7,64)	0
Conv2d_transpose	(None, 14,14,64)	36928
LeakyRelu_2	(None, 14,14,64)	0
Batch_normalization_2	(None, 14,14,64)	256
Conv2d_transpose_1	(None, 28,28,32)	18464
LeakyRelu_3	(None, 28,28,32)	0
Batch_normalization_3	(None, 28,28,32)	128
Conv2d_transpose_2	(None, 28,28,1)	289
Activation	(None, 28,28,1)	0

Table 3: Search results of five methods.

	Precision		Recall	
	Back to 10	Back to 20	Back to 10	Back to 20
RGB	39%	36%	2%	4%
Sift	53%	45%	3%	6%
Gabor	58%	56%	3%	9%
Vgg16	90%	82%	4%	8%
ConvAutoencoder	93%	78%	5%	10%

After training the constructed autoencoder, the reconstruction image shows that our autoencoder is doing a good job in reconstructing the input image (Figure 6).

Using the trained automatic encoder to achieve image retrieval, first of all, the images of the image database are sequentially input to the autoencoder to construct the feature vector that generates the index. These feature vectors represent the features of the image. During the retrieval, the retrieved image is input, and the same feature vector is extracted. Compared with the data in the index database and return similar images through the result of the Euclidean distance measurement [21-23].

Evaluation index

The evaluation of retrieval performance mainly includes retrieval accuracy, sorting effect and retrieval speed. The accuracy rate reflects the performance of the feature extraction algorithm and the similarity matching algorithm, the ranking effect and retrieval speed reflect the image feature indexing effect and the complexity of the similarity matching algorithm. Since this article focuses on the performance and accuracy of the textile fabric image retrieval process, the precision and recall indicators are used [24].

This article uses Precision and Recall evaluating the performance of textile fabric retrieval. Recall rate refers to the percentage of correct images in the query results returned by the retrieval system to all correct images in the image database. The precision rate has nothing to do with the number of images in the database, it mainly refers to the correct rate of the retrieved images. The formula for calculating the precision rate and the recall rate is as follows.

$$P = \frac{N_A}{N_B + N_A} X100\% \qquad (7)$$

$$R = \frac{N_A}{N_A + N_C} X100\% \qquad (8)$$

Among them, N_A and N_B respectively represent the number of correct images and incorrect images in the returned images, and NC represents the number of correct images that are not returned in the image database.

Experimental results and analysis

Vgg16 training process: Figure 7

Autoencoder training process: Figure 8

In the figure above, it can be found that the loss value declined faster in the first few times, and tended to be flat in the subsequent times, indicating that the results of the experiment have stabilized after multiple iterations.

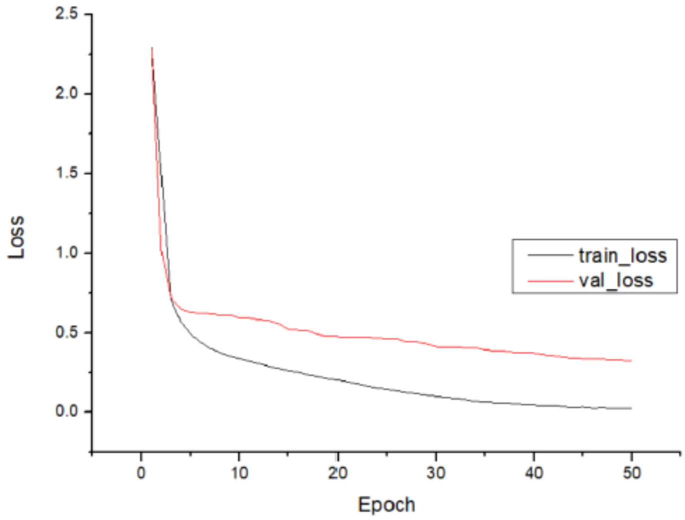

Figure 7: vgg16 training process diagram.

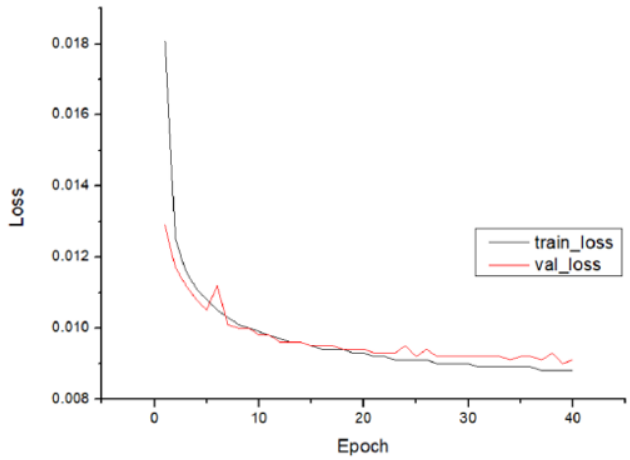

Figure 8: Autoencoder training process diagram.

The results of five retrieval algorithms were compared and analyzed (Table 3).

Compared with the image retrieval method of the traditional feature engineering extraction algorithm based on the underlying features, the textile fabric image retrieval based on neural network has a significant improvement in precision and recall. When returning 10 images, the accuracy of color feature, shape feature and texture feature are 39%, 53% and 58%, the accuracy of convolutional neural network and autoencoder is 90% and 93%. When returning 20 images, the accuracy of color feature, shape feature, and texture feature are 36%, 45%, and 56%, while the accuracy of convolutional neural network and autoencoder are 82%, 78%. The precision rate has been greatly improved (Figure 9).

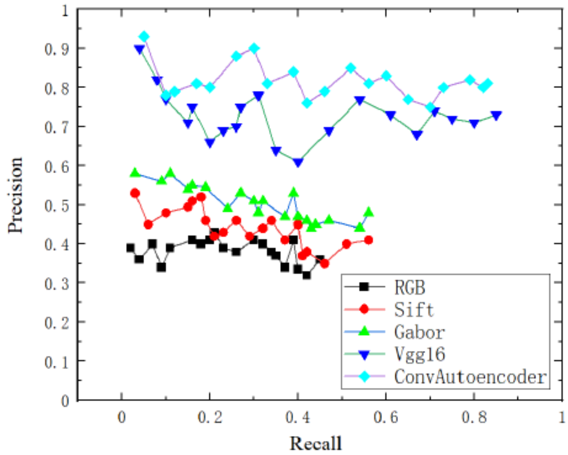

Figure 9: P-R diagram of five kinds of image retrieval.

In terms of recall, combined with the trend of P-R line graph of five image retrievals, as the number of similar images returned increases, the image retrieval method based on neural network is better than the image retrieval method based on the underlying feature as the extraction algorithm. The recall rate has also improved significantly.

From the textile fabric image retrieval results, it can be seen that the deep features extracted by the deep neural network can fully express the image compared to the underlying features such as color, shape, and texture, and the retrieval effect is better. Of course, for all different retrieval methods, the recall rate has increased while the precision rate has a clear downward trend [25].

Conclusion

In order to accurately retrieve the fabrics that meet the requirements from the massive textile fabric image database, this paper uses five content-based image retrieval algorithms to apply to the textile fabric image retrieval. First of all, the current image retrieval methods based on content-based feature extraction are summarized. Secondly, three image retrieval methods based on underlying features are used, and two image retrieval methods based on neural network feature extraction are used. The comparative analysis of experimental results can prove that the method proposed in this paper is feasible in practical applications. Compared with image retrieval based on low-level features, neural networks can better express image features, and at the same time avoid the subjectivity of image identification based on text search. The experimental results show that: (1) Compared with the image retrieval method based on underlying features, the textile fabric image retrieval based on neural network has a significant improvement in precision and recall and can achieve better retrieval results. (2) From the textile fabric image retrieval results, it can be seen that the deep features extracted by the deep neural network can fully express the features of the image compared to the underlying features such as color, shape, and texture, so that the retrieval effect is better. (3) For all different retrieval methods, the recall rate has increased while the precision rate has a downward trend.

Acknowledgement

None.

Conflict of Interest

Author declare no conflict of interest.

References

1. Jin W (2014) Image retrieval based on fusion of color features and shape features . Xidian University, China.
2. Zheng Q (2015) Research on image retrieval technology based on deep learning. Fujian Normal University, China.
3. Chen Lili (2018) Clothing image classification and retrieval based on deep learning. Beijing Institute of Fashion Technology, China.
4. Bao Q (2017) Clothing image classification and retrieval based on deep learning. Zhejiang University, China.
5. Poultney C, Chopra S, Cun YL (2006) Efficient learning of sparse representations with an energy-based model. Advances in neural information processing systems: 1137-1144.

6. Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders.Proceedings of the 25th international conference on Machine learning. ACM: 1096-1103.
7. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA (2010) Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research 11: 3371-3408.
8. Rifai S, Vincent P, Muller X, Glorot X, Bengio Y (2011) Contractive auto-encoders: Explicit invariance during feature extraction. Proceedings of the 28th international conference on machine learning (ICML-11): 833-840.
9. Wang B, Li Z, Li M, Ma WY (2006) Large-scale duplicate detection for web image search. IEEE International Conference on Multimedia and Expo (ICME). IEEE: 353–356.
10. Zhang S, Yang M, Cour T, Yu K, Metaxas DN (2012) Query specific fusion for image retrieval. European Conference on Computer Vision (ECCV).
11. Lowe DG (2004) Distinctive image features from scale invariant keypoints. International Journal of Computer Vision 60(2): 91-110.
12. Lowe DG (1999) Object recognition from local scale-invariant features. IEEE International Conference on Computer Vision 2: 1150-1157.
13. Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10): 761-767.
14. Mikolajczyk K, Schmid C (2004) Scale & affine invariant interest point detectors. International Journal of Computer Vision 60(1): 63-86.
15. Xie H, Gao K, Zhang Y, Tang S, Li J, et al. (2011) Efficient feature detection and effective post-verification for large scale near-duplicate image search. IEEE Transactions on Multimedia (TMM) 13(6): 1319-1332.
16. Rosten E, Porter R, Drummond T (2010) Faster and better: A machine learning approach to corner detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(1): 105-119.
17. Strieker M, Dimai A (1997) Spectral covariance and fuzzy regions forimage indexing. Machine Vision and Applications10(2): 66-73.
18. Pass G, Zabih R, Miller J (1996) Comparing images using color coherence vectors. Proceedings of the 4th ACM International Conference on Multimedia (MM), ACM: 65-73.
19. Chen WT, Liu WC, Chen MS (2010) Adaptive color feature extraction based on image color distributions. IEEE Transactions on Image Processing 19(8): 2005- 2016.
20. Shao T (2004) Research on image retrieval technology based on multi feature DS fusion strategy. Harbin Engineering University, China.
21. Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recognition 37(1):1-19.
22. livarinen J, Peura M, Srel J, Visa A (1997) Comparison of combined shape descriptors for irregular objects. Proceedings of the 8th British Machine Vision Conference (BMVC). Essex, Great Britain: 430-439.
23. Rangayyan RM, El-Faramawy NM, Desautels JE, Alim OA (1997) Measures of acutance and shape for classification of breast tumors. IEEE Trans Med Imaging 16(6): 799-810.
24. Berretti S, Bimbo A, Pala P (2000) Retrieval by shape similarity with perceptual distance and effective indexing. IEEE Transactions on Multimedia 2(4): 225-239.
25. Roussopoulos N, Kelley S, Vincent F (1995) Nearest neighbor queries. Acm Sigmod Record 24(2): 71-79.

Citation: Zha Tao. Application Comparison of Textile Fabric Image Retrieval Algorithms Based on Content. 7(2): 2020. JTSFT. MS.ID.000659. DOI: 10.33552/JTSFT.2020.07.000659.