Application of Image Matching Algorithm Based on Lifted Structured in Large Scale Database of Furniture

Yong Wang, Wenwen Gao[a], Ying Wang
School of Computer, Guangdong University of Technology, Guangdong, Guangzhou 51006, China
[a]382471541@qq.com

Keywords: furniture image search; deep learning; convolution neural network;

Abstract. Based on the Lifted Structured Feature Embedding (LSFE) proposed by H.O. Song, it is extended to the two million-level furniture image database retrieval application, which proves that it is excellent content-based image matching capabilities on large scale database.

1. Introduction

With the development of big data and artificial intelligence, deep learning applications began to penetrate into all aspects of life. Image as a major component of media transmission contains a wealth of information. It is also of great practical value and research significance to retrieve images from large-scale image resources, and can be widely used in medical, e-commerce, transportation, social security and so on. Many researchers and scholars have done a lot of research on content-based image retrieval. At present, the mainstream algorithms are Siamese Network and Triplet Network, and these algorithms need to construct pairs of training samples. The construction of sample pairs often results in a sharp increase of training samples, resulting in inefficient algorithms. Moreover, the convergence of algorithms often depends on the construction strategy of sample pairs, so it is difficult to apply them in practice. In addition, most of the research work only uses the order of one hundred thousand of the public database to verify the performance of the algorithm, and in practical applications, the image database often reaches one million or even tens of millions. In a larger database, Whether the matching algorithm can keep its accuracy needs to be verified in the application. Based on Lifted Structured Feature Embedding (LSFE), a metric learning algorithm proposed by H.O.Song [1], this paper applies it to a furniture database containing 2 million images and compares it with Siamese Network and Triplet Network proved this algorithm outperforms the latter two in large databases.

2. Related Work

2.1 Image Retrieval Based on Lifted Structured Feature Embedding

H. Song [1] proposed a metric learning algorithm called Lifted Structured Feature Embedding (LSFE) for image content retrieval. Based on the traditional Siamese Network (SN), this algorithm greatly improves the training efficiency by additionally constructing positive and negative sample pairs in the batch of training samples, and obtains higher retrieval accuracy than the traditional SN in the small sample databas.

The traditional SN manual designates positive and negative sample pairs for training. Triplet Network [2-7] (TN) also need to pre-specified triple input (query samples, positive samples and negative samples). In LSFE, the positive sample pair is specified in advance as shown in Fig. 1. For the negative sample, the LSFE calculates the distance between the sample and each of the remaining samples in a batch, and select a negative sample with the smallest distance from the sample. For each batch, the loss function is defined as follows,

$$L(\mathrm{X},\mathrm{y}) = \frac{1}{|p|}\sum_{(i,j)\in p}[D_{i,j}^2 + a - D_{i,k^*(i,j)}^2]_+ \tag{1}$$

where: $k^*(i,j) = \arg\min_{k:y[k]\neq y[i]} D_{i,k}^2 s.t. D_{i,k}^2 > D_{i,j}^2$

p is the number of positive samples in a batch. If none of these negative samples satisfy such a constraint, we pick the negative sample closest to the batch as follows: $k^*(i,j) = \arg\max_{k:y[k]\neq y[i]} D_{i,k}^2$

To ensure accuracy, batch need be as possible as big.it can make the best use of batch to find the best negative samples.

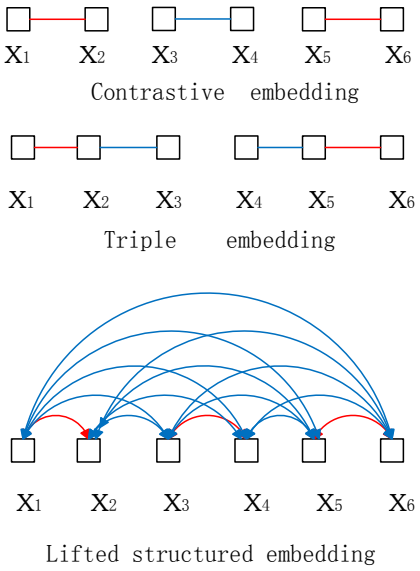

Fig. 1. LSFE

However, the author only validates at the laboratory (100,000) database. Due to its excellent algorithmic principle, it is of great significance to verify whether it can still maintain excellent matching ability in a large database of one million . This thesis will build a two-million-level furniture picture database. Experiments show that the LSFE algorithm still has excellent matching ability in large-scale databases [8-10].

3. Experiment

3.1Database

Using the crawler from Jingdong E-commerce website, we got about 2.02 million furniture images and set up our database. In the database, there are 7 types of tables, chairs, couches, beds, tea sets, cupboards, and others. Each type of product has an average of 6 images. The total size of the database is about 1 million 600 thousand

3.2 Parameter

We use the Caffe [10] package for training and testing. In the experiment, we use our self-built JD database, in which 800,000 data samples are used for training and the remaining 100,000 samples are tested. The 800,000 samples constitute about 9.18 million positive and negative samples. All training and test images size uniformly cropped of 256 × 256. To increase training data, all images are cropped randomly to 227 × 227 and a random image is generated horizontally. For all experiments, the maximum number of training iterations set to 500000. Set the edge parameter α to 1.0. The batch size set to 128. We use a random gradient descent method. All convolutional initialization weights come from GoogleNet's[11]best training model in the ImageNet ILSVRC database, and the weights of fully connected layers are initialized randomly. Multiply the learning rate of fully connected layers by 10.0 for faster convergence. We conducted different training and testing under the Embedding dimension of 128,256,512 respectively. For training, we make every effort to use all similar pairs and non-similar pairs that are approximately as many as similar pairs.

4. Experimental Results and Evaluation

For the test, we get the embedding of the test set in different dimensions ({128,256,512}). Follow the practice of [10]. We use the standard Recall @ metric [9] search to evaluate the experimental results, that is, the test gets the closest k products with the queried product, and measures whether the labeled quasi products appear in the nearest k results. If Yes , recall set to 1. On the contrary, recall set to 0.

Figure 2 shows the effect of Embedding dimensions on R @ 1, R @ 10, and R @ 100 for furniture datasets. Due to the memory limitations of the GPU, the maximum batch size in the GoogLeNet model is 128. Figure 4-2 shows Recall @ 1,10,100 in the same database using the Siamese Network, Triplet Network, and Lifted Structured methods respectively. Figure 3 shows some example query success stories using our embedding (dimension 512). Despite drastic changes in perspective, configuration and lighting in the database, our method can still successfully retrieve the same class in the sample. Because some product categories contain only a small number of samples (2 or 3) and product categories do not completely exclude non-related samples. These have a great influence on the image training and testing. Figure 4 and Figure 5 shows some of the failed cases in the sample query using our embedding (dimension 512), that is, only images similar to the original are found.

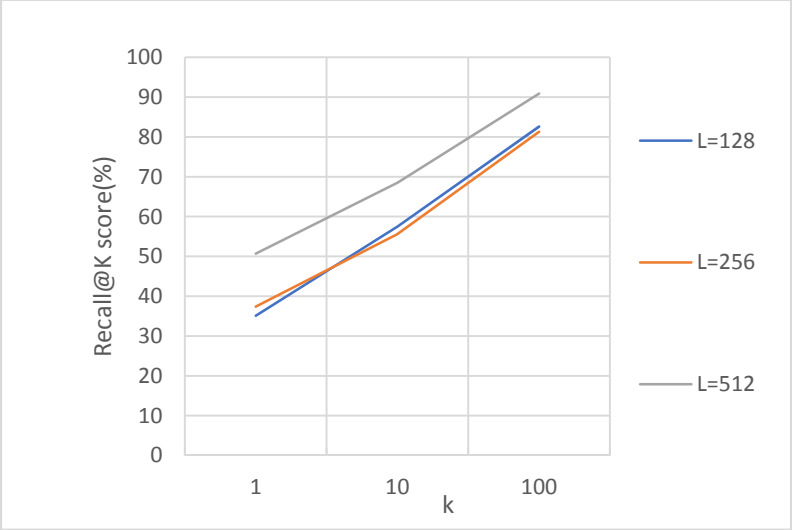

Fig. 2 Effect of embedding dimensions

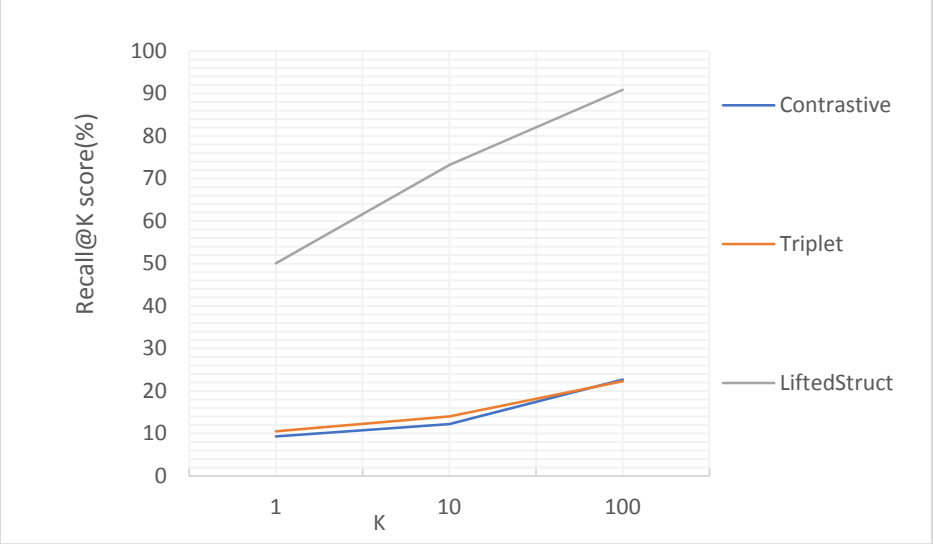

Fig. 3 Accuracy using different matching algorithms

Fig. 4 Example of successful furniture database queries Images in the first column are query images and rest are three nearest neighbors.

Fig. 5 Examples of failed furniture database queries. Images in the first column are query images and rest are three nearest neighbors

5. Conclusion

In this paper, the deep learning algorithm is successfully applied to the furniture image search. It also proves that LSFE also has excellent matching ability in big data. We can successfully search the same products and similar style of similar products through this application. However, despite the initial preprocessing of the database, the quality of the image data is still not high. So for the image training and testing have a certain of impact.

Future work will mainly focus on improving the image quality of the furniture database and then improve search accuracy.

1). Manually remove the remaining irrelevant image;

2). Image detection, extraction of specific icons in the furniture images (there is some images in the database which have different types of product icons).

References

[1]. Song H O, Xiang Y, Jegelka S, Savarese S. Deep metric learning via lifted structured feature embedding. In IEEE Conference on Computer Vision and Pattern Recognition,2016[C] [S.l.] :4004-4012.

[2]. Nister D, Stewenius H. Scalable recognition with a vocabulary tree. In IEEE Conference on Computer Vision and Pattern Recognition,2006[C] [S.l.] :2161-2168.

[3]. Philbin J, Chum O, Isard M, Sivic J, Zisserman A.Object retrieval with large vocabularies and fast spatial matching. In IEEE Conference on Computer Vision and Pattern Recognition,2007[C] [S.l.] :1-8.

[4]. Krizhevsky A, Sutskeve R, Hinton G E.ImageNet classification with deep convolutional neural network in Advances. In Neural Information Processing Systems,2012[C].

[5]. Deng J, Dong , Socher R, et al.Image Net: a large-scale hierarchical image database. In IEEE Conference on Computer Vision and Pattern Recognition, 2009[C][S.l.] : 248 -255.

[6]. Razavian A S, Azizpour R H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition. In IEEE Conference on Computer Vision and Pattern Recognition Workshops,2014[C][S.l.]: 512 -519.

[7]. Wan J, Wang D, Hoi SCH, et al. Deep learning for content-based image retrieval: a comprehensive study. In ACM International Conference on Multimedia In Acm International Conference on Multimedia ,2014[C]: 157 -166.

[8]. Hoffer E, Ailon N: Deep metric learning using triplet network. In: SIMBAD. (2015) .

[9]. Song H O, Jegelka S, Rathod V, Murphy K.Learnable structured clustering framework for deep metric learning. In IEEE Conference on Computer Vision and Pattern Recognition,2016[C] [S.l.].

[10]. Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering.In IEEE Conference on Computer Vision and Pattern Recognition, 2015 [C] [S.l.].

[11]. Christian Szegedy,Wei Liu.Going deeper with convolutions In Computer Vision and Pattern Recognition, 2014 [C].