PERSON RE-IDENTIFICATION WITH DEEP DENSE FEATURE REPRESENTATION AND JOINT BAYESIAN

Shengke Wang, Lianghua Duan, Na Yang, Junyu Dong

Department of Computer Science and Technology, Ocean University of China, China

ABSTRACT

Person re-identification that aims at matching individuals across multiple camera views has become indispensable in intelligent video surveillance systems. It remains challenging due to the large variations of pose, illumination, occlusion and camera viewpoint. Feature representation and metric learning are the two fundamental components in person re-identification.

In this paper, we present a Special Dense Convolutional Neural Network ***(SD-CNN) to extract the feature*** and apply ***Joint Bayesian to measure the similarity*** of pedestrian image pairs. The SD-CNN can preserve more horizontal information to against viewpoint changes, maximize the feature reuse and ensure feature distributing discriminative. Joint Bayesian models the extracted feature representation as the sum of inter- and intra-personal variations, and the joint probability of two images being a same person can be obtained through log-likelihood ratio. Experiments show that our approach significantly outperforms state-of-the-art methods on several benchmarks of person re-identification.

METHODS

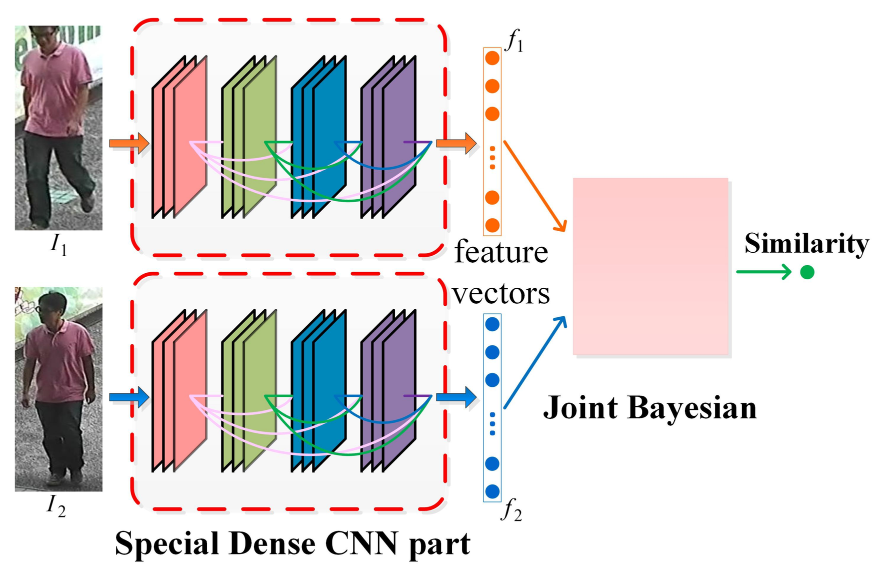


Fig. 1. The overview of our framework

We propose a novel **framework** (See Fig. 1) to solve person re-identification problem that **including a SD-CNN feature extractor and a Joint Bayesian model for distance metric.**

The proposed Special Dense Convolutional Neural Network (**SD-CNN**) architecture can **outperform majority** of existing deep learning extractor for person re-identification. Basing on our efficient SD-CNN feature, we applied **Joint Bayesian** to person Re-ID problem for the first time and get a **2%** performance **improvement.**

SD-CNN STRUCTURE FOR FEATURE EXTRACTION

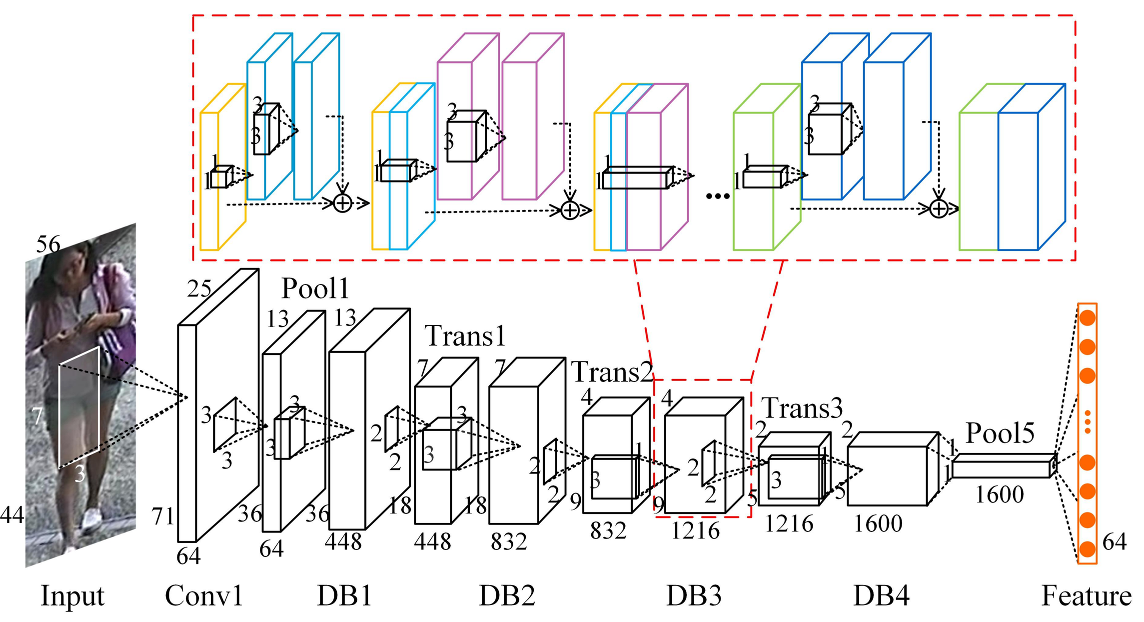


Fig. 2. The SD-CNN structure for feature extraction. Note that the red box is the example of dense block(DB).

An illustration of the SD-CNN structure is shown in Fig. 2. Compared with DenseNet[1], our SD-CNN has the following characteristics:

① Due to the pose variations of one pedestrian across different views, the local features appearing in one views may not exactly at the same position in the other view while its very similar along the same horizontal region. Therefore, **we use asymmetric filtering in some critical convolution layers to preserve the horizontal features.**

② **Mmount of 1 × 1 convolution are employed to reduce the dimension, interact and integrate the information across channels, and increase the nonlinear characteristics.**

③ Softmax loss function can supervise feature distributing discriminative, for propagating this constraint better **we abandon the activation function between the Feature layer and last full connected layer.**

The network is trained to minimize the cross-entropy loss by using the stochastic gradient descent search, combined with error back-propagation.

$$L(f, y, \theta_{id}) = -\sum_{i=1}^{n} p_i \log q_i$$

where f is the feature vector, y is the target class, and θ_{id} denotes the softmax layer parameters. p_i is the true probability distribution that $p_i = 0$ for all i except $p_y = 1$ for the target class y, q_i is the predicted probability distribution.

JOINT BAYESIAN FOR DISTANCE METRIC

We learned the Joint Bayesian model for distance metric based on the extracted SD-CNN feature. Follow the Joint Bayesian [2], **the feature of a pedestrian image can be represent as the sum of inter- and intra-personal variations:**

$$x = \mu + \varepsilon$$

where μ and ε follow two Gaussian distributions $N(0, S_\mu)$ and $N(0, S_\varepsilon)$ can be estimated from the training data.

Given the features of image pairs $\{x_1, x_2\}$, extracted by SD-CNN from two images, Let H_I represents the intrapersonal (same) hypothesis that two images belong to the same person, and H_E is the extra-personal (not same) hypothesis, then the person re-id problem amounts to classifying the difference $\Delta = x_1 - x_2$ as intra-personal variation or extra-personal variation. Based on the MAP(Maximum a Posterior) rule, **the distance is made by testing a log-likelihood ratio :**

$$d(x_1, x_2) = \log\frac{P(\Delta \mid H_I)}{P(\Delta \mid H_E)} = x_1^T A x_1 + x_2^T A x_2 - 2x_1^T G x_2$$

Where A and G can be estimated by the algorithm in Table 1

Tabel 1. The Joint Bayesian learning algorithm. Assume there are n identities and each identity has mi images.

While not converge **do**

$t \leftarrow t + 1.$

$F = S_\varepsilon^{-1}, G = -(x_i S_\mu + S_\varepsilon)^{-1} S_\mu S_\varepsilon^{-1}$

$\mu_i = \sum_{j=1}^{m_i} S_u (F + m_i G) x_j, \varepsilon_{ij} = x_j + \sum_{j=1}^{m_i} S_\varepsilon G x_j$

Update the parameters S_μ by $S_\mu = \frac{1}{n}\sum_i \mu_i \mu_i^T$

Update the parameters S_ε by $S_\varepsilon = \frac{1}{n}\sum_i \sum_j \varepsilon_{ij}\varepsilon_{ij}^T$

end while

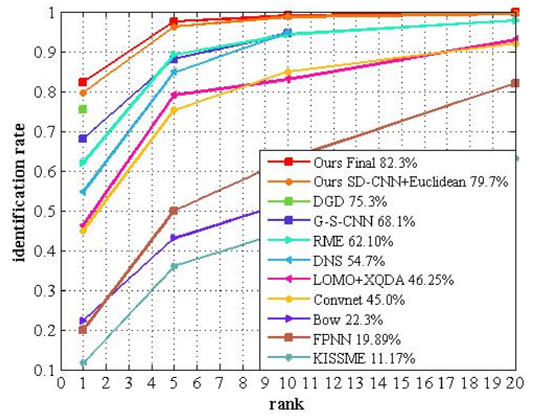
$F = S_\varepsilon^{-1}, G = -(2S_\mu + S_\varepsilon)^{-1} S_\mu S_\varepsilon^{-1}$

$A = (S_\mu + S_\varepsilon)^{-1} - (F + G)$

Output A, G

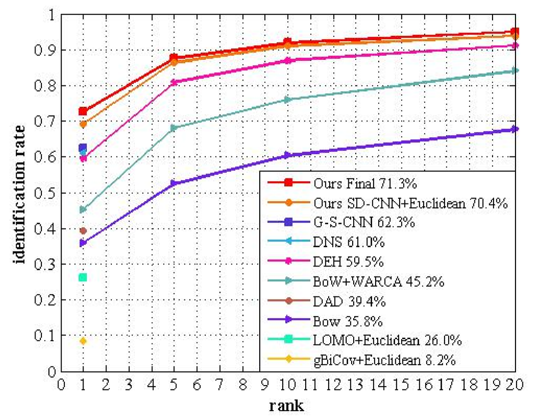
RESULTS

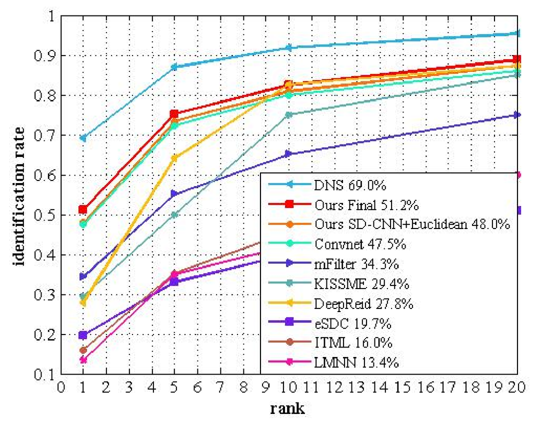
We present a comprehensive evaluation of our framework by comparing it against the baseline SD-CNN feature with Euclidean distance as well as other state-of-art methods for person re-identification. All evaluations is based on the Cumulative Matching Characteristics (CMC).



Compared with deep learning methods, our SD-CNN features with simple Euclidean distance easily bests the other methods in **CUHK03**[3] dataset, and with the improvements of our Joint Bayesian, our framework achieves the **state-of-the-art result with rank-1 accuracy up to 82.3%.**

For **Market-1501**[4] dataset, our final framework gains 72.6% rank-1 accuracy, both Euclidean distance and Joint Bayesian with our SD-CNN feature are **greatly better than the previous methods results .**





However, for small dataset **CUHK01**[5], it would be insufficient to learn such a large capacity network from scratch, our framework with SD-CNN and Joint **Bayesian fail to achieve the best result (compared with DNS [6]). But compared with others our method get the best result.**

CONCLUSIONS

This paper proposed a novel framework for person reidentification, it consist of a convolutional neural network extractor named SD-CNN and metric measure named Joint Bayesian. Experiments shown that our framework achieved state-of-the-art result in several dataset.

REFERENCES

[1] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger, "Densely connected convolutional networks," 2016.

[2] Dong Chen, Xudong Cao, Liwei Wang, Fang Wen, and Jian Sun, "Bayesian face revisited: a joint formulation," in European Conference on Computer Vision, 2012, pp. 566–579.

[3] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159.

[4] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian, "Scalable person reidentification: A benchmark," in IEEE International Conference on Computer Vision, 2015, pp. 1116–1124.

[5] Junlin Hu, Jiwen Lu, and Yap Peng Tan, "Deep transfer metric learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 325–333.

[6] Li Zhang, Tao Xiang, and Shaogang Gong, "Learning a discriminative null space for person re-identification," pp. 1239–1248, 2016.

CONTACT

Department of Computer Science and Technology, Ocean University of China, China

neverme@ouc.edu.cn

lianghua_duan@outlook.com

dongjunyu@ouc.edu.cn

Code is avaliable at https://github.com/duanLH/