J. Vis. Commun. Image R. 55 (2018) 756–760

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A method of multi-criteria set recognition based on deep feature representation ☆

Caiyou Zhang *, Man Zhou, Hongzhen Yang, Xiaojun Shen, Yanbo Wang

State Grid Zhejiang Electric Power Company Information & Telecommunication Branch, Hangzhou, China

ARTICLE INFO

Article history:
Received 24 June 2018
Revised 6 August 2018
Accepted 14 August 2018
Available online 16 August 2018

Keywords:
Person re-identification
Convolutional neural network
Multiple metric ensembles

ABSTRACT

The large variations in the angle of different camera views and illumination can change the appearance of a lot of people, which makes human re identification is still a challenging problem. Therefore, the development of robust feature descriptors and the design of discriminative distance metrics to measure similarity between pedestrian images are two key aspects of human re identification. In this paper, we propose a method to improve the performance of the re identification using depth learning and multiple metric ensembles. First, we use a variety of data sets to train the general convolutional neural network (CNN), which is used to extract the features of the training and test set after deep level. Deep architecture makes it possible for people to learn more abstract and internal features that are robust to changes in viewpoint and illumination. Then, we utilize the deep features of the training set to learn a specific distance metric and combine it with the cosine distance metric. Multi metric sets can be used to measure the similarity between different images. Finally, a large number of experiments show that our method can effectively improve the recognition performance compared to the state-of-the-art methods.

1. Introduction

Person recognition is an important task in the identification task, an intersection with the individual in the camera view [1], which is in the video surveillance system and the human-computer interaction system. The huge changes in the appearance of the same person in different camera views of the pose, illumination, occlusion and complex changes caused by camera settings are still a challenging problem.

A representative feature descriptor [2–6] has been frequently used in the re identification of people, usually including color, texture, shape, gradient and local patch. In [2], a set of local features (ELF) proposed a subset of color and texture features of choice representation; in [4], intermediate filter (intermediate) is proposed, which based on coherent appearance by hierarchical cluster tree pruning achieved patch clusters and view invariant discriminant feature; color and texture features of symmetry different views of slander and asymmetric o et al. [5] of the people re identification (sdalf); Liao et al. [6] by using sliding window to describe the characters and the local characteristics of HSV and SILTP local histogram levels, obtained a strong objections to change feature representation. Although a variety of functional descriptors have been proposed, and has made impressive improvements, people re identification, or learning how to design a powerful function of complex changes, both in the lighting and view is still a challenging problem. In [28], the atuhors discussed some widely-used deep learning architectures and their practical applications. An up-to-date overview is provided on four deep learning architectures, namely, autoencoder, convolutional neural network, deep belief network, and restricted Boltzmann machine.

Metric learning [6–12] is also widely used for human recognition and has obtained good performance in some cases. Mignon et al. [8] presents a pairwise constraint component analysis (PCCA) learning distance metric for sparse 22 similarity/dissimilarity constraints in high dimensional spaces. In [9], a large nearest neighbor (LMNN) classifier is proposed to learn Mahalanobis distance metric for the K- nearest neighbor (KNN) classification using semidefinite programming. Two discriminant analysis (xqda) () [Abstract] a new method of learning QDA metric to discriminate the low dimensional subspace from the perspective of the two discriminant analysis is proposed. Although these methods have achieved remarkable performance, people re identification, most of them work only two or three data sets and a single distance measure is one-sided, find the difference between the different images and connection.

☆ This article is part of the Special Issue on TIUSM.

* Corresponding author.
E-mail address: 634423778@qq.com (C. Zhang).

https://doi.org/10.1016/j.jvcir.2018.08.011
1047-3203/

Recently, in-depth study of several people re identified [13–16,33–38] method. Deep learning has been paid more and more attention in the field of computer vision. The depth learning method can obtain more abstract and internal features compared with the above manual features automatically, and the deep structure makes it possible to learn more complex geometric and photometric transform gain robust features. Li et al. [13] proposed a new filter for neural network (FPNN) encoding conversion and learning optimal feature automatic photometry. The texture features, color features and metrics of the "conjoined" deep neural network to learn the set of data sets are completely set up by one person [14]. Ahmad et al. [15] proposed a deep convolution structure specifically designed to capture the middle layer based on the relationship between the two views from each view. Shaw et al. [16] proposes a domain oriented drop out algorithm to determine whether a particular neuron works or does not work on different data sets. Although these deep learning methods based on the performance of people in a certain extent improve the re identification, due to the large training set, the number of parameters it requires learning a large, small data sets usually cannot get significant results [13,15]. Different sets of data need to learn different network parameters to better performance in general, taking advantage of the change of multiple data sets, but spent a lot of time [13–15].

In order to solve these problems, we propose a method to extract the distance between images by using depth learning to extract features and multiple metric ensembles. We first combine a variety of benchmark data sets to train our convolution neural network (CNN) as a generic feature extraction for all data sets. Set in the CNN model is trained by using a combination of various data, the training set becomes more diversity and learning function become more powerful, in addition, small data sets can also be applied to deep learning. Then, we use the xqda [6] to learn different data sets, different distance measure, can change a plurality of data sets and the advantage of fast, many CNN compared to fine-tuning to adapt to different data sets. Finally, we put the similarity between images learned distance and cosine distance metric, and the experimental results show that by adding a simple cosine distance metric, we have good performance, verify that the two different distance measure is complementary to the good. Note that the individuals we choose for the training set are the same depth learning and distance metric learning. Experiments show that our proposed method achieves superior performance compared to state-of-the-art works.

2. Related work

Person re-identification or person tracking is an important application in intelligent systems. With the development of video surveillance, person re-identification requires a higher speed and accuracy. But multi-view camera lead to different shape or appearance of the same person in different angle. So in this paper, we propose a novel method to improve the performance of the re-identification using depth learning and multiple ensembles.

In the next section, we model the system structure of the CNN are described, then we introduced the measurement ensemble of section third, section fourth shows a large number of experiments to verify the effectiveness of our approach, and the fifth part is the conclusion of our paper.

3. Convolutional neural network architecture

We build our model inspired by CNN [16–18], the architecture of the proposed network is described in Table 1.

Table 1
The architecture of proposed network.

Name	Patch size/stride or remarks	Input size
conv1	$3 \times 3/1$	$144 \times 56 \times 3$
conv2-conv 3	$3 \times 3/1$	$144 \times 56 \times 32$
Pool 3	$2 \times 2/2$	$144 \times 56 \times 32$
Inception 4a	As in Fig. 1(a), stride 1	$72 \times 28 \times 32$
Inception 4b	As in Fig. 1(a), stride 2	$72 \times 28 \times 256$
Inception 5a	As in Fig. 1(a), stride 1	$36 \times 14 \times 384$
Inception 5b	As in Fig. 1(a), stride 2	$36 \times 14 \times 512$
Inception 6a	As in Fig. 1(b), stride 1	$18 \times 7 \times 786$
Inception 6b	As in Fig. 1(b), stride 2	$18 \times 7 \times 1024$
global pool	$9 \times 4/1$	$9 \times 4 \times 1536$
fc7	logits	$1 \times 1 \times 1536$
fc8	logits	$1 \times 1 \times 1536$
softmax	classifier	$1 \times 1 \times 2633$

The structure of our CNN is the same as [16] is expected for the last two start modules and the two fully connected layers. Fig. 1 shows the structure of the self starting module that is used, and the two boot modules are presented in [17] for image classification, Fig. 1(b) extended filter bank output to facilitate high dimensional representation with Fig. 1(a). After that, two fully connected layers are applied, first of all, with 1536 channels and the number of individuals in the training set containing the set of second channels, which are set to 2633.

4. Multiple metric ensembles

At the end of the training the above network output layer, FC7 extracted from the deep features of pedestrian images, we use multi-scale integration to measure the similarity between these image features. As the essence of distance metric learning is to obtain a measurement matrix, can be in a more discriminatory manner projection function, we first used the deep characteristic of CNN training set, learning specific distance metric for different data sets. Considering the function of general information, we measure the similarity between different images by combining learning distance measure and cosine distance measure. Here is a presentation of our distance metric learning method.

4.1. Distance metric learning

In this paper, we use the xqda algorithm [6] to learn discriminant space while an effective distance metric. The xqda algorithm is an extension of the kiss I [11] method.

Kiss my algorithm so that the optimal statistical decision is similar to the image or not by the log likelihood ratio test from

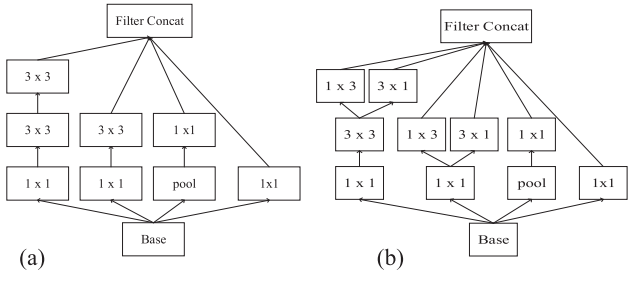

Fig. 1. In the structure of our CNN starting module, module (a) and (b) are proposed in [17] image classification, the execution module and memory module and calculate the budget even in the strict constraints (b) is an expanded (a) to promote the coarse grid.ieee high dimensional representation of the final format of your paper will do the. If your paper is for the meeting, please comply with the restrictions on the conference page.