Deep Metric Learning: A Frontal Report

Hongwei Fan

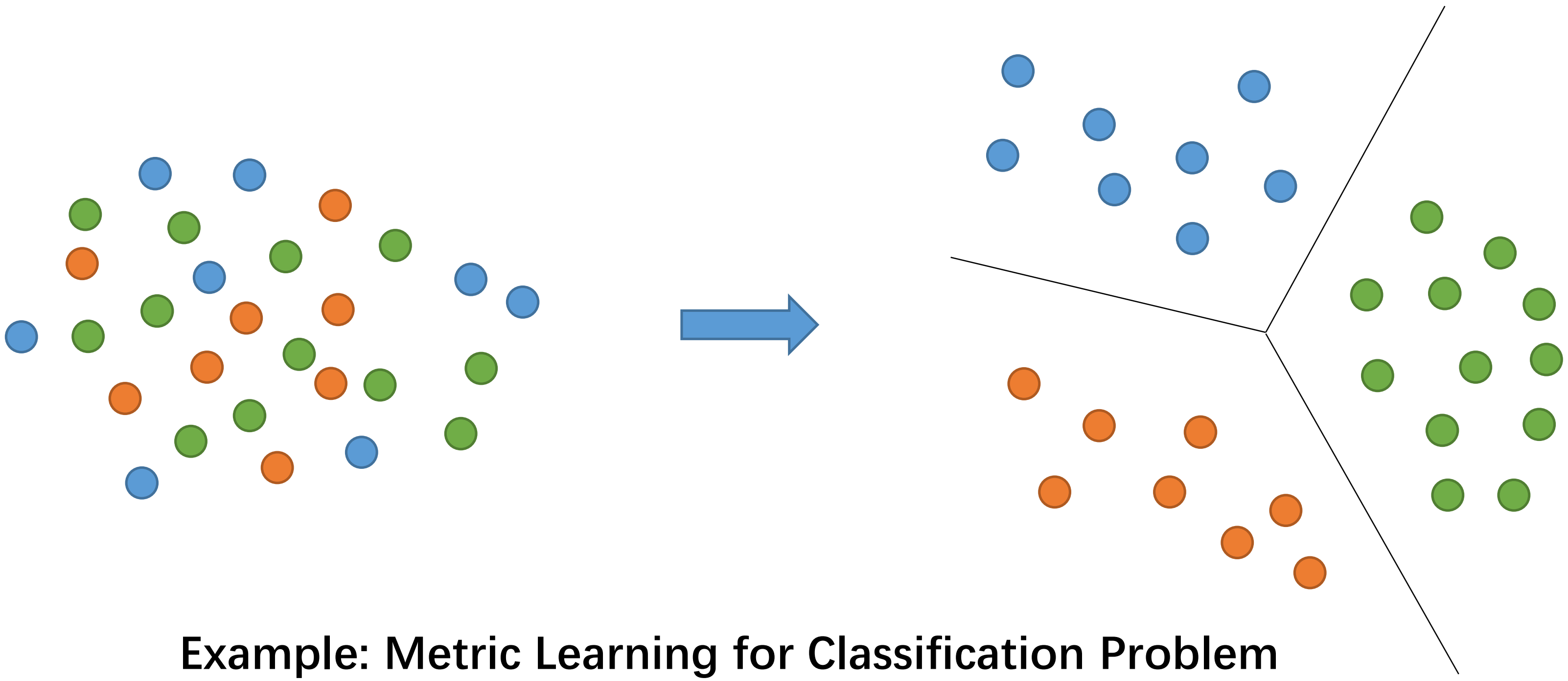
Lab of Pattern Recognition and Intelligent Systems, BUPT

About Me

- Hongwei Fan(范弘炜)
- Educational Background
  - 2015.9~2019.6: Bachelor degree of the School of Information and Communication Engineering, BUPT.
  - 2019.9~: Master candidate of the Lab of Pattern Recognition and Intelligent Systems, BUPT. (Under the instruction of Prof. Weihong Deng)
- Research Interests
  - Metric Learning(Face Recognition/**Person Re-identification**)
  - Transfer Learning(**Unsupervised Domain Adaptation**)

Metric Learning

- **Metric**: In mathematics, a metric or distance function is a function that defines a **distance between each pair of elements of a set**.



Example: Metric Learning for Classification Problem

AMC
AI ML Club

Why Metric Learning

- Fine-grained Recognition: **Need for space between classes**

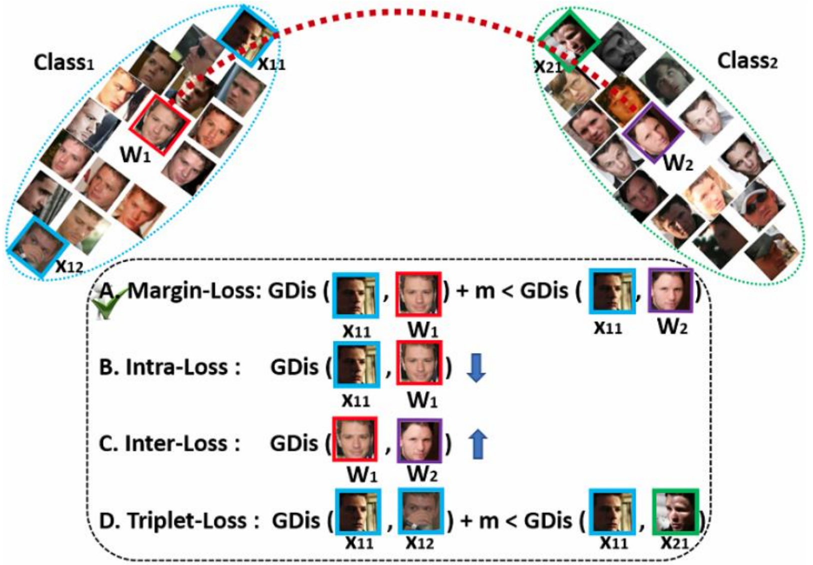
Different classes with high distance

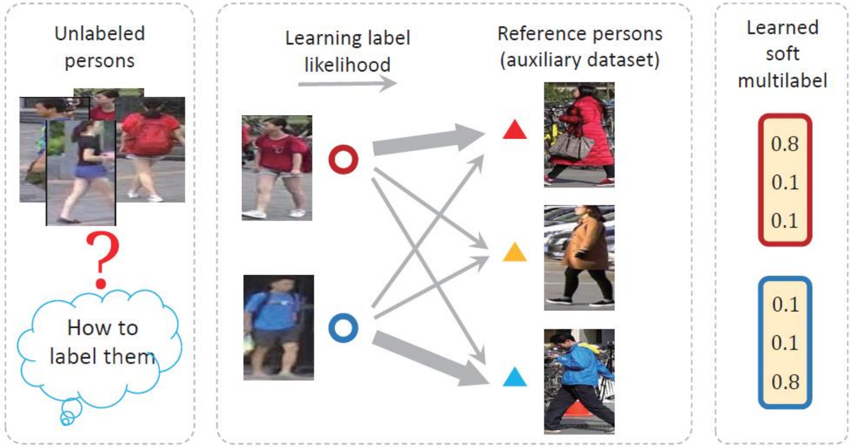
Different classes with low distance

Example: Object Classification vs. Face Recognition

Frontal of Metric Learning

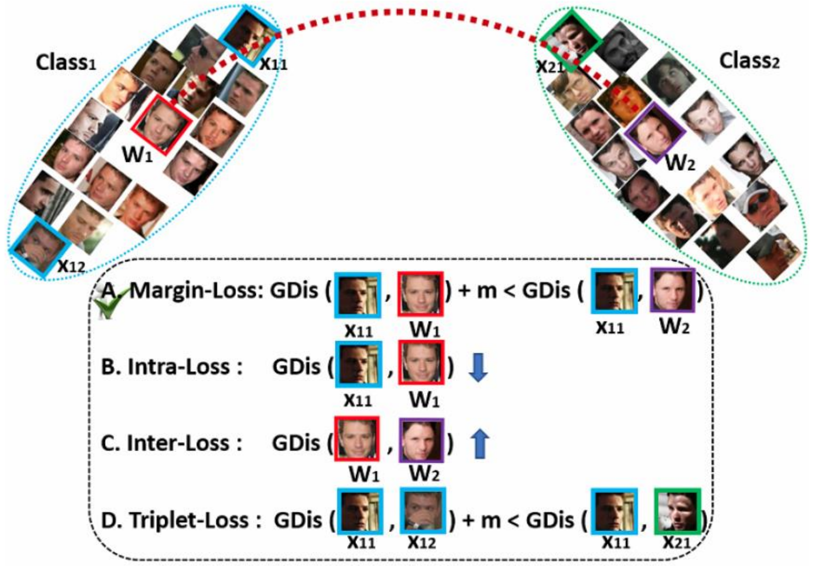
- **Supervised Metric Learning**
  - **Euclidean**: Triplet Loss[3]/Center Loss[4]
  - **Cosine**: SphereFace[5]/CosFace[6]/**ArcFace**
  - **ArcFace[1]: Additive Angular Margin Loss for Deep Face Recognition**
- **Unsupervised Metric Learning**
  - **Data Adaptation**: PTGAN[7]/StarGAN[8]/Camstyle[9]
  - **Pseudo-Label**: MAR[10]/IMAN[11]/**ECN**
  - **(ECN) Invariance Matters[2]: Exemplar Memory for Domain Adaptive Person Re-identification**

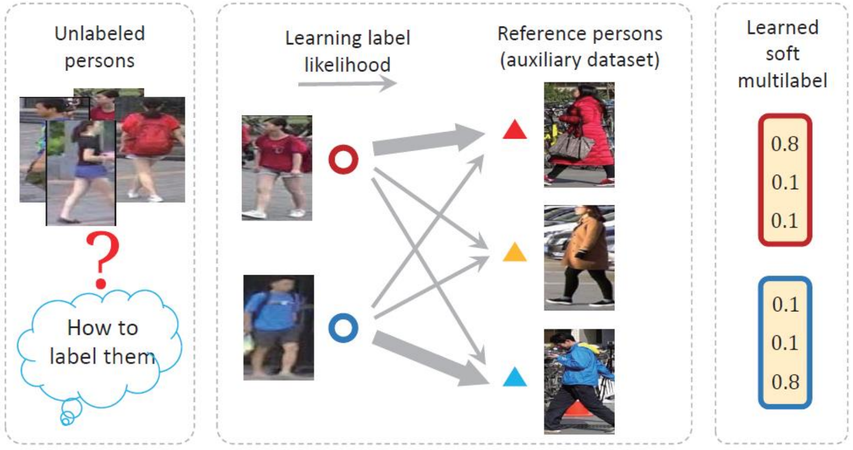




Frontal of Metric Learning

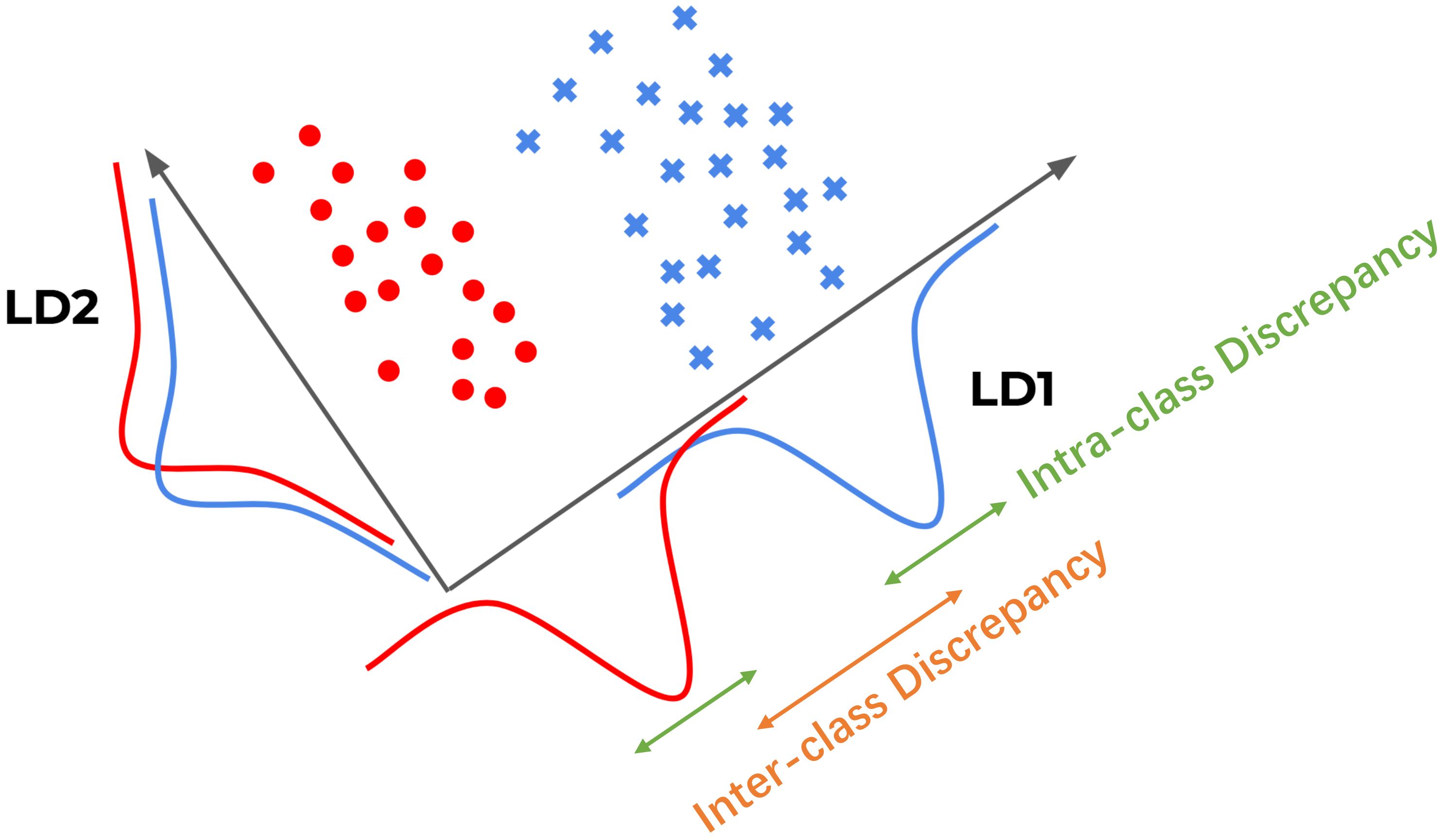
- **Supervised Metric Learning**
  - **Euclidean**: Triplet Loss[3]/Center Loss[4]
  - **Cosine**: SphereFace[5]/CosFace[6]/**ArcFace**
  - **ArcFace[1]: Additive Angular Margin Loss for Deep Face Recognition**
- **Unsupervised Metric Learning**
  - **Data Adaptation**: PTGAN[7]/StarGAN[8]/Camstyle[9]
  - **Pseudo-Label**: MAR[10]/IMAN[11]/**ECN**
  - **(ECN) Invariance Matters[2]: Exemplar Memory for Domain Adaptive Person Re-identification**





ArcFace: Motivation

- **Linear Discriminant Analysis**



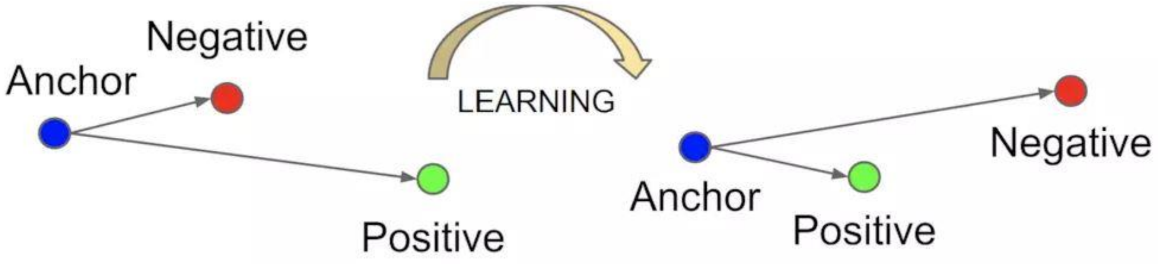
ArcFace: Motivation

- **Euclidean Method**

Constrastive Loss: $y_i d_i^2 + (1 - y_i)\max(margin - d_i, 0)^2$

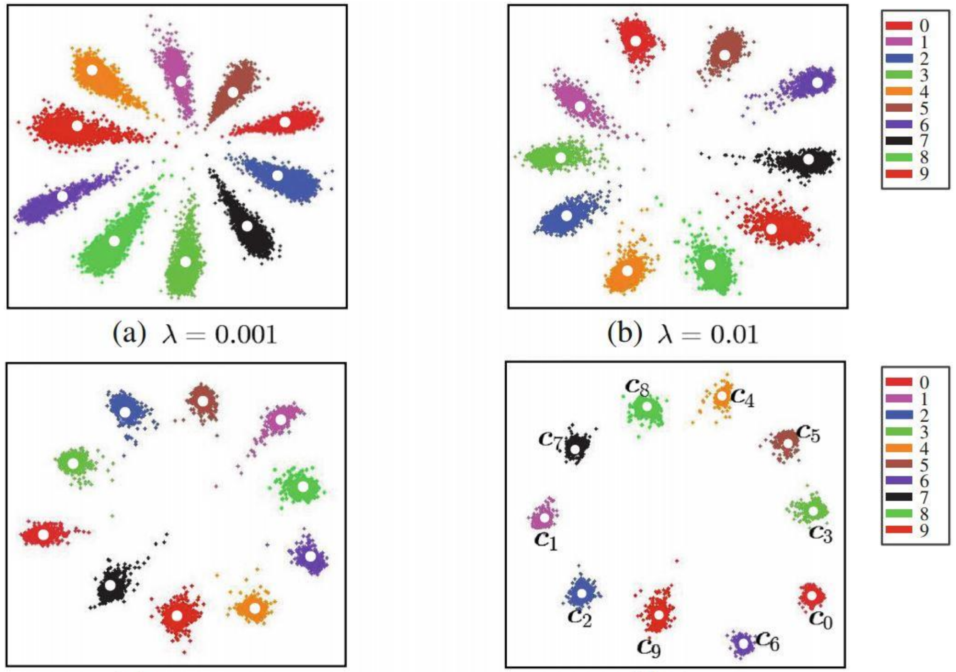
Positive Pairs **Negative Pairs**

Triplet Loss: $\max(d(a_i, p_i) - d(a_i, n_i) + margin, 0)$



Sample Center

Center Loss: $\left\|\boldsymbol{x_i} - \boldsymbol{c_{y_i}}\right\|^2$



(a) $\lambda = 0.001$ (b) $\lambda = 0.01$

(c) $\lambda = 0.1$ (d) $\lambda = 1$

ArcFace: Methodology

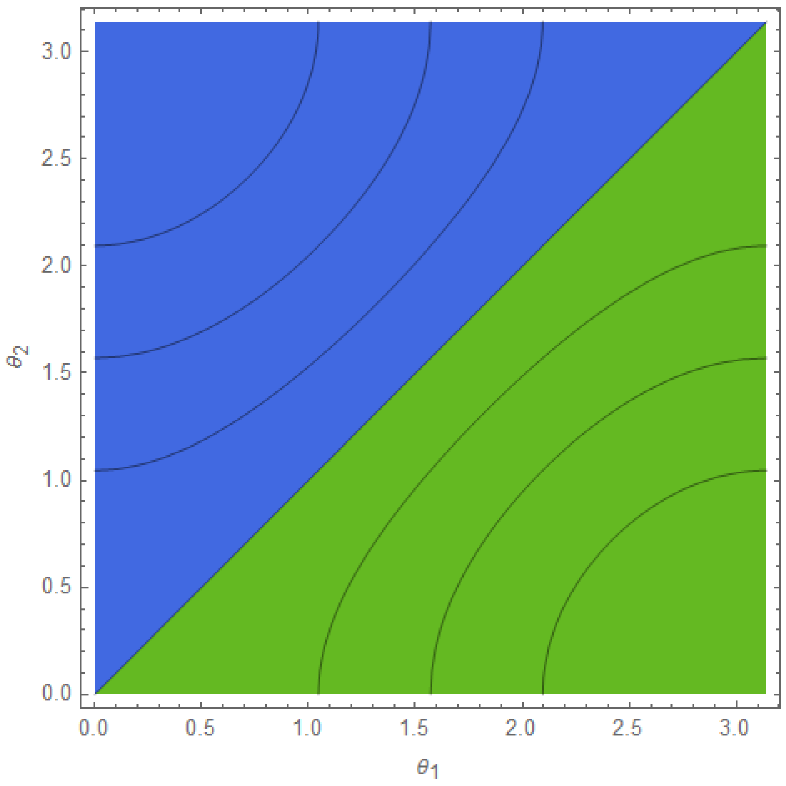
- ~~Euclidean~~ Cosine Method

$$p_i(x) = \frac{e^{w_i^T x + b_i}}{\sum_i e^{w_i^T x + b_i}}$$

$$b_i = 0$$
$$\|w\| = 1$$
$$\|x\| = 1$$

$$p_i(x) = \frac{e^{\cos\theta_i}}{\sum_i e^{\cos\theta_i}}$$

Softmax Interface

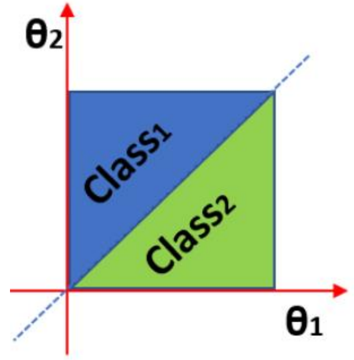


ArcFace: Methodology

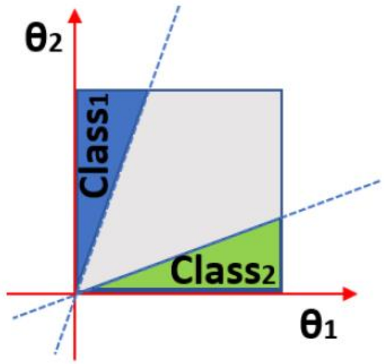
- **Advanced Cosine Classification Interface**

$$p_i(x) = \frac{e^{\cos\theta_i}}{\sum_i e^{\cos\theta_i}} \longrightarrow p_i(x) = \frac{e^{\varphi(\theta_i)}}{e^{\varphi(\theta_i)} + \sum_{j\neq i} e^{\cos\theta_j}}$$

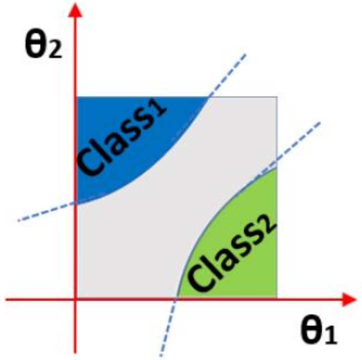
Loss Function	$\varphi(\theta)$	Binary Classification Interface
Softmax	$\cos\theta$	$\theta_1 - \theta_2 = 0$
Sphereface	$\cos(m\theta)$	$\cos(m\theta_1) - \cos\theta_2 = 0$
Cosface	$\cos\theta - m$	$\cos\theta_1 - m - \cos\theta_2 = 0$
Arcface	$\cos(\theta + m)$	$\cos(\theta_1 + m) - \cos\theta_2 = 0$



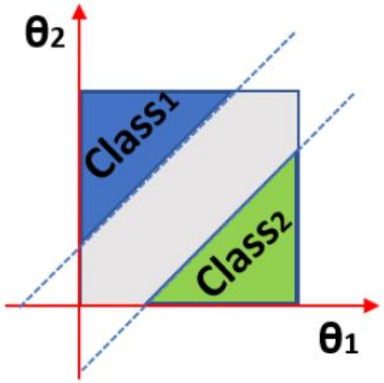
Softmax



SphereFace



CosFace



ArcFace

ArcFace: Experiments

AMC
AI ML Club

Method	#Image	LFW	YTF
DeepID [32]	0.2M	99.47	93.20
Deep Face [33]	4.4M	97.35	91.4
VGG Face [24]	2.6M	98.95	97.30
FaceNet [29]	200M	99.63	95.10
Baidu [16]	1.3M	99.13	-
Center Loss [38]	0.7M	99.28	94.9
Range Loss [46]	5M	99.52	93.70
Marginal Loss [9]	3.8M	99.48	95.98
SphereFace [18]	0.5M	99.42	95.0
SphereFace+ [17]	0.5M	99.47	-
CosFace [37]	5M	99.73	97.6
MS1MV2, R100, ArcFace	5.8M	**99.83**	**98.02**

Table 4. Verification performance (%) of different methods on LFW and YTF.

	NS	ArcFace	IntraL	InterL	TripletL
W-EC	44.26	14.29	8.83	46.85	-
W-Inter	69.66	71.61	31.34	75.66	-
Intra1	50.50	38.45	17.50	52.74	41.19
Inter1	59.23	65.83	24.07	62.40	50.23
Intra2	33.97	28.05	12.94	35.38	27.42
Inter2	65.60	66.55	26.28	67.90	55.94

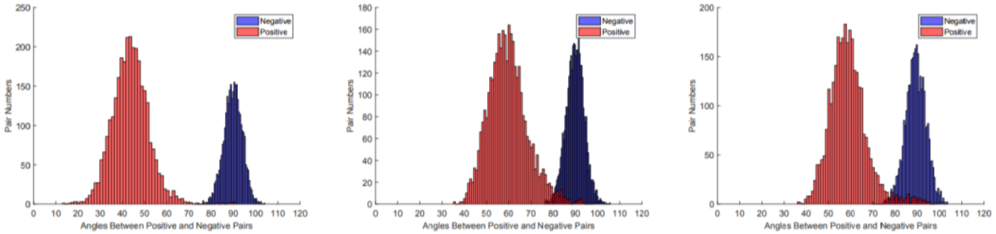
Table 3. The angle statistics under different losses ([CASIA, ResNet50, loss*]). Each column denotes one particular loss. "W-EC" refers to the mean of angles between W_j and the corresponding embedding feature centre. "W-Inter" refers to the mean of minimum angles between W_j's. "Intra1" and "Intra2" refer to the mean of angles between x_i and the embedding feature centre on CASIA and LFW, respectively. "Inter1" and "Inter2" refer to the mean of minimum angles between embedding feature centres on CASIA and LFW, respectively.

AMC
AI ML Club

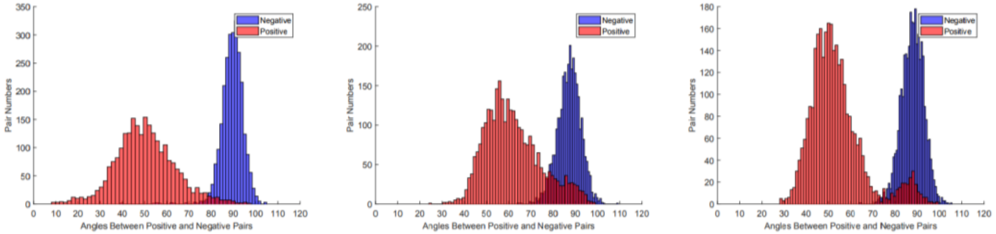
ArcFace: Experiments

Method	LFW	CALFW	CPLFW
HUMAN-Individual	97.27	82.32	81.21
HUMAN-Fusion	99.85	86.50	85.24
Center Loss [38]	98.75	85.48	77.48
SphereFace [18]	99.27	90.30	81.40
VGGFace2 [6]	99.43	90.57	84.00
MS1MV2, R100, ArcFace	**99.82**	**95.45**	**92.08**

Table 5. Verification performance (%) of open-sourced face recognition models on LFW, CALFW and CPLFW.



(a) LFW (99.83%) (b) CFP-FP (98.37%) (c) AgeDB (98.15%)



(d) YTF (98.02%) (e) CPLFW (92.08%) (f) CALFW (95.45%)

Methods	Id (%)	Ver (%)
Softmax [18]	54.85	65.92
Contrastive Loss[18, 32]	65.21	78.86
Triplet [18, 29]	64.79	78.32
Center Loss[38]	65.49	80.14
SphereFace [18]	72.729	85.561
CosFace [37]	77.11	89.88
AM-Softmax [35]	72.47	84.44
SphereFace+ [17]	73.03	-
CASIA, R50, ArcFace	77.50	92.34
CASIA, R50, ArcFace, R	91.75	93.69
FaceNet [29]	70.49	86.47
CosFace [37]	82.72	96.65
MS1MV2, R100, ArcFace	81.03	96.98
MS1MV2, R100, CosFace	80.56	96.56
MS1MV2, R100, ArcFace, R	98.35	98.48
MS1MV2, R100, CosFace, R	97.91	97.91

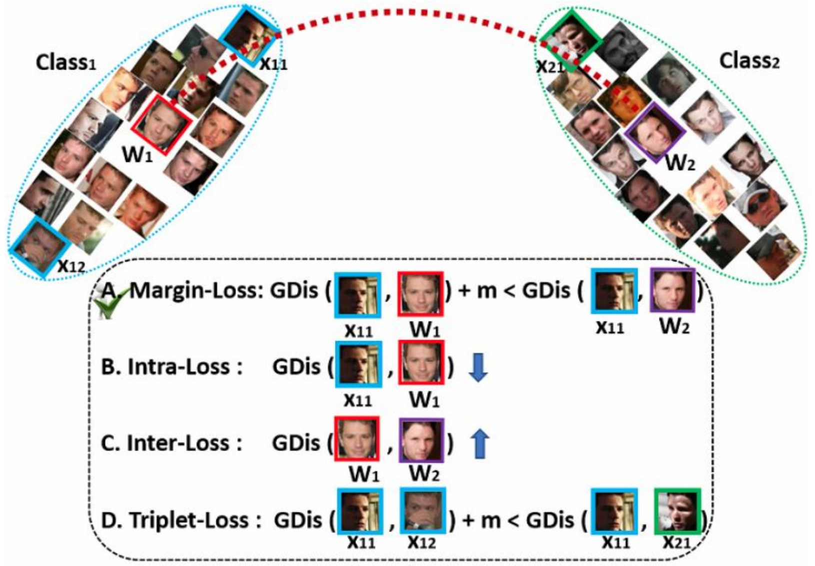
Table 6. Face identification and verification evaluation of different methods on MegaFace Challenge1 using FaceScrub as the probe set. "Id" refers to the rank-1 face identification accuracy with 1M distractors, and "Ver" refers to the face verification TAR at 10^{-6} FAR. "R" refers to data refinement on both probe set and 1M distractors. ArcFace obtains state-of-the-art performance under both small and large protocols.

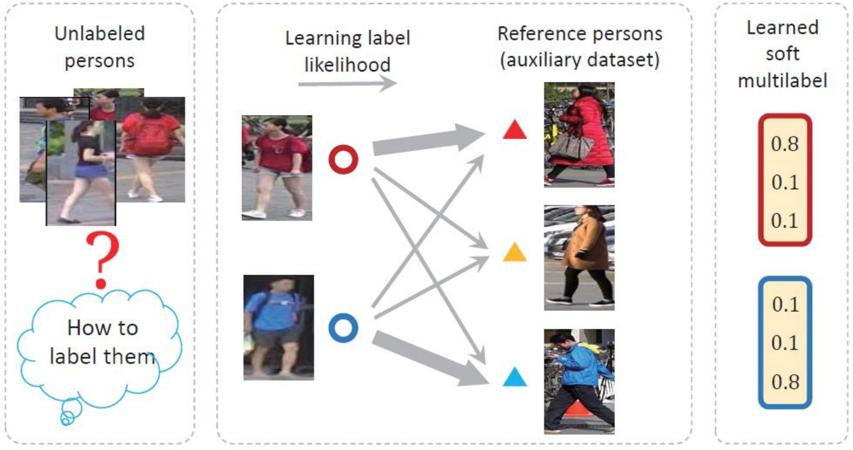
ArcFace: Discussion

- **Settings of parameters: manually to automatically?**
  - Bingyu Liu, Weihong Deng, et al. Fair Loss: Margin-aware Reinforcement Learning for Deep Face Recognition. ICCV 2019.
- **Noisy dataset: a more robust loss function is needed**
  - Yaoyao Zhong, Weihong Deng, et al. Unequal-training for deep face recognition with long-tailed noisy data. CVPR 2019.
- **Universal margin-based metric learning?**
  - Xing Fan, Wei Jiang, Hao Luo, et al. SphereReID: Deep Hypersphere Manifold Embedding for Person Re-Identification. Journal of Visual Communication and Image Representation (2019).

Frontal of Metric Learning

- **Supervised Metric Learning**
  - **Euclidean**: Triplet Loss[3]/Center Loss[4]
  - **Cosine**: SphereFace[5]/CosFace[6]/**ArcFace**
  - **ArcFace[1]: Additive Angular Margin Loss for Deep Face Recognition**
- **Unsupervised Metric Learning**
  - **Data Adaptation**: PTGAN[7]/StarGAN[8]/Camstyle[9]
  - **Pseudo-Label**: MAR[10]/IMAN[11]/**ECN**
  - **(ECN) Invariance Matters[2]: Exemplar Memory for Domain Adaptive Person Re-identification**





ECN: Motivation

Transfer/Adaptation

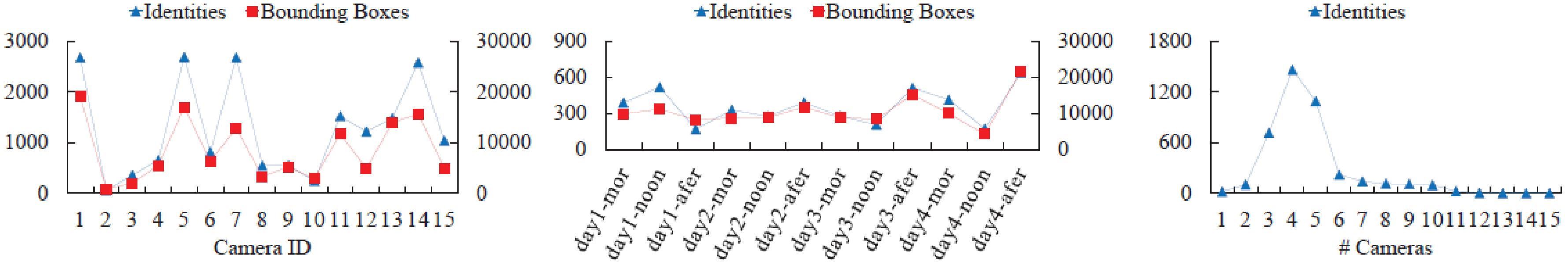
Big Data

Still naïve!

Small(naïve) Data

Table 1: Comparison between *MSMT17* and other person ReID datasets.

Dataset	*MSMT17*	*Duke* [41, 27]	*Market* [39]	*CUHK03* [20]	*CUHK01* [19]	*VIPeR* [8]	*PRID* [10]	*CAVIAR* [3]
BBoxes	**126,441**	36,411	32,668	28,192	3,884	1,264	1,134	610
Identities	**4,101**	1,812	1,501	1,467	971	632	934	72
Cameras	**15**	8	6	2	10	2	2	2
Detector	**Faster RCNN**	hand	DPM	DPM, hand	hand	hand	hand	hand
Scene	**outdoor, indoor**	outdoor	outdoor	indoor	indoor	outdoor	outdoor	indoor



(a) The number of identities and bounding boxes on each camera

(b) The number of identities and bounding boxes in each time slot

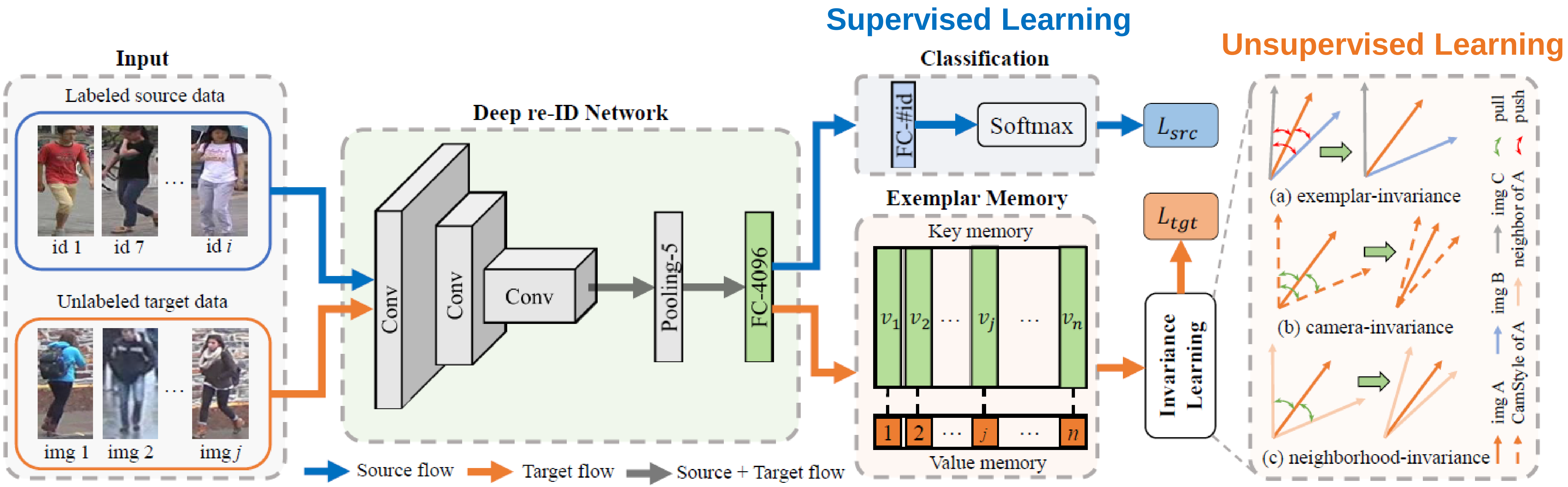
(c) The number of identities across different, *i.e.*, 1-15, cameras

Figure 3: Statistics of *MSMT17*.

Longhui Wei, Shiliang Zhang, et. al., Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. CVPR 2018.

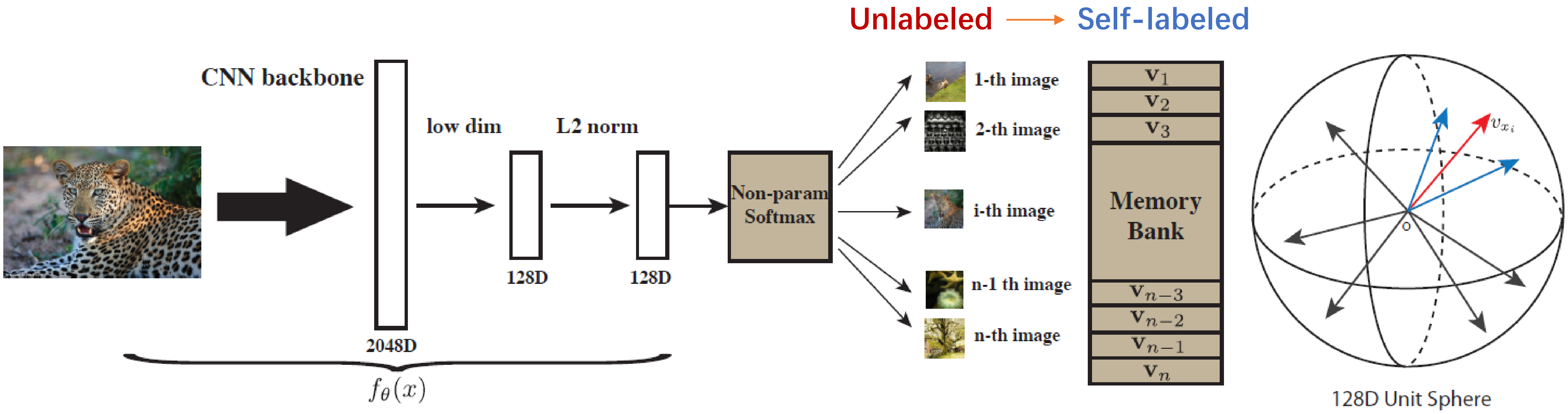
ECN: Methodology

- **Overall Structure**



ECN: Methodology

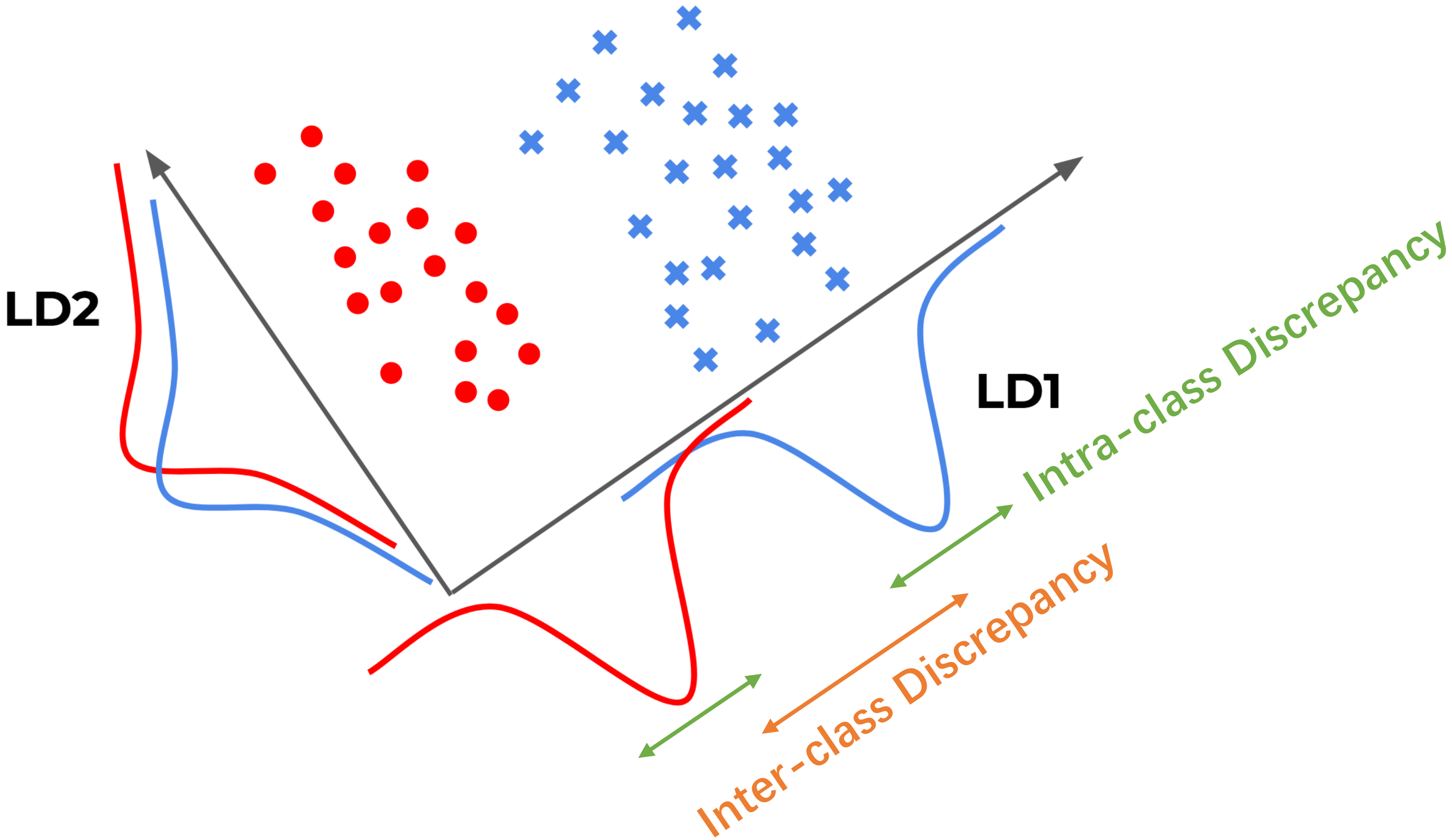
- **Unsupervised Learning**



Zhirong Wu, Yuanjun Xiong, et. al., Unsupervised Feature Learning via Non-Parametric Instance Discrimination. CVPR 2018.

ECN: Methodology

- **Linear Discriminant Analysis**



ECN: Methodology

- **Unsupervised Learning**

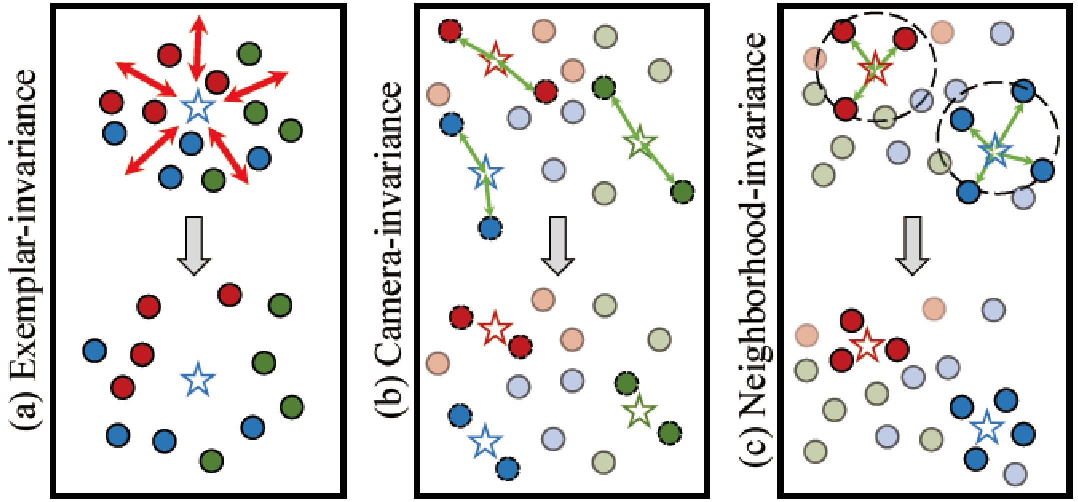
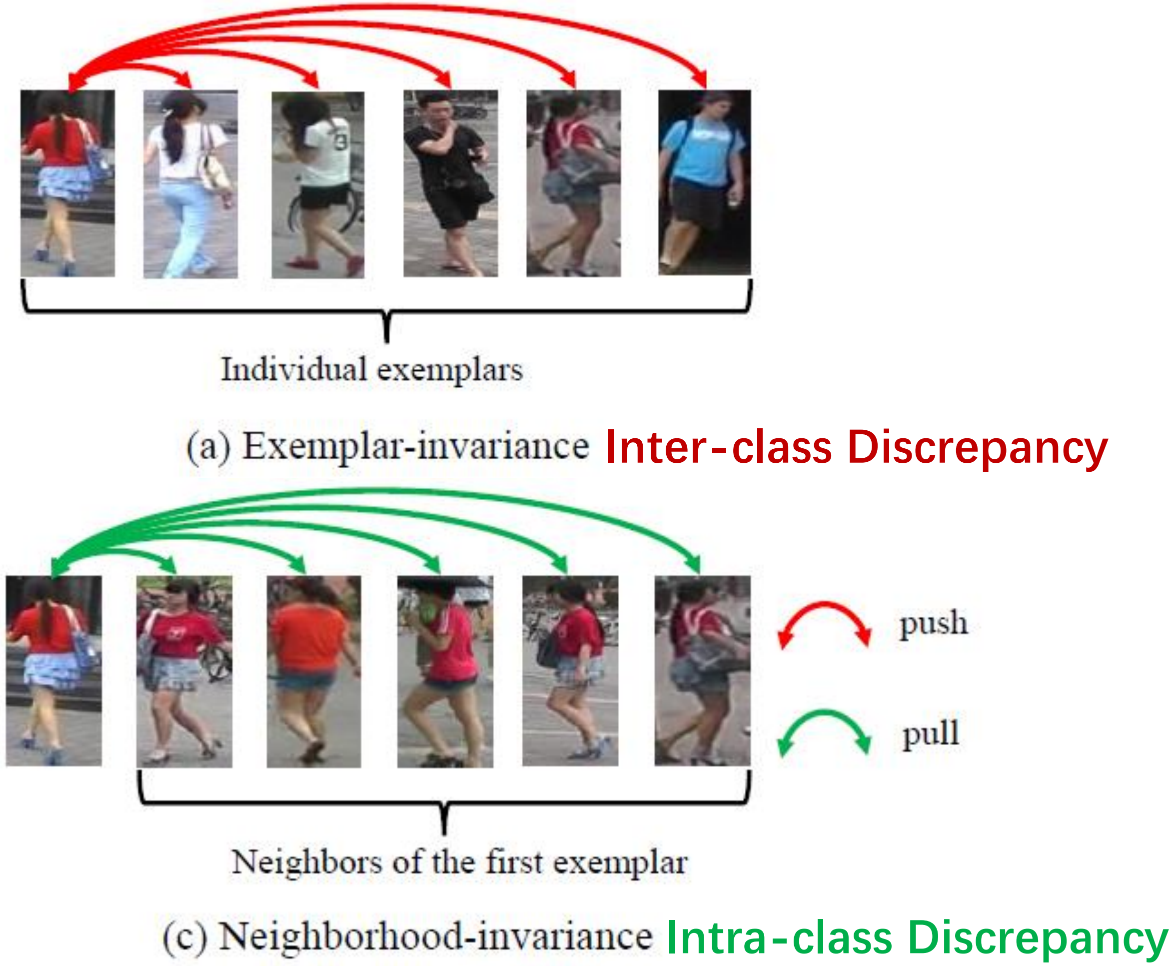
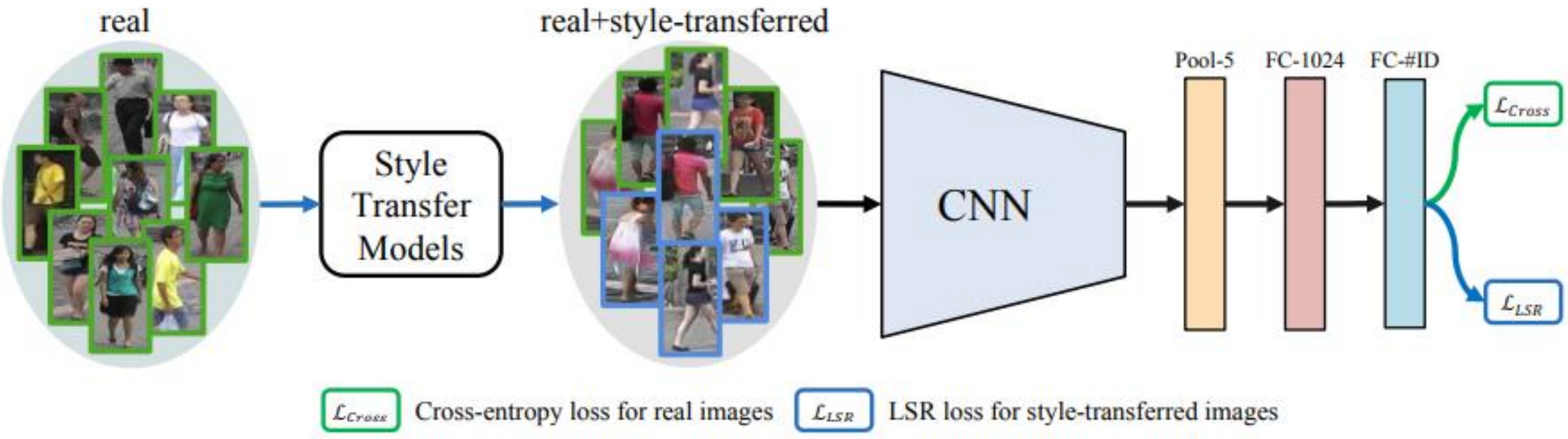


Fig. 1. Examples of three underlying properties of invariance. Colors indicate identities. (a) Exemplar-invariance: an input exemplar (denoted by ⋆) is enforced to be away from others. (b) Camera-invariance: an input exemplar (denoted by ⋆) and its CamStyle transferred images (with dashed outline) are encouraged to be close to each other. (c) Neighborhood-invariance: an input exemplar (denoted by ⋆) and its reliable neighbors (highlighted in dashed circle) are forced to be close to each other. Best viewed in color.



ECN: Methodology

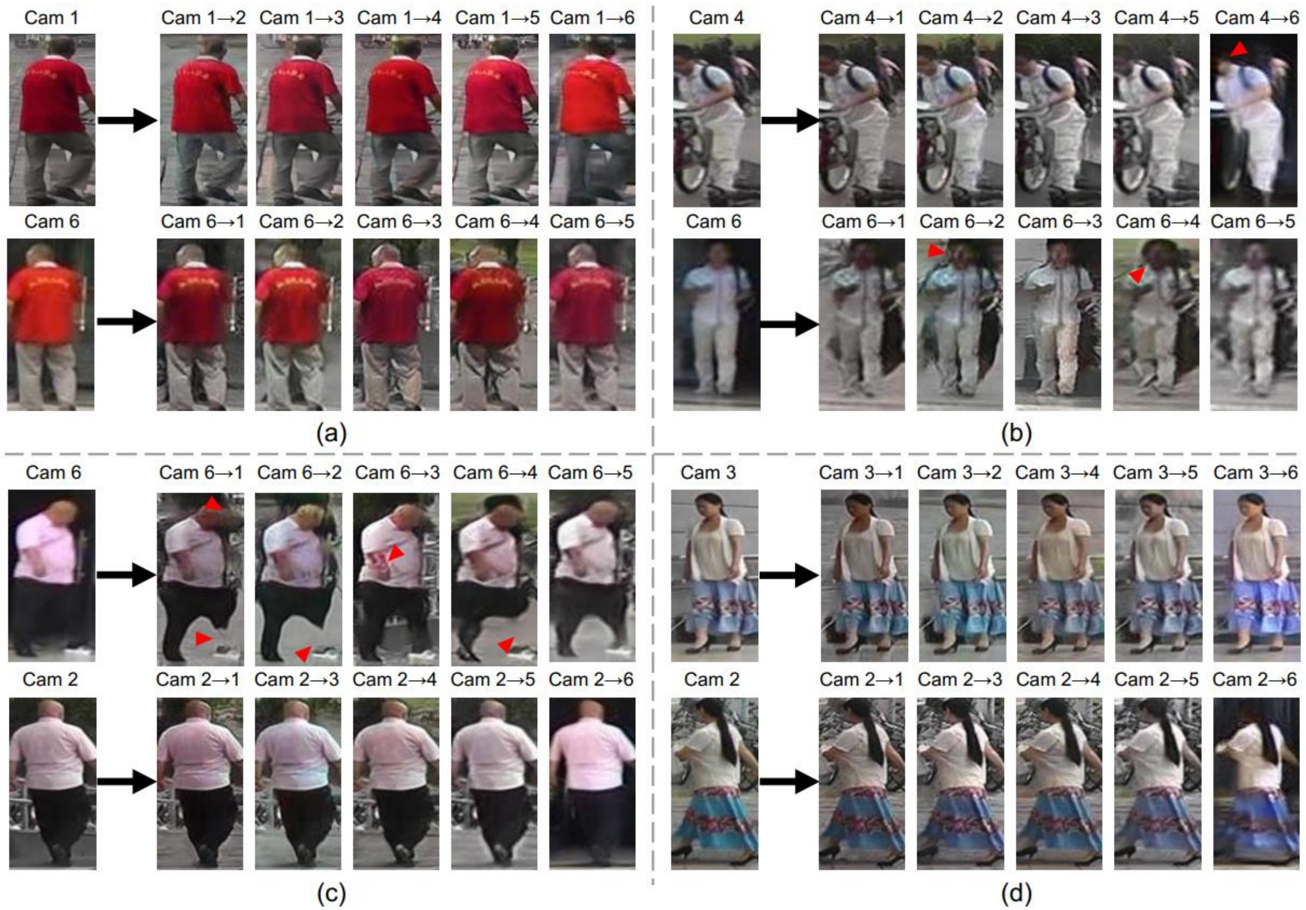
- **Camera Factor: CamStyle(StarGAN)**



Camera Style Adaptation for Person Re-identification. Zhun Zhong, Liang Zheng, et. al., Camera Style Adaptation for Person Re-identification. CVPR 2018.

ECN: Methodology

- **Camera Factor: CamStyle(StarGAN)**



(b) Camera-invariance

Inter-Camera Discrepancy

AMC AI ML Club

ECN: Methodology

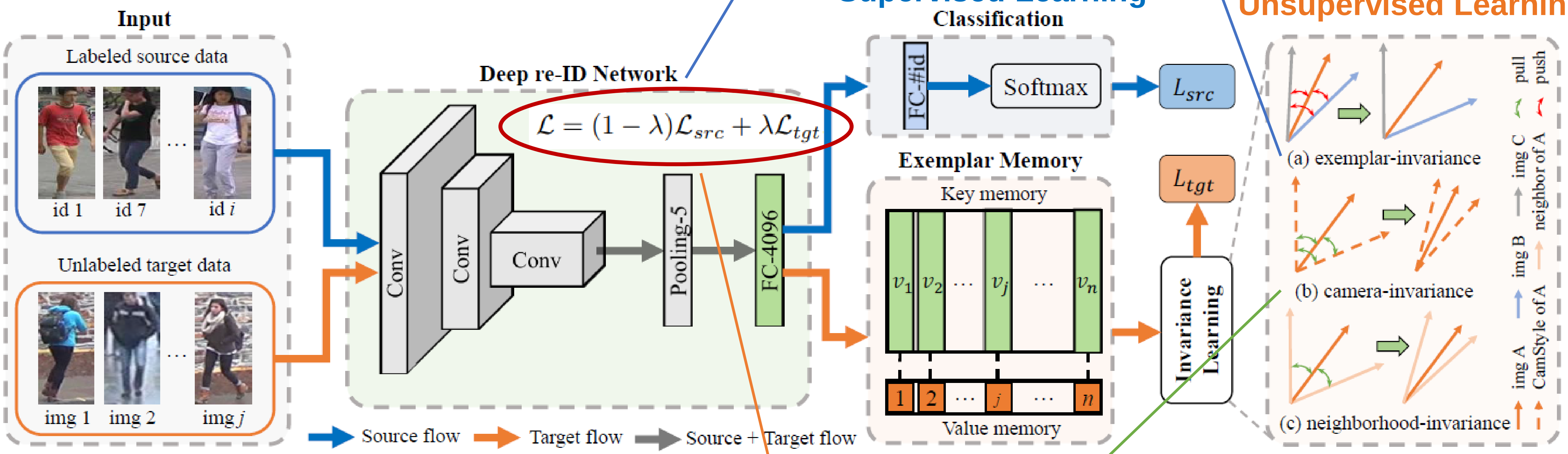
- **Overall Structure**

$$\mathcal{L}_{src} = -\frac{1}{n_s}\sum_{i=1}^{n_s} \log p(y_{s,i}|x_{s,i})$$

$$\mathcal{L}_{ei} = -\log \frac{\exp(\mathcal{K}[i]^{\mathrm{T}} f(x_{t,i})/\beta)}{\sum_{j=1}^{N_t} \exp(\mathcal{K}[j]^{\mathrm{T}} f(x_{t,i})/\beta)}$$

Supervised Learning

Unsupervised Learning



$$\mathcal{L}_{ci} = -\log p(i|\hat{x}_{t,i})$$

$$\mathcal{L}_{ni} = -\sum_{j\neq i} w_{i,j} \log p(j|x_{t,i})$$

$$w_{i,j} = \begin{cases} \frac{1}{k}, & j \neq i \\ 1, & j = i \end{cases}, \forall j \in \mathcal{M}(x_{t,i}, k)$$

Neighbors of $x_{t,i}$

$$\mathcal{L}_{tgt} = -\frac{1}{n_t}\sum_{i=1}^{n_t}\sum_j w_{i,j} \log p(j|x^*_{t,i})$$

ECN: Experiments

AMC
AI ML Club

Methods	Market-1501						DukeMTMC-reID					
	Src.	R-1	R-5	R-10	R-20	mAP	Src.	R-1	R-5	R-10	R-20	mAP
Supervised Learning	N/A	87.6	95.5	97.2	98.3	69.4	N/A	75.6	87.3	90.6	92.9	57.8
Source Only	DukeMTMC	43.1	58.8	67.3	74.3	17.7	Market-1501	28.9	44.0	50.9	57.5	14.8
Ours w/ E		48.7	67.4	74.0	80.2	21.0		34.2	51.3	58	64.2	18.7
Ours w/ E+C		63.1	79.1	84.6	89.1	28.4		53.9	70.8	76.1	80.7	29.7
Ours w/ E+N		58.0	69.9	75.6	80.4	27.7		39.7	53.0	58.1	62.9	23.6
Ours w/ E+C+N		**75.1**	**87.6**	**91.6**	**94.5**	**43.0**		**63.3**	**75.8**	**80.4**	**84.2**	**40.4**

Table 2. Methods comparison when tested on Market-1501 and DukeMTMC-reID. **Supervised Learning**: Baseline model trained with labeled target data. **Source Only**: Baseline model trained with only labeled source data. **E**: Exemplar-invariance. **C**: Camera-invariance. **N**: Neighborhood-invariance. **Src.**: Source domain.

Methods	Market-1501				DukeMTMC-reID			
	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
LOMO [15]	27.2	41.6	49.1	8.0	12.3	21.3	26.6	4.8
Bow [37]	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
UMDL [20]	34.5	52.6	59.6	12.4	18.5	31.4	37.6	7.3
PTGAN [30]	38.6	-	66.1	-	27.4	-	50.7	-
PUL [9]	45.5	60.7	66.7	20.5	30.0	43.4	48.5	16.4
SPGAN [7]	51.5	70.1	76.8	22.8	41.1	56.6	63.0	22.3
CAMEL [36]	54.5	-	-	26.3	-	-	-	-
MMFA [16]	56.7	75.0	81.8	27.4	45.3	59.8	66.3	24.7
SPGAN+LMP [7]	57.7	75.8	82.4	26.7	46.4	62.3	68.0	26.2
TJ-AIDL [29]	58.2	74.8	81.1	26.5	44.3	59.6	65.0	23.0
CamStyle [45]	58.8	78.2	84.3	27.4	48.4	62.5	68.9	25.1
HHL [43]	62.2	78.8	84.0	31.4	46.9	61.0	66.7	27.2
Ours (ECN)	**75.1**	**87.6**	**91.6**	**43.0**	**63.3**	**75.8**	**80.4**	**40.4**

Table 4. Unsupervised person re-ID performance comparison with state-of-the-art methods on Market-1501 and DukeMTMC-reID.

ECN: Experiments

β	Duke → Market-1501		Market-1501 → Duke	
	Rank-1	mAP	Rank-1	mAP
0.01	47.3	20.0	29.1	13.2
0.03	72.3	40.3	59.7	35.7
0.05	**75.1**	**43.0**	**63.3**	**40.4**
0.1	71.4	36.8	59.3	35.8
0.5	52.3	23.1	45.4	24.2
1.0	47.8	20.8	40.2	19.3

Table 1. Evaluation with different values of β in Eq. 3.

Methods	Src.	MSMT17			
		R-1	R-5	R-10	mAP
PTGAN [30]	Market	10.2	-	24.4	2.9
Ours (ECN)	Market	**25.3**	**36.3**	**42.1**	**8.5**
PTGAN [30]	Duke	11.8	-	27.4	3.3
Ours (ECN)	Duke	**30.2**	**41.5**	**46.8**	**10.2**

Table 5. Performance evaluation when tested on MSMT17.

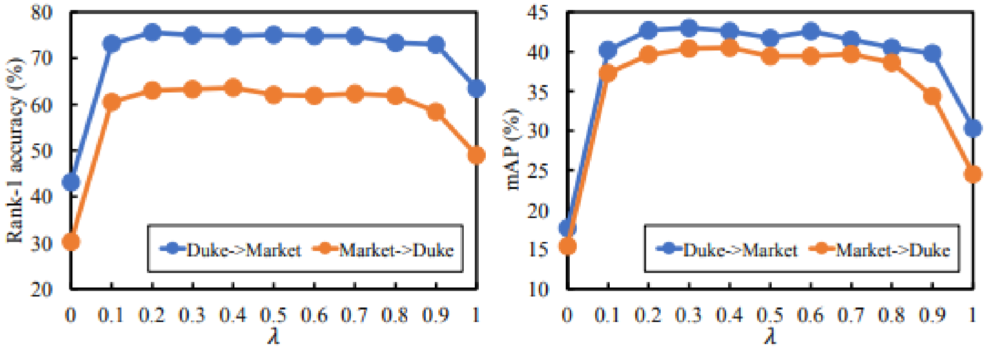


Figure 3. Evaluation with different values of λ in Eq. 9.

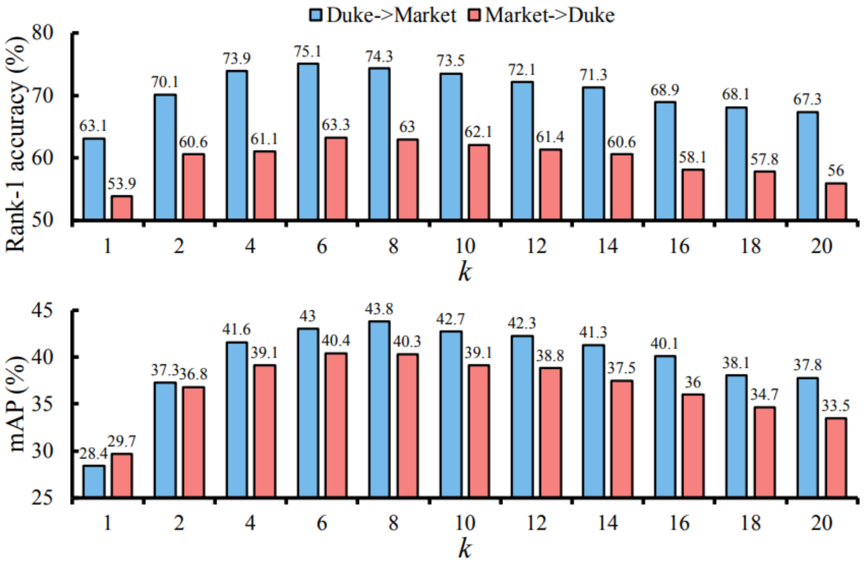


Figure 4. Evaluation with different number of candidate positive samples in neighborhood-invariance learning.

ECN: Discussion

AMC
AI ML Club

- **Better Clustering Method: Graph Network?**
  - Zhun Zhong, Liang Zheng, et. al. Learning to Adapt Invariance in Memory for Person Re-identification. arXiv:1908.00485, 2019.
- **Memory Bank: What if big data?**
  - Xiaohang Zhan, Ziwei Liu, et. al. Consensus-Driven Propagation in Massive Unlabeled Data for Face Recognition. ECCV 2018.
- **Association with classical domain adaptation method?**
  - Mei Wang, Weihong Deng, et. al. Racial Faces in-the-Wild: Reducing Racial Bias by Information Maximization Adaptation Network. ICCV 2019.
  - Kihyuk Sohn, Wenling Shang, et. al. Unsupervised Domain Adaptation for Distance Metric Learning. ICLR 2019.

References

- **[1] Jiankang Deng, Jia Guo, et. al. ArcFace: Additive Angular Margin Loss for Deep Face Recognition, CVPR 2019.**
- **[2] Zhun Zhong, Liang Zheng, et. al. Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-identification, CVPR 2019.**
- [3] Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. CVPR 2015.
- [4] Yandong Wen, Kaipeng Zhang, et. al. A Discriminative Feature Learning Approach for Deep Face Recognition. ECCV 2016.
- [5] Weiyang Liu, Yandong Wen, et. al. Large-margin softmax loss for convolutional neural networks. ICML 2016.
- [6] Hao Wang, Yitong Wang, et. al. Cosface: Large margin cosine loss for deep face recognition. CVPR 2018.
- [7] Longhui Wei, Shiliang Zhang, et. al. Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. CVPR 2018.
- [8] Yunjey Choi, Minje Choi, et. al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. CVPR 2018.
- [9] Zhun Zhong, Liang Zheng, et. al. Camera Style Adaptation for Person Re-identification. CVPR 2018.
- [10] Hong-Xing Yu, Wei-Shi Zheng, et. al. Unsupervised Person Re-identification by Soft Multilabel Learning. CVPR 2019.
- [11] Mei Wang, Weihong Deng, et. al. Racial Faces in-the-Wild: Reducing Racial Bias by Information Maximization Adaptation Network. ICCV 2019.

Thank you for listening

Hongwei Fan
Lab of Pattern Recognition and Intelligent Systems, BUPT