Supplementary Material for "Weakly Supervised Discriminative Feature Learning with State Information for Person Identification"

Hong-Xing Yu
Sun Yat-sen University, China
xkoven@gmail.com

Wei-Shi Zheng
Sun Yat-sen University, China
wszheng@ieee.org

1. Algorithm description and details

We summarize our model in Algorithm 1. The equations in Algorithm 1 are defined in the main manuscript.

We used standard data augmentation (random crop and horizontal flip) during training. We used the spherical feature embedding [4, 3], i.e., we enforced $||x||_2 = 1$ and $||\mu_k||_2 = 1, \forall k$. To address the gradient saturation in the spherical embedding [4], we followed the method introduced by [4] to scale every inner product $x^{\mathrm{T}}\mu$ up to 30. We updated $\{p(k)\}_{k=1}^{K}$ every $T = 40$ iterations, as we found it not sensitive in a broad range. We maintained a buffer for m and σ as a reference, whereas m_j and σ_j were estimated within each batch to obtain the gradient. We updated the buffer with a momentum $\alpha = B/N$ for each batch where B denoted the batch size and N denotes the training set size.

2. Probablistic interpretation of the WDBR

In this section we first show that the weakly supervised decision boundary rectification (WDBR) is the maximum a posteriori (MAP) optimal estimation of the surrogate label $\hat{y}$ under suitable assumptions. Specifically, if we assume that a surrogate class is modeled by a normal distribution [5, 2, 1] parameterized by a mean vector μ and covariance matrix $\mathbf{\Sigma}$, the likelihood that x is generated by the y-th surrogate class is:

$$p(x|y) = \frac{\exp(-||x-\mu_y||^2/2)}{\Sigma_{k=1}^{K}\exp(-||x-\mu_k||^2/2)}, \tag{S1}$$

where we assume the identity covariance matrix [5, 2, 1], i.e. $\mathbf{\Sigma}_k = I, \forall k$. Since we enforce $||x||_2 = 1$ and $||\mu_k||_2 = 1, \forall k$, we have $-||x-\mu_y||^2/2 = x^{\mathrm{T}}\mu_y + 1$. Then Eq. (S1) is equivalent to:

$$p(x|y) = \frac{\exp(x^{\mathrm{T}}\mu_y)}{\Sigma_{k=1}^{K}\exp(x^{\mathrm{T}}\mu_k)}. \tag{S2}$$

From Eq. (S2) we can see that the basic surrogate classification in Eq. (1) in the main manuscript is the Maximum Likelihood Estimation (MLE) of model parameters θ and $\{\mu_k\}$. And the assignment

$$\hat{y} = \arg\max_k \exp(x^{\mathrm{T}}\mu_k) = \arg\max_k p(x|k) \tag{S3}$$

is the MLE optimal assignment. If we further consider the prior information of each surrogate class, i.e., which surrogate classes are more preferable to assign, we can improve the assignment to the Maximum a Posteriori (MAP) optimal assignment:

$$\hat{y} = \arg\max_k p(k)\exp(x^{\mathrm{T}}\mu_k) = \arg\max_k p(k|x), \tag{S4}$$

where

$$p(k|x) = \frac{p(y)\exp(x^{\mathrm{T}}\mu_y)}{\Sigma_{k=1}^{K}p(k)\exp(x^{\mathrm{T}}\mu_k)} \tag{S5}$$

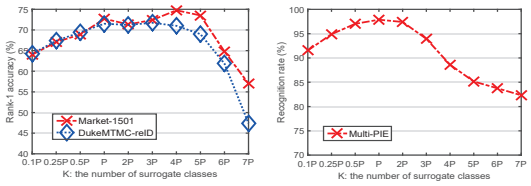
is the posterior probability. Eq. (S4) is identical to the rectified assignment in Eq. (6) in the main manuscript. Hence, we can interpret the weakly supervised rectifier function as a prior probability that specifies our preference on the surrogate class. In particular, when we use the hard rectification, we actually specify that we dislike severely unbalanced surrogate classes. When we use the soft rectification, we specify that we favor the more balanced surrogate classes.

Derivation of the decision boundary. Here we consider the simplest two-surrogate class case. It is straightforward to extend it to multi-surrogate class cases. From Eq. (S4) we can see that the decision boundary between two surrogate class μ_1 and μ_2 is:

$$\begin{aligned} & p(1)\exp(x^{\mathrm{T}}\mu_1) = p(2)\exp(x^{\mathrm{T}}\mu_2) \qquad (S6)\\ \Rightarrow\ & \exp(x^{\mathrm{T}}\mu_1 + \log p(1)) = \exp(x^{\mathrm{T}}\mu_2 + \log p(2))\\ \Rightarrow\ & (\mu_1-\mu_2)^{\mathrm{T}}x + \log p(1) - \log p(2) = 0\\ \Rightarrow\ & (\mu_1-\mu_2)^{\mathrm{T}}x + \log\frac{p(1)}{p(2)} = 0. \end{aligned}$$

Algorithm 1: Weakly supervised discriminative learning

Input: Training set $\mathcal{U} = \{u_i\}$, state information $\mathcal{S} = \{s_i\}$, pretrained model $f(\cdot, \theta^{(0)})$
Output: Learned model $f(\cdot, \theta)$
Initialization:
Obtain the initial feature space $\mathcal{X}_{init} = f(\mathcal{U}, \theta^{(0)})$.
Initialize surrogate classifiers $\{\mu_k\}_{k=1}^K$ as the centroids obtained by performing standard K-means clustering on $\mathcal{X}_{init}$.
Initialize the surrogate rectifiers $p(k) = 1, k = 1, \cdots, K$.
Initialize total distribution vectors m/σ on $\mathcal{X}_{init}$.
Training:
for *the i-th batch* $\{\mathcal{U}^{(i)}, \mathcal{S}^{(i)}\}$ **do**
    Obtain the features in the batch $\mathcal{X}^{(i)} = f(\mathcal{U}^{(i)}, \theta)$.
    Assign every $x \in \mathcal{X}^{(i)}$ to a surrogate class $\hat{y}$ by Eq. (6) in the main manuscript.
    Estimate the state sub-distributions $\{m_j, \sigma_j\}_{j=1}^J$ in this batch.
    Compute the loss (Eq. (11) in the main manuscript) and update the model.
    Estimate the total distribution in this batch: $m^{(i)}/\sigma^{(i)}$.
    Update m by $m \leftarrow (1-\alpha)m + \alpha m^{(i)}$ and σ by $\sigma \leftarrow (1-\alpha)\sigma + \alpha\sigma^{(i)}$.
    Obtain $\{R_k\}_{k=1}^K$ by Eq. (3) and update $\{p(k)\}_{k=1}^K$ by Eq. (7) every T batches.
end
Testing:
Discard $\{\mu_k\}_{k=1}^K$ and use $f(\cdot|\theta)$ to extract the discriminative feature.



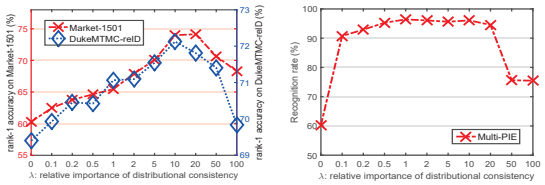
(a) RE-ID (b) PIFR

Figure S1. Evaluation for the number of surrogate classes K. P is the precise number of classes in the training set.

3. Hyperparameter evaluation

In order to provide insight and guidance on choosing the hyperparameter values, in this section we show evaluation results of the hyperparameters to reveal some behaviors and characteristics of our model. For person re-identification (RE-ID) we evaluated on the widely-used Market-1501 and DukeMTMC-reID datasets, and for pose-invariant face recognition (PIFR) we evaluated on the large-scale Multi-PIE dataset. For easier interpretation and more in-depth analysis, we used the hard rectification function on *all* datasets. This was because the hard rectification function could be interpreted as nullification of high maximum predominance index (and thus likely to be dominated by the extrinsic state) surrogate classes.

K: Number of surrogate classes. In each task (i.e. RE-ID or PIFR), we varied the number of surrogate classes K by setting it to a multiple of the precise number of classes in the training set P (e.g. $P = 750$ for Market-1501 and $P = 200$ for Multi-PIE). We show the results in Figure S1. From Figure S1(a) and S1(b) we can see that the optimal performances could be achieved when $K = P$ or $K > P$. This might be because the dynamic nullification (i.e. hard rectification) reduced the effective K in training, so that a larger K could also be optimal. In a practical perspective, we might estimate an "upper bound" of P and set K to it according to some prior knowledge.



(a) RE-ID (b) PIFR

Figure S2. Evaluation for the feature drift regularization weight λ.

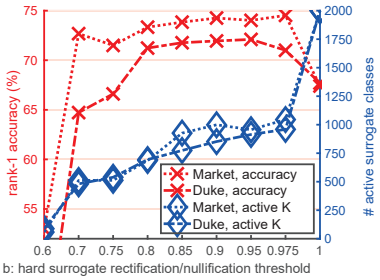


Figure S3. Evaluation for the rectification threshold b (hard threshold here). We also show the active number of surrogate classes in the final convergence epoch (denoted as "active K" in the legend).

λ: Weight of feature drift regularization. We show the evaluation results in Figure S2. Here we removed the surrogate decision boundary rectification for PIFR to better understand the characteristic of λ. From Figure S2(a) and S2(b), we found that while the performances on RE-ID were optimal around $\lambda = 10$, for PIFR it was near optimal within the range of $[0.5, 20]$.

Hard surrogate rectification/nullification threshold. We show the evaluation results on RE-ID datasets in Figure S3. The performances were optimal when b was not too low, e.g. $b \in [0.8, 0.9]$ was optimal for both RE-ID datasets. A major reason was that it was difficult to form sufficient surrogate classes when the threshold was too low. To see this, in Figure S3 we also show the number of active (i.e. not nullified) surrogate classes in the final convergence epoch. Clearly, a lack of surrogate classes was harmful to the discriminative learning.

Soft surrogate rectification. We show the evaluation results on Multi-PIE in Figure S4. As analyzed in the main manuscript, soft rectification consistently improved over the hard rectification due to the balanced classes on Multi-PIE.

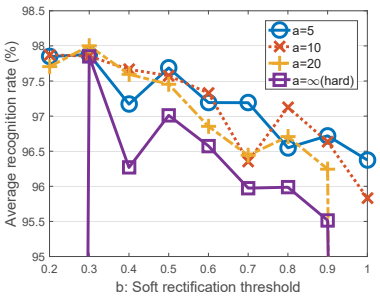


Figure S4. Soft surrogate rectification evaluation on Multi-PIE. a denotes the rectification strength.

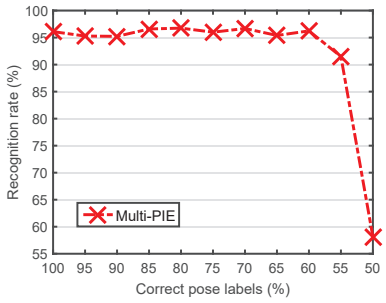


Figure S5. Evaluation for noisy pose labels on Multi-PIE.

The optimal value of the soft rectification threshold was around 0.3 because on Multi-PIE the five poses were evenly distributed and thus the optimal MPI shall be around slightly above $1/5$. In a practical perspective, when we have prior knowledge of the unlabelled data, we might be able to estimate the soft rectification threshold. Nevertheless, even when we do not have reliable prior knowledge, the robust conservative hard rectification could also be effective.

4. Simulating using estimated pose labels

In this supplementary material we present the evaluation on our method's robustness for pose label perturbation. This is a simulation for the more challenging real-world PIFR setting, where the pose labels are obtained by pose estimation models, and thus there might be incorrect pose labels. We note that this is not the case for person re-identification (RE-ID), because in RE-ID every image comes from a certain camera view of the surveillance camera network, so that no estimation is involved.

To simulate the pose label noise, we add perturbation to the groundtruth pose labels. We randomly reset some pose labels to incorrect values (e.g. we reset a randomly chosen pose label to 15°which is actually 60°). The randomly reset pose labels were equally distributed in every degree. For example, when we reset 20% pose labels, there were 20% of 60°pose labels were reset incorrect, 20% of 45°pose labels were incorrect, and so forth for other degrees. We vary the incorrect percents and show the results in Figure S5.

From Figure S5 we observed that the performance on PIFR did not drop significantly until less than 60% pose labels were correct. This observation indicated that our model could tolerate a moderate extent of pose label noise. A major reason was that when a few pose labels were incorrect, a highly affected surrogate class (whose members were mostly of the same pose) would still have a high Maximum Predominance Index for it to be nullfied. In addition, when most pose labels were correct the estimation of the manifestation sub-distributions should approximate the correct manifestation sub-distributions. Therefore, our model should be robust for the unsupervised PIFR task when a few pose labels were perturbated.

References

[1] David Berthelot, Thomas Schumm, and Luke Metz. Began: boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*, 2017. 1

[2] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan. Wasserstein cnn: Learning invariant features for nir-vis face recognition. *TPAMI*, 2018. 1

[3] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *CVPR*, 2017. 1

[4] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: l 2 hypersphere embedding for face verification. In *ACMMM*, 2017. 1

[5] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jianhuang Lai. Unsupervised person re-identification by soft multilabel learning. In *CVPR*, 2019. 1