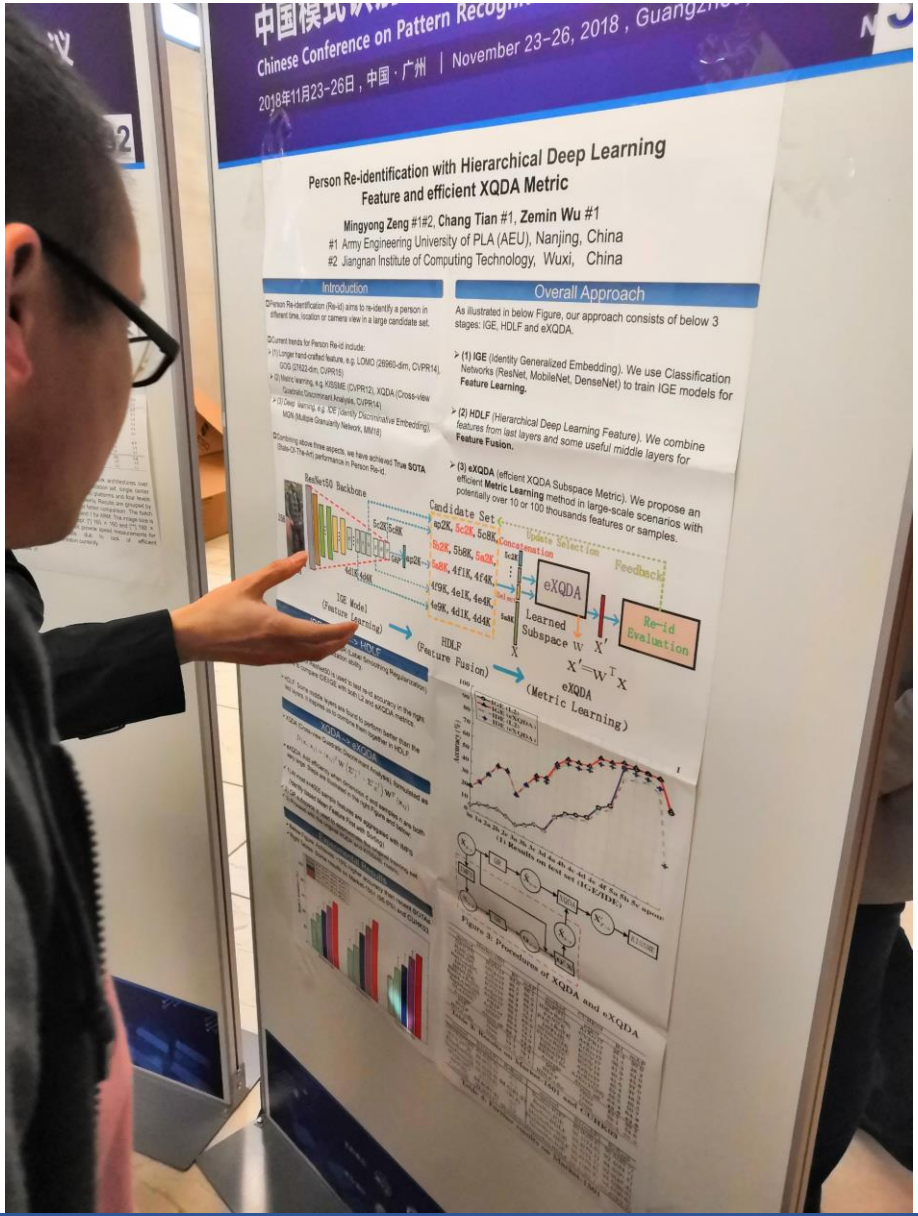
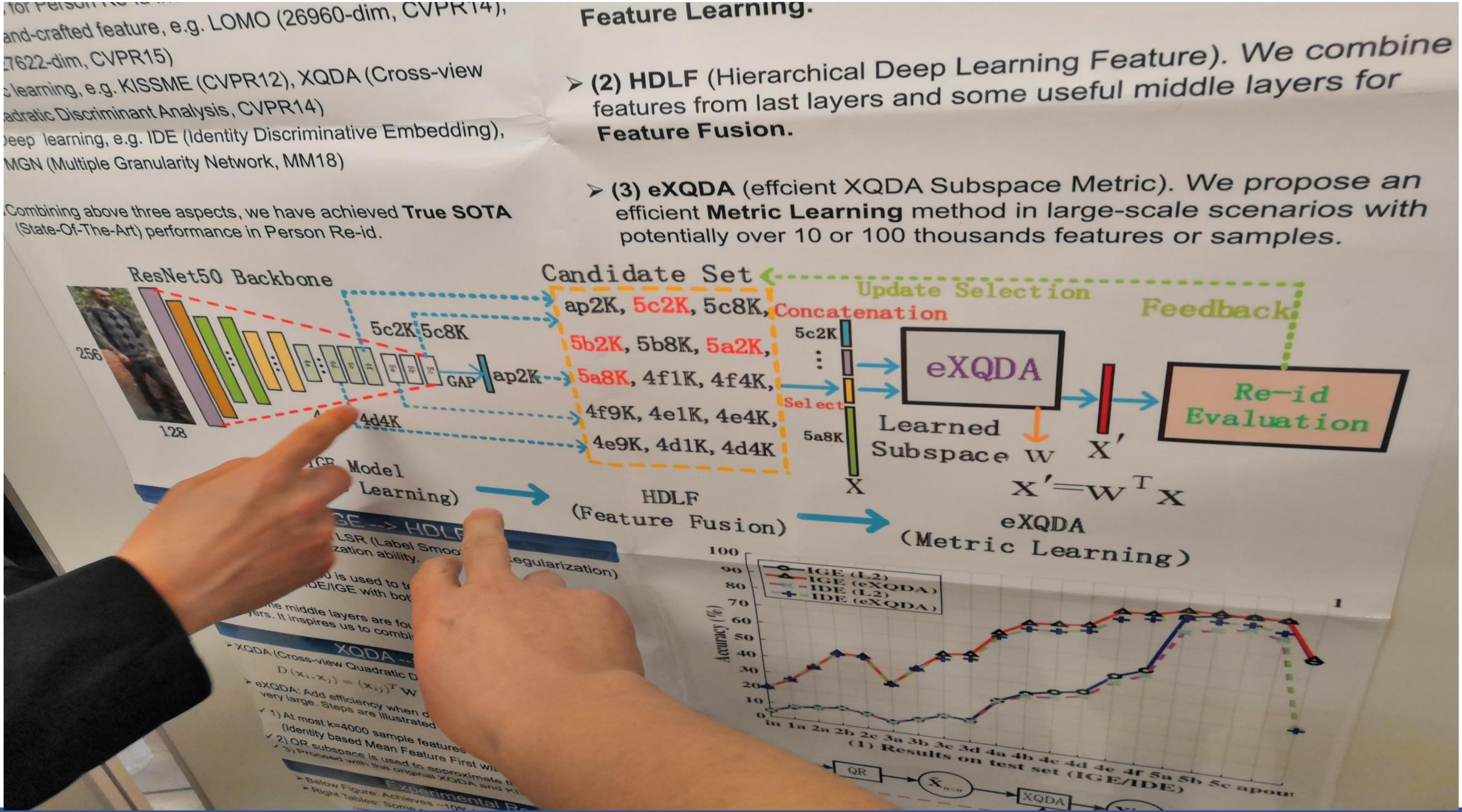
PRCV 相关poster

Intelligent Information Fusion Research Group

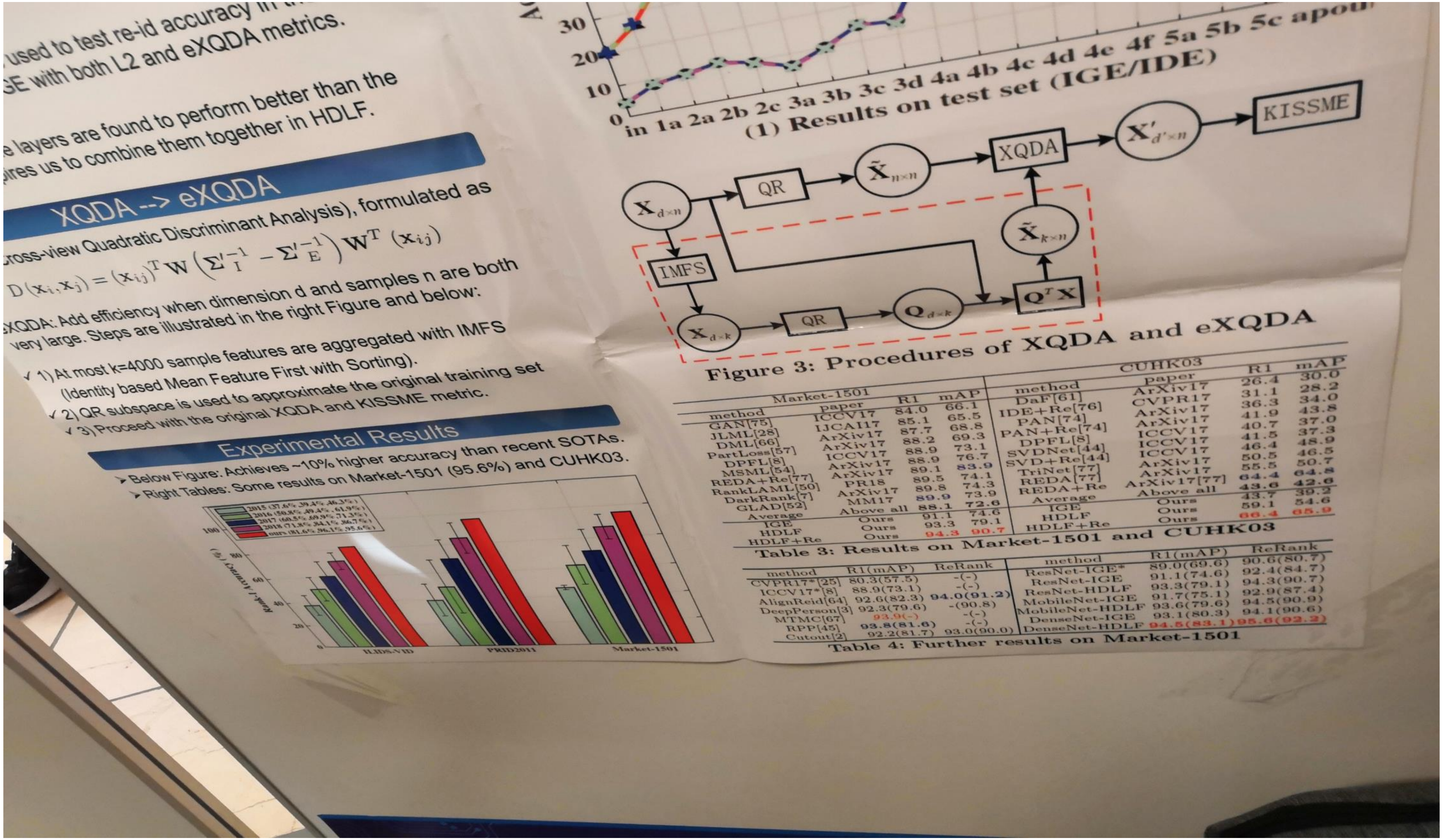
Person Re-identification with Hierarchical Deep Learning Feature and efficient XQDA Metric

MM 2018





and-crafted feature, e.g. LOMO (26960-dim, CVPR14),
7622-dim, CVPR15)
learning, e.g. KISSME (CVPR12), XQDA (Cross-view
adratic Discriminant Analysis, CVPR14)
eep 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 (effcient XQDA Subspace Metric). We propose an efficient Metric Learning method in large-scale scenarios with potentially over 10 or 100 thousands features or samples.
ResNet50 Backbone
256
128
GAP
ap2K
5c2K
5c8K
4d4K
Candidate Set
ap2K, 5c2K, 5c8K,
5b2K, 5b8K, 5a2K,
5a8K, 4f1K, 4f4K,
4f9K, 4e1K, 4e4K,
4e9K, 4d1K, 4d4K
Concatenation
Select
5c2K
5a8K
X
Update Selection
Feedback
eXQDA
Learned Subspace W
X'
Re-id Evaluation
X'=W^T X
Model
Learning)
HDLF
(Feature Fusion)
eXQDA
(Metric Learning)
IGE (L2)
IGE (eXQDA)
IDE (L2)
IDE (eXQDA)
Accuracy (%)
(1) Results on test set (IGE/IDE)
QR
XQDA



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

e layers are found to perform better than the
pires us to combine them together in HDLF.

XQDA --> eXQDA

Cross-view Quadratic Discriminant Analysis), formulated as

$$D(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_{ij})^T \mathbf{W} \left(\Sigma'^{-1}_I - \Sigma'^{-1}_E \right) \mathbf{W}^T (\mathbf{x}_{ij})$$

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.

(1) Results on test set (IGE/IDE)

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

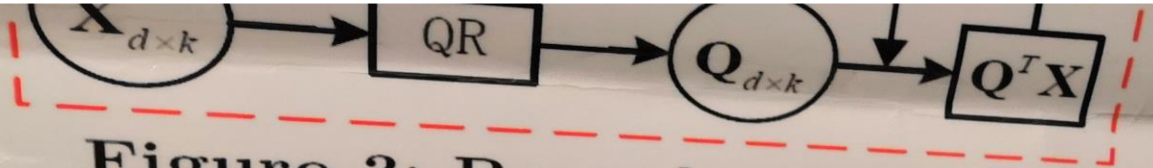


Figure 3: Procedures of XQDA and eXQDA

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

Mask-guided Contrastive Attention Model for Person Re-Identification

CVPR 2018

CVPR 2018

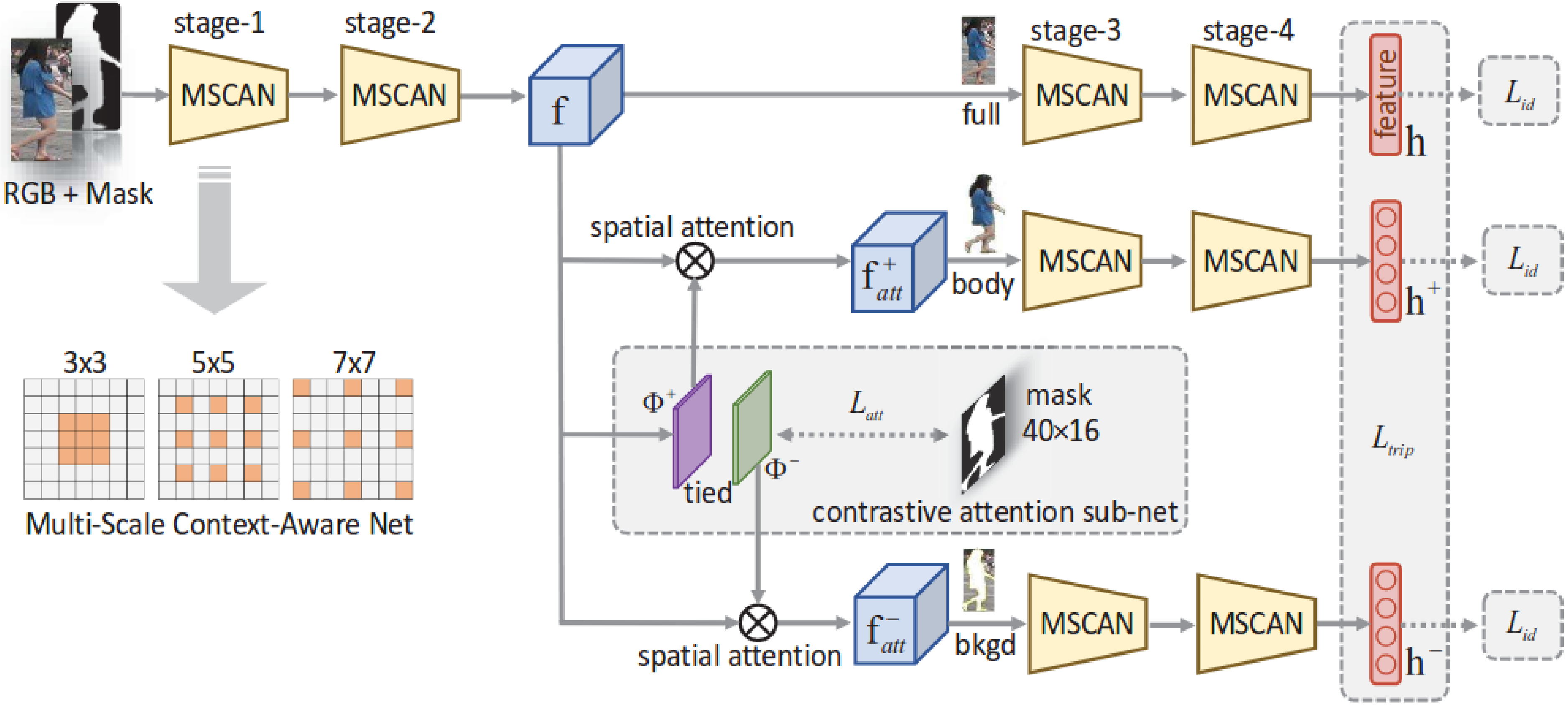
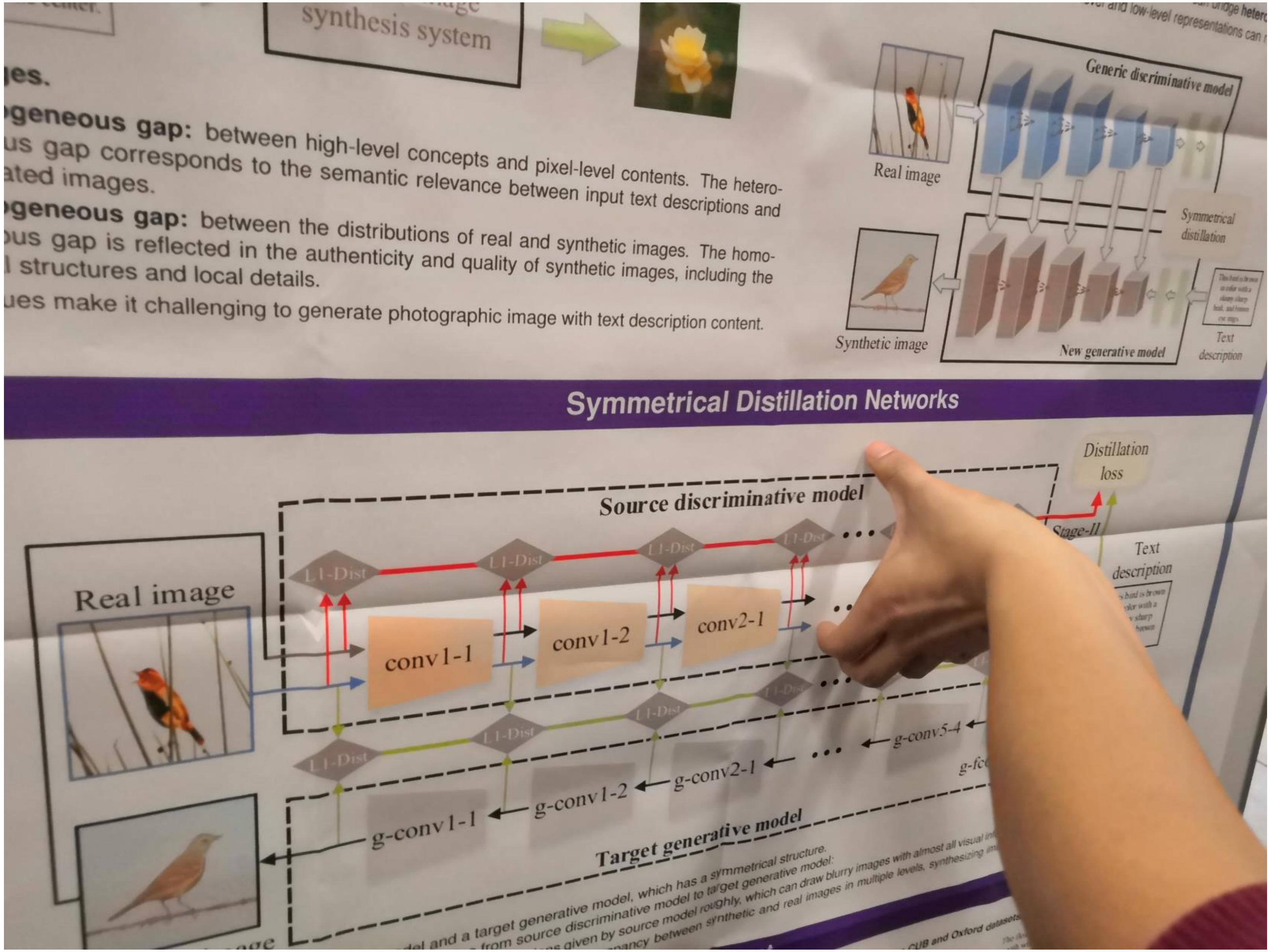


Figure 2. Framework of proposed Mask-guided Contrastive Attention Model (MGCAM) for person ReID. It contains four multi-scale context-aware stages and a fully-connected layer to learn final features. There are three main streams, i.e., the full-stream, the body-stream and the background-stream. In the 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



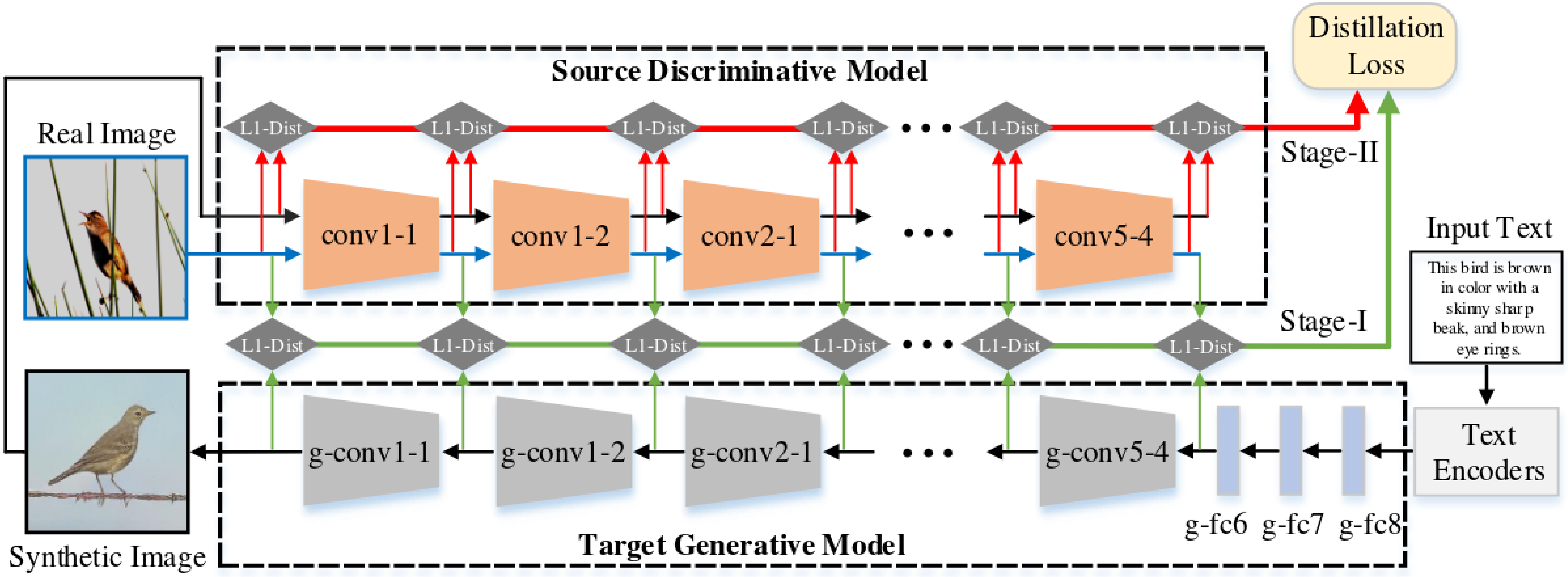


Figure 2: The architecture of proposed Symmetrical Distillation Networks (SDN), which consists of a source discriminative model and a target generative model. The source model receives images as input and produces multi-level representations as guidance for the training of target model. The target model 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.

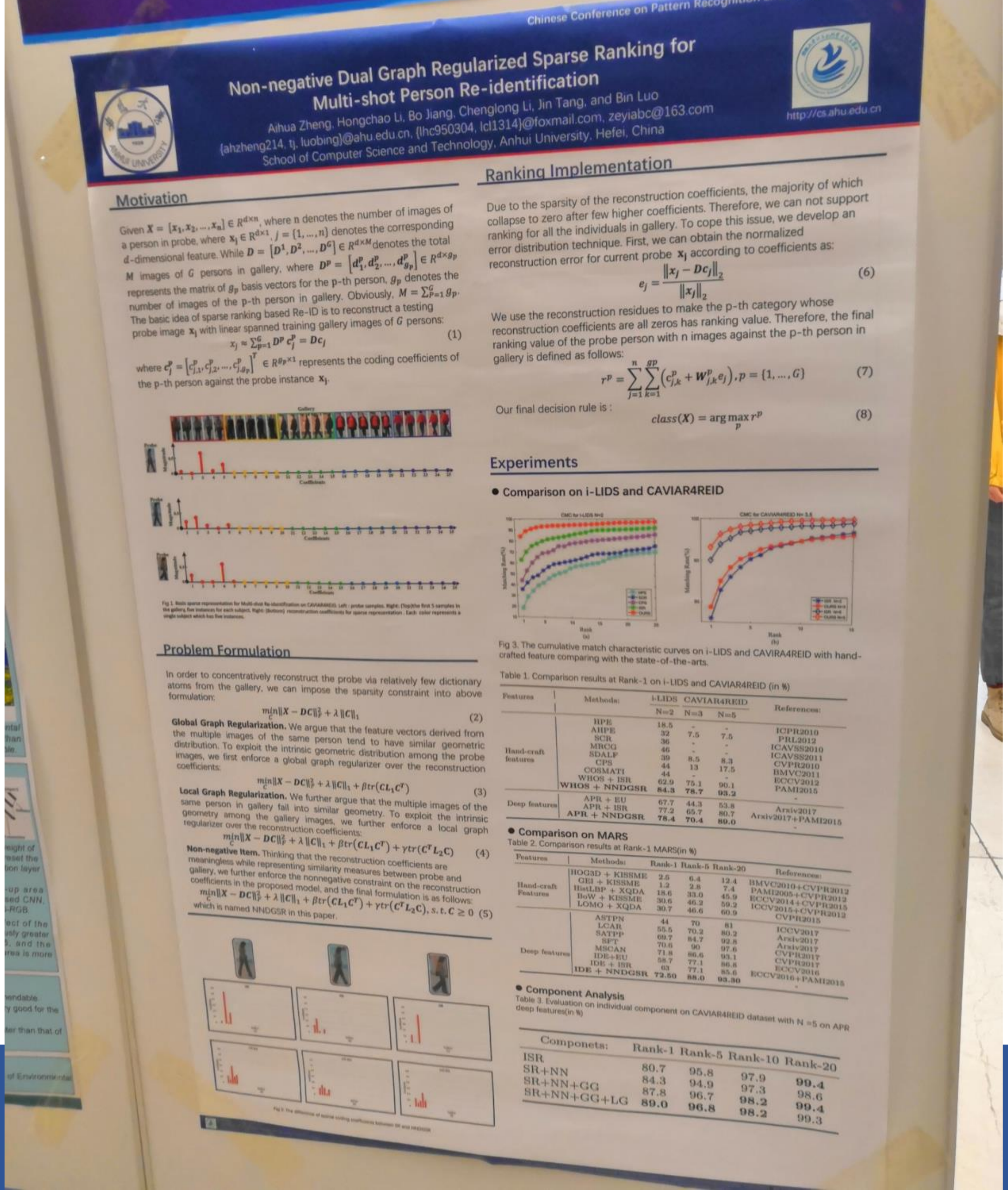
Table 1: Inception, SSIM and FSIM scores of our SDN and compared methods. Higher scores mean better results.

Datasets	Methods	Inception	SSIM	FSIM
CUB-200-2011	**our SDN**	**6.89 ± 0.06**	**0.3160**	**0.6264**
	StackGAN	4.95 ± 0.04	0.2812	0.5869
	GAWWN	5.22 ± 0.08	0.2370	0.5653
	GAN-INT-CLS	5.08 ± 0.08	0.2934	0.6082
Oxford-Flower-102	**our SDN**	**4.28 ± 0.09**	**0.2174**	**0.6227**
	StackGAN	3.54 ± 0.07	0.1837	0.6009
	GAN-INT-CLS	4.17 ± 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。



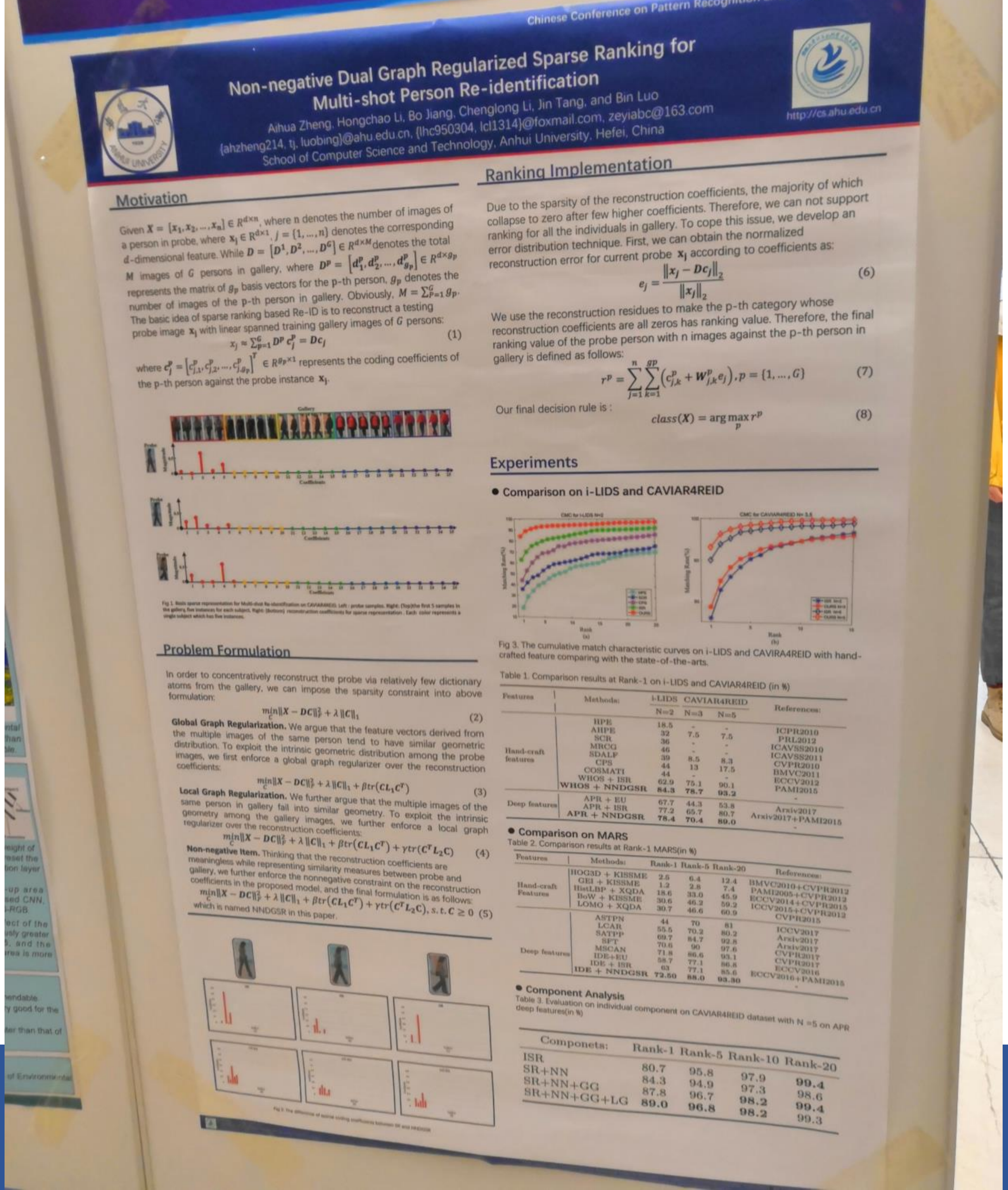
Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

被推荐到IEEE ACCES

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

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

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

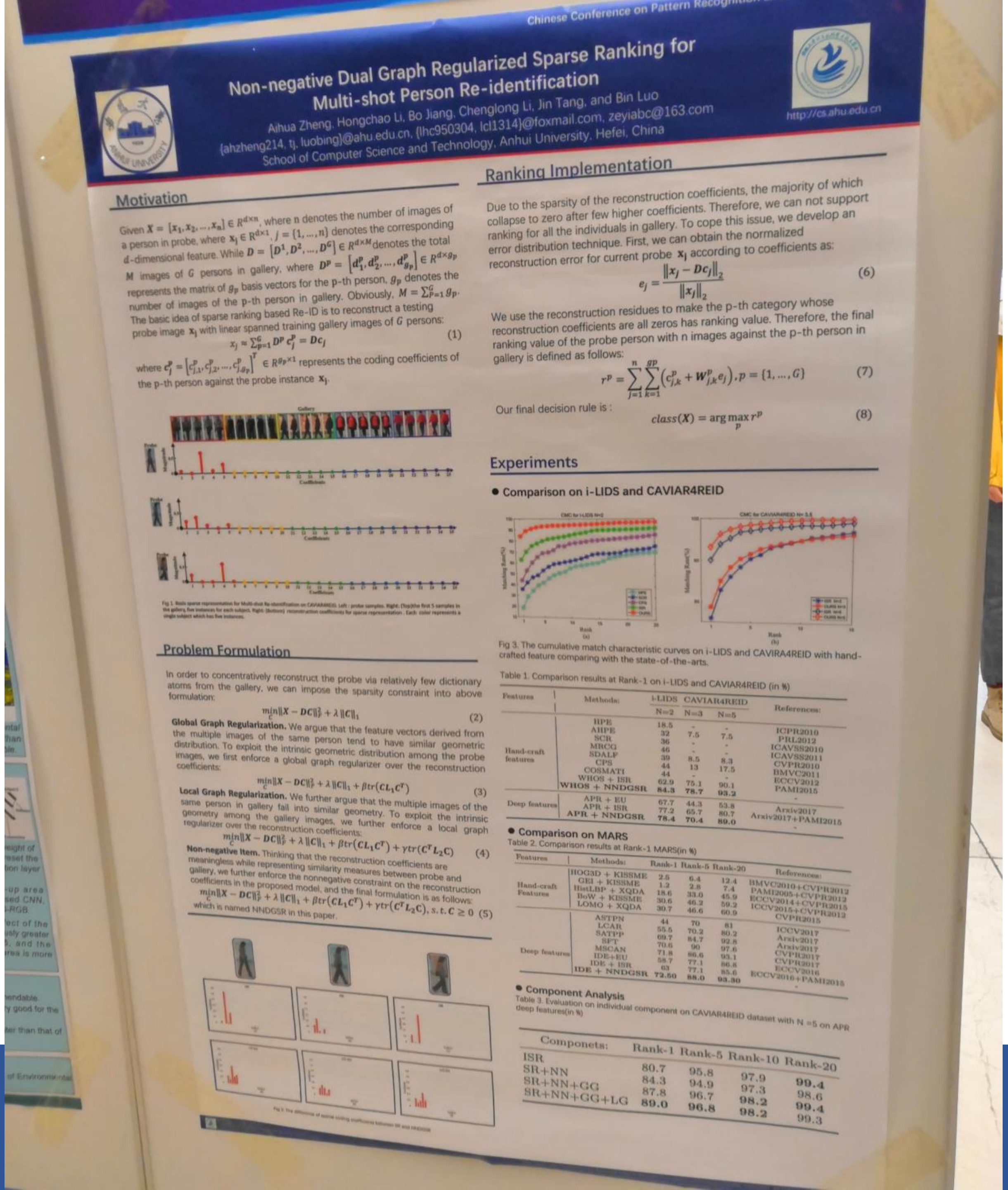


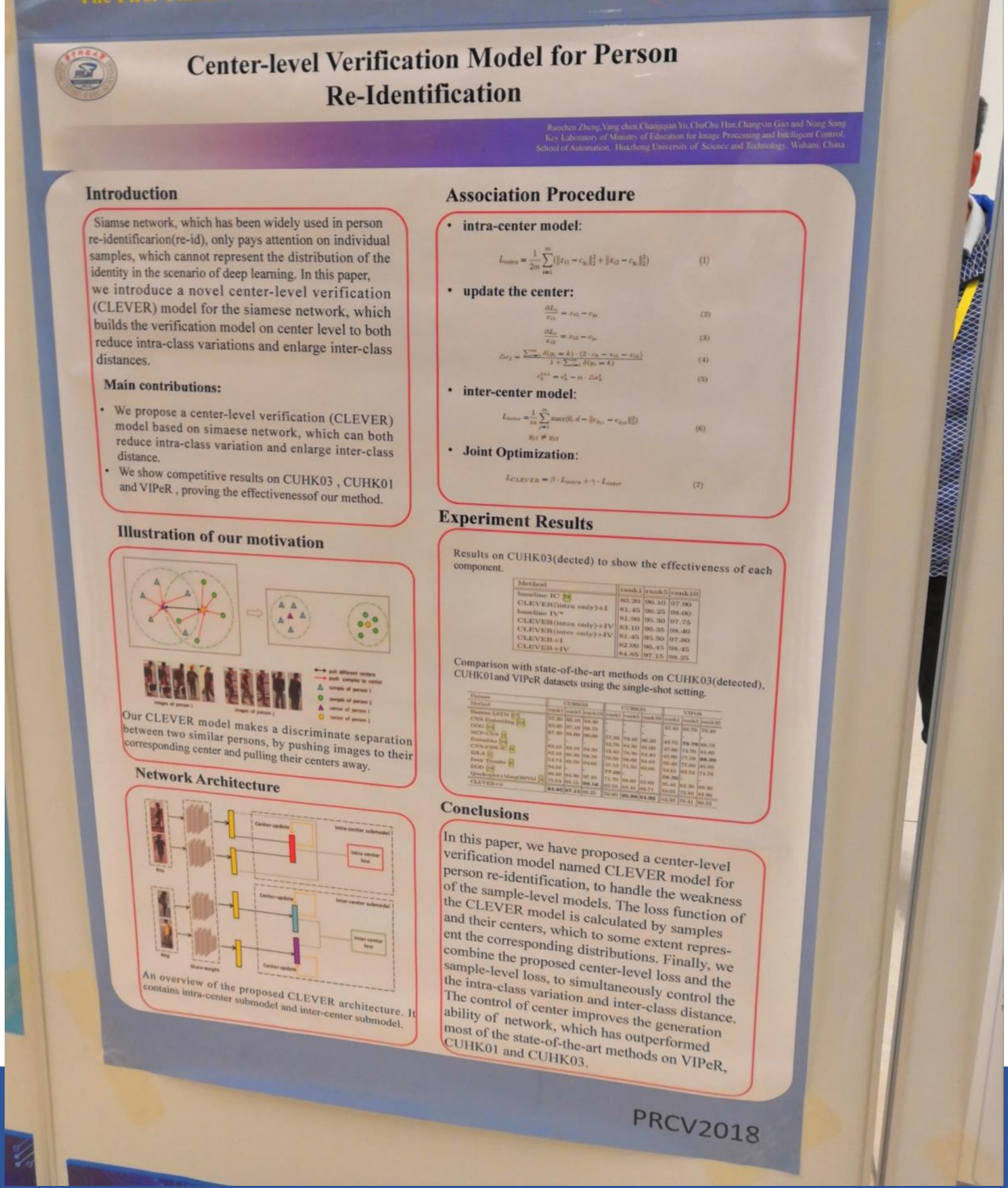
Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

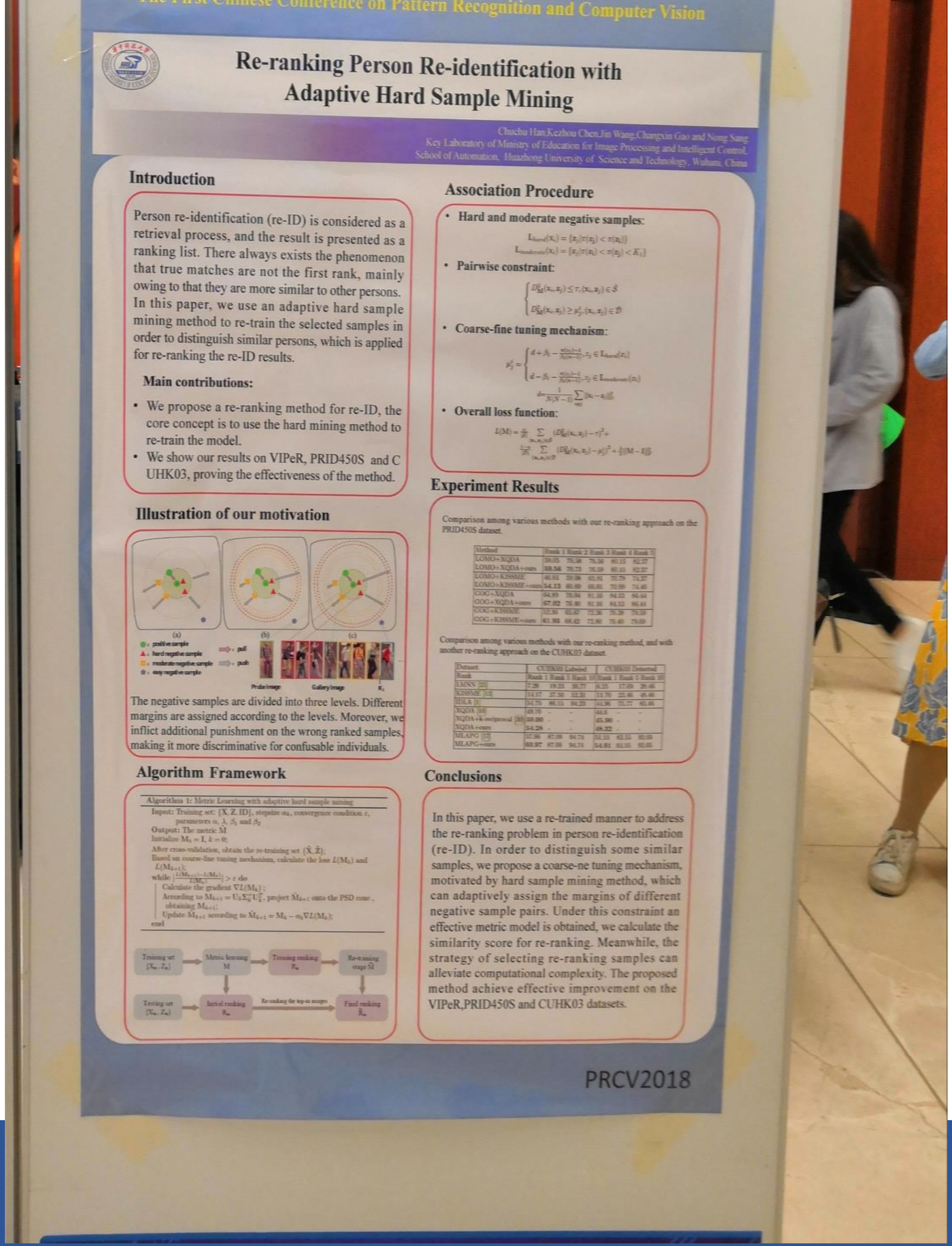
被推荐到IEEE ACCES

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

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




Center-level Verification Model for Person Re-Identification
Ruochen Zheng, Yang chen, Changqian Yu, ChuChu Han, 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
Siamse network, which has been widely used in person re-identificarion(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 simaese network, which can both reduce intra-class variation and enlarge inter-class distance.
We show competitive results on CUHK03 , CUHK01 and VIPeR , proving the effectivenessof our method.
Illustration of our motivation
Our CLEVER model makes a discriminate separation between two similar persons, by pushing images to their corresponding center and pulling their centers away.
Network Architecture
An overview of the proposed CLEVER architecture. It contains intra-center submodel and inter-center submodel.
Association Procedure
intra-center model:
update the center:
inter-center model:
Joint Optimization:
Experiment Results
Results on CUHK03(dected) to show the effectiveness of each component.
Comparison with state-of-the-art methods on CUHK03(detected), CUHK01and VIPeR datasets using the single-shot setting.
Conclusions
In this paper, we have proposed a center-level verification model named CLEVER model for person re-identification, to handle the weakness of 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 sample-level loss, to simultaneously control the 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.
PRCV2018


The First Chinese Conference on Pattern Recognition and Computer Vision
Re-ranking Person Re-identification with Adaptive Hard Sample Mining
Chuchu Han,Kezhou 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
Association Procedure
Hard and moderate negative samples:
Pairwise constraint:
Coarse-fine tuning mechanism:
Overall loss function:
Experiment Results
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-ne 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.
PRCV2018

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.

(a) (b) (c)

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.

(a) (b) (c)

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, RID450s 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
kLFDA	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**	**92.24**	**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