Learn by Guessing: Multi-Step Pseudo-Label Refinement for Person Re-Identification

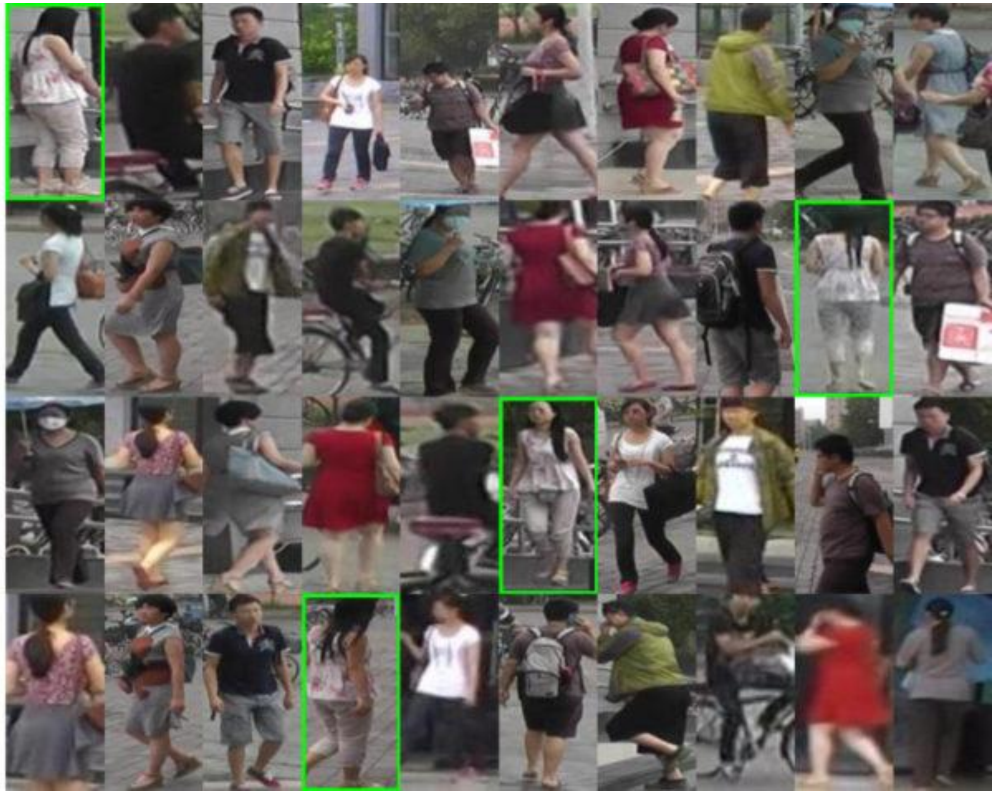
Tiago de C. G. Pereira and Teofilo E. de Campos

Departamento de Ciência da Computação, Universidade de Brasília, Brasília - DF

February 2022

Definition

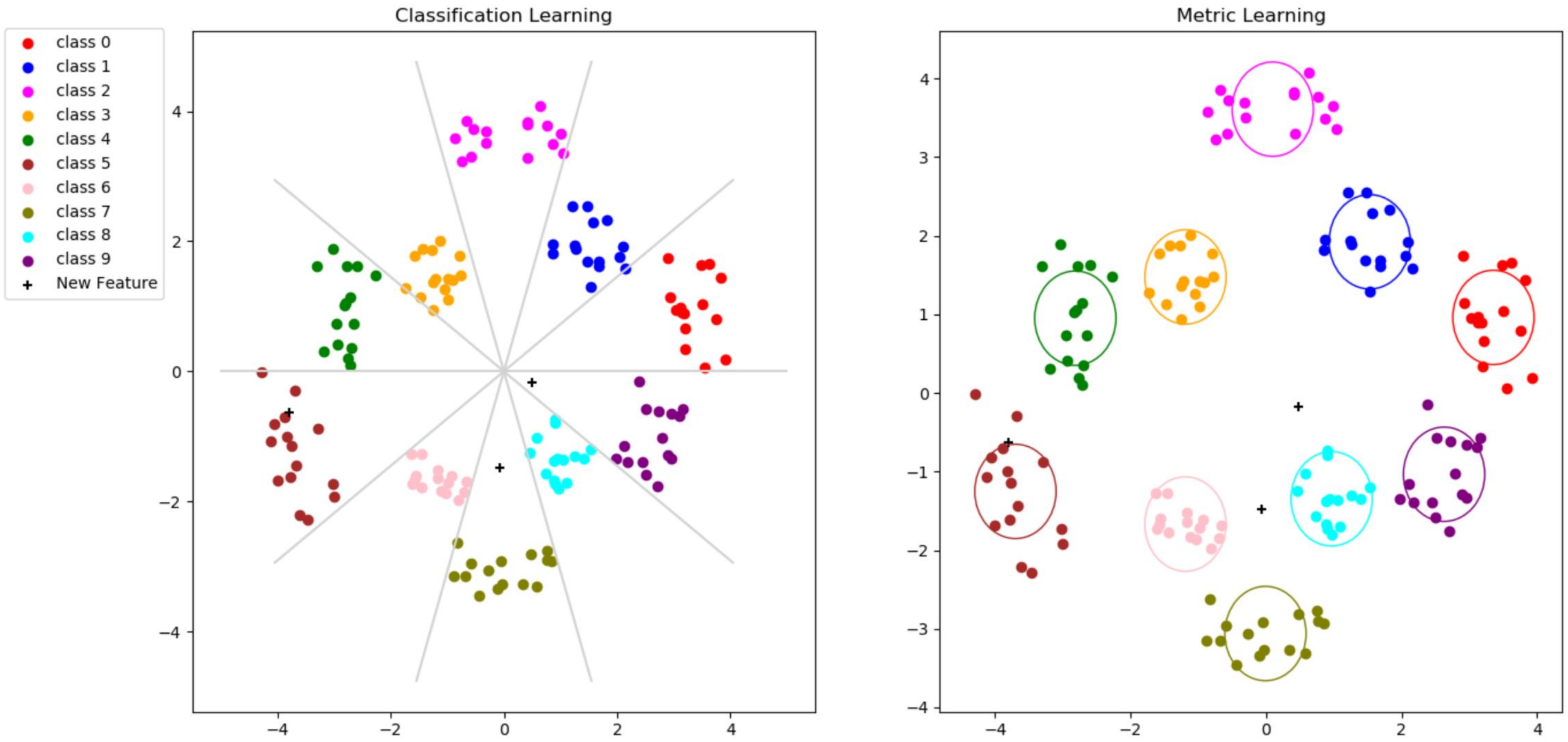
Person Re-Identification is an image retrieval task, where the object in the images are people.



Motivation

Regardless of the scenario or camera, the goal of person Re-ID is **Matching person images from different non-overlapping cameras views**. However, the addition of a new camera view normally has a direct impact in the algorithm performance, and this is a roadblock for diverse real-world applications.

Metric Learning vs. Classification Learning

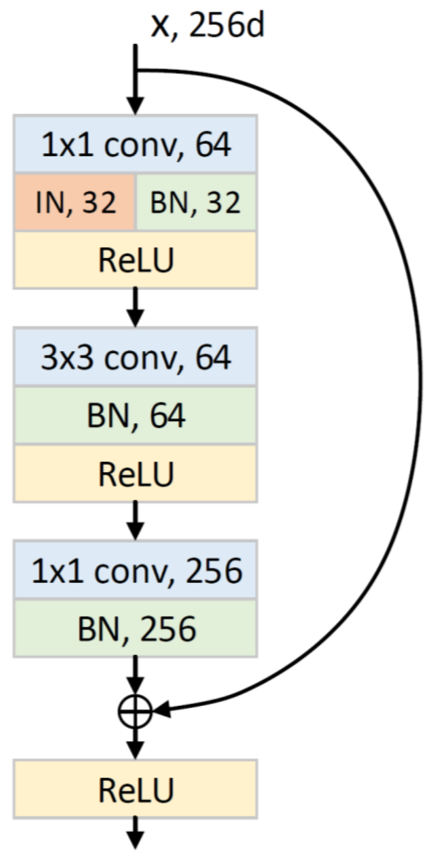


Model Architecture – IBN Net-50 a

For the person Re-ID challenge, we need a feature extractor that can encode person information while disregarding camera variations and background noise. Therefore, architectures designed for image classification, like Resnet-50, are excellent starting points.

We believe that Resnet-50 [1] is a great choice, because the residual blocks are capable of efficiently propagating information of multiple semantic levels.

Furthermore, we use the IBN [2] Net version of the Resnet-50 to enhance its generalization capacity.



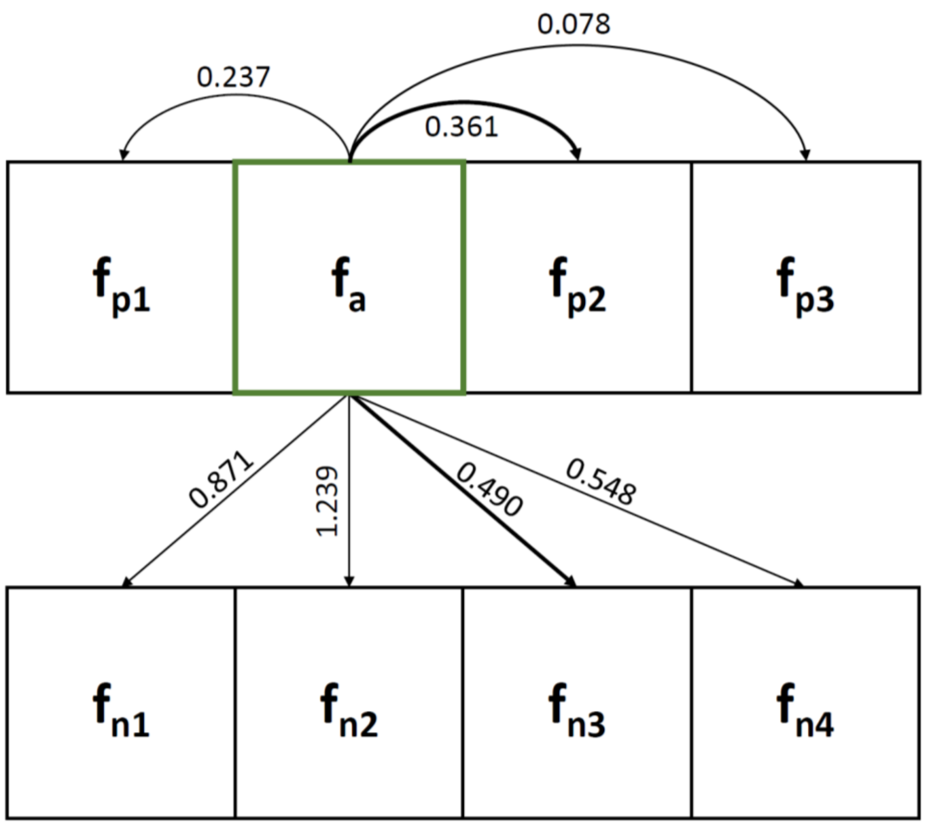
[1] He, K. et al.: Deep Residual Learning for Image Recognition. CVPR, 2016.
[2] Pan, X. et al.: Two at Once: Enhancing Learning and Generalization Capacities via IBN-Net. ECCV, 2018.

Triplet Loss & Batch hard

- The Triplet Loss is responsible for producing output vectors that belong to a Euclidean feature space;

- It is better than the contrastive loss, once it can push pairs from different people away while pulling feature pairs from same people together;

- **Challenge:** How to choose the best triplets? Based on Hermans et al.'s work [3], batch hard is the best approach.



[3] Hermans, A., Beyer, L., and Leibe, B. In defense of the triplet loss for person re-identification. arXiv 2017.

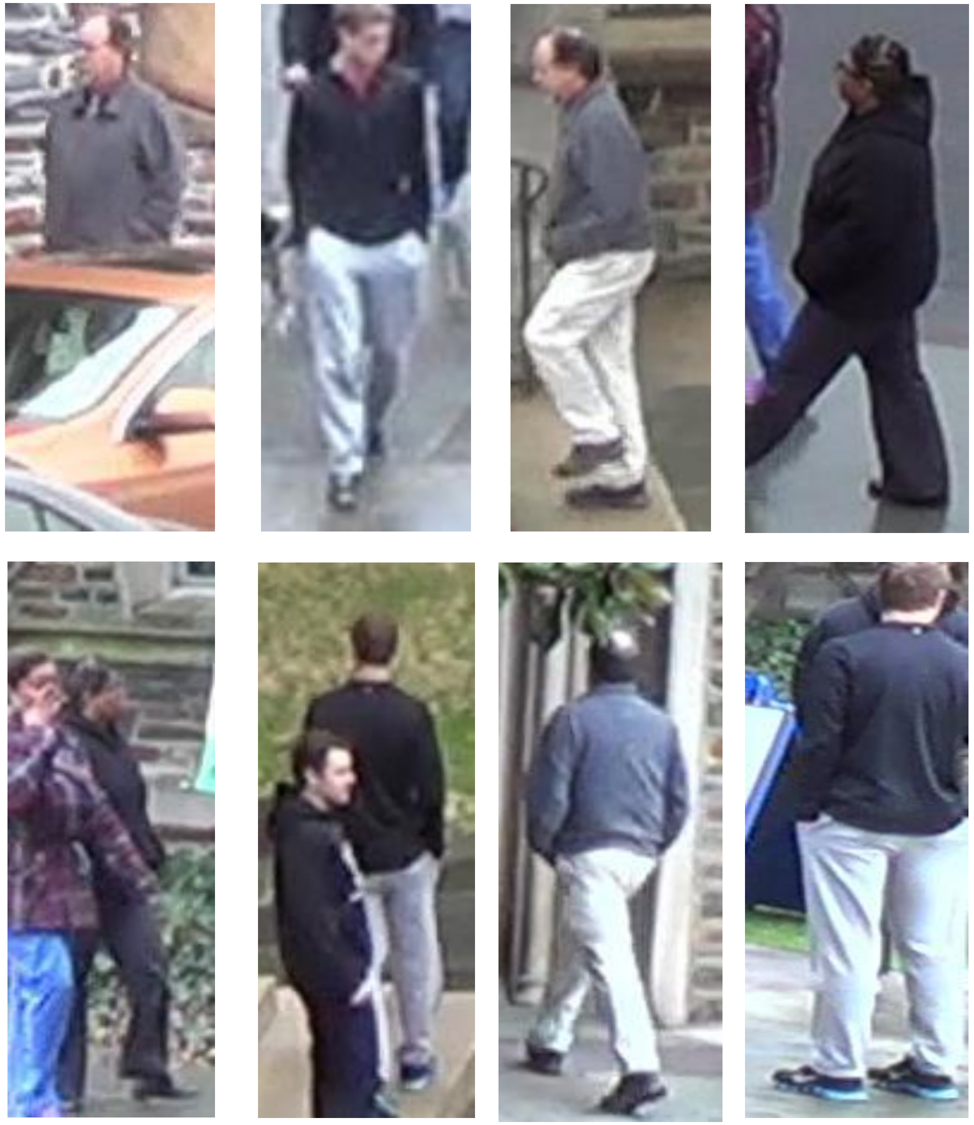
Market1501 Dataset [4]

	Market1501
Release Year	2015
Samples	32668
Identities	1501
Cameras	6
Avg Number of Cameras Passed per Identity	4.42
Scene	outdoor

[4] Zheng, et al.: Scalable Person Re-identification: A Benchmark. ICCV, 2015.

DukeMTMC Dataset [5]

	DukeMTMC
Release Year	2016
Samples	36411
Identities	1812
Cameras	8
Avg Number of Cameras Passed per Identity	2.67
Scene	outdoor



[5] Zheng, Z. et al.: Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro. ICCV, 2017.

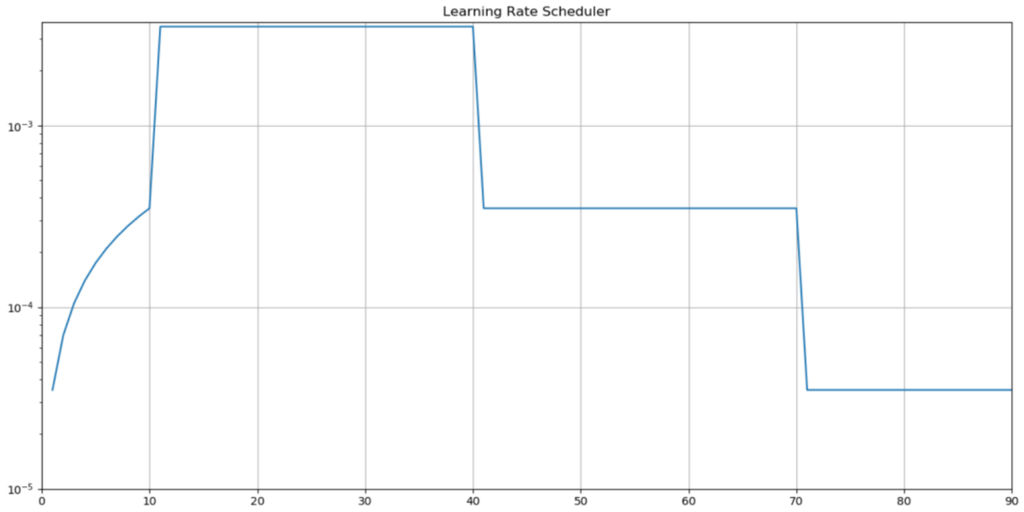
Overview

	Market1501	DukeMTMC
Release Year	2015	2016
Samples	32668	36411
Identities	1501	1812
Cameras	6	8
Avg Number of Cameras Passed per Identity	4.42	2.67
Scene	outdoor	outdoor

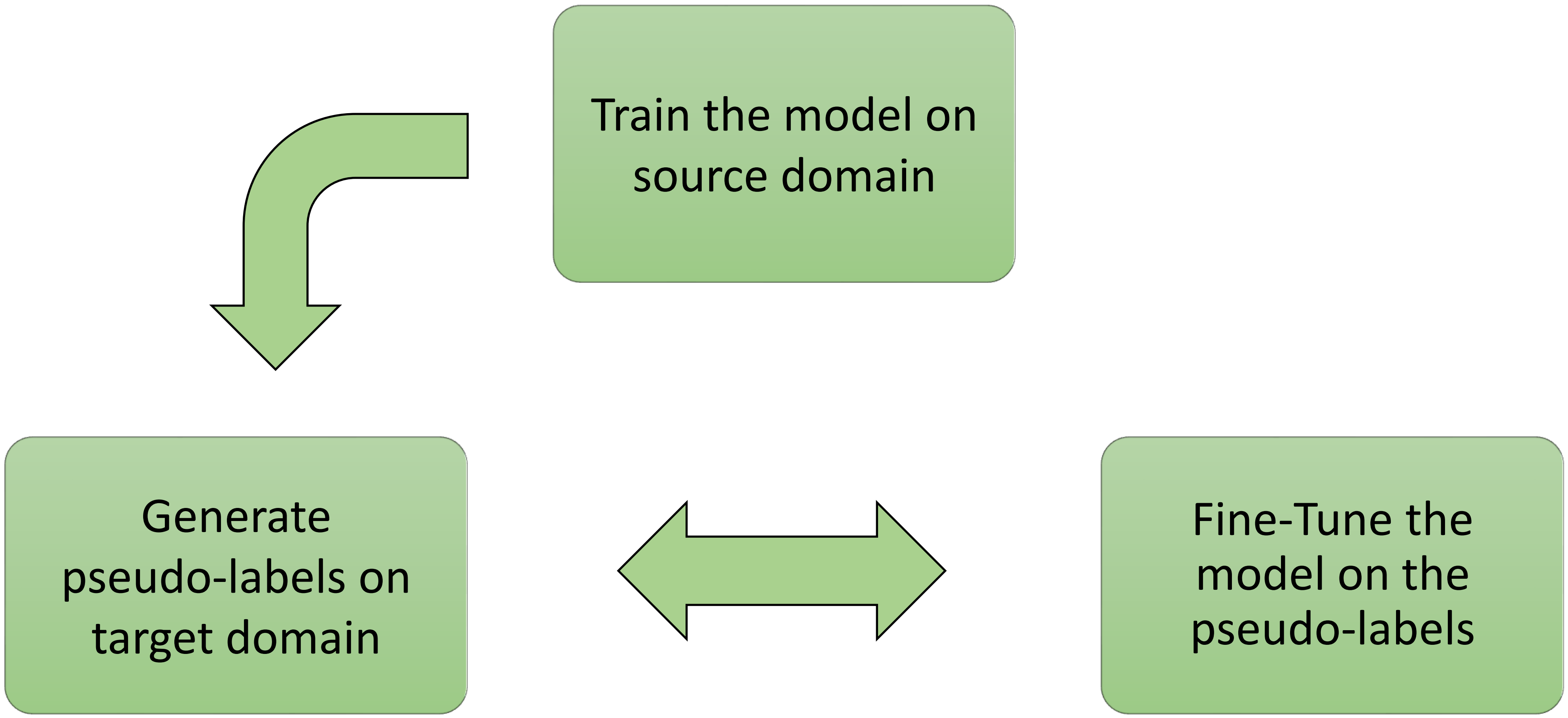
Step 1 – Architecture + Training Strategy

Our general training strategy had the following configurations:

- IBN Net-50 a
- Adam optimizer
- A three-factor loss function given by $\mathcal{L} = \mathcal{L}_{triplet} + \mathcal{L}_{ID} + 0.005 * \mathcal{L}_{center}$ where:
  - $\mathcal{L}_{triplet}$ is the triplet Loss responsible for the metric leaning,
  - $\mathcal{L}_{ID}$ is a label smooth cross entropy loss for person ID classification
  - $\mathcal{L}_{center}$ is a center loss to enforce cluster compactness
- A learning rate scheduler for the 90 training epochs defined by:

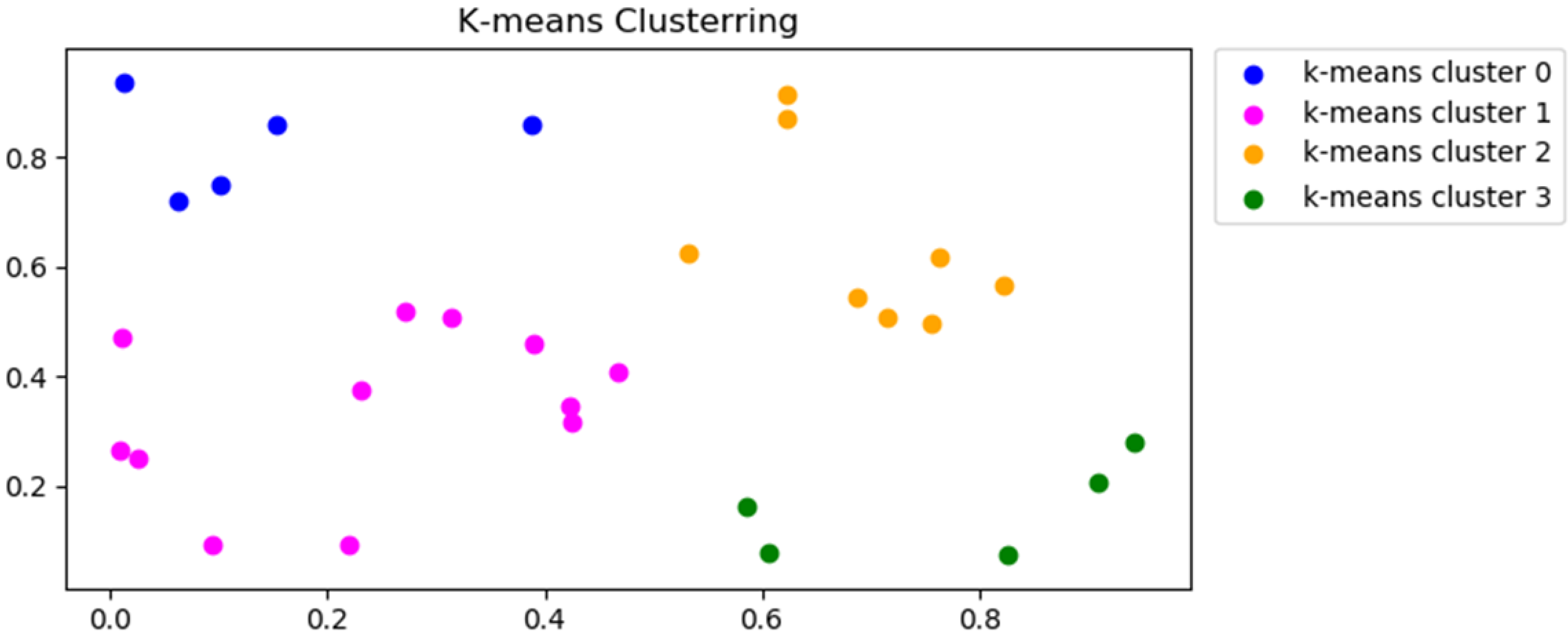


Step 2 – Progressive Learning [6]

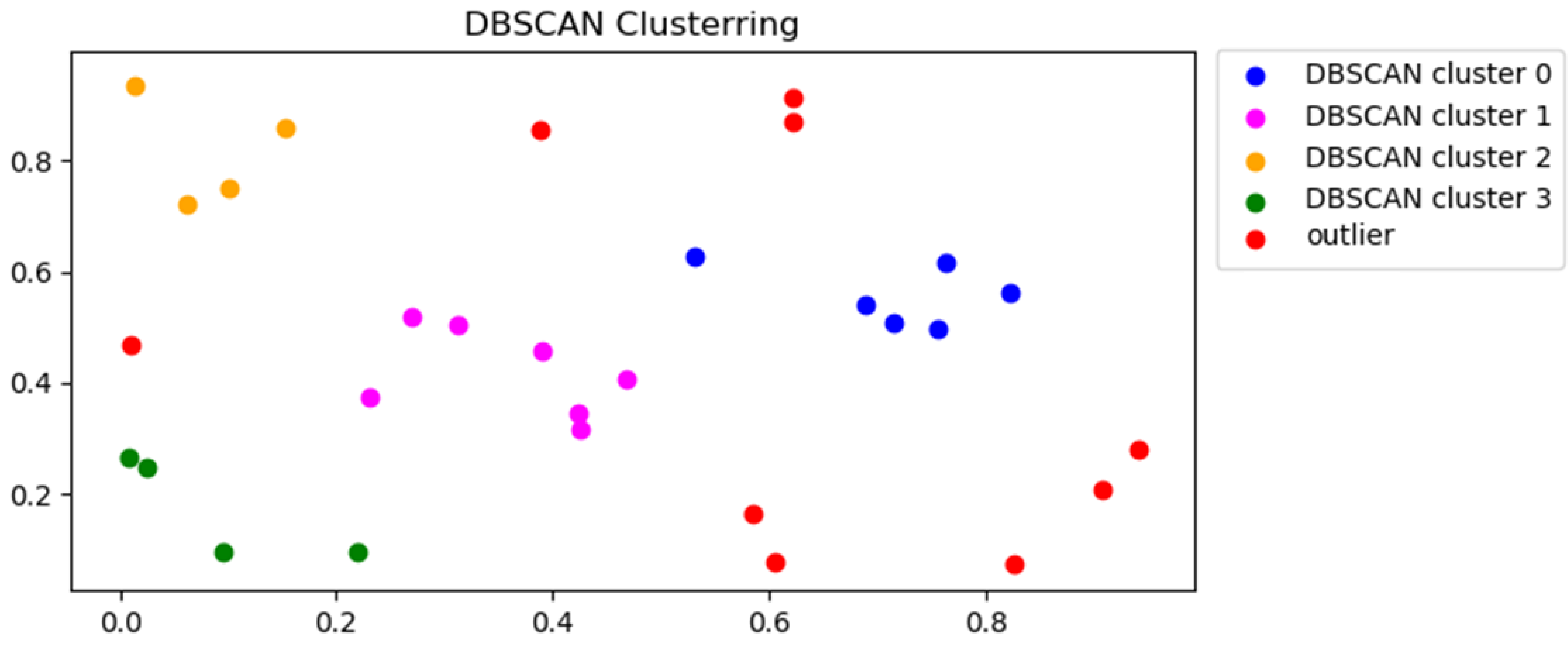


[6] Fan, H. et al.: Unsupervised Person Re-identification: Clustering and Fine-tuning. TOMM, 2018.

Step 3 – Clustering Techniques

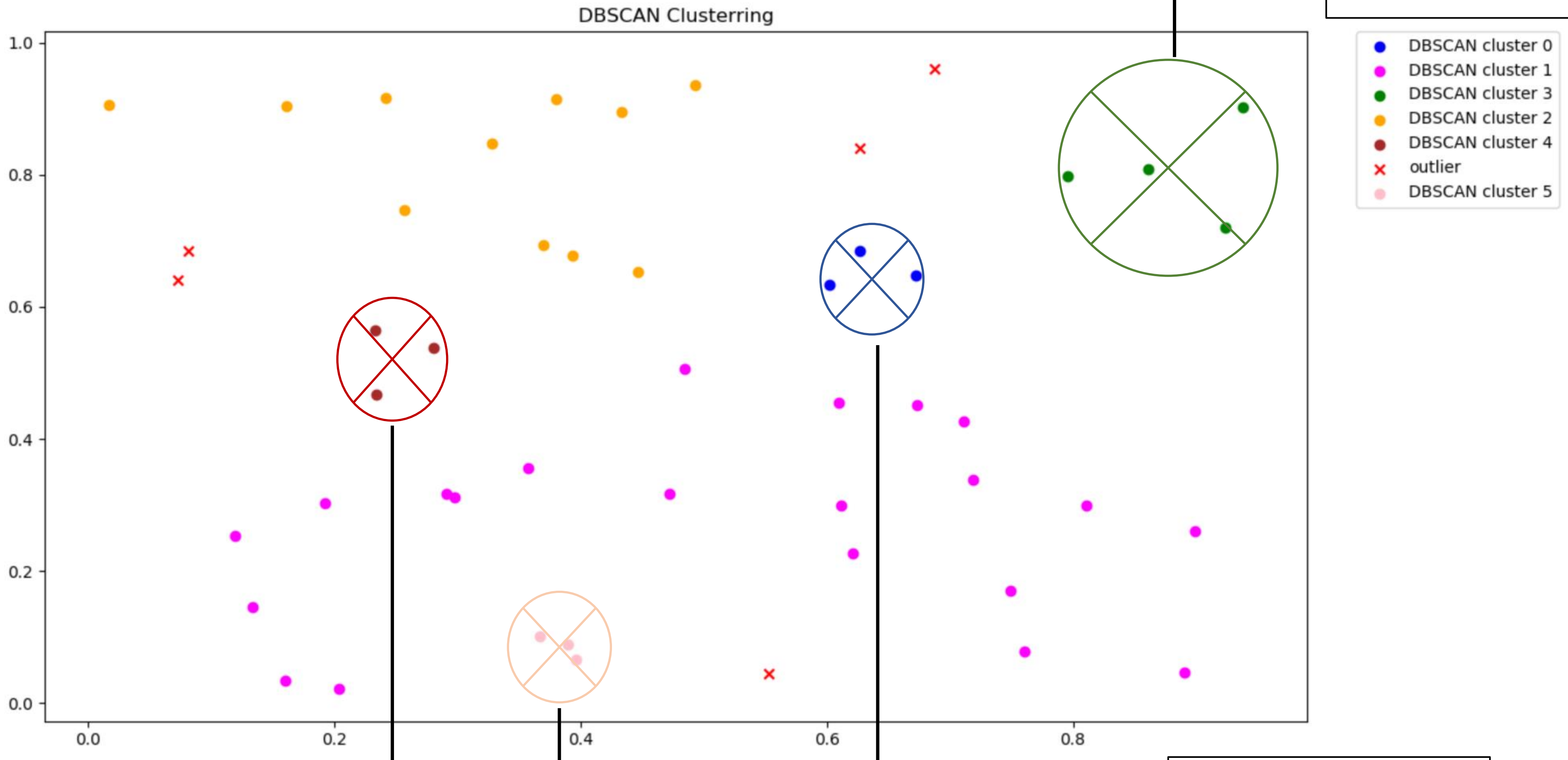


Step 3 – Clustering Techniques



Step 4 – Cluster Selection

All samples from same camera



Less than K samples in cluster

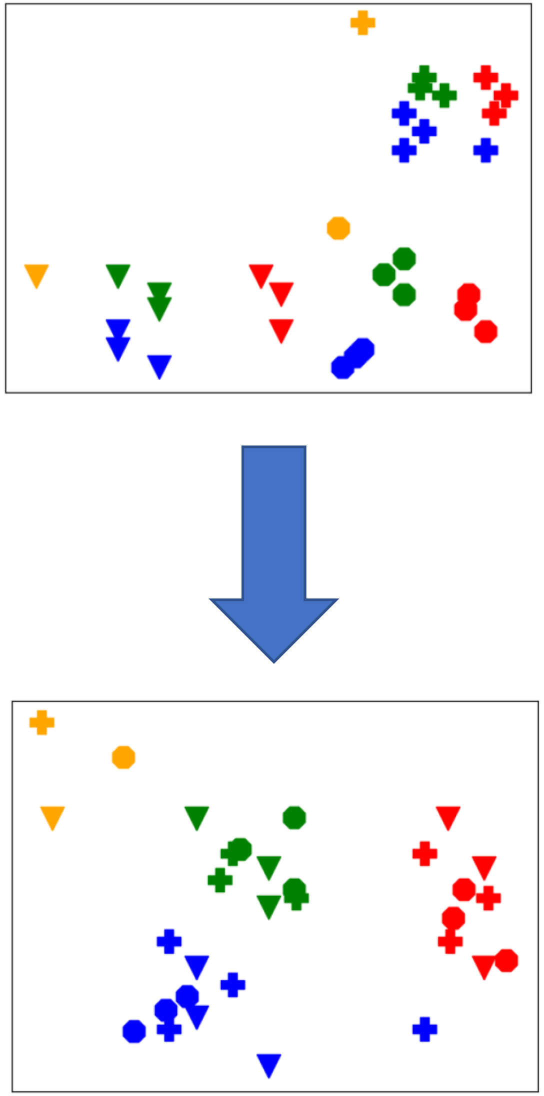
Step 5 – Camera-Guided Feature Normalization

The high variance present in person Re-ID is mainly caused by different camera views, as each camera has its own characteristics.

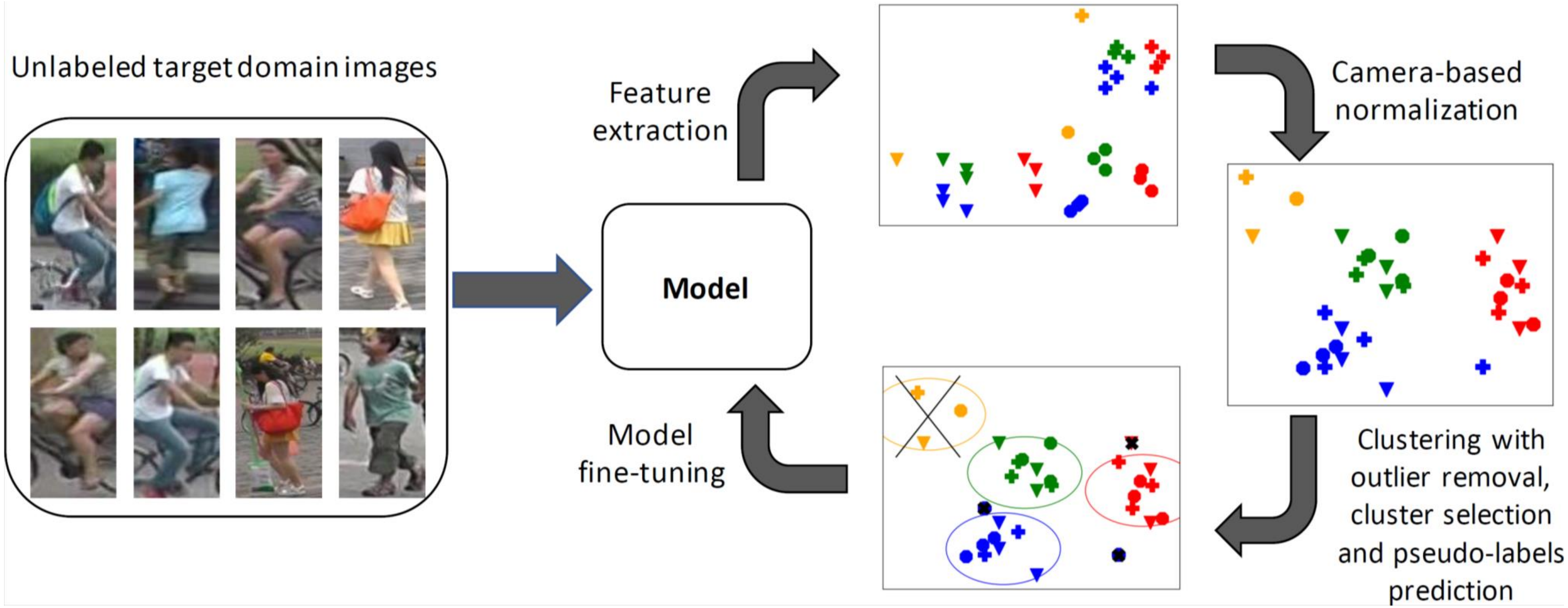
Therefore, the model tends to cluster images by cameras rather than clustering images from the same person in different views.

A camera guided normalization step is then necessary to reduce this variance and allow the clustering step to create better clusters. The normalization is done by:

$$\bar{f}_{v_j} = \frac{f_{v_j} - \mu_{v_j}}{\sigma_{v_j}}$$



Step 6 – Unsupervised Domain Adaptation



Results

Methods	Market1501 -> DukeMTMC				DukeMTMC -> Market1501			
	Rank – 1	**Rank – 5**	**Rank - 10**	**mAP**	**Rank – 1**	**Rank – 5**	**Rank - 10**	**mAP**
SPGAN	46.9	62.6	68.5	26.4	58.1	76.0	82.7	26.9
UCDA-CCE	55.4	-	-	36.7	64.3	-	-	34.5
ARN	60.2	73.9	79.5	33.4	70.3	80.4	86.3	39.4
MAR	67.1	79.8	-	48.0	67.7	81.9	-	40.0
ECN	63.3	75.8	80.4	40.4	75.1	87.6	91.6	43.0
PDA-Net	63.2	77.0	82.5	45.1	75.2	86.3	90.2	47.6
EANet	67.7	-	-	48.0	78.0	-	-	51.6
CBN + ECN	68.0	80.0	83.9	44.9	81.7	91.9	94.7	52.0
Theory	68.4	80.1	83.5	49.0	75.8	89.5	93.2	53.7
CR-GAN	68.9	80.2	84.7	48.6	77.7	89.7	92.7	54.0
PCB-PAST	72.4	-	-	54.3	78.4	-	-	54.6
AD Cluster	72.6	82.5	85.5	54.1	86.7	94.4	96.5	68.3
SSG	76.0	85.8	89.3	60.3	86.2	94.6	96.5	68.7
DG-Net++	78.9	87.8	90.4	63.8	82.1	90.2	92.7	61.7
MMT	*79.3*	*89.1*	*92.4*	*65.7*	<u>90.9</u>	**96.4**	**97.9**	<u>76.5</u>
Ours	<u>82.7</u>	<u>90.5</u>	**93.5**	<u>69.3</u>	*89.1*	<u>95.8</u>	<u>97.2</u>	*73.6*
Ours + RR	**84.8**	**90.8**	<u>93.2</u>	**81.2**	**92.0**	*95.3*	*96.6*	**88.1**

Final Observations

1. The initial pseudo-labels are noisy, but with the iterative process we can have a better representation of the real labels;

2. Using DBSCAN with the outlier detector plays a crucial role to continuously enhance the model performance;

3. Camera-Guided Normalization is essential when applying the model to a new set of cameras.

Thank You!

For more details about our works, please check https://teodecampos.github.io/tiago/

We acknowledge the support of Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) through Convênio 07/2019 - project KnEDLe - with support of Finatec (project 6429).