Coarse-to-Fine Person Re-Identification with Auxiliary-Domain Classification and Second-Order Information Bottleneck (Supplementary Material)

Anguo Zhang[1,2] Yueming Gao[1,2] Yuzhen Niu*[3] Wenxi Liu[3]
Yongcheng Zhou[1]
1. College of Physics and Information Engineering, Fuzhou University
2. Key Laboratory of Medical Instrumentation and Pharmaceutical Technology of Fujian Province
3. College of Mathematics and Computer Science, Fuzhou University

1. Basic Information Theory

To better understand the theoretical derivation in Section 3.2 of the paper, we need to introduce some basic information theory. Due to the length limitation of the paper, we put it in the supplementary material.

Let X and Y be two random variables that take values x and y from two finite sets with distribution $p(x)$ and $p(y)$, respectively. The *Shannon entropy* $H(X)$ measures the average stochasticity of X, which is defined by

$$H(X) = -\sum_{x} p(x) \log p(x). \tag{1}$$

If not specified, all the logarithms are base 2 and entropy is expressed in bits. The conditional entropy $H(Y|X)$ measures the average stochasticity associated with Y under a known outcome of X, which is defined by

$$H(Y|X) = -\sum_{x,y} p(x)\, p(y|x) \log p(y|x). \tag{2}$$

In general, it holds $H(Y|X) \neq H(X|Y)$, and $H(X) \geq H(X|Y) \geq 0$.

Further, the mutual information between X and Y measures the shared information of X and Y, which is defined by

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= \sum_{x,y} p(x)\, p(y|x) \log \frac{p(y|x)}{p(y)} \\ &\geq 0. \end{aligned} \tag{3}$$

Let T be the intermediately compressed representation of X in the way of Markov chain $Y \leftrightarrow X \leftrightarrow T$. For a given encoding map from X to the variable T, and two conditional probability $q(t|x)$ and $q(y|t)$, we can compute the joint distribution

$$q(x,t) = q(t|x)\, p(x), \tag{4}$$

$$q(y,t) = \sum_{x} q(t|x)\, p(x,y), \tag{5}$$

and

$$q(t|x,y) = q(t|x). \tag{6}$$

Subsequently,

$$q(x,y,t) = p(x,y)\, q(t|x,y) = p(x,y)\, q(t|x), \tag{7}$$

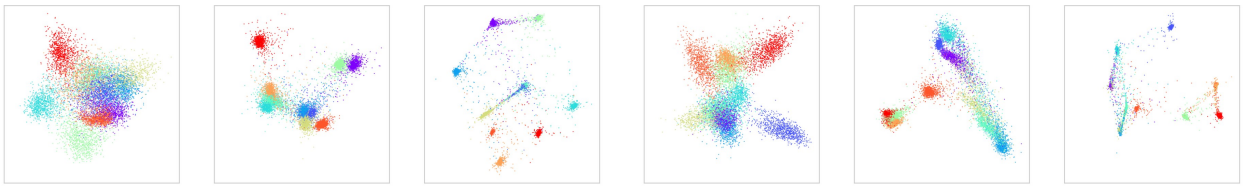
$$\begin{aligned} q(t|y) &= \frac{1}{p(y)} \sum_{x} q(x,y,t) = \frac{1}{p(y)} \sum_{x} p(x,y)\, q(t|x) \\ &= \sum_{x} p(x|y)\, q(t|x), \end{aligned} \tag{8}$$

$$\begin{aligned} q(y|t) &= \frac{1}{q(t)} \sum_{x} q(x,y,t) = \frac{1}{q(t)} \sum_{x} p(x,y)\, q(t|x) \\ &= \sum_{x} p(y|x)\, q(x|t). \end{aligned} \tag{9}$$

Further,

$$q(t) = \sum_{x} p(x)\, q(t|x). \tag{10}$$

* Yuzhen Niu is the corresponding author (e-mail: yuzhenniu@gmail.com).



(a) MNIST (b) Fashion-MNIST

Figure 1. Principal component analysis (PCA) transformed two-dimensional projections of neuronal activities of bottleneck layer (on test data, no noise) for models trained with regular cross-entropy loss (left), Variational IB (middle), and the proposed 2O-IB (right), respectively. Top row: tested on MNIST dataset; Bottom row: tested on Fashion-MNIST dataset.

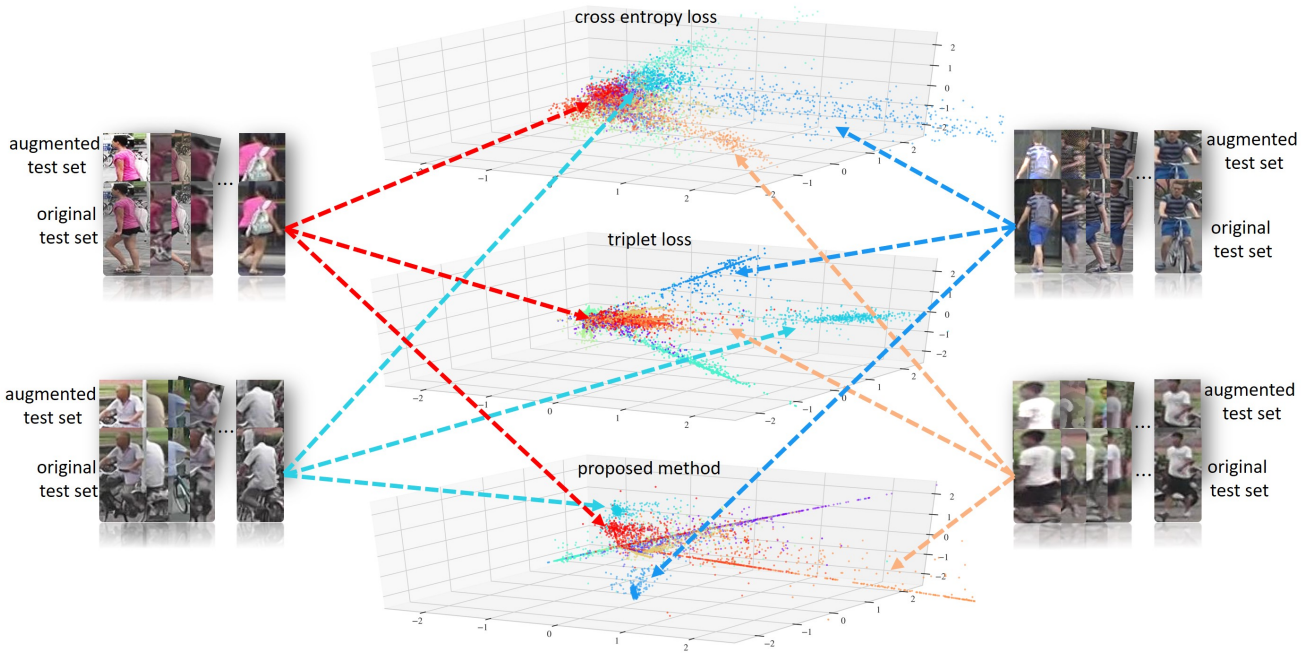


Figure 2. Three-dimensional PCA projection of the output embedding vector for 10 different pedestrians. The projection points of each pedestrian are marked with different colors. For more intuitive presentation, input images of four pedestrians are shown and they are pointing to their corresponding projection points, where *original test set* denotes the original images of a pedestrian from test set, *augmented test set* denotes the augmented images.

2. Effectiveness of the Proposed Second-Order Information Bottleneck

In order to illustrate the effectiveness of our proposed second-order information bottleneck (2O-IB) more clearly and intuitively, we first compare the compression performance of intra-class variations of 2O-IB with the variational IB [1] and cross entropy optimizer on two simple and commonly used image classification datasets, i.e., MNIST [2] and Fashion-MNIST [3]. We record the neuronal activation of 10-neuron information bottleneck layer on the test dataset (if cross entropy loss is used as the optimization target of the neural network, the information bottleneck layer will be replaced with the basic fully connected layer), and project it to a 2-dimensional plane by principal component analysis (PCA).

As shown in Figure 1, the differences in neuron representation for the two datasets are shown, where different colors represent different sample classes. Training with variational IB and 2O-IB objectives causes inputs corresponding to different classes to fall into well-separated clusters, unlike training with cross-entropy loss. Moreover, the cluster with 2O-IB is tighter than that with variational IB, which indicates that the 2O-IB states carry information that most related to class identity, thus reduces intra-class variations.

Furthermore, in Figure 2, we show the three-dimensional PCA projection of the output embedding vectors for 10 different pedestrians computed by using cross entropy loss,

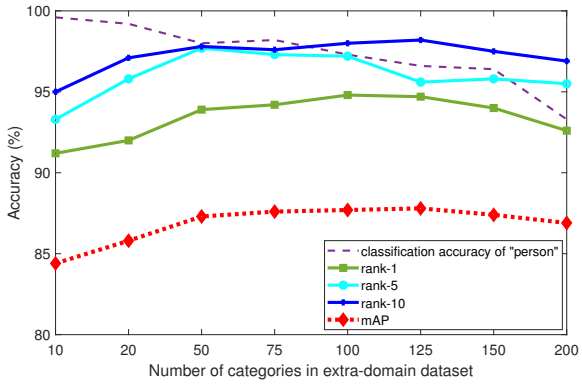


Figure 3. Re-ID performance on the cases of different number of categories in extra-domain dataset.

triplet loss, and the proposed framework. Among them, test images of four different pedestrians and their links to the corresponding projection points are shown. The label "cross entropy loss" for the top projection subfigure means the network (ResNet50) is trained using the cross entropy loss function on conventional Market1501 [4], the label "triplet loss" for the central subfigure denotes the network (ResNet50) is trained using the triplet loss with hard positive/negative mining on Market1501, and "proposed method" denotes our proposed framework that trained by using the proposed ADC and 2O-IB methods.

From Figure 2, we can see that compared to "cross entropy loss" and "triplet loss" training, our proposed method can make the projection point cluster for each category of pedestrians much tighter, which means that it can more effectively handle the intra-class variances and make the output embedding vectors more consistent for each class.

3. Ablation Study With the Extra-Domain Dataset

Our proposed framework uses an extra-domain dataset for the auxiliary-domain classification (ADC) task. In our submission, we show that the ADC task can improve the ReID performance in Table 6. Besides, to validate the robustness of the proposed framework when using different extra-domain datasets, we also experiment with the extra-domain datasets using a different number of object categories and show the experimental results in Figure 3.

As the number of categories in the extra-domain dataset increases, the classification accuracy of "person" gradually decreases. This is because in the ADC task we proposed, the depth of the coarse-grained feature extraction (CGFE) module is not very deep, so its computational power is limited. With the continuous increase of image categories, CGFE needs to maximize the overall classification accuracy on the extra-domain dataset plus Re-ID dataset, so it will balance the classification performance of person and other categories. Therefore, the classification accuracy for "person" category only is slightly reduced to optimize the performance on the entire dataset.

Furthermore, the increase in the number of categories in the extra-domain dataset does not necessarily increase the rank and mAP performance of the Re-ID task. The performance can be optimized only when the number is appropriate. As can be seen in Figure 3, when the number of extra image categories is between 50 and 100, the performance of both rank and mAP is relatively highest. Among them, when the number of categories in extra-domain dataset is 100 and 125, the rank-1 of our proposed model is the largest (94.8% and 94.7%, respectively); when the number of categories is 50, the rank-5 is the largest (97.7%). While the values of rank-10 and mAP vary much smaller when the number of categories is between 50 and 100.

In summary, the proposed framework can achieve robust and significant performance improvements when the number of categories is between 50 and 100. And in the experiments, the extra-domain dataset has 100 categories.

References

[1] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep Variational Information Bottleneck. In *International Conference on Representation Learning*, pages 1–19, 2017. 2

[2] Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2

[3] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *Arxiv*, 2017. 2

[4] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE International Conference on Computer Vision*, 2015. 3