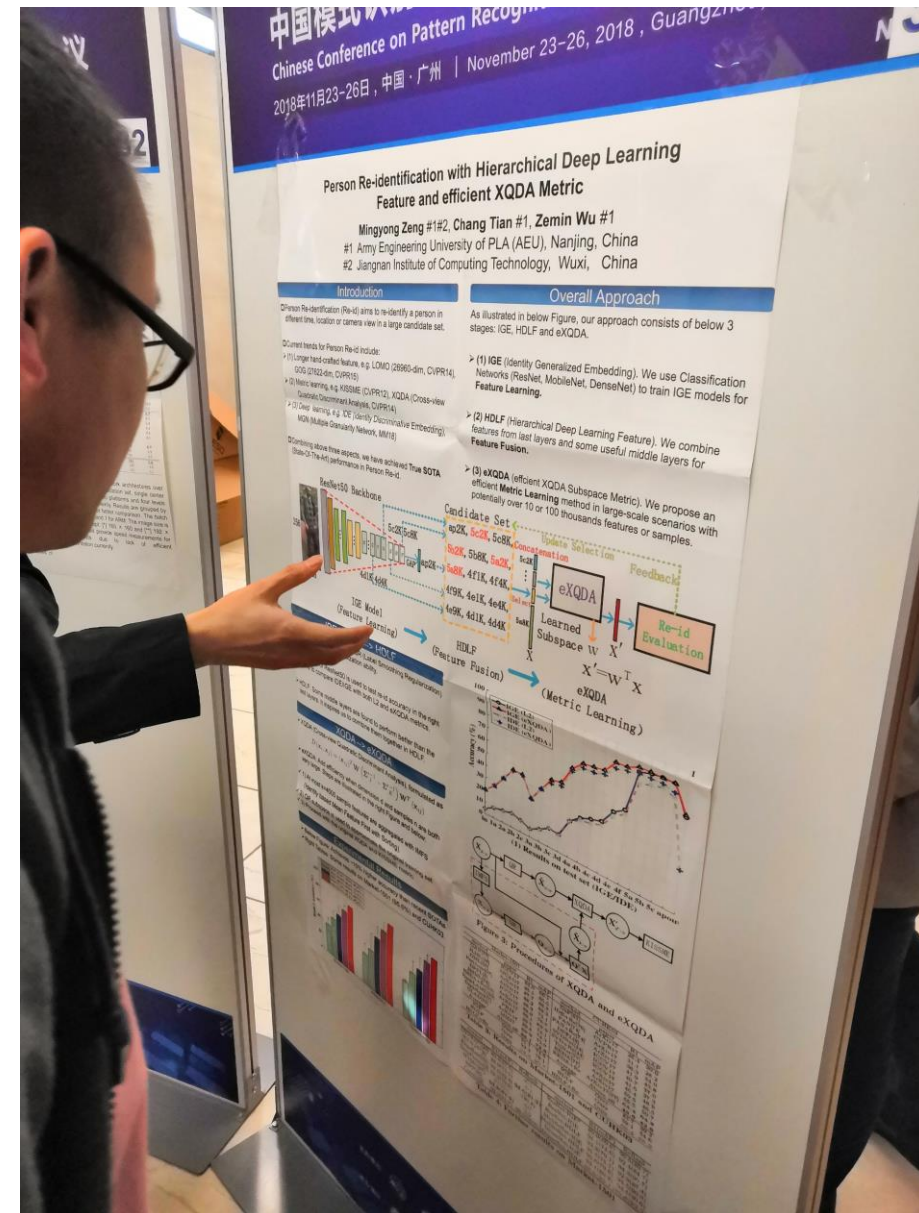


# PRCV 相关poster

Intelligent Information Fusion Research Group

# Person Re-identification with Hierarchical Deep Learning Feature and efficient XQDA Metric

## MM 2018

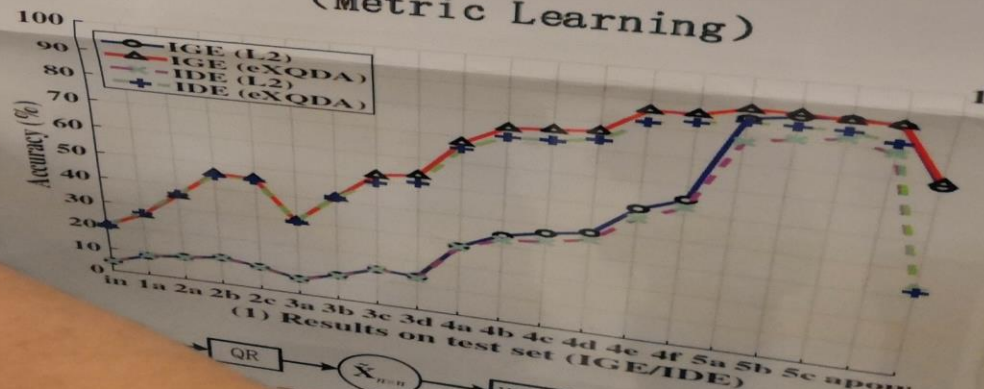
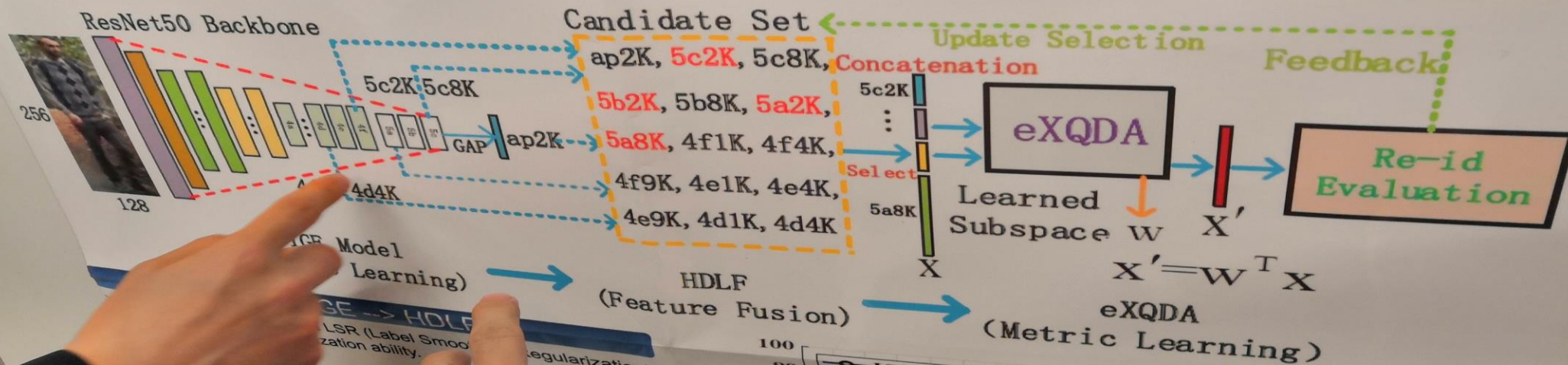


and-crafted feature, e.g. LOMO (26960-dim, CVPR14),  
 7622-dim, CVPR15)  
 learning, e.g. KISSME (CVPR12), XQDA (Cross-view  
 Quadratic Discriminant Analysis, CVPR14)  
 deep learning, e.g. IDE (Identity Discriminative Embedding),  
 MGN (Multiple Granularity Network, MM18)

Combining above three aspects, we have achieved **True SOTA**  
 (State-Of-The-Art) performance in Person Re-id.

## Feature Learning.

- (2) **HDLF** (Hierarchical Deep Learning Feature). We combine features from last layers and some useful middle layers for **Feature Fusion**.
- (3) **eXQDA** (efficient XQDA Subspace Metric). We propose an efficient **Metric Learning** method in large-scale scenarios with potentially over 10 or 100 thousands features or samples.





used to test re-id accuracy in HDLF with both L2 and eXQDA metrics.

layers are found to perform better than the HDLF. This suggests that combining them together in HDLF.

## XQDA --> eXQDA

Cross-view Quadratic Discriminant Analysis, formulated as

$$D(x_i, x_j) = (x_i - x_j)^T W (\Sigma_I^{-1} - \Sigma_E^{-1}) W^T (x_i - x_j)$$

eXQDA: Add efficiency when dimension  $d$  and samples  $n$  are both very large. Steps are illustrated in the right Figure and below:

- ✓ 1) At most  $k=4000$  sample features are aggregated with IMFS (Identity based Mean Feature First with Sorting).
- ✓ 2) QR subspace is used to approximate the original training set.
- ✓ 3) Proceed with the original XQDA and KISSME metric.

## Experimental Results

- Below Figure: Achieves ~10% higher accuracy than recent SOTAs.
- Right Tables: Some results on Market-1501 (95.6%) and CUHK03.

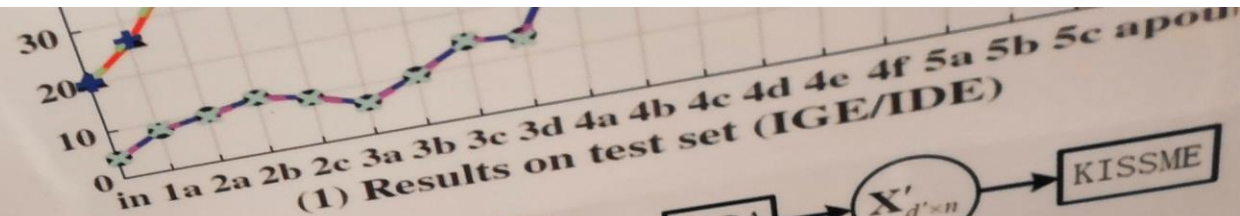
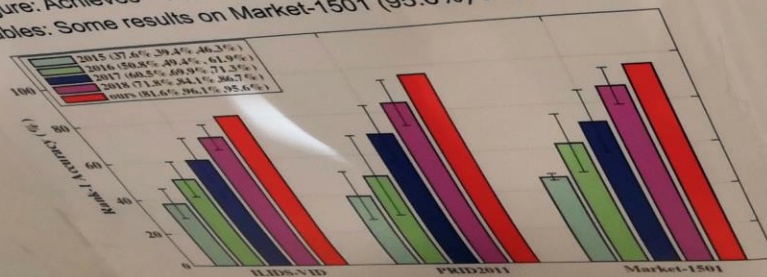


Figure 3: Procedures of XQDA and eXQDA

Figure 3: Procedures of XQ

Market-1501				CUHK03			
method	paper	R1	mAP	method	paper	R1	mAP
GAN[75]	ICCV17	84.0	66.1	IDE+Re[76]	ArXiv17	26.4	30.0
JLML[28]	IJCAI17	85.1	65.5	PAN[74]	CVPR17	31.1	28.2
DML[66]	ArXiv17	87.7	68.8	PAN+Re[74]	ArXiv17	36.3	34.0
PartLoss[57]	ArXiv17	88.2	69.3	DPFL[8]	ArXiv17	41.9	43.8
DPFL[8]	ICCV17	88.9	73.1	SVDNet[44]	ICCV17	40.7	37.0
MSML[54]	ArXiv17	88.9	76.7	SVD+Re[44]	ICCV17	41.5	37.3
REDA+Re[77]	PR18	89.5	83.9	TriNet[77]	ArXiv17	46.4	48.9
RankLAML[50]	ArXiv17	89.8	74.1	REDA[77]	ArXiv17	50.5	46.5
DarkRank[7]	MM17	89.9	73.9	REDA+Re	ArXiv17[77]	55.5	50.7
GLAD[52]	Above all	88.1	72.6	Average	Above all	64.4	64.8
Average	Ours	91.1	74.6	IGE	Ours	43.7	39.2
IGE	Ours	93.3	79.1	HDLF	Ours	59.1	54.6
HDLF	Ours	94.3	90.7	HDLF+Re	Ours	66.4	65.9
HDLF+Re	Ours	94.3	90.7				

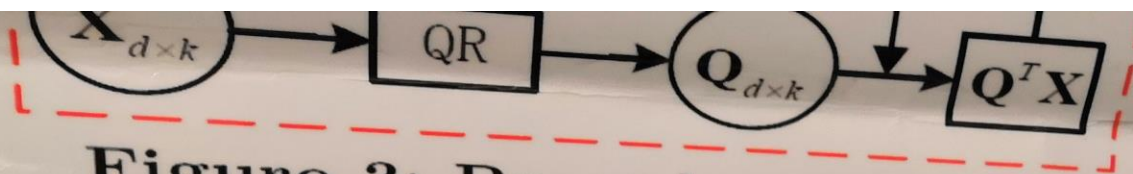
Market-1501 and CUHK03

Table 3: Results on Market-1501 and CUHK03

method	R1(mAP)	ReRank	method	R1(mAP)	ReRank
CVPR17*[25]	80.3(57.5)	-(-)	ResNet-IGE*	89.0(69.6)	90.6(80.7)
ICCV17*[8]	88.9(73.1)	-(-)	ResNet-IGE	91.1(74.6)	92.4(84.7)
AlignReid[64]	92.6(82.3)	94.0(91.2)	ResNet-HDLF	93.3(79.1)	94.3(90.7)
DeepPerson[3]	92.3(79.6)	-(90.8)	MobileNet-IGE	91.7(75.1)	92.9(87.4)
MTMC[67]	93.9(-)	-(-)	MobileNet-HDLF	93.6(79.6)	94.5(90.9)
RP[45]	93.8(81.6)	-(-)	DenseNet-IGE	93.1(80.3)	94.1(90.6)
Cutout[2]	92.2(81.7)	93.0(90.0)	DenseNet-HDLF	94.5(83.1)	95.6(92.2)

Table 4: Further results on Market-1501





**Figure 3: Procedures of XQDA and eXQDA**

Market-1501				CUHK03			
method	paper	R1	mAP	method	paper	R1	mAP
GAN[75]	ICCV17	84.0	66.1	DaF[61]	ArXiv17	26.4	30.0
JLML[28]	IJCAI17	85.1	65.5	IDE+Re[76]	CVPR17	31.1	28.2
DML[66]	ArXiv17	87.7	68.8	PAN[74]	ArXiv17	36.3	34.0
PartLoss[57]	ArXiv17	88.2	69.3	PAN+Re[74]	ArXiv17	41.9	43.8
DPFL[8]	ICCV17	88.9	73.1	DPFL[8]	ICCV17	40.7	37.0
MSML[54]	ArXiv17	88.9	76.7	SVDNet[44]	ICCV17	41.5	37.3
REDA+Re[77]	ArXiv17	89.1	83.9	SVD+Re[44]	ICCV17	46.4	48.9
RankLAML[50]	PR18	89.5	74.1	TriNet[77]	ArXiv17	50.5	46.5
DarkRank[7]	ArXiv17	89.8	74.3	REDA[77]	ArXiv17	55.5	50.7
GLAD[52]	MM17	89.9	73.9	REDA+Re	ArXiv17[77]	64.4	64.8
Average	Above all	88.1	72.6	Average	Above all	43.6	42.6
IGE	Ours	91.1	74.6	IGE	Ours	43.7	39.2
HDLF	Ours	93.3	79.1	HDLF	Ours	59.1	54.6
HDLF+Re	Ours	94.3	90.7	HDLF+Re	Ours	66.4	65.9

**Table 3: Results on Market-1501 and CUHK03**

method	R1(mAP)	ReRank	method	R1(mAP)	ReRank
CVPR17*[25]	80.3(57.5)	-(-)	ResNet-IGE*	89.0(69.6)	90.6(80.7)
ICCV17*[8]	88.9(73.1)	-(-)	ResNet-IGE	91.1(74.6)	92.4(84.7)
AlignReid[64]	92.6(82.3)	94.0(91.2)	ResNet-HDLF	93.3(79.1)	94.3(90.7)
DeepPerson[3]	92.3(79.6)	-(90.8)	MobileNet-IGE	91.7(75.1)	92.9(87.4)
MTMC[67]	93.9(-)	-(-)	MobileNet-HDLF	93.6(79.6)	94.5(90.9)
RPP[45]	93.8(81.6)	-(-)	DenseNet-IGE	93.1(80.3)	94.1(90.6)
Cutout[2]	92.2(81.7)	93.0(90.0)	DenseNet-HDLF	94.5(83.1)	95.6(92.2)

**Table 4: Further results on Market-1501**



# Unsupervised Cross-dataset person Re-identification by Transfer Learning of Spatial-temporal Patterns

## CVPR 2018



### UNSUPERVISED CROSS-DATASET PERSON RE-IDENTIFICATION BY TRANSFER LEARNING OF SPATIAL-TEMPORAL PATTERNS

Jianming Lv (jmlv@scut.edu.cn), Weihang Chen (csscut@mail.scut.edu.cn), Qing Li, Can Yang



#### PROBLEM

Most of the proposed person re-identification algorithms conduct supervised training on single small labeled dataset, so directly deploying these trained models to a large-scale real-world camera network may lead to poor performance due to underfitting. It is challenging to incrementally optimize the models by using the abundant unlabeled data collected from the target domain:

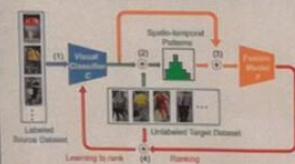
1. Usually expensive and impractical to label the massive online surveillance videos to support supervised learning.
2. The performance of unsupervised algorithms is typically poor.

#### CONTRIBUTIONS

We propose a high-performance unsupervised transfer learning algorithm, *TFusion*, for incremental learning in unlabeled target datasets, which includes:

1. A novel method to learn pedestrians' spatio-temporal patterns in unlabeled target datasets.
2. A Bayesian fusion model to combine the spatio-temporal patterns and the visual features for high performance person Re-ID.
3. A learning-to-rank based scheme for incremental optimization based on unlabeled data.

#### METHOD



The *TFusion* model consists of 4 steps:

1. Train the visual classifier  $\mathcal{C}$  in the labeled source dataset;
2. Using  $\mathcal{C}$  to learn the pedestrians' spatio-temporal patterns in the unlabeled target dataset;
3. Construct the Bayesian fusion model  $\mathcal{F}$ ;
4. Incrementally optimize  $\mathcal{C}$  by using the ranking results of  $\mathcal{F}$  in the unlabeled target dataset.

#### RESEARCH GROUP

Intelligent Information Fusion Research Group,  
School of Computer Science and Engineering,  
South China University of Technology.  
<http://io-link.org>



#### BAYESIAN FUSION MODEL

The fusion model is based on the conditional probability:

$$Pr(T(S_i) = T(S_j) | \bar{v}_i, \bar{v}_j, \Delta_{ij}, c_i, c_j)$$

Here  $S_i$  and  $S_j$  are any pair of surveillance images from the target dataset.  $S_i$  is taken at the camera  $c_i$  at the time  $t_i$ , and  $S_j$  is taken at the camera  $c_j$  at the time  $t_j$ . Their visual feature vectors are denoted as  $\bar{v}_i$  and  $\bar{v}_j$ . The timing interval between them is  $\Delta_{ij} = t_j - t_i$ .  $T(\cdot)$  denotes the ID of the pedestrian in an input image.

According to the Bayesian rule, we have:

$$Pr(T(S_i) = T(S_j) | \bar{v}_i, \bar{v}_j, \Delta_{ij}, c_i, c_j) = \frac{Pr(T(S_i) = T(S_j) | \bar{v}_i, \bar{v}_j) \cdot Pr(\Delta_{ij}, c_i, c_j | T(S_i) = T(S_j))}{Pr(\Delta_{ij}, c_i, c_j)}$$

Here  $Pr(\Delta_{ij}, c_i, c_j | T(S_i) = T(S_j))$  indicates the spatio-temporal pattern of pedestrians, which can be approximated by the spatio-temporal distribution of pedestrian images which seem to contain the same person judged by  $\mathcal{C}$ .

#### A FUTURE DIRECTION

We will extend *TFusion* by integrating the fusion model with multi-source spatio-temporal patterns to improve the Re-ID performance, e.g. the gait sequence, spatio-temporal signals collected from the pedestrians' mobile phones, etc.

#### RESULTS

Table 2: Compare the precision of *TFusion* with the state-of-art unsupervised transfer learning methods.

Method	Source	Target	Performance		
			rank-1	rank-5	rank-10
UMER[11]	Market1501	GRID	3.97	7.36	9.91
	CUHK01	GRID	3.58	7.56	9.50
	VIPeR	GRID	3.97	8.14	10.73
	GRID	Market1501	36.46	45.07	52.38
	CUHK01	Market1501	29.89	44.33	51.49
TFusion-uns	VIPeR	Market1501	30.14	44.92	52.14
	Market1501	GRID	40.40	57.30	67.40
	CUHK01	GRID	50.90	70.40	80.30
	VIPeR	GRID	62.70	81.70	89.20
	GRID	Market1501	58.27	72.37	76.84

Our model outperforms the state-of-art cross-dataset unsupervised transfer algorithms by a big margin, and can achieve comparable or even better performance than the state-of-art supervised algorithms using the same datasets.

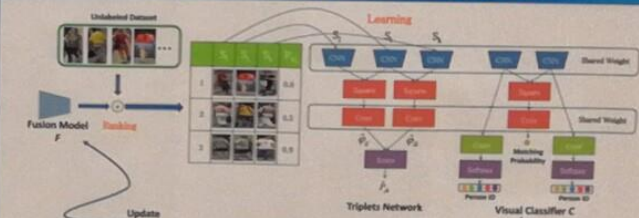
Table 3: Compare the precision of *TFusion* with the supervised methods on GRID.

Method	Performance		
	rank-1	rank-5	rank-10
GOO + XQDA[25]	34.80	-	58.40
HIERARCHICAL-CRAFT[11]	26.00	50.00	62.80
SSIM[23]	27.20	-	61.12
RLML[24]	37.5	61.4	69.4
TFusion-uns (Market1501 + GRID)	40.40	57.30	67.40
TFusion-sup	64.10	81.90	90.30

Table 4: Compare the precision of *TFusion* with the supervised algorithms on Market1501.

Method	Performance		
	rank-1	rank-5	rank-10
SSIM[23]	31.40	-	-
LDHRL[1]	29.47	60.73	80.94
S-CNN[2]	65.88	-	-
FA-CNN[7]	79.51	90.81	94.09
SVFNet[22]	82.3	-	-
RLML[24]	88.5	-	-
TFusion-uns (CUHK01 + Market1501)	80.98	92.48	95.53
TFusion-sup	93.13	98.43	99.46

#### INCREMENTAL OPTIMIZATION BY LEARNING-TO-RANK



The fusion model  $\mathcal{F}$  may be more accurate than the pure visual classifier  $\mathcal{C}$ , so we can utilize  $\mathcal{F}$  to optimize  $\mathcal{C}$  by teaching it with the ranking results in the unlabeled target dataset. Subsequently, the improvement of  $\mathcal{C}$  may also derive a better fusion model  $\mathcal{F}$ . In this mutual promotion procedure, both of the classifiers  $\mathcal{C}$  and  $\mathcal{F}$  can get incremental optimization in the unlabeled target dataset.

#### SOURCE CODE

The source code and compiled executables with an interactive interface are available at <https://github.com/ahangchen/TFusion>



# Mask-guided Contrastive Attention Model for Person Re-Identification

CVPR 2018

### Background of ReID

Camera Input → Query → Other Cameras → Output Results → Applications (Access Control, Video Retrieval)

### Challenges

- View: Different person with same background
- Pose: Same person with different backgrounds
- Detection Error
- Illumination

A core challenge is the background clutterers!

### Motivation

Guess who?

In a 'hard' mode

Mask is useful for ReID:  
a) contains **body shape** information, b) helps to remove the **background clutterers**.

### Our Method

RGB, Mask, Masked-RGB, RGB+Mask → body, full-image

For Motivation-1: Mask as extra input. For Motivation-2: Mask can guide the feature learning in a 'soft' mode.

### Framework of MGCAM

Input → Stage 1 → Stage 2 → Stage 3 → Stage 4 → Output

### Objective Function

The **Region-Level Triplet Loss**:

$$L_{trip} = \|h_{full} - h_{body}\|_2^2 + \max\{m - \|h_{full} - h_{body}\|_2^2, 0\}$$

The **Instance-Level Siamese Loss**:

$$L_{sia} = \begin{cases} \|h(p) - h(g)\|_2^2, & p = g \\ \max\{m - \|h(p) - h(g)\|_2^2, 0\}, & p \neq g \end{cases}$$

The **Attention Loss**:

$$L_{att} = \sum_{i=1}^I \sum_{j=1}^J \|M_{(i,j)} - \Phi_{(i,j)}\|_2^2$$

The **Overall Loss**:

$$L_{all} = L_{id(p,g)} + \lambda \cdot L_{sia} + \alpha \cdot L_{trip(p,g)} + \beta \cdot L_{att(p,g)}$$

MGCAM + Siamese Net

### Experimental Results

Mask examples of MARS, Market1501 and CUHK03.

Inputs Evaluation

Methods	Inputs	Rank-1	mAP
MSCAN-body [24]	RGB	69.70	52.41
	Mask	29.34	12.83
	Masked RGB	68.13	51.49
Ours(hard)	RGB-M	71.26	55.44
Ours	RGB-M	70.40	54.27
Ours	RGB	72.83	57.39
Ours	RGB-M	74.19	59.13

MGCAM Evaluation

Methods	Ref	Labeled	Detected	
Ours	Rank1	mAP	Rank1	mAP
	7.93	9.25	6.36	6.39
	14.8	13.6	12.8	11.5
	27.5	31.5	26.4	30.0
	38.1	40.3	34.7	37.4
Ours-Siamese	Rank1	mAP	Rank1	mAP
	40.93	46.89	46.39	46.74
	41.0	40.5	40.7	37.0
	46.29	46.89	46.39	46.74
	58.14	58.21	46.71	46.87

CUHK03

Methods	Distance Metric	Rank-1	mAP
MSCAN-body [24]	Euclidean	71.21	54.92
	KISSME	69.23	47.47
	Re-ranking	71.26	55.03
Ours	Euclidean	74.20	59.50
Ours	KISSME	70.06	51.26
Ours	XQDA	74.19	58.13
Ours	Re-ranking	76.01	70.13
Ours-Siamese	Euclidean	73.66	61.29
Ours-Siamese	KISSME	72.43	59.11
Ours-Siamese	XQDA	75.33	60.34
Ours-Siamese	Re-ranking	77.17	71.17

Market1501

Methods	Ref	Rank1	mAP	
Ours	Rank1	mAP	Rank1	mAP
	65.3	47.6	65.3	47.6
	68.23	51.82	68.23	51.82
	70.6	50.7	70.6	50.7
	70.51	55.12	70.51	55.12
Ours-Siamese	Rank1	mAP	Rank1	mAP
	71.77	56.05	71.77	56.05
	73.94	62.45	73.94	62.45
	76.61	70.13	76.61	70.13
	77.17	71.17	77.17	71.17

MARS

Methods	Ref	Rank1	mAP	
Ours	Rank1 <td>mAP</td> <td>Rank1 <td>mAP</td> </td>	mAP	Rank1 <td>mAP</td>	mAP
	65.3	47.6	65.3	47.6
	68.23	51.82	68.23	51.82
	70.6	50.7	70.6	50.7
	70.51	55.12	70.51	55.12
Ours-Siamese	Rank1 <td>mAP</td> <td>Rank1 <td>mAP</td> </td>	mAP	Rank1 <td>mAP</td>	mAP
	71.77	56.05	71.77	56.05
	73.94	62.45	73.94	62.45
	76.61	70.13	76.61	70.13
	77.17	71.17	77.17	71.17

Visualization of the attention maps.

CMC curves on MARS.

### Conclusions

- Mask as an extra input is helpful.
- The body **shape** from the mask contains **identity** information.
- Mask can guide the generation of **attention maps**.
- Produce body-aware and background-aware features.
- Mask guides the **region-level triplet loss**.
- Help to learn features invariant to background clutterers.

The mask and code are released at <https://github.com/developeng/MGCAM>

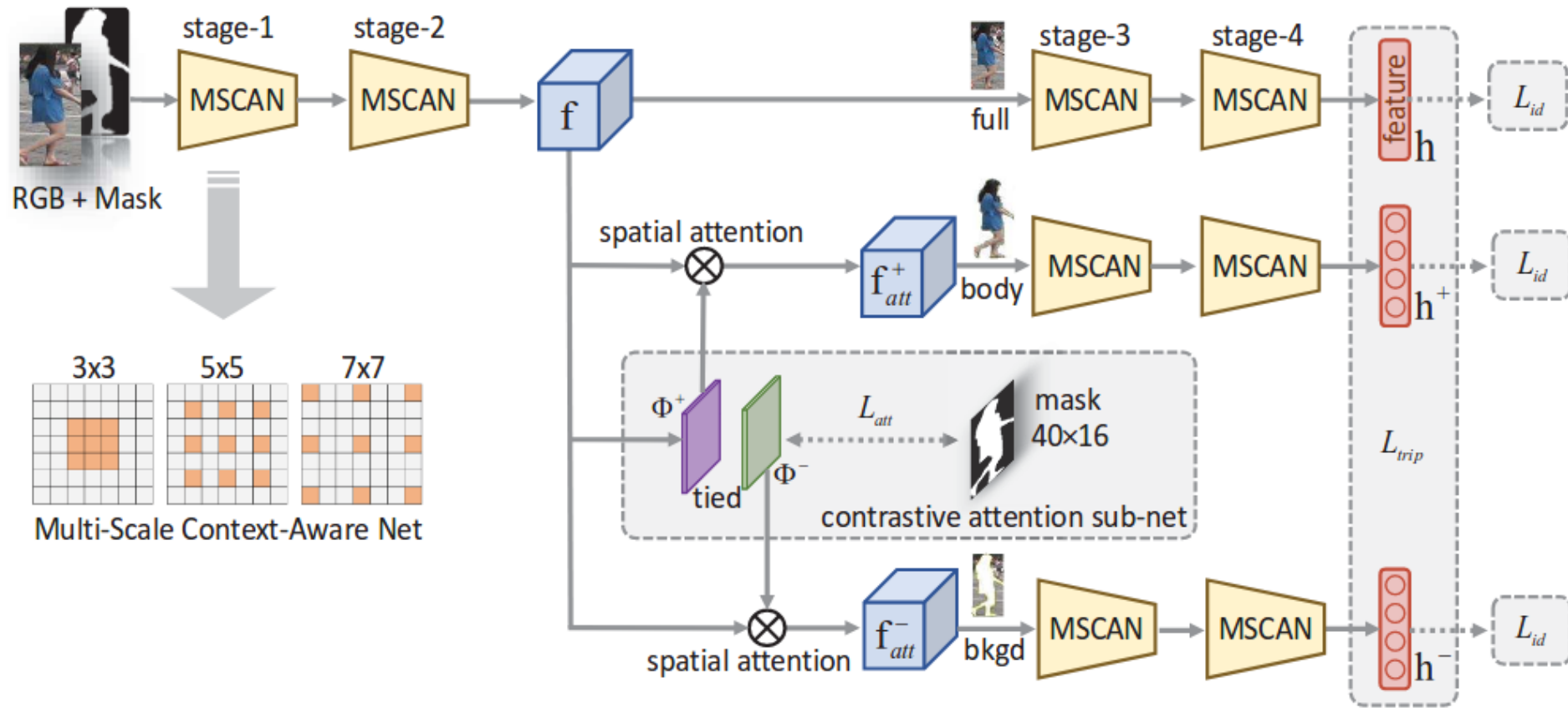
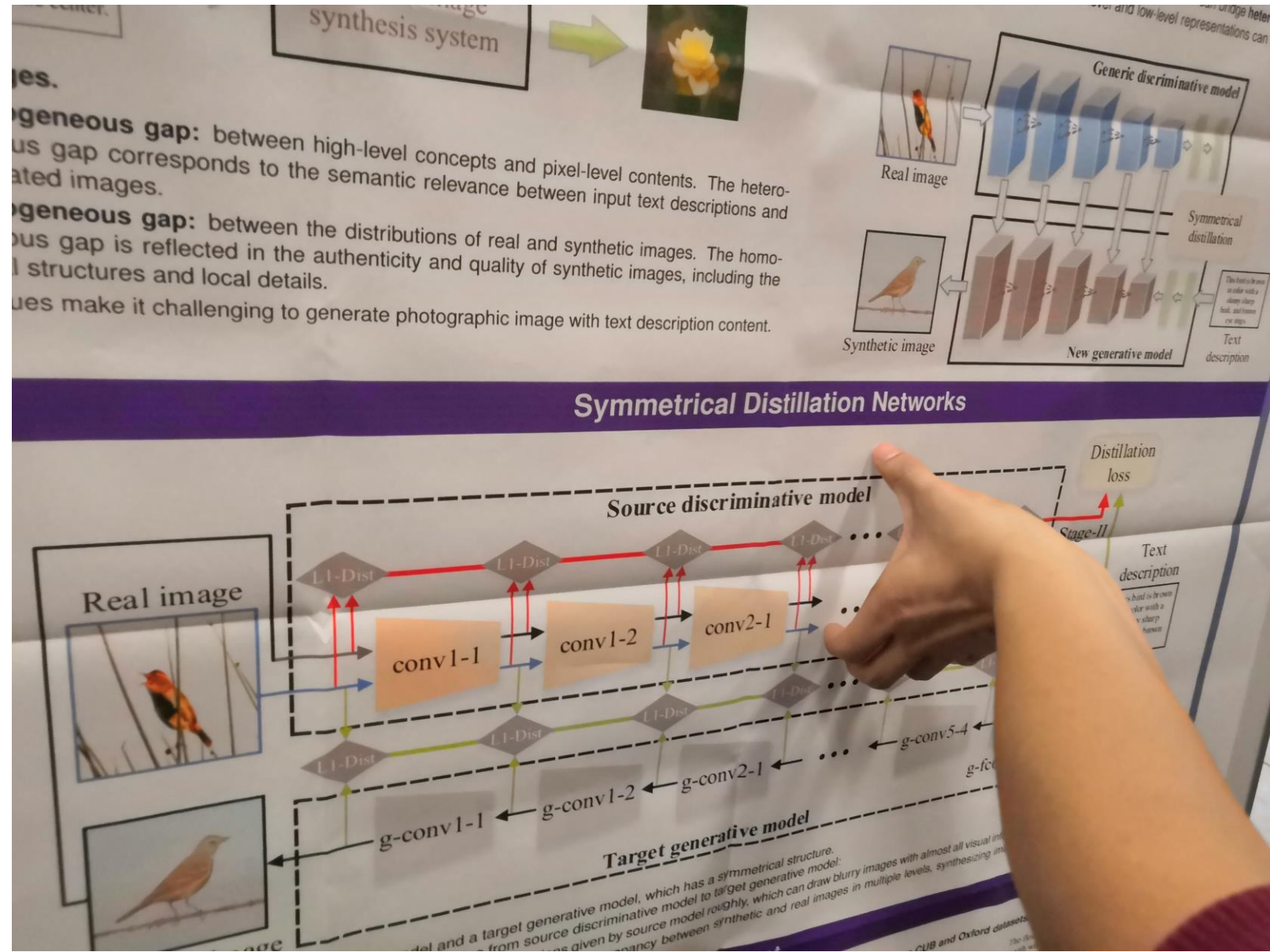


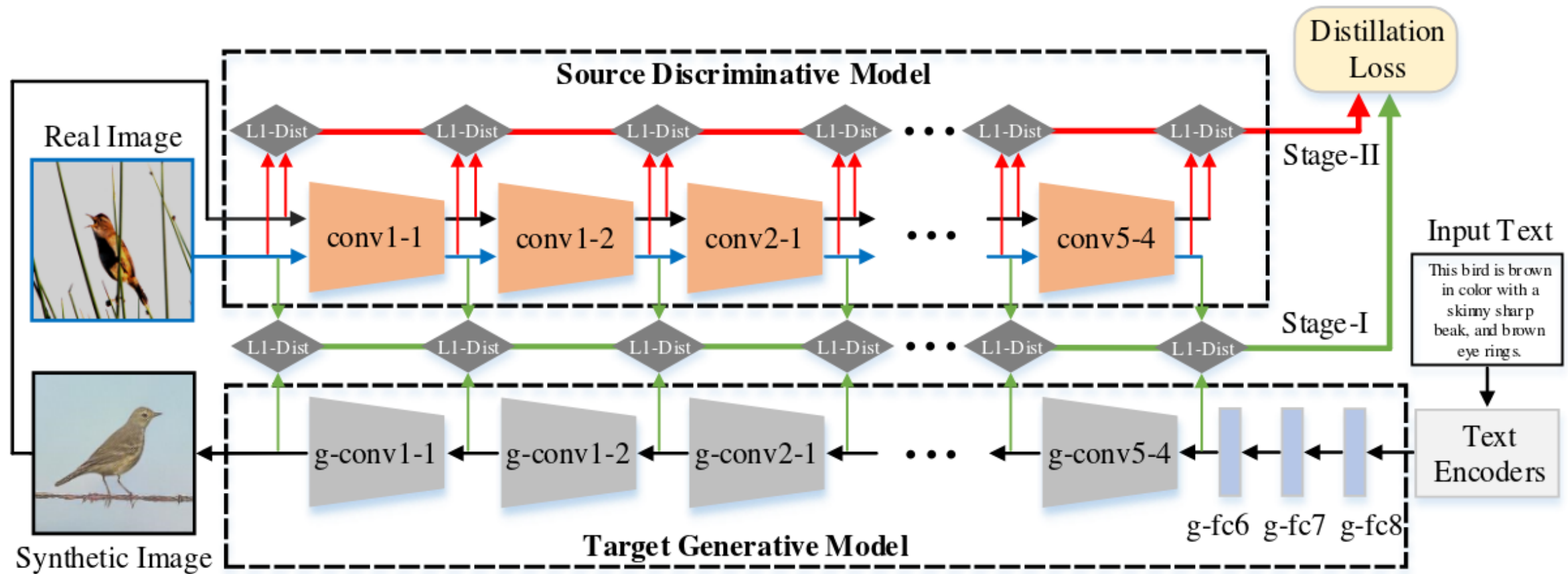
Figure 2. Framework of proposed Mask-guided Contrastive Attention Model (MGCAM) for person ReID. It contains four multi-scale context-aware stages and a fully-connected layer to learn final features. There are three main streams, i.e., the full-stream, the body-stream and the background-stream. In the middle is the contrastive attention sub-net which can generate a pair of body-aware and background-aware attention maps under the guide of binary mask. A region-level triplet loss is implemented on the features learnt from three streams.



# Text-to-image Synthesis via Symmetrical Distillation Networks

MM 2018 Oral





**Figure 2: The architecture of proposed Symmetrical Distillation Networks (SDN), which consists of a source discriminative model and a target generative model. The source model receives images as input and produces multi-level representations as guidance for the training of target model. The target model generates images conditioned on the text embedding produced by text encoders. The SDN applies two kinds of distillation loss in different stage to transfer hierarchical knowledge from the source model to the target model.**





**Table 1: Inception, SSIM and FSIM scores of our SDN and compared methods. Higher scores mean better results.**

Datasets	Methods	Inception	SSIM	FSIM
CUB-200-2011	<b>our SDN</b>	<b><math>6.89 \pm 0.06</math></b>	<b>0.3160</b>	<b>0.6264</b>
	StackGAN	$4.95 \pm 0.04$	0.2812	0.5869
	GAWWN	$5.22 \pm 0.08$	0.2370	0.5653
	GAN-INT-CLS	$5.08 \pm 0.08$	0.2934	0.6082
Oxford-Flower-102	<b>our SDN</b>	<b><math>4.28 \pm 0.09</math></b>	<b>0.2174</b>	<b>0.6227</b>
	StackGAN	$3.54 \pm 0.07$	0.1837	0.6009
	GAN-INT-CLS	$4.17 \pm 0.07$	0.1948	0.6214

# Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

被推荐到IEEE ACCES

single-shot v.s. multi-shot reid: 前者输入为两张图片; 后者输入为两个序列 (sequences, tracks)。multi-shot 比 single shot 有更丰富的信息, 然而这些信息中会有很多 noisy information。



Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification  
Aihua Zheng, Hongchao Li, Bo Jiang, Chenglong Li, Jin Tang, and Bin Luo  
(ahzheng214, tj, luobing)@ahu.edu.cn, (lhc950304, lcl1314)@foxmail.com, zeyiabc@163.com  
School of Computer Science and Technology, Anhui University, Hefei, China



## Motivation

Given  $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n}$ , where  $n$  denotes the number of images of a person in probe, where  $x_j \in \mathbb{R}^{d \times 1}$ ,  $j = [1, \dots, n]$  denotes the corresponding  $d$ -dimensional feature. While  $D = [D^1, D^2, \dots, D^G] \in \mathbb{R}^{d \times M}$  denotes the total  $M$  images of  $G$  persons in gallery, where  $D^p = [d_1^p, d_2^p, \dots, d_{g_p}^p] \in \mathbb{R}^{d \times g_p}$  represents the matrix of  $g_p$  basis vectors for the  $p$ -th person,  $g_p$  denotes the number of images of the  $p$ -th person in gallery. Obviously,  $M = \sum_{p=1}^G g_p$ . The basic idea of sparse ranking based Re-ID is to reconstruct a testing probe image  $x_i$  with linear spanned training gallery images of  $G$  persons:

$$x_i \approx \sum_{p=1}^G D^p c_i^p = D c_i \quad (1)$$

where  $c_i^p = [c_{i1}^p, c_{i2}^p, \dots, c_{ig_p}^p] \in \mathbb{R}^{g_p \times 1}$  represents the coding coefficients of the  $p$ -th person against the probe instance  $x_i$ .

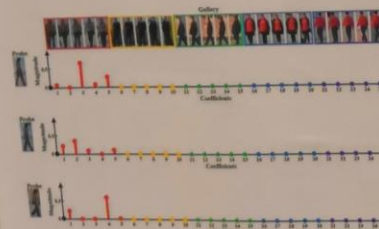


Fig. 1. Rank sparse representation for Multi-shot re-identification on CAVIAR4REID. Left: probe samples. Right: top five samples in the gallery for each probe. Bottom: gallery reconstruction coefficients for sparse representation. Each color represents a single subject which has five instances.

## Problem Formulation

In order to concentratively reconstruct the probe via relatively few dictionary atoms from the gallery, we can impose the sparsity constraint into above formulation:

$$\min \|X - DC\|_2 + \lambda \|C\|_1 \quad (2)$$

**Global Graph Regularization.** We argue that the feature vectors derived from the multiple images of the same person tend to have similar geometric distribution. To exploit the intrinsic geometric distribution among the probe images, we first enforce a global graph regularizer over the reconstruction coefficients:

$$\min \|X - DC\|_2 + \lambda \|C\|_1 + \beta \text{tr}(C L_1 C^T) \quad (3)$$

**Local Graph Regularization.** We further argue that the multiple images of the same person in gallery fall into similar geometry. To exploit the intrinsic geometry among the gallery images, we further enforce a local graph regularizer over the reconstruction coefficients:

$$\min \|X - DC\|_2 + \lambda \|C\|_1 + \beta \text{tr}(C L_1 C^T) + \gamma \text{tr}(C^T L_2 C) \quad (4)$$

**Non-negative Item.** Thinking that the reconstruction coefficients are non-negative while representing similarity measures between probe and gallery, we further enforce the nonnegative constraint on the reconstruction coefficients in the proposed model, and the final formulation is as follows:

$$\min \|X - DC\|_2 + \lambda \|C\|_1 + \beta \text{tr}(C L_1 C^T) + \gamma \text{tr}(C^T L_2 C), \text{ s.t. } C \geq 0 \quad (5)$$

which is named NNDGSR in this paper.

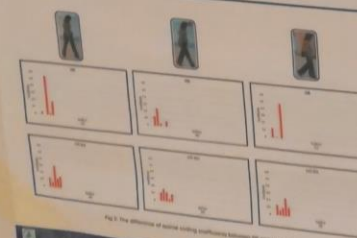


Fig. 2. The difference of sparse coding coefficients between the two methods.

## Ranking Implementation

Due to the sparsity of the reconstruction coefficients, the majority of which collapse to zero after few higher coefficients. Therefore, we can not support ranking for all the individuals in gallery. To cope this issue, we develop an error distribution technique. First, we can obtain the normalized reconstruction error for current probe  $x_i$  according to coefficients as:

$$e_j = \frac{\|x_i - D c_j\|_2}{\|x_i\|_2} \quad (6)$$

We use the reconstruction residues to make the  $p$ -th category whose reconstruction coefficients are all zeros has ranking value. Therefore, the final ranking value of the probe person with  $n$  images against the  $p$ -th person in gallery is defined as follows:

$$r^p = \sum_{j=1}^n \sum_{k=1}^{g_p} (c_{jk}^p + w_{jk}^p e_j), p = [1, \dots, G] \quad (7)$$

Our final decision rule is:

$$\text{class}(X) = \arg \max_p r^p \quad (8)$$

## Experiments

### Comparison on i-LIDS and CAVIAR4REID

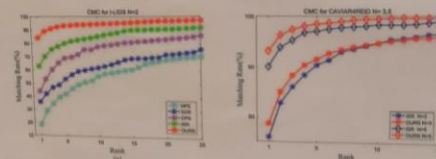


Fig. 3. The cumulative match characteristic curves on i-LIDS and CAVIAR4REID with hand-crafted feature comparing with the state-of-the-arts.

Table 1. Comparison results at Rank-1 on i-LIDS and CAVIAR4REID (in %)

		CAVIAR4REID (in %)				
Features	Methods	i-LIDS		CAVIAR4REID		References
		N=2	N=3	N=5		
Hand-craft features	HPE	18.5				ICPR2010
	AHPE	32	7.5	7.5		PRL2012
	SCR	36				ICAVS2010
	MRCCG	46				ICAVS2011
	BDALP	39	8.5	8.3		CVPR2010
	CPS	44	13			BMVC2011
	COSMATI	44		17.5		ECCV2012
	WHOS + ISR	62.9	5.1	96.1		PAMI2015
	WHOS + NNDGSR	84.3	78.7	93.2		
	APR + EU	67.7	44.3	53.8		Arxiv2017
Deep features	APR + ISR	77.3	65.7	80.7		Arxiv2017 + PAMI2015
	APR + NNDGSR	78.4	70.4	89.0		

### Comparison on MARS

Table 2. Comparison results at Rank-1 MARS (in %)

Features	Methods	Rank-1	Rank-5	Rank-20	References
Hand-craft features	BOGD + KISSME	2.6	6.4	12.4	BMVC2010 + CVPR2012
	GRE + KISSME	1.2	2.8	7.4	PAMI2005 + CVPR2012
	BoW + KISSME	18.6	33.0	45.9	ECCV2014 + CVPR2015
	LOMO + XQDA	30.7	46.3	59.2	ICCV2015 + CVPR2015
	ASTPN	44	70	81	CVPR2015
	LCAR	55.5	70.2	80.2	ICCV2017
Deep features	SATTP	69.7	84.7	92.8	Arxiv2017
	SPT	70.8	90	92.8	Arxiv2017
	MSCAN	71.8	86.6	93.1	CVPR2017
	IDR + IDR	58.7	77.1	86.4	CVPR2017
	IDR	61	77.1	85.6	ECCV2016
	IDR + NNDGSR	73.50	88.0	93.90	ECCV2016 + PAMI2015

### Component Analysis

Table 3. Evaluation on individual component on CAVIAR4REID dataset with N=5 on APR deep features (in %)

Components	Rank-1	Rank-5	Rank-10	Rank-20
ISR	80.7	95.8	97.9	99.4
SR+NN	84.3	94.9	97.3	98.6
SR+NN+GG	87.8	96.7	98.2	99.4
SR+NN+GG+LG	89.0	96.8	98.2	99.3



# Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

被推荐到IEEE ACCES

$$\mathbf{x}_j \approx \sum_{p=1}^G \mathbf{D}^p \mathbf{c}_j^p$$

$$= \mathbf{D} \mathbf{c}_j$$

$$\min_{\mathbf{c}_j} \|\mathbf{x}_j - \mathbf{D} \mathbf{c}_j\|_2^2 + \lambda \|\mathbf{c}_j\|_1$$



Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification  
Aihua Zheng, Hongchao Li, Bo Jiang, Chenglong Li, Jin Tang, and Bin Luo  
(ahzheng214, tj\_luobing)@ahu.edu.cn, (lhc950304, lcl1314)@foxmail.com, zeyiabcc@163.com  
School of Computer Science and Technology, Anhui University, Hefei, China



## Motivation

Given  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ , where  $n$  denotes the number of images of a person in probe, where  $\mathbf{x}_j \in \mathbb{R}^{d \times 1}$ ,  $j = \{1, \dots, n\}$  denotes the corresponding  $d$ -dimensional feature. While  $\mathbf{D} = [\mathbf{D}^1, \mathbf{D}^2, \dots, \mathbf{D}^G] \in \mathbb{R}^{d \times M}$  denotes the total  $M$  images of  $G$  persons in gallery, where  $\mathbf{D}^p = [\mathbf{d}_1^p, \mathbf{d}_2^p, \dots, \mathbf{d}_{g_p}^p] \in \mathbb{R}^{d \times g_p}$  represents the matrix of  $\mathbf{D}^p$  basis vectors for the  $p$ -th person,  $\mathbf{D}^p$  denotes the number of images of the  $p$ -th person in gallery. Obviously,  $M = \sum_{p=1}^G g_p$ . The basic idea of sparse ranking based Re-ID is to reconstruct a testing probe image  $\mathbf{x}_j$  with linear spanned training gallery images of  $G$  persons:

$$\mathbf{x}_j \approx \sum_{p=1}^G \mathbf{D}^p \mathbf{c}_j^p = \mathbf{D} \mathbf{c}_j \quad (1)$$

where  $\mathbf{c}_j^p = [c_{j1}^p, c_{j2}^p, \dots, c_{jg_p}^p]^T \in \mathbb{R}^{g_p \times 1}$  represents the coding coefficients of the  $p$ -th person against the probe instance  $\mathbf{x}_j$ .

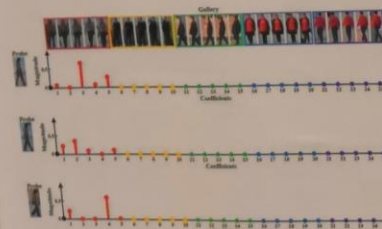


Fig. 1. Rank sparse representation for Multi-shot Re-identification on CAVIA4REID. Left: probe samples. Right: Top five rank 5 samples in the gallery for samples in the test set. Each color represents a single subject within the test set.

## Problem Formulation

In order to concentratively reconstruct the probe via relatively few dictionary atoms from the gallery, we can impose the sparsity constraint into above formulation:

$$\min \|\mathbf{x} - \mathbf{D} \mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 \quad (2)$$

**Global Graph Regularization.** We argue that the feature vectors derived from the multiple images of the same person tend to have similar geometric distribution. To exploit the intrinsic geometric distribution among the probe images, we first enforce a global graph regularizer over the reconstruction coefficients:

$$\min \|\mathbf{x} - \mathbf{D} \mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 + \beta \text{tr}(\mathbf{C}^T \mathbf{L} \mathbf{C}) \quad (3)$$

**Local Graph Regularization.** We further argue that the multiple images of the same person in gallery fall into similar geometry. To exploit the intrinsic geometry among the gallery images, we further enforce a local graph regularizer over the reconstruction coefficients:

$$\min \|\mathbf{x} - \mathbf{D} \mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 + \beta \text{tr}(\mathbf{C}^T \mathbf{L}_l \mathbf{C}) + \gamma \text{tr}(\mathbf{C}^T \mathbf{L}_g \mathbf{C}) \quad (4)$$

**Non-negative Item.** Thinking that the reconstruction coefficients are non-negative while representing similarity measures between probe and gallery, we further enforce the nonnegative constraint on the reconstruction coefficients in the proposed model, and the final formulation is as follows:

$$\min \|\mathbf{x} - \mathbf{D} \mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 + \beta \text{tr}(\mathbf{C}^T \mathbf{L}_l \mathbf{C}) + \gamma \text{tr}(\mathbf{C}^T \mathbf{L}_g \mathbf{C}), \text{ s.t. } \mathbf{c} \geq 0 \quad (5)$$

which is named NNDGSR in this paper.



Fig. 2. The difference of local coding coefficients between the test samples.

## Ranking Implementation

Due to the sparsity of the reconstruction coefficients, the majority of which collapse to zero after few higher coefficients. Therefore, we can not support ranking for all the individuals in gallery. To cope this issue, we develop an error distribution technique. First, we can obtain the normalized reconstruction error for current probe  $\mathbf{x}_j$  according to coefficients as:

$$e_j = \frac{\|\mathbf{x}_j - \mathbf{D} \mathbf{c}_j\|_2}{\|\mathbf{x}_j\|_2} \quad (6)$$

We use the reconstruction residues to make the  $p$ -th category whose reconstruction coefficients are all zeros has ranking value. Therefore, the final ranking value of the probe person with  $n$  images against the  $p$ -th person in gallery is defined as follows:

$$r^p = \sum_{j=1}^n \sum_{k=1}^{g_p} (c_{jk}^p + w_{jk}^p e_j), p = \{1, \dots, G\} \quad (7)$$

Our final decision rule is:

$$\text{class}(\mathbf{X}) = \arg \max_p r^p \quad (8)$$

## Experiments

### Comparison on i-LIDS and CAVIA4REID

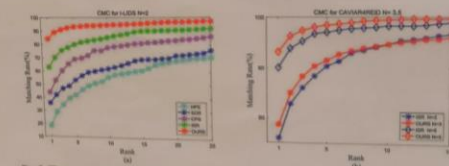


Fig. 3. The cumulative match characteristic curves on i-LIDS and CAVIA4REID with hand-crafted feature comparing with the state-of-the-arts.

Table 1. Comparison results at Rank-1 on i-LIDS and CAVIA4REID (in %)

Features	Methods	i-LIDS N=2	CAVIA4REID N=3	References
Hand-craft Features	HPE	18.5	-	ICPR2010
	AHPE	32	7.5	ICPR2012
	SCR	36	-	ICAVS2010
	MRCCG	46	-	ICAVS2011
	BDALP	39	8.5	CVPR2010
	CPS	44	13	BMVC2011
	COSMATI	44	-	ECCV2012
	WIOS + ISR	62.9	5.1	PAMI2015
	WIOS + NNDGSR	84.3	78.7	93.2
	APR + EU	67.7	44.3	53.8
Deep features	APR + ISR	77.3	65.7	80.7
	APR + NNDGSR	78.4	70.4	89.0

### Comparison on MARS

Table 2. Comparison results at Rank-1 MARS (in %)

Features	Methods	Rank-1	Rank-5	Rank-20	References
Hand-craft Features	BOGD + KISME	2.6	6.4	12.4	BMVC2010+CVPR2012
	GBE + KISME	1.2	2.8	7.4	PAMI2005+CVPR2012
	BoW + KISME	18.6	33.0	45.9	ECCV2014+CVPR2015
	LOMO + XQDA	30.6	46.2	59.2	ICCV2015+CVPR2012
	LOMO + XQDA	30.7	46.6	60.9	CVPR2015
	ASTPN	44	70	81	ICCV2017
Deep features	LCAR	55.5	70.2	80.2	Arxiv2017
	SATTP	69.7	84.7	92.8	Arxiv2017
	SPT	70.8	90	97.8	CVPR2017
	MSCAN	71.8	86.6	93.1	CVPR2017
	IDR+ISR	58.7	77.1	86.8	ECCV2016
	IDR + NNDGSR	72.50	85.0	93.90	ECCV2016+PAMI2015

### Component Analysis

Table 3. Evaluation on individual component on CAVIA4REID dataset with N=5 on APR deep features (in %)

Components	Rank-1	Rank-5	Rank-10	Rank-20
ISR	80.7	95.8	97.9	99.4
SR+NN	84.3	94.9	97.3	98.6
SR+NN+CG	87.8	96.7	98.2	99.4
SR+NN+CG+LG	89.0	96.8	98.2	99.3

# Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

被推荐到IEEE ACCES

$$\min_{\mathbf{c}_j} \|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2^2 + \lambda \|\mathbf{c}_j\|_1$$

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta \text{tr}(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma \text{tr}(\mathbf{C}^T\mathbf{L}_2\mathbf{C}).$$



## Motivation

Given  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ , where  $n$  denotes the number of images of a person in probe, where  $\mathbf{x}_j \in \mathbb{R}^{d \times 1}$ ,  $j = \{1, \dots, n\}$  denotes the corresponding  $d$ -dimensional feature. While  $\mathbf{D} = [\mathbf{D}^1, \mathbf{D}^2, \dots, \mathbf{D}^G] \in \mathbb{R}^{d \times M}$  denotes the total  $M$  images of  $G$  persons in gallery, where  $\mathbf{D}^p = [\mathbf{d}_1^p, \mathbf{d}_2^p, \dots, \mathbf{d}_{g_p}^p] \in \mathbb{R}^{d \times g_p}$  represents the matrix of  $\mathbf{D}^p$  basis vectors for the  $p$ -th person,  $\mathbf{D}^p$  denotes the number of images of the  $p$ -th person in gallery. Obviously,  $M = \sum_{p=1}^G g_p$ . The basic idea of sparse ranking based Re-ID is to reconstruct a testing probe image  $\mathbf{x}_j$  with linear spanned training gallery images of  $G$  persons:

$$\mathbf{x}_j \approx \sum_{p=1}^G \mathbf{D}^p \mathbf{c}_j^p = \mathbf{D} \mathbf{c}_j \quad (1)$$

where  $\mathbf{c}_j^p = [c_{j1}^p, c_{j2}^p, \dots, c_{jg_p}^p]^T \in \mathbb{R}^{g_p \times 1}$  represents the coding coefficients of the  $p$ -th person against the probe instance  $\mathbf{x}_j$ .

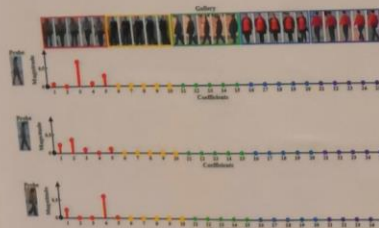


Fig. 1. Rank sparse representation for Multi-shot Re-identification on CAVIAR4REID. Left: probe samples. Right: Top five rank 5 samples in the gallery for each probe. Bottom: gallery reconstruction coefficients for sparse representation. Each color represents a single subject within the test instance.

## Problem Formulation

In order to concentratively reconstruct the probe via relatively few dictionary atoms from the gallery, we can impose the sparsity constraint into above formulation:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_F^2 + \lambda \|\mathbf{C}\|_1 \quad (2)$$

**Global Graph Regularization.** We argue that the feature vectors derived from the multiple images of the same person tend to have similar geometric distribution. To exploit the intrinsic geometric distribution among the probe images, we first enforce a global graph regularizer over the reconstruction coefficients:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta \text{tr}(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) \quad (3)$$

**Local Graph Regularization.** We further argue that the multiple images of the same person in gallery fall into similar geometry. To exploit the intrinsic geometry among the gallery images, we further enforce a local graph regularizer over the reconstruction coefficients:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta \text{tr}(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma \text{tr}(\mathbf{C}^T\mathbf{L}_2\mathbf{C}) \quad (4)$$

**Non-negative Item.** Thinking that the reconstruction coefficients are non-negative while representing similarity measures between probe and gallery, we further enforce the nonnegative constraint on the reconstruction coefficients in the proposed model, and the final formulation is as follows:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta \text{tr}(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma \text{tr}(\mathbf{C}^T\mathbf{L}_2\mathbf{C}), \text{ s.t. } \mathbf{C} \geq 0 \quad (5)$$

which is named NNDGSR in this paper.

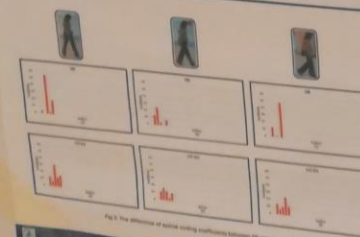


Fig. 2. The difference of sparse coding coefficients between the test instance and gallery.

## Ranking Implementation

Due to the sparsity of the reconstruction coefficients, the majority of which collapse to zero after few higher coefficients. Therefore, we can not support ranking for all the individuals in gallery. To cope this issue, we develop an error distribution technique. First, we can obtain the normalized reconstruction error for current probe  $\mathbf{x}_j$  according to coefficients as:

$$e_j = \frac{\|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2}{\|\mathbf{x}_j\|_2} \quad (6)$$

We use the reconstruction residues to make the  $p$ -th category whose reconstruction coefficients are all zeros has ranking value. Therefore, the final ranking value of the probe person with  $n$  images against the  $p$ -th person in gallery is defined as follows:

$$r^p = \sum_{j=1}^n \sum_{k=1}^{g_p} (c_{jk}^p + w_{jk}^p e_j), p = \{1, \dots, G\} \quad (7)$$

Our final decision rule is:

$$\text{class}(\mathbf{X}) = \arg \max_p r^p \quad (8)$$

## Experiments

### Comparison on i-LIDS and CAVIAR4REID

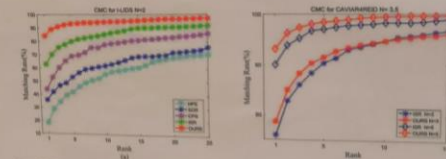


Fig. 3. The cumulative match characteristic curves on i-LIDS and CAVIAR4REID with hand-crafted feature comparing with the state-of-the-arts.

Table 1. Comparison results at Rank-1 on i-LIDS and CAVIAR4REID (in %)

Features	Methods	i-LIDS CAVIAR4REID (in %)			References
		N=2	N=3	N=5	
Hand-craft features	HPE	18.5			ICPR2010
	AHPE	32	7.5	7.5	PRIL2012
	SCR	36			ICAVS2010
	MRCCG	46			ICAVS2011
	BDALP	39	8.5	8.3	CVPR2010
	CPS	44	13		BMVC2011
	COSMATI	44		17.5	ECCV2012
	WHOS + ISR	62.9	5.1	96.1	PAMI2015
	WHOS + NNDGSR	84.3	78.7	93.2	
	APR + EU	67.7	44.3	53.8	Arxiv2017
Deep features	APR + ISR	77.2	65.7	80.7	Arxiv2017 + PAMI2015
	APR + NNDGSR	78.4	70.4	89.0	

### Comparison on MARS

Table 2. Comparison results at Rank-1 MARS (in %)

Features	Methods	Rank-1	Rank-5	Rank-20	References
Hand-craft features	HOCD + KISME	2.6	6.4	12.4	BMVC2010 + CVPR2012
	GEI + KISME	1.2	2.8	7.4	PAMI2005 + CVPR2012
	HuLBP + XQDA	18.6	33.0	45.9	ECCV2014 + CVPR2015
	BoW + KISME	30.6	46.2	59.2	ICCV2015 + CVPR2015
	LOMO + XQDA	30.7	46.6	60.9	CVPR2015
	ASTPN	44	70	81	ICCV2017
Deep features	SATTP	55.5	70.2	80.2	Arxiv2017
	SPT	69.7	84.7	92.8	Arxiv2017
	MSCAN	70.8	90	97.8	CVPR2017
	IDE+EU	58.7	86.6	93.1	CVPR2017
	IDE + ISR	61	77.1	86.4	ECCV2016
	IDE + NNDGSR	73.50	88.0	93.80	ECCV2016 + PAMI2015

### Component Analysis

Table 3. Evaluation on individual component on CAVIAR4REID dataset with N=5 on APR deep features (in %)

Components	Rank-1	Rank-5	Rank-10	Rank-20
ISR	80.7	95.8	97.9	99.4
SR+NN	84.3	94.9	97.3	98.6
SR+NN+GG	87.8	96.7	98.2	99.4
SR+NN+GG+LG	89.0	96.8	98.2	99.3





# Center-level Verification Model for Person Re-Identification

Ruo Chen, Zheng Yang, Chen Changpan, Yu Chu, Chu Han, Changxin Gao, and Ning Song  
Key Laboratory of Ministry of Education for Image Processing and Intelligent Control,  
School of Automation, Huazhong University of Science and Technology, Wuhan, China

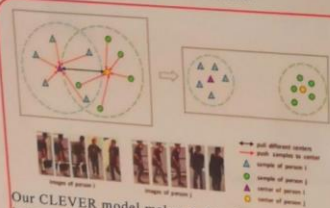
## Introduction

Siamese network, which has been widely used in person re-identification(re-id), only pays attention on individual samples, which cannot represent the distribution of the identity in the scenario of deep learning. In this paper, we introduce a novel center-level verification (CLEVER) model for the siamese network, which builds the verification model on center level to both reduce intra-class variations and enlarge inter-class distances.

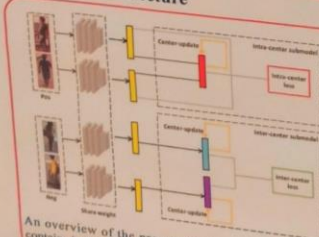
### Main contributions:

- We propose a center-level verification (CLEVER) model based on siamese network, which can both reduce intra-class variation and enlarge inter-class distance.
- We show competitive results on CUHK03, CUHK01 and VIPeR, proving the effectiveness of our method.

## Illustration of our motivation



## Network Architecture



## Association Procedure

### intra-center model:

$$L_{intra} = \frac{1}{2n} \sum_{i=1}^n (\|x_{i1} - c_{i1}\|_2^2 + \|x_{i2} - c_{i2}\|_2^2) \quad (1)$$

### update the center:

$$\frac{\partial L_{intra}}{\partial c_{i1}} = x_{i1} - c_{i1} \quad (2)$$

$$\frac{\partial L_{intra}}{\partial c_{i2}} = x_{i2} - c_{i2} \quad (3)$$

$$c_{i1} = \frac{\sum_{j=1}^n x_{j1}}{1 + \sum_{j=1}^n \mathbb{I}(x_{j1} = c_{i1})} \quad (4)$$

$$c_{i2} = \frac{\sum_{j=1}^n x_{j2}}{1 + \sum_{j=1}^n \mathbb{I}(x_{j2} = c_{i2})} \quad (5)$$

### inter-center model:

$$L_{inter} = \frac{1}{2n} \sum_{i=1}^n \max(0, d - \|c_{i1} - c_{i2}\|_2) \quad (6)$$

### Joint Optimization:

$$L_{CLEVER} = \beta \cdot L_{intra} + \gamma \cdot L_{inter} \quad (7)$$

## Experiment Results

Results on CUHK03(detected) to show the effectiveness of each component.

Method	rank1	rank5	rank10
baseline IC	80.3	96.1	97.0
CLEVER(intra only)+I	81.4	96.2	98.4
baseline IV	81.3	96.3	97.7
CLEVER(intra only)+IV	83.1	96.3	98.4
CLEVER+I	81.4	96.3	97.8
CLEVER+IV	82.0	96.4	98.4
CLEVER+IV	84.8	97.5	98.9

Comparison with state-of-the-art methods on CUHK03(detected), CUHK01 and VIPeR datasets using the single-shot setting.

Dataset	Method	rank1	rank5	rank10
CUHK03 (detected)	Baseline	80.3	96.1	97.0
	CLEVER	84.8	97.5	98.9
	DeepVid	81.4	96.2	98.4
	DeepVid	81.4	96.2	98.4
CUHK01	Baseline	79.8	95.8	97.0
	CLEVER	81.4	96.2	98.4
	DeepVid	81.4	96.2	98.4
	DeepVid	81.4	96.2	98.4
VIPeR	Baseline	79.8	95.8	97.0
	CLEVER	81.4	96.2	98.4
	DeepVid	81.4	96.2	98.4
	DeepVid	81.4	96.2	98.4

## Conclusions

In this paper, we have proposed a center-level verification model named CLEVER model for the sample-level models. The loss function of the CLEVER model is calculated by samples and their centers, which to some extent represent the corresponding distributions. Finally, we combine the proposed center-level loss and the intra-class variation and inter-class distance. The control of center improves the generation ability of network, which has outperformed most of the state-of-the-art methods on VIPeR, CUHK01 and CUHK03.



# Re-ranking Person Re-identification with Adaptive Hard Sample Mining

Chuchu Han, Xizhou Chen, Jin Wang, Changxin Gao and Nong Sang  
Key Laboratory of Ministry of Education for Image Processing and Intelligent Control,  
School of Automation, Huazhong University of Science and Technology, Wuhan, China

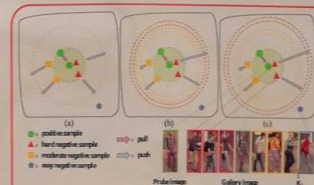
## Introduction

Person re-identification (re-ID) is considered as a retrieval process, and the result is presented as a ranking list. There always exists the phenomenon that true matches are not the first rank, mainly owing to that they are more similar to other persons. In this paper, we use an adaptive hard sample mining method to re-train the selected samples in order to distinguish similar persons, which is applied for re-ranking the re-ID results.

### Main contributions:

- We propose a re-ranking method for re-ID, the core concept is to use the hard mining method to re-train the model.
- We show our results on VIPeR, PRID450S and CUHK03, proving the effectiveness of the method.

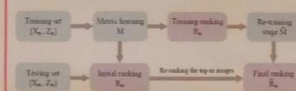
## Illustration of our motivation



The negative samples are divided into three levels. Different margins are assigned according to the levels. Moreover, we inflict additional punishment on the wrong ranked samples, making it more discriminative for confusable individuals.

## Algorithm Framework

**Algorithm 1:** Metric Learning with adaptive hard sample mining  
Input: Training set:  $\{X, Z, ID\}$ , stepsize  $\alpha$ , convergence condition  $\epsilon$ , parameters  $\alpha, \lambda, \beta_1$  and  $\beta_2$   
Output: The metric  $M$   
Initialize  $M_0 = I, i = 0$   
After cross-validation, obtain the re-training set  $\{X, Z\}$ .  
Based on coarse-fine tuning mechanism, calculate the loss  $L(M_0)$  and  $L(M_{i+1})$ .  
while  $|L(M_{i+1}) - L(M_i)| > \epsilon$  do  
    Calculate the gradient  $\nabla L(M_i)$ .  
    According to  $M_{i+1} = U_i \nabla L(M_i) U_i^T$ , project  $M_{i+1}$  onto the PSD cone, obtaining  $\tilde{M}_{i+1}$ .  
    Update  $M_{i+1}$  according to  $M_{i+1} = M_i - \alpha_i \nabla L(M_i)$ .  
end



## Association Procedure

- Hard and moderate negative samples:

$$L_{hard}(x_i, x_j) = \{x_i, x_j | d(x_i, x_j) < \tau\}$$
$$L_{moderate}(x_i, x_j) = \{x_i, x_j | d(x_i, x_j) < \tau_i\}$$

- Pairwise constraint:

$$D_{ij}(x_i, x_j) \leq \tau, (x_i, x_j) \in \mathcal{D}$$
$$D_{ij}(x_i, x_j) \geq \tau_j, (x_i, x_j) \in \mathcal{D}$$

- Coarse-fine tuning mechanism:

$$\mu_j^i = \begin{cases} d + \beta_1 - \frac{\tau_j - \tau_i}{\tau_j - \tau_i} \cdot \tau_j \in L_{moderate}(x_i) \\ d - \beta_1 - \frac{\tau_j - \tau_i}{\tau_j - \tau_i} \cdot \tau_j \in L_{moderate}(x_i) \end{cases}$$
$$d = \frac{1}{|N| - 1} \sum_{i=1}^N |x_i - x_j|$$

- Overall loss function:

$$L(M) = \frac{1}{N} \sum_{i=1}^N (D_{ij}(x_i, x_j) - \tau_j)^2 + \frac{1}{|N|} \sum_{(x_i, x_j) \in \mathcal{D}} (D_{ij}(x_i, x_j) - \mu_j^i)^2 + \frac{1}{2} (M - I)$$

## Experiment Results

Comparison among various methods with our re-ranking approach on the PRID450S dataset.

Method	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
LOMO+VQA	26.05	70.38	70.38	60.53	62.27
LOMO+VQA+ours	38.58	79.72	79.72	69.53	62.27
LOMO+K-SIM	46.51	79.72	79.72	70.79	71.27
LOMO+K-SIM+ours	54.13	80.40	80.40	70.89	71.40
FOCC+VQA	56.49	79.72	79.72	70.44	70.44
FOCC+VQA+ours	67.82	79.72	79.72	64.53	64.53
FOCC+K-SIM	72.36	82.67	79.72	70.38	70.38
FOCC+K-SIM+ours	81.93	84.42	79.72	70.40	70.40

Comparison among various methods with our re-ranking method, and with another re-ranking approach on the CUHK03 dataset.

Method	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
CUHK03 (baseline)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00
CUHK03 (ours)	14.17	27.50	37.50	45.00	45.00

## Conclusions

In this paper, we use a re-trained manner to address the re-ranking problem in person re-identification (re-ID). In order to distinguish some similar samples, we propose a coarse-fine tuning mechanism, motivated by hard sample mining method, which can adaptively assign the margins of different negative sample pairs. Under this constraint an effective metric model is obtained, we calculate the similarity score for re-ranking. Meanwhile, the strategy of selecting re-ranking samples can alleviate computational complexity. The proposed method achieve effective improvement on the VIPeR, PRID450S and CUHK03 datasets.





# Feature Fusion and Ellipse Segmentation for Person Re-identification

Meibin Qi, Junxian Zeng, Jianguo Jiang, and Cuiqun Chen  
School of Computer and Information, Hefei University of Technology

## 1 Introduction

Person re-identification matches persons across non-overlapping camera views at different time. It is applied to criminal investigation, pedestrian search, and multi-camera pedestrian tracking, etc. And person re-identification plays a crucial role in the field of video surveillance. Actually, the pedestrian images come from different cameras, and the appearance of pedestrians will change greatly when the lighting, background and visual angle vary. In order to solve the above problems, many of the previous works mainly focus on two aspects: extracting features from images and measuring the similarity between images. Our contributions can be summarized as follows:

- (1) We propose an effective feature representation that uses the fusion of LOMO and GOG features as the global feature and then combine the global and local features to form the final feature.
- (2) We present a new and simple segmentation method called ellipse segmentation, which can effectively reduce the impact of background interference.
- (3) We operate in-depth experiments to analyze various aspects of our approach, and the final results outperform the state-of-the-art over three benchmarks.

## 2 Methods

This paper uses the ellipse segmentation and extracts the LOMO and GOG features from the segmented images, then fuses them as global feature, then combines the local features proposed in SCSP to form the final feature. In terms of metric learning, this paper uses the metric function combining the bilinear similarity metric and the Mahalanobis distance, and finally adopts the ADMM (Alternating Direction Method of Multipliers) optimization algorithm to obtain the optimal metric matrix.

### 2.1 Ellipse Segmentation

Because pedestrians are generally in the center of the rectangular box, and the four right-angled areas of the rectangular box are basically background information. In order to tackle this problem, this paper proposes a new segmentation method called ellipse segmentation. It can preserve the effective information of pedestrians and reduce the impact of background interference. The specific segmentation method is shown in Fig. 1.

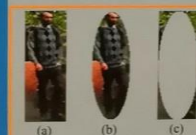


Fig. 1: Ellipse segmentation of image.  
(a)Original image; contains all the information for the entire image. (b) Ellipse area: retains valid pedestrian information after ellipse splitting and contains a small amount of background information. (c)Background area: contains background information and a small amount of pedestrian information.



Fig. 2: LOMO feature composition: LOMO(a+b+c).  
(a)We extract the LOMO(a) feature from the whole picture. (b) We extract the LOMO(b) feature from the elliptical area. (c) We extract the LOMO(c) feature from the elliptical area.

### 2.2 Feature Extraction and Fusion

#### Extracting LOMO feature.

We perform the ellipse segmentation operation on the image and then extract the LOMO feature, and denote it as LOMO(b), as shown in Fig. 2. we also extract the LOMO feature from the original image to supplement the information, denote it as LOMO(a), as well as the improved mean LOMO (LOMO mean) to reduce the background noise in the elliptical region, which is denoted as LOMO(c). we combine three LOMO features as LOMO(a+b+c).

#### Extracting GOG feature.

We extract the GOG feature from the whole image as GOG(a) to ensure the integrity of the information. Simultaneously, we also extract the GOG feature from the ellipse region as GOG(b). we combine two features as GOG(a+b).

## 3 Result

Three widely used datasets are selected for experiments, including VIPeR, PRID450s and CUHK01. Finally, we take the average results of the 10 experiments.

### Results on VIPeR.

From the results in Tab.1, we can conclude that our algorithm, based on SCSP, has significantly improved the matching rates in comparison with other algorithms. The recognition rate is 9% higher than SCSP on Rank1. At the same time, Rank5, Rank10 and Rank20 have been improved. The Tab.1 shows that our method has stronger expression ability and better recognition effect.

Table 1: Matching rates (%) of different methods on VIPeR.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
LOMO+XQDA	40.00	68.13	80.51	91.08
S-SVM	42.66	-	84.27	91.93
SSDAL	43.50	71.80	81.50	89.00
ME	45.89	77.40	88.87	95.84
LRP	49.05	74.08	84.43	93.10
NFST	51.17	82.09	90.51	95.92
SCSP	53.54	82.59	91.49	96.65
Ours	62.56	87.53	93.89	97.97

### Results on PRID450s.

From the experimental data in Tab.2, we can see that the algorithm in the PRID450s dataset has the highest recognition rate over the state-of-the-art methods. The best Rank1 identification rate of comparison methods is 68.47%, while we has achieved 73.29%, with an improvement by nearly 5%.

Table 2: Matching rates (%) of different methods on PRID450s.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
KISSME	33.0	59.8	71.0	79.0
SCNCD	41.6	68.9	79.4	87.8
DRML	56.4	-	82.2	90.2
LSSCDL	60.5	-	88.6	93.6
LOMO+XQDA	62.60	85.60	92.00	96.60
FFN	66.6	86.8	92.8	96.9
GOG	68.47	88.80	94.50	97.80
Ours	73.29	91.78	95.11	97.73

### Results on CUHK01.

Tab.3 shows the recognition rates of the proposed algorithm and the existing algorithm on the CUHK01 dataset. It can be seen that the algorithm still has significant improvements in the recognition rates in comparison with the existing algorithms on large datasets. Compared with LRP(Local Region Partition), the algorithm of this paper improves about 6% on Rank1. Moreover, our method is 9% higher than the GOG.

Table 3: Matching rates (%) of different methods on CUHK01.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
KISSME	17.9	42.4	55.9	69.1
kLFD	29.1	55.2	66.4	77.3
Semantic	31.5	52.5	65.8	77.6
FFN	55.5	78.4	83.7	92.6
LOMO+XQDA	63.2	-	90.8	94.9
GOG	67.3	86.9	91.8	95.9
LRP	70.45	87.92	92.67	96.34
Ours	76.19	92.34	95.58	98.09

## 4 Conclusions

In this paper, the proposed method fuses LOMO(a+b+c) and GOG(a+b) features as the global feature, and combines them with local features, thus forming more robust feature for the changes of illumination and visual angle. Meanwhile, the algorithm of ellipse segmentation reduces background noise. Furthermore, it can increase the proportion of effective area for pedestrians and enhance the robustness of final joint features. Experimental results show that the proposed algorithm significantly improves the recognition rate of pedestrian re-identification. The recognition rate on Rank10 in the VIPeR, PRID450s, and CUHK01 datasets all reach over 90%, which has practical application of great value.

THANKS