Neurocomputing 207 (2016) 387–397

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep sketch feature for cross-domain image retrieval

Xinggang Wang, Xiong Duan, Xiang Bai *

School of Electronic Information and Communications, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei Province 430074, PR China

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form
30 March 2016
Accepted 29 April 2016
Communicated by Zhang Zhaoxiang
Available online 13 May 2016

Keywords:
Sketch recognition
Image retrieval
Deep learning

ABSTRACT

Deep learning has been proven to be very effective for various image recognition tasks, e.g., image classification, semantic segmentation, image retrieval, shape classification, etc. However, existing works on deep learning for image recognition mainly focus on either natural image data or binary shape data. In this paper, we show that deep convolutional neural networks (DCNN) is also suitable for cross-domain image recognition, i.e., using sketch as query to retrieve natural images in a large dataset. To solve this kind of cross-domain problem, we propose to train CNN jointly using image data and sketch data in a novel way. The learned deep feature is effective for cross-domain image retrieval – using simple Euclidean distance on the learned feature can significantly outperform the previous state-of-the-arts. In addition, we find that pre-training and a feasible data-argumentation for DCNN can largely surpass human-level performance in the standard sketch classification benchmark.

1. Introduction

Drawing sketch is a simple way for people to show ideas intuitively. It has been widely used in human computer interaction (HCI) systems. In computer vision, sketch recognition is an important yet challenging problem, since sketch is different from both binary shape and natural image, does not contain textual information and has large variation in shape. Sketch recognition is useful in real many applications. For example, a very interesting application in sketch recognition is sketch-to-image retrieval, that is using a sketch as query to find the images with the similar semantic content in a large image set. This is a cross-domain image retrieval problem. Based on sketch-to-image retrieval, Microsoft has developed a popular product called MindFinder.[1] The core of Mindfinder is to use chamfer distance and inverted file algorithm to fast compare the query sketch with the edge-maps of millions images collected from web [1]. However, due to the simplicity of the sketch feature in MindFinder and the difficulty of extracting true edges in images, it is impossible to recognize complex sketch and images.

To study the sketch recognition problem, in the previous research works, both binary shape descriptors, such as chamfer matching [38] and shape context [5], and natural image descriptors, such as SIFT [26] and HOG [12], have been applied for sketch recognition. As mentioned above, in [1] a popular shape descriptor, chamfer matching, is used in MindFinder. Besides shape descriptor, some image descriptors have been adopted. For example, the classic image descriptor, SIFT, has been directly applied for sketch retrieval [16]. For sketch having strong gradient information which could be captured by SIFT, SIFT based methods show good performance in sketch recognition applications. However, SIFT was designed for natural image and it is only able to describe local information in sketch, which makes SIFT not a perfect sketch descriptor. More methods for sketch recognition will be given in the following related work section.

Recently, deep learning methods, such as deep convolutional neural networks (DCNN), deep auto-encoder (DAE), and deep recurrent neural networks (DRNN), have been proven to be very successful for image recognition. The key advantage of deep learning methods is that those methods can learn features directly from image, including low-, middle- and high-level image features, rather than feature engineering.

Following the path of learning feature using deep networks, we propose to learn sketch feature for sketch classification and sketch to image retrieval using DCNN. As we all know, sketch and image[2] have large appearance differences even when they belong to the

* Corresponding author.
E-mail address: xbai@hust.edu.cn (X. Bai).

[1] http://research.microsoft.com/en-us/projects/mindfinder

[2] For simplicity, in this paper, image means natural image and does not contain sketch.

http://dx.doi.org/10.1016/j.neucom.2016.04.046
0925-2312/

same semantic category. This is the semantic gap between sketch and image. Due to the semantic gap, hand-crafted features are not able to get good performance on sketch recognition. The key idea of the proposed method is to learn a feature extractor which is able to extract similar features for sketches/images if they have the same semantic label. We choose DCNN as the feature extractor. The label information of sketch/image is used as supervision to guide the feature learning via back-propagation as training DCNN. The learned feature is named as deep sketch feature (DSF).

In practice, there are several challenges in learning deep sketch feature: (1) There are limited number of sketches for training while DCNN requires a large number of data to learn millions of parameters inside. (2) As mentioned above, image and sketch are in different domains and there is a big semantic gap between image and sketch even when they belong to the same subject. To deal with these problems, we carefully investigate the steps of data argumentation, fine-tuning, etc. in training DCNN. To better understand the learned neural features, we visualize the processes of feature extraction. And, we carry out extensive experiments to validate the effectiveness of the learned features for both sketch classification and cross-domain image retrieval.

In the following parts of the paper, we review related works of sketch recognition and deep learning in Section 2, describe the proposed framework for sketch feature learning and applications in Section 3, carry out experiments for sketch classification and retrieval in Section 4, and finally conclude the paper in Section 6.

2. Related work

Since the touch-based smart computing devices became popular and lots of people use them instead of pen to draw simple things come from minds, the research of sketch/shape recognition attracts more attention, and produce some interesting applications [10] and is potentially be applied in other problems [13]. Two types of features have been used in sketch recognition.

On one hand, some researchers use binary shape features in sketch recognition. Shape matching and recognition have been studied for a long time. Previously, sketch recognition problem is deemed to a branch of shape problems, and naturally many works use the classic binary shape descriptors, such as shape context [5], chamfer matching [38], shape band [4], bag of contour fragments [40], two layer coding [3], shape-index feature [41], to solve sketch/shape recognition problems. Those descriptors work well on the conventional shape recognition datasets which contain shapes with clean structure. For natural image, sketch can be extracted using Canny edge detector or advanced deep learning edge detector [35].

However, these shape descriptors do not work well with sketch since sketch has larger appearance variation, especially, it usually has many internal contours [9,11]. In addition, sketches are drawn by different people and different people have different drawing style. Hence, only a few attempts using binary shape features succeed. As far as we know, Olsen et al. [30] made a survey that their method uses points and curves of a sketch to match the 2D templates from 3D model. Eitz et al. [15] developed a descriptor that is based on structure sketched feature lines. Cao et al. [7] built a large-scale sketch-based image search system using an indexable oriented chamfer matching method. Parui et al. [31] proposed a sketch-based image retrieval approach for large datasets using a middle descriptor based on sketch chains. Though these methods show their efficiencies due to simplified matching procedures by an indexing mechanism, they are only able to handle simple images with clean background.

On other hand, a large number of studies employ natural image features or their variants to represent sketches. Eitz et al. [16] proposed a benchmark of sketch-based image retrieval and used a lot of features as well as a BOF presentation model to compare those features performances. It is found that shape descriptors are not directly suitable for the task of SBIR and HOG outperforms others. Later, Eitz et al. [14] collected a dataset of 20,000 unique sketches evenly distributed over 250 object categories, which are totally completed by no-expert free hands, and they investigated the sketch recognition issue in both the humans and the proposed method using SIFT-variant descriptor. Recently, Ma et al. [28] used the key-point-based detector to localize the local features. They described the detected stroke features by a quantized HOG of sampled stroke points and then used a codebook organized in a hierarchical vocabulary tree to maintain the structure information of visual words. Li et al. [20,21] gave up shape matching techniques and chose a star graph based on ensemble matching of structured features strategy to recognize sketch and achieved a good performance in classification accuracy. Furthermore, they proposed a multiple kernel learning (MKL) framework to fuse several features to sketch and further they got four percentage improvement in classification accuracy. Schneider et al. [34] also used SIFT features and received a state-of-the-art classification accuracy in traditional ways. Overall, those works directly apply the hand-designed local natural image features for sketch recognition.

Lots of researches have given up the low level descriptors design for sketches, and most of them have a common sense that natural image features indeed outperform than shape features, although low-level descriptors have achieved impressive performance in various recognition tasks [44,45,39,23–25,42]. However, their methods are simple and nothing more than three steps: (a) selecting the most informational areas or points; (b) computing the best local features; (c) building a structure model as well as a evaluation scheme by spatial cues (structural information) or local feature similarity (matching score). These frameworks seriously depend on parameters of the chosen local features and models selected. It means that we should try various features and models, and ensemble them to find a best state. What's more, sketch is easy to draw but hard to recognize so that models based on specific local features are not robust enough for the abstract sketch. None of the above methods can exceed the humans' recognition in sketch.

Nowadays, deep convolutional neural network makes breakthroughs and shows strong performances in computer vision tasks. Since it was firstly used in the Imagenet classification [19] and made a great improvement in the image classification task, DCNN has been extended to many traditional areas such as face recognition [37], object detection [18], and image retrieval [2]. And most of them have achieved better results than traditional methods. Unlike conventional hand-crafted features, DCNN is an end-to-end model trained from image to class label. When we extract different layers output from the learnt model, we actually receive different level features. Recently, Babenko et al. [2] investigated the use of that different layers' features within image retrieval and achieved an improvement in the retrieval performance of neural codes. However, we noticed that few research use deep feature in the sketch based image retrieval. As far as we know, Chen et al. [8] present a system that converts a simple freehand sketch automatically into a photo-realistic picture by seamlessly composing multiple images discovered online. Sketch is always treated as low level concepts such as shape, edge map, contour or strokes that prompted large researchers focusing on low-level features design for sketch. While a total sketch is associated with a real object or scene, this makes the sketch a high-level concept which is far more difficult than natural images. Recently, there are some methods use DCNN for sketch recognition, such as [33,43]. Compared to these methods, the proposed sketch classification method