Research on the Technology of Searching for Fashion Trend Image Based on ResNet50 Model

Yujing Tian¹, ZeYu Zheng²*, Ge Li², Chengrui Xu¹, WenXiang Shang² and Xiaogang Liu¹

¹ Donghua University, Fashion and Design College, No.1882 West Yan’an Road, Changning District, Shanghai, P.R. China
² Hangzhou Zhiyi Technology Co., Ltd., Room 1201, Block F, Hangzhou Bay Information Port, Xiaoshan District, Hangzhou, P.R. China

* zeyu@zhiyitech.cn

Abstract. With the further development of technology, providing technical support for the forecast and development of fashion design trends through artificial intelligence has gradually become a reality. In order to enable users to obtain more accurate design trend materials in a shorter time, this paper presents a ‘searching by image’ model of clothing trends based on the ResNet50 model. Through the targeted collection of data and the rational construction of algorithms, the system can retrieve and output more image materials with relevant design features for the existing image materials for the development of design trends as well as the development and design of styles. By comparing the system with Google's and Taobao's ‘searching by image’ system, it is concluded that this system has high efficiency, high accuracy and high correlation.

1. Introduction

In the Internet era, with the popularity of websites and apps such as Google, Taobao and Instagram, unstructured data such as images and videos is exploding every day at an alarming rate. For these massive images with rich visual information, how to easily, quickly and accurately search for the images that users need or are interested in from the vast image database of the fashion industry is a difficult problem in the field of multimedia information search, and ‘searching by image’ is the key technology to solve this problem.

With the further development and improvement of all aspects of the apparel industry, trend prediction and product development can also become more convenient and scientific through the participation of artificial intelligence. Shi Yingjie et al. [1] combined with the characteristics of clothing fashion data and proposed a basic framework for clothing fashion analysis based on machine learning. Trend forecasting and product development not only need to aggregate similar content to demonstrate the popularity of elements, but also need to expand on the basis of existing materials to find more research and application space. ‘Searching by image’ is widely used on the current Internet, and is often used to retrieve similar pictures through existing picture materials to achieve the effect of updating picture quality, retrieving picture sources or collecting relevant text introductions.

Image feature extraction algorithms are generally divided into global features and local features. In recent years, it is more common to use local features for classification and retrieval [2]. The local features commonly used in images are Harris corner [3], SURF feature [4], SIFT feature [5], and
Msers region of maximum stable extreme value region [6]. On this basis, Dong Junjie et al. [7] used HOG feature and SVM classifier to establish an online clothing similarity retrieval system. Chen Qijin et al. [8] proposed a BUNDLED feature composed of SIFT features and regions of extremely stable extreme values for clothing image retrieval. However, in the field of clothing fashion trend prediction and product development using a large number of pictures, there has not been a ‘searching by image’ system model for the garment market.

Under the current research situation, through the establishment of a targeted ‘searching by image’ database, you can effectively integrate resources through deep learning, screen irrelevant data, and provide accurate guarantee for image retrieval and output. By optimizing the model structure, you can improve the efficiency of the operation, reduce labor costs and quickly aggregate effective information so as to provide material evidence for the selection of trend elements.

2. Technical analysis of searching by image

Based on the basic principle, general processes and algorithm of "searching by image", the ResNet50 model structure can be further optimized.

2.1. Basic principles

The ‘searching by image’ system based on deep learning generally uses the deep learning model for image feature extraction, and then a specific feature vector can be obtained after the extraction is completed. The feature vector can represent the image, and the essence of searching by image is to calculate the distance between the image feature vectors. The closer the distance is, the more similar the images are. Similarly, the similarity between images is the distance between feature vectors. This distance is usually the Euclidean distance or the cosine distance.

2.2. General process and algorithm

The commonly used searching by image system can be divided into two steps.

- The first step is to train a deep learning model so that the model can extract features from the image. The figure above shows the training process of the feature extraction model. In this process, the ‘searching by image’ algorithm will use a training method similar to the face recognition algorithm. First, we will organize a training gallery, which is composed of multiple sets of similar images. We all images inside the same set as similar, and give each set an ID, which will be used in the ID loss. As shown in Figure 1, each time of training will use the combination of Triplet loss and ID loss to train the model. The following Ltpl gives the calculation formula of Triplet loss and Lsoftmax gives the calculation formula of ID loss. In

![Figure 1. Training process of feature extraction model](image-url)
each training example, 3 images are selected. The first image is called the anchor image and the feature corresponding to this image is represented as $x_{a}^{i}$. The second image is called the positive example and the corresponding feature is represented as $x_{+}^{i}$. This image should be selected from the same set of the first image, i.e. the anchor image. The third image is called the negative example and the corresponding feature is represented as $x_{-}^{i}$. This image should be selected from the different set of the first image.

$$L_{triplet} = \sum_{i=1}^{N} \max(0, m + D(x_{a}^{i}, x_{+}^{i}) - D(x_{a}^{i}, x_{-}^{i}))$$

$$L_{softmax} = -\frac{1}{N} \sum_{i=1}^{N} \left( \log \frac{e^{\frac{W^{x} x_{a}^{i} + b_{y}^{a}}}{\sum_{j=1}^{C} e^{W^{x} x_{j}^{i} + b_{y}^{j}}} + \log \frac{e^{\frac{W^{x} x_{+}^{i} + b_{y}^{+}}}{\sum_{j=1}^{C} e^{W^{x} x_{j}^{i} + b_{y}^{j}}} + \log \frac{e^{\frac{W^{x} x_{-}^{i} + b_{y}^{-}}}{\sum_{j=1}^{C} e^{W^{x} x_{j}^{i} + b_{y}^{j}}}}}{2} \right)$$

The model trained by Triplet loss and ID loss can extract texture, color and shape features from the image as much as possible so as to represent a picture well, and the features extracted by similar pictures will also be similar.

- In the second step, after obtaining the features of each image, by comparing the similarity between the features of the given target image and the features of other images in the server, the final results are sorted according to the similarity to find similar images. This process is shown in Figure 2.

**Figure 2.** Ranking results based on similarity comparison

The general searching by image system usually uses the fully connected layer of the VGG16 model for image feature representation, but because the VGG16 model uses a general category for training, the granularity of the extraction of fashion features is not enough.

2.3. **Optimization of ResNet50 model structure**

In order to better extract fashion points, especially to extract clothing design points, the ResNet50 model is used as a Backbone, and effective feature extraction is achieved by designing a multi-branch model structure. As shown in Figure 3, we propose a model structure of searching by image system.

**Figure 3.** Model structure of searching by image system
As can be seen from Figure 3, the model is divided into multiple branches, including a global branch and multiple local branches. Global connects softmax and triplet loss at the same time, and performs classification and measuring learning tasks at the same time; other branches of local only perform classification learning tasks; the global branch is used to extract global features; all other branches of local are used to extract detailed features of clothing. Figure 4 shows the sensitivity of different branches to clothing design points.

![Figure 4. Sensitivity of different branches to clothing design points](image)

3. Fashion industry database

In order to enable the retrieval system to provide users with better picture materials and resources based on the ‘searching by image’ technology, while optimizing the deep learning of ‘searching by image’ model, we have also established a professional fashion industry database to enable the system to output more professional results and better serve fashion trend analysis.

In the construction of the database, as Figure 5, the source of the original data can be divided into two parts, which are respectively obtained based on product data and consumer behavior. The data can be continuously and efficiently constructed by artificially crawling the data.

3.1. Product Image

- Raw material exhibition
  Based on the perspective of surface material developers, it will display the textures and patterns that may be popular in the new development quarter. Well-known exhibitions at home and abroad include Première Vision Paris, Lineapelle London, Intertextile Shanghai, etc.

- Fashion brand image
  The brand company's official website has detailed product picture data for the season. These data include product pictures of different angles and wearing pictures of models, as well as Lookbook pictures that are carefully photographed for distribution. By constantly collecting product pictures, you can accumulate a large amount of picture material as the research object of the popular styles in the current quarter.

- Fashion exhibition image
  Different types of finished product exhibitions can also be used as platforms for mobile phones of styles, and the data obtained from them are more referenceable than raw material exhibitions. World-renowned exhibitions include MICAM Milano, Milano Moda, Premiere Classe, etc.

- Fashion show image
  Paying attention to the color performance of the fashion week show can intuitively understand the changes in design elements in different quarters. Due to the development of technology, there are more and more channels and ways to pay attention to the fashion week show. Among
them, Vogue Runway has a wealth of image data and fashion reviews for the show throughout the year. Its labeling and sorting of brands, designers and different quarters can also make the data crawl clearer and more accurate.

- Fashion market data
  The product data of shopping platforms at home and abroad is an intuitive reflection of the sales situation, which includes images from department stores, shopping malls, specialty stores, buyer shops, e-commerce platforms and second-hand markets; the product display reflects the preferences of specific consumer groups in different markets for styles and elements. Foreign e-commerce websites include, for example, Farfetch, Ssense, Browns Fashion, etc., and domestic ones include, for example, Tmall, Siku, etc.

3.2. Consumer Image

- Social media image
  Many Internet celebrities and We-Media bloggers from different social platforms attract attention and flow by sharing fashion outfits. By collecting fashion pictures shared by influential accounts on social media at home and abroad, it can be used as an important reference for consumers in the database. These platforms include Instagram, twitter, Weibo, Bilibili, etc. As more and more brands open social media accounts, social media is also an important channel for paying attention to the release of products by fashion brands.

- Street snap image
  The source of street photography is divided into two main parts. On the one hand, it includes street photography columns of fashion magazines, such as VOGUE street style, HARPERS BAZAAR street style, ELLE street style, GQ street style, W MAGAZINE street style, and on the other hand, it includes some professional street shooting websites, such as Schuman, Atelierdore, Street Peeper, etc.

![Figure 5. Fashion industry database system](image)

4. Experiment

The author uses the same fashion picture for ‘searching by image’ in Google, Taobao and ResNet50 optimization model systems, checks and analyzes the results generated by each channel, and finally makes a comprehensive comparison.

4.1. Experimental process

- Choose case picture as Figure 6.
- ‘Searching by image’ in Google, Taobao and ResNet50 optimization model systems
- Checks and analyzes the results
- Comparative analysis
4.2. Experimental results

4.2.1. Results and analysis of Google's searching by image. Google's searching by image algorithm, while extracting image features, will also associate a keyword with each image, and find the most similar images through the combination of keywords and image features. As can be seen from Figure 7, Google will first associate it with the keyword girl. According to the content characteristics of keywords and images, Google will return some women's clothing, including shirts, T-shirts, suspenders, coats, short skirts and other category clothing. However, the search results are not precise enough, and at the same time, no attention is paid to the design features of the clothing.

4.2.2. Results and analysis of Taobao's searching by image. Figure 8 shows the result of searching by image for the same picture in Taobao. Taobao first determines which category the target image belongs to through category prediction, and then performs target detection on the target image to reduce the influence of the background and multiple subjects in the target image. Next, Taobao then extracted features from the detected target image, and at the same time extracted features from the category data corresponding to the target image. The features include the deep features extracted from the deep learning model, and the local features are used as supplements to the deep features. Finally, Taobao maps the target feature to the database feature, calculates the distance between the target feature and the database feature respectively, and sorts according to the distance to obtain the search results. As can be seen from the figure, in the search of fashion data, Taobao's effect is still significantly better than Google. However, Taobao can only search for the products being sold on it and can't find images with higher fashion degree, so it is difficult to directly use them in trend research.

Figure 6. Case picture of searching by image

Figure 7. Searching results on Google

Figure 8. Searching results on Taobao
4.2.3. Results and analysis of searching by image based on ResNet50 model. In order to solve the problem of professional fashion's 'searching by image', we provide a professional 'searching by image' system for trend analysis. Figure 9 shows the results returned by the same picture in our searching by image system. Relatively speaking, the results are not limited to the exact same style, but more diverse to filter the local characteristics of the clothing in the database. The final result is more abundant, and for designers, it has high referenceable degree and is more applicable to the study of clothing trends.

Table 1. Comparative analysis in the three systems.

| Database comparison | Principle comparison | Result comparison |
|----------------------|----------------------|------------------|
| Google (Network comprehensive data) | Keyword and image feature joint search | No attention is paid to the design features of clothing |
| Taobao (Seller's product) | Category prediction judgment, target image extraction feature, target feature and database feature comparison and distance calculation ranking | Users can only search for products sold on Taobao, but they can't find more fashionable images |
| ResNet50 model (Data with influence on the fashion industry) | Target image multi-branch feature extraction, multi-target feature and database feature comparison and distance calculation ranking | Meet the design features of clothing. High degree of fashion, diverse pictures, suitable for clothing trend analysis |
4.3. Comparative analysis
After passing the model experiment based on three kinds of ‘searching by image’, Table 1 shows the three systems can be compared and analyzed according to the database construction, system principles and experimental results. Compared with the purpose of obtaining professional design materials through system operation, the database of Google platform is more extensive. It analyzes and imports the features of pictures and searches through related keywords. The search results are more comprehensive, but there is no development and application for the clothing market. The database of Taobao platform is formed based on the aggregation of the sellers’ product data. Due to the limitation of the scope, the search results are only for Taobao's existing products. The search result coincidence rate is high, and it is impossible for users to find trend information that is helpful for design and development. The ResNet50 model proposed in this paper actively collects data from professional channels during the database construction stage, and the output results are more in line with the principles of clothing design with higher fashion and diversity, which are suitable for trend researchers and clothing design developers in the clothing industry.

5. Conclusions
The ResNet50 model proposed in this paper is based on the theoretical model of ‘searching by image’. Through the construction of databases and system models, images with higher accuracy can be output. Because its database is more focused on the clothing industry, and the retrieval principle pays more attention to the details of the clothing itself, by arranging and presenting the pictures in layers and sensitivity, it is more in line with the actual market performance obtained by the clothing industry during research and development demand. After comparing with Google and Taobao, which are also based on the principle of ‘searching by image’, we can get the conclusion that the system proposed in this article is more efficient and accurate.

Acknowledgments
Project Support by the Fundamental Research Funds for the Central Universities, Shanghai Style Fashion Design & Value Creation Collaborative Innovation Center & the Fundamental Research Funds for the Central Universities & Shanghai Summit Discipline in Design (DB18212) , Shanghai Education Commission & Shanghai Education Development Foundation Chenguang Plan (13CG74).

References
[1] Shi Y, Yang T, Huang H and Zhang Y 2020 J. A Survey of Clothing Fashion Analysis Based on Machine Learning Journal of Beijing Institute of Fashion Technology (Natural Science Edition) 40 p 100-110
[2] Bao Q 2017 D. Clothing Image Classification and Retrieval Based on Deep Learning Zhejiang University P 13
[3] Harris C and Stephens M 1988 A Combined Corner and Edge Detector In Proc. of Fourth Alvey Vision Conference vol 6 (Manchester: University of Manchester) p 147
[4] Bay H, Tuytelaars T, Gool L 2006 Surf: Speeded up robust features Proc. of 9th. European Conference on Computer Vision vol 3 p 404
[5] David G 2004 J. Distinctive Image Features from Scale-Invariant Keypoints International Journal of Computer Vision 60 91-110
[6] J Matas, O Chum, M Urban, T Pajdla 2004 J. Robust wide-baseline stereo from maximally stable. extremal regions Image and Vision Computing 22 761-767
[7] Dong J 2014 D. Design and Implementation of Clothing Image Retrieval System Based on HOG. and SVM Sun Yat-sen University P 36
[8] Chen Q 2013 D. Research on Clothing Retrieval and Matching Technology Based on Image. Content Zhejiang University P 11