Supplementary Material for Efficient Online Local Metric Adaptation via Negative Samples for Person Re-Identification

Jiahuan Zhou, Pei Yu, Wei Tang and Ying Wu
Electrical Engineering and Computer Science, Northwestern University, US
{jzt011, pyi980, wtt450, yingwu}@eecs.northwestern.edu

1. More Analyses of Proposed OL-MANS

1.1. Learned Metric Rank Analysis of OL-MANS

Feature descriptors used in PRID are generally high dimensional in order to handle the complex appearance variations. In practice, most existing methods apply PCA blindly to reduce the feature dimension without clear justification and effectiveness. In contrast, our OL-MANS can be performed in the original high dimensional space while allowing the selection of a low rank local metric. The effectiveness of the low rank metric is also verified in [8] and [6].

As described in the paper, in our proposed OL-MANS, we solve the original time-consuming positive semidefinite (PSD) problem by solving an efficient kernel SVM instead, as in Eqn. 1.

$$\begin{aligned} &\mathbf{M}_L^i = \arg\min_{\mathbf{M}_L^i} \frac{1}{2}\left\|\mathbf{M}_L^i\right\|^2 \\ &sub\ \ to: \ \ \zeta_j\left(\tilde{y}_j^T\mathbf{M}_L^i\tilde{y}_j - 1\right) \geq 1, \forall\ \ 0 \leq j \leq k \end{aligned} \tag{1}$$

The obtained local metric $\mathbf{M}_L^i$ is formed as Eqn. 2:

$$\mathbf{M}_L^i = \sum_{j=0}^{k} \alpha_j\zeta_j\tilde{y}_j\tilde{y}_j^T = \sum_{j=1}^{k} \alpha_j\tilde{y}_j\tilde{y}_j^T = \sum_{j=1}^{k'} \alpha_j^r(\tilde{y}_j^r)(\tilde{y}_j^r)^T \succeq 0 \tag{2}$$

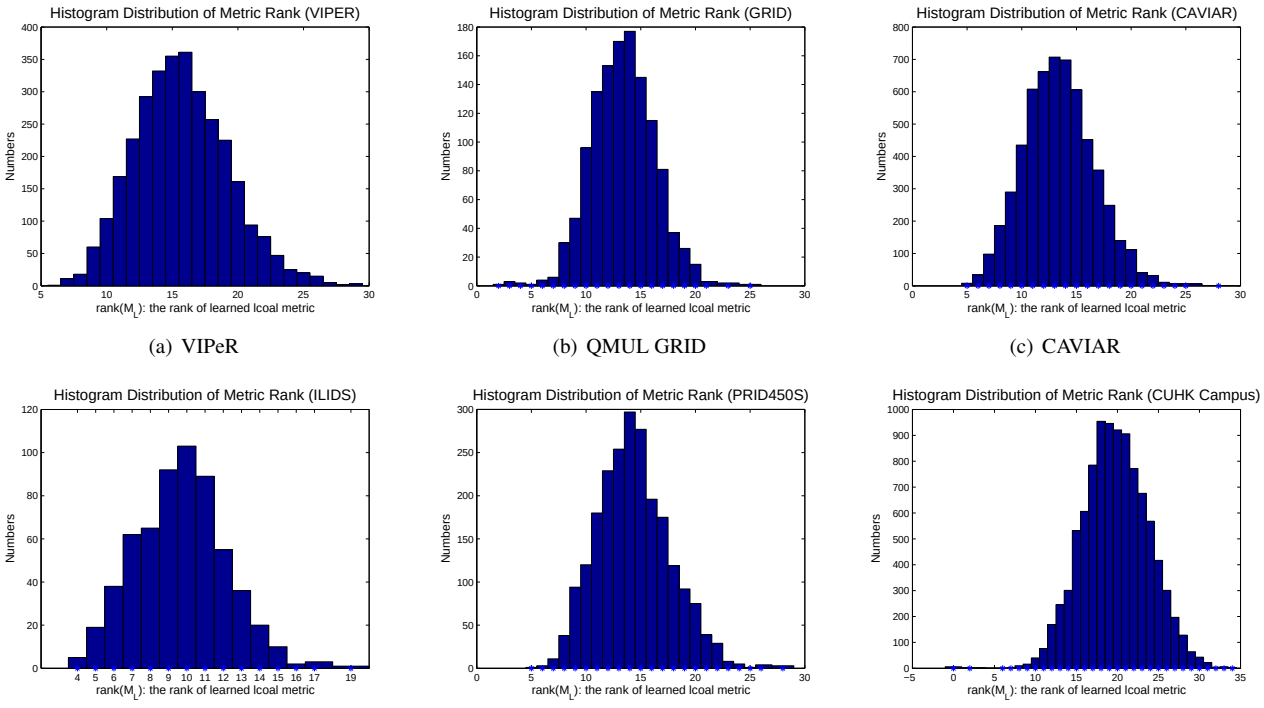
where $\alpha_j^r \neq 0$. It is obvious that $\mathbf{M}_L^i$ is the linear combination of all the support vectors of $\{\tilde{y}_j^r\}_{j=1}^{k'}$. Therefore, the rank of $\mathbf{M}_L^i$ is bounded by the number of support vectors, k'. In practice, the local metric is constrained by the strong negative samples (the hard negatives). In other words, the coefficient vector $A = [\alpha_1, \alpha_2, ..., \alpha_k]$ should be sparse.

To validate this, we have conducted an experiment that we compute the ranks of all the learned local metrics for all the probes in different benchmarks (VIPeR, GRID, CAVIAR, iLIDS, PRID 450S and CUHK Campus). The result is presented in Fig. 1, where it is evident that almost all the learned local metrics are pretty low rank, even though the size of the negative database (NDB) is large. This negative database has over 10,000 negative samples, and more than 500 strong negative samples are generally selected for each datum to learn its local metric. This clearly shows an advantage of our proposed method, as it allows us to work in a high dimensional space while most existing methods do not.

2. Proof of Theorem. 2

Theorem 2 *For an input x, its nearest neighbor (NN) is x' in the augmented dataset $\mathcal{D}^a$. This augmented dataset is obtained by adding hard negative samples to the original dataset $\mathcal{D}$. Define the probability that the nearest neighbor of x' is an augmented data x_a, i.e., $x' \sim x_a$ by $P(x' \sim x_a) = q$; otherwise, the nearest neighbor of x' is not an augmented data x_a, i.e., $x' \neg x_a$, $P(x' \neg x_a) = 1 - q$, where $0 \leq q \leq 1$. The asymptotic error $\mathcal{P}^a(e|x)$ by using $\mathcal{D}^a$ is:*

$$\mathcal{P}^a(e|x) = \frac{(2-q)\mathcal{P}(e|x)}{2 - 2q\mathcal{P}(e|x)} \leq \mathcal{P}(e|x) \tag{3}$$



(a) VIPeR (b) QMUL GRID (c) CAVIAR

(d) iLIDS (e) PRID 450S (f) CUHK Campus

Figure 1. Histogram Distributions of Metric Rank for All the Learned Local Metrics

Proof 2 *Denote by $\mathcal{D}$ the original data set (i.e., the gallery set), and by $\mathcal{D}^a$ the augmented data set by adding hard negatives. Except for the augmented hard negative data (denoted by x_a), the rest in $\mathcal{D}^a$ are the same as $\mathcal{D}$.*

Let's consider the two-class 1-NN classification, without losing the generality. The asymptotic error for 2-class 1-NN using $\mathcal{D}$ is

$$\mathcal{P}(e|x) = 2P(\omega_+|x)P(\omega_-|x)$$

Let's consider the asymptotic error for 2-class 1-NN using $\mathcal{D}^a$. We denote it by $P^a(e|x)$. Our goal is to prove:

$$\mathcal{P}^a(e|x) \leq \mathcal{P}(e|x) \tag{4}$$

The prove is the following. For an input x, its nearest neighbor x' in $\mathcal{D}^a$ (denoted by $x \sim x'$) has two cases:

- *case 1: the nearest neighbor of x' is an augmented data x_a, i.e., $x' \sim x_a$. Its probability $P(x' \sim x_a) = q$;*
- *case 2: the nearest neighbor of x' is not an augmented data x_a, i.e., $x' \neg x_a$. Its probability $P(x' \neg x_a) = 1 - q$.*

We denote the nearest neighbor of x' in $\mathcal{D}$ by $x''(x')$. There are two cases. If x' is x_a, then its nearest neighbor in $\mathcal{D}$ is $x''(x_a)$ whose class label is ω_+ (because x_a are all hard negative samples). If x' is not x_a, then its nearest neighbor in $\mathcal{D}$ is $x''(x') = x'$.

Now we consider the asymptotic probability of assigning ω_+ to x'. In case 1, we need to guarantee both x and $x''(x')$ to be ω_+ (i.e., the hard negative data is actually useful). In case 2, we only need to guarantee x to be ω_+ (i.e., the hard negative data is not effective). So we have:

$$\phi(\omega_+|x') \propto P^2(\omega_+|x)q + P(\omega_+|x)(1-q)$$

Similarly, the asymptotic probability of assigning ω_- to x' is:

$$\phi(\omega_-|x') \propto P^2(\omega_-|x)q + P(\omega_-|x)(1-q)$$

Because $\phi(\omega_+|x') + \phi(\omega_-|x') = 1$*, we have:*

$$\begin{aligned}\phi(\omega_+|x') &= \frac{P^2(\omega_+|x)q + P(\omega_+|x)(1-q)}{(1-q) + q[1 - 2P(\omega_+|x)P(\omega_-|x)]} \\ \phi(\omega_-|x') &= \frac{P^2(\omega_-|x)q + P(\omega_-|x)(1-q)}{(1-q) + q[1 - 2P(\omega_+|x)P(\omega_-|x)]}\end{aligned} \tag{5}$$

Therefore, we can compute the 2-class 1-NN asymptotic error for x *on* $\mathcal{D}^a$*:*

$$\begin{aligned}\mathcal{P}^a(e|x) &= \phi(\omega_+|x')P(\omega_-|x) + \phi(\omega_-|x')P(\omega_+|x) \\ &= \frac{[2(1-q)+q]P(\omega_+|x)P(\omega_-|x)}{(1-q) + q[1 - 2P(\omega_+|x)P(\omega_-|x)]} \\ &= \frac{(2-q)\mathcal{P}(e|x)}{2(1-q) + 2q(1-\mathcal{P}(e|x))} \\ &= \frac{(2-q)\mathcal{P}(e|x)}{2 - 2q\mathcal{P}(e|x)}\end{aligned} \tag{6}$$

Because $0 \leq q \leq 1$*, it is easy to see:*

$$\begin{aligned}if\ \ \mathcal{P}^a(e|x) \leq \mathcal{P}(e|x) &\Leftrightarrow \frac{(2-q)P(e|x)}{2 - 2qP(e|x)} \leq P(e|x) \\ &\Leftrightarrow (2-q) \leq 2 - 2qP(e|x) \\ &\Leftrightarrow 2qP(e|x) \leq q \\ &\Leftrightarrow P(e|x) \leq \frac{1}{2}\end{aligned} \tag{7}$$

Since the error rate $P(e|x) \leq \frac{1}{2}$ *is always true, the proof in Eqn. 7 is true, and Theorem. 2 holds.*

3. Proof of Theorem. 3 and Theorem. 4

Theorem 3 *Let* $\phi^\lambda(\boldsymbol{M}_L, s_i)$ *be a distance-based loss function that is* λ*-Lipschitz in the first argument. Then with probability at least* $1-\delta$ *over* $\{s_1, ..., s_k\}$ *from an unknown B-bounded-support (each* $(x, l) \sim \mathbb{D}, \|x\| \leq B$*) distribution* $\mathbb{D}$*, we have:*

$$\underset{\boldsymbol{M}_L \in \mathcal{M}}{Sup} \left[Err^\lambda(\boldsymbol{M}_L, \mathbb{D}) - Err^\lambda(\boldsymbol{M}_L, S_k^{pair}) \right] \leq O\left(\lambda B^2 \sqrt{D \ln(1/\delta)/k}\right) \tag{8}$$

Proof 3 *Let* $\mathcal{P}$ *be the probability measure induced by the random variable* $(X; \mathcal{L})$*, where* $X := (x, x')$*,* $\mathcal{L} := 1[l = l']$*. Define function class:*

$$\mathcal{F} := \{X \mapsto \|x - x'\|_{\boldsymbol{M}_L}\}$$

and consider our loss function $\phi^\lambda(\boldsymbol{M}_L, s_i) = \lambda[\zeta_i \left((x_i - x_0)^T \boldsymbol{M}_L (x_i - x_0)\right) - \gamma_{\zeta_i}]_+$ *which is* λ*-Lipschitz in the first argument. Then, we are interested in bounding the quantity*

$$\underset{(X;\mathcal{L}) \in \mathcal{P}}{Sup} \left[\phi^\lambda(f_{\boldsymbol{M}_L}(X), \mathcal{L}) - \frac{1}{k} \sum_{i=1}^{k} \phi^\lambda(f_{\boldsymbol{M}_L}(X_i), \mathcal{L}_i) \right]$$

Define $\hat{x}_i := x_0 - x_i$ *for each pair* s_i*, then the Rademacher complexity*[1] *of our function class* $\mathcal{F}$ *(with respect to the distribution* $\mathcal{P}$*) is bounded, since (let* $\sigma_1, \sigma_2, ..., \sigma_k$ *denote independent uniform* ± 1*-valued random variables):*

[1] See the formal definition of the Rademacher complexity in [2]

$$
\begin{aligned}
&\mathcal{R}(\mathcal{F},\mathcal{P}) := \mathbb{E}_{X_i,\sigma_i}\left[\underset{f_{\boldsymbol{M}_L}\in\mathcal{F}}{Sup}\frac{1}{k}\sum_{i=1}^{k}\sigma_i f_{\boldsymbol{M}_L}(X_i)\right]\\
&=\frac{1}{k}\mathbb{E}_{X_i,\sigma_i}\underset{\boldsymbol{M}_L\in\mathcal{F}}{Sup}\left[\sum_{i=1}^{k}\sigma_i\hat{x}_i^T\boldsymbol{M}_L\hat{x}_i\right]\\
&=\frac{1}{k}\mathbb{E}_{X_i,\sigma_i}\underset{\boldsymbol{M}_L\in\mathcal{F},[a^{jk}]_{jk}=\boldsymbol{M}_L}{Sup}\left[\sum_{j,k}a^{jk}\sum_{i=1}^{k}\sigma_i\hat{x}_i^j\hat{x}_i^k\right]\\
&\le\frac{1}{k}\mathbb{E}_{X_i,\sigma_i}\underset{\boldsymbol{M}_L\in\mathcal{F}}{Sup}\left[\|\boldsymbol{M}_L\|_F\left(\sum_{j,k}\left(\sum_{i=1}^{k}\sigma_i\hat{x}_i^j\hat{x}_i^k\right)^2\right)^{1/2}\right]\\
&\le\frac{\sqrt{D}}{k}\mathbb{E}_{X_i,i\in[k]}\left(\mathbb{E}_{\sigma_i,i\in[k]}\sum_{j,k}\left(\sum_{i=1}^{k}\sigma_i\hat{x}_i^j\hat{x}_i^k\right)^2\right)^{1/2}\\
&=\frac{\sqrt{D}}{k}\mathbb{E}_{X_i,i\in[k]}\left(\sum_{j,k}\sum_{i=1}^{k}(\hat{x}_i^j)^2(\hat{x}_i^k)^2\right)^{1/2}\\
&=\frac{\sqrt{D}}{k}\mathbb{E}_{X_i,i\in[k]}\left(\sum_{i=1}^{k}\|\hat{x}_i\|^4\right)^{1/2}\\
&=\frac{\sqrt{D}}{k}\mathbb{E}_{(x_0,x_i)\in(\mathbb{D}\times\mathbb{D}),i\in[k]}\left(\sum_{i=1}^{k}\|x_i-x_0\|^4\right)^{1/2}\\
&\le\sqrt{\frac{D}{k}}\mathbb{E}_{(x_0,x_i)\in(\mathbb{D}\times\mathbb{D}),i\in[k]}\left(\|x_i-x_0\|^4\right)^{1/2}\\
&\le 4B^2\sqrt{\frac{D}{k}}
\end{aligned}
\tag{9}
$$

Recall that $\mathbb{D}$ has bounded support (with bound B). Thus, by noting that ϕ^λ is $8B^2$ bounded function that is λ-Lipschitz in the first argument, we can readily apply the Theorem.8 in [2] to obtain the desired uniform deviation bound.

Theorem 4 *Let $\boldsymbol{M}_L$ be any class of weighting metrics on the feature space $X=\mathbb{R}^D$, and define $d := Sup_{\boldsymbol{M}_L\in\mathcal{M}}\|\boldsymbol{M}_L\|_F^2$. Following the same parameter setting in Theorem. 3, we have:*

$$
\underset{\boldsymbol{M}_L\in\mathcal{M}}{Sup}\left[Err^\lambda(\boldsymbol{M}_L,\mathbb{D})-Err^\lambda(\boldsymbol{M}_L,S_k^{pair})\right]\le O\left(\lambda B^2\sqrt{d\ln(1/\delta)/k}\right) \tag{10}
$$

Proof 4 *Let $\mathcal{P}$ be the probability measure induced by the random variable $(X;\mathcal{L})$, where $X := (x,x')$, $\mathcal{L} := 1[l=l']$. Define function class:*

$$
\mathcal{F} := \{X\mapsto\|x-x'\|_{\boldsymbol{M}_L}\}
$$

Following the same steps in the proof of Theorem. 3, we can conclude that the Rademacher complexity of $\mathcal{F}$ is bounded. In particular,

$$
\mathcal{R}_k(\mathcal{F})\le 4B^2\sqrt{\frac{Sup_{\boldsymbol{M}_L\in\mathcal{M}}\|\boldsymbol{M}_L\|_F^2}{k}}
$$

Finally, we note that ϕ^λ is λ-Lipschitz in the first argument, so that we can readily apply Theorem.8 in [2].

4. More Experiment Results

4.1. Influence of the Weighting Parameter λ

The parameter λ in $d(x^p_{v_i}, x^g_{v_j}) = (x^p_{v_i} - x^g_{v_j})^T \mathbf{W}^T (\mathbf{I} + \lambda \mathbf{M}^i_L) \mathbf{W} (x^p_{v_i} - x^g_{v_j})$ is used to balance the underlying global metric and the learned local metric adaptation. Different λ have different influences to the identification performances. We have conducted an experiment on the VIPeR dataset to determine the value of λ, the results of which are shown in Fig. 2.

To read the results in Fig. 2, let's explain λ. When $\lambda = 0$, it is the baseline from [8] without our local metric adaptation; when $\lambda = max$, it represents that λ is set to be:

$$\lambda = \max_{1 \le j \le m} \left(d_{\mathbf{M}_G}(x^p_{v_i}, y^g_{v_j}) \right) / \max_{1 \le j \le m} \left(d_{\mathbf{M}_L}(x^p_{v_i}, y^g_{v_j}) \right) \tag{11}$$

As we can see from this experiment, setting λ as Eqn. 11 achieves the best result because it normalizes the norm scales of the global and local metric distances.

4.2. Influence of Negative Sample Database

For our OL-MANS, a negative sample database (NDB) is used to provide the negative training data. Because there are various strategies to collect NDB, we conduct the following experiments to investigate the influences of different NDB choices. The first two experiments are conducted on the VIPeR dataset [5] and the challenging CAVIAR [4] is used in the third experiment. Moreover, the global metric learning method proposed in [8] is adopted as the baseline method for the global metric learning $\mathbf{M}_G$.

- **Using the training data $\mathbf{X}_{train}$ from the same benchmark as negative sample**: Here the training samples $\mathbf{X}_{train}$ in VIPeR which have different identities from P_i(the training data for global metric learning) are used as negative samples. It guarantees that the obtained NDB is clearly meaningful. The accuracy in PRID is given in Table.1 as **Our-SAME**.

- **Using different benchmark datasets as the NDB**: In this experiment, we utilize other benchmarks as the NDB. The QMUL GRID [10] and CAVIAR [4] are combined into one dataset then used as the NDB in this experiment, so that we can guarantee that the identities of all the negative samples in the NDB are different from P_i. For each identity P_i, the k nearest negative samples are found in the NDB (under $\mathbf{M}_G$) and used for our OL-MANS. Different values of k (50, 100, 500) are chosen for further comparisons. The experiment results **Our-D-50/100/500** are shown in Table.1. Moreover, an additional experiment **Our-D-RAM** that uses 50 random negative samples from the NDB for OL-MANS is compared. This experiment validates the insight of our method that the effective negative samples are those that are close to the probe in the feature space (e.g., strong false positives).

- **The NDB includes the false negative samples**: We investigate how the "contamination" in the NDB impacts our proposed method. In this situation, some negative samples in the NDB are deliberately collected from the same identity to P_i. We call them *false negative samples*, in addition to the use of the probe image set of CAVIAR [4] as the rest of the NDB. Since there are multiple images of the same identity in CAVIAR, they can be considered as the false negative samples. The experiment results **Our-NoFN, Our-FN** are shown in Table. 1. **Our-NoFN** refers to a "clean" NDB with no false negative samples in it, and **Our-FN** refers to a "contaminated" NDB that includes false negative samples for all the probe images.

From Table. 1, it can be observed that **Our-SAME** performs the best because the negative data from the same benchmark dataset are most discriminative. Results on **Our-D-50/100/500** also largely outperform the baseline by consistent improvements. Moreover, the false negative sample may influence and degrade the performance of **Our-NoFN**, but not significantly. Nevertheless, a clean NDB with hard negatives is useful and effective.

4.3. Full CMC Curve Plot and More Comparison Results on Benchmarks

First, in Table. 2, we show more comparison results with state-of-the-arts PRID methods on VIPeR dataset. One interesting observation is that our performances at rank-20 are a slightly lower than the latest TSRPR [13] method. This is expected as our local metric becomes less effective when the true positive gallery image is far from the probe in the feature space. Other than that, our new method still outperforms all the other methods at rank-20.

Besides, the CMC curves of the state-of-the-art methods on various benchmarks are plotted and compared in Fig. 3, from which it is evident that our new method significantly outperforms all other methods on all different benchmarks.

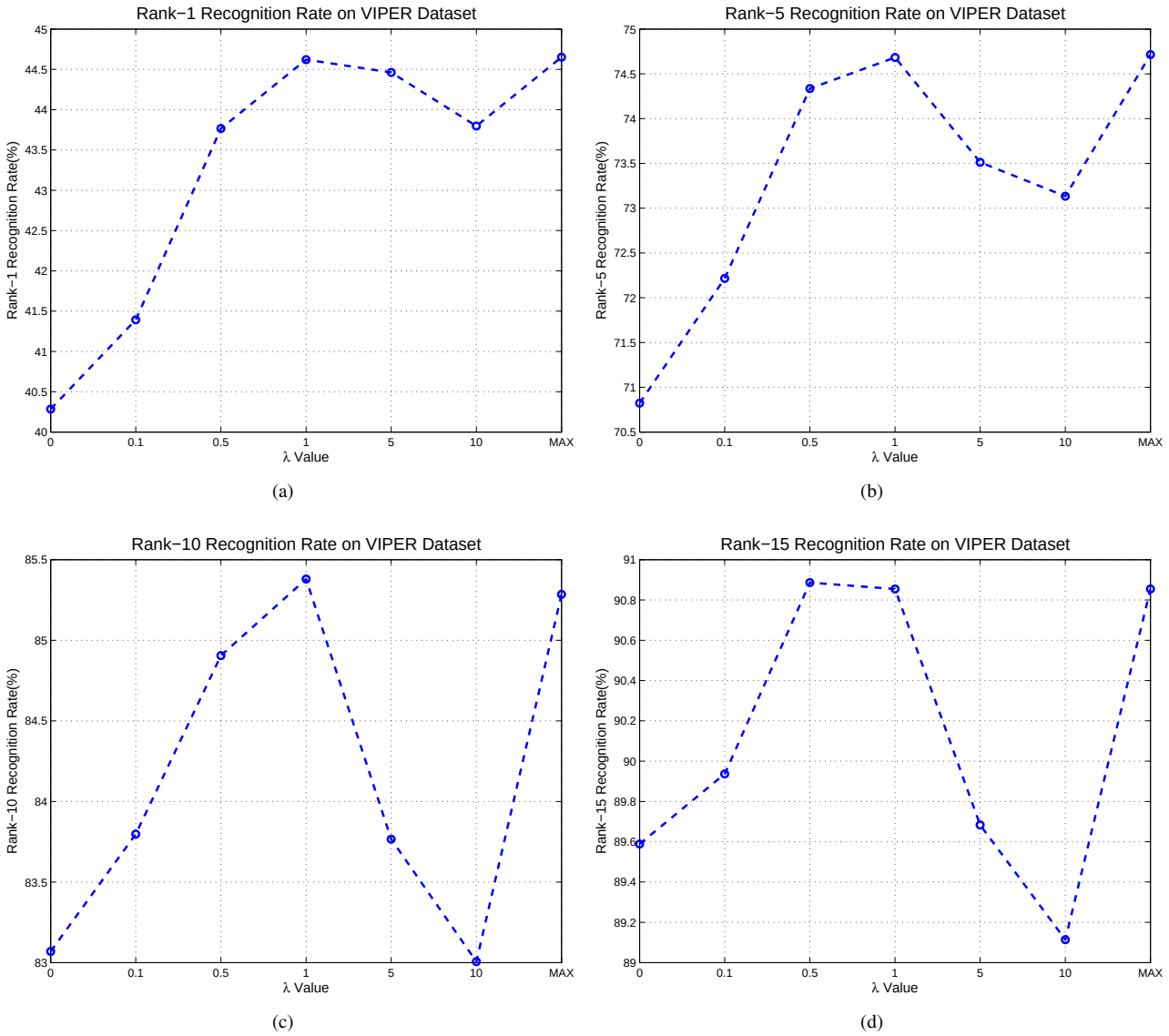


Figure 2. We conducted the experiment to determine the influence of λ. The x-axis means the value of λ and the y-axis is the identification rate. From (a) to (d), there are the identification rates at Rank-1, Rank-5, Rank-10 and Rank-15 on VIPeR dataset.

5. Source Codes

The codes of our proposed OL-MANS algorithm and some demos can be found at the project website: http://www.ece.northwestern.edu/~jzt011/project/ICCV17_OLMANS/ICCV17_OLMANS.html

Benchmark Dataset	Method	R=1	R=5	R=10	R=20
VIPeR	Baseline [8]	40.73	69.94	82.34	92.37
	Our-SAME	44.97	74.43	84.97	93.64
	Our-D-050	42.63	73.63	84.81	93.54
	Our-D-100	43.04	73.86	84.30	93.42
	Our-D-500	42.53	73.89	84.15	93.35
	Our-D-RAM	39.87	70.51	82.28	91.77
CAVIAR	Baseline [8]	40.63	71.72	83.34	95.67
	Our-NoFN	51.68	76.36	86.38	96.55
	Our-FN	50.34	74.83	85.72	96.03

Table 1. Comparison of different NDBs on the VIPeR dataset (P = 316) and CAVIAR dataset (P = 36).

Method	R=1	R=5	R=10	R=20
Ours	44.97	74.43	84.97	93.64
SCNCD[15]	37.80	68.50	81.20	90.40
EPKFM[3]	36.80	70.40	83.70	91.70
K-Ensb2[14]	36.10	68.70	80.10	85.60
IDLA[1]	34.81	-	-	-
TSRPR[13]	31.10	68.60	82.80	94.90
kBiCov[11]	31.11	58.33	70.71	82.44
LADF[7]	30.22	64.70	78.92	90.44
SalMatch[16]	30.16	-	65.54	79.15
Mid-L-F[18]	29.11	-	65.95	79.87
MMCML[12]	28.83	59.34	75.82	88.51
eSDC[17]	26.74	50.70	62.37	76.36
SSCDL[9]	25.60	53.70	68.10	83.60
PRDC[19]	15.66	38.40	53.86	70.09

Table 2. More comparison results with state-of-the-arts on VIPeR (P = 316).

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. *Differences*, 2015. 7

[2] P. L. Bartlett and S. Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *JMLR*, 2002. 3, 4

[3] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang. Similarity learning on an explicit polynomial kernel feature map for person re-identification. In *CVPR*, 2015. 7

[4] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011. 5

[5] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *PETS*, 2007. 5

[6] N. Jiang, W. Liu, and Y. Wu. Order determination and sparsity-regularized metric learning adaptive visual tracking. In *CVPR*, 2012. 1

[7] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013. 7

[8] S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *ICCV*, 2015. 1, 5, 7

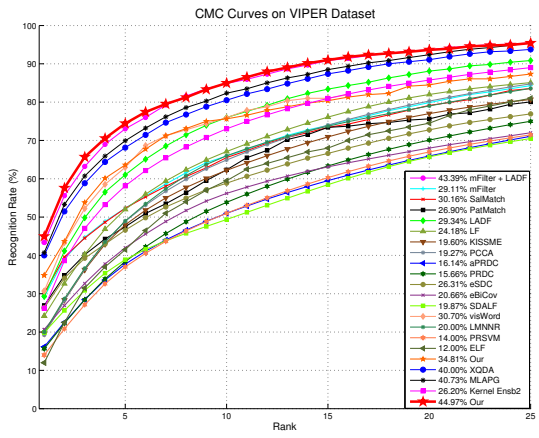
[9] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-supervised coupled dictionary learning for person re-identification. In *CVPR*, 2014. 7

[10] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *CVPR*, 2009. 5

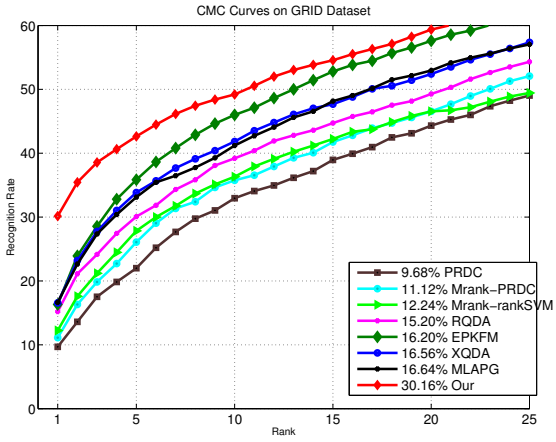
[11] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 2014. 7

[12] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE TIP*, 2014. 7

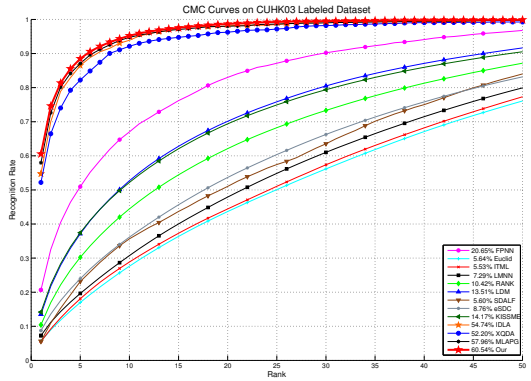
[13] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *CVPR*, 2015. 5, 7



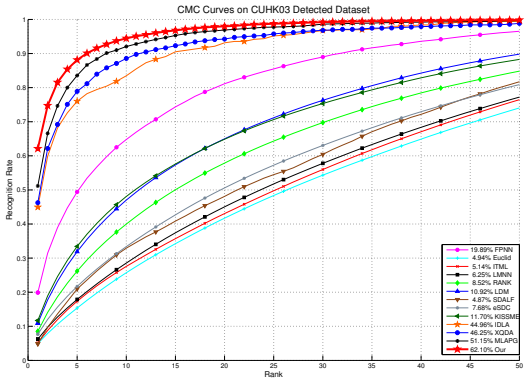
(a) VIPeR Dataset



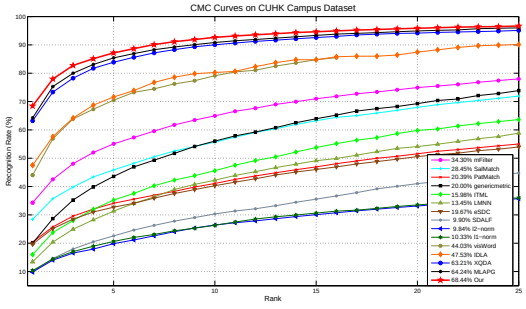
(b) QMUL GRID Dataset



(c) CUHK03 Labeled Dataset



(d) CUHK03 Detected Dataset



(e) CUHK Campus Dataset

Figure 3. Comparison of CMC curves and rank-1 identification rates on benchmark **(a)** VIPeR, **(b)** QMUL GRID, **(c)** CUHK Campus, **(d)** CUHK03-Labeled and **(e)** CUHK03-Detected datasets.

[14] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*. 2014. 7

[15] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *ECCV*. 2014. 7

[16] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013. 7

[17] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013. 7

[18] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014. 7
[19] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE TPAMI*, 2013. 7