Part-based Pseudo Label Refinement for Unsupervised Person Re-identification

윤성의

SGVR Lab, KAIST

SGVR Lab
KAIST

Project Guidelines: Project Topics

- Any topics related to the course theme are okay
  - You can find topics by browsing recent papers

Expectations

- **Mid-term project presentation**
  - **Introduce problems and explain why it is important**
  - **Give an overall idea on the related work**
  - **Explain what problems those existing techniques have**
  - **(Optional) explain how you can address those problems**
  - **Explain roles of each member**

Expectations

- Final-term project presentation
  - Cover all the materials that you talked for your mid-term project
  - Present your ideas that can address problems of those state-of-the-art techniques
  - Give your qualitatively (or intuitive) reasons how your ideas address them
  - Also, explain expected benefits and drawbacks of your approach
  - (Optional) backup your claims with quantitative results collected by some implementations
  - Explain roles of each members

A few more comments

- **Start to implement a paper, if you don't have any clear ideas**
  - **While you implement it, you may get ideas about improving it**

Speaker	Novelty of the project and idea (1 ~ 5)	Practical benefits of the method (1 ~ 5)	Completeness level of the project (1 ~ 5)	Total score (3 ~ 15)	Role of each student is clear and well balanced? (Yes or No)
XXX					
YYY					

Class Objectives

- Person Re-identification
- Unsupervised Approaches
- Part-based Pseudo Label Refinement for Unsupervised Person Re-ID (CVPR 2022)

Slide Ack: TA 조윤기

Person Re-identification (Person Re-ID)

- Person re-ID aims to **retrieve a person corresponding to a given query** across disjoint camera views or different time stamps.
- Applications: Surveillance system, Finding a missing person, etc.

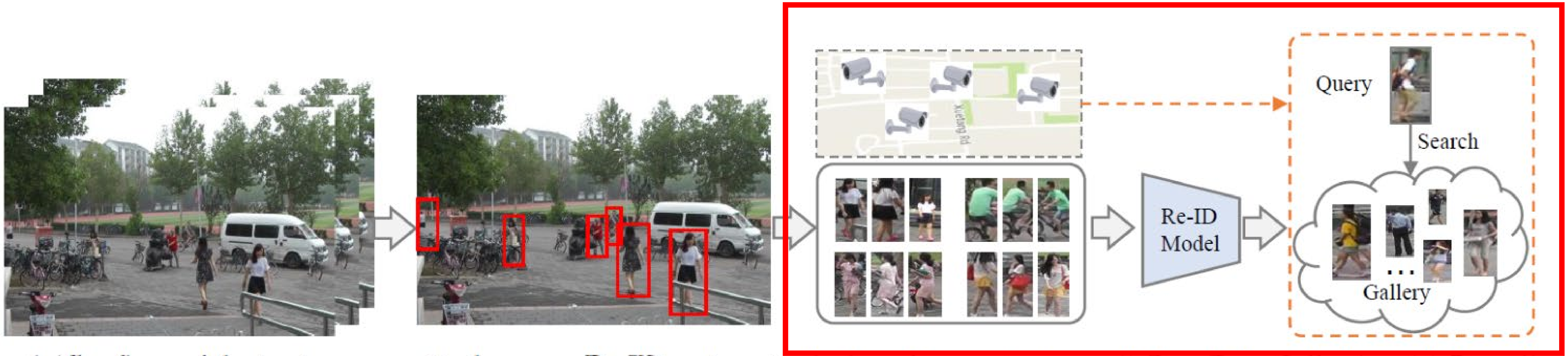


Fig. 1: The flow of designing a practical person Re-ID system, including five main steps: 1) *Raw Data Collection*, (2) *Bounding Box Generation*, 3) *Training Data Annotation*, 4) *Model Training* and 5) *Pedestrian Retrieval*.

SGVR Lab KAIST

Ye et al. Deep Learning for Person Re-identification: A Survey and Outlook. In TPAMI 2021.

Person Re-identification (Person Re-ID)

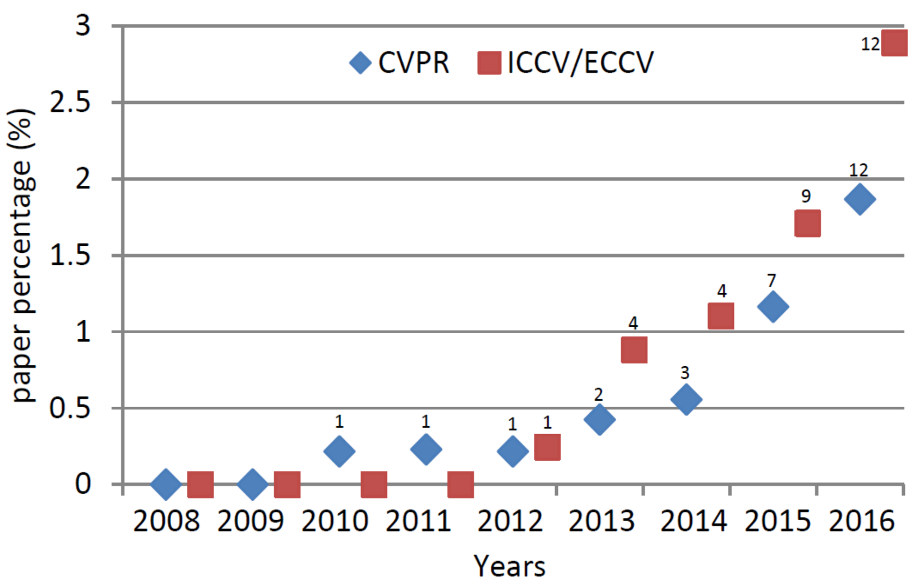
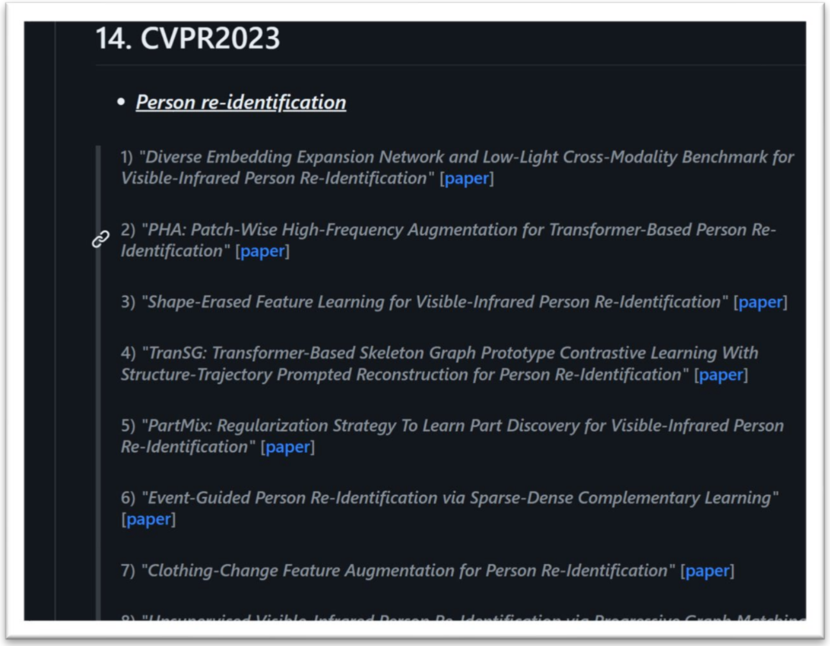


Fig. Percentage of person re-ID papers on top conferences over the years. Numbers above the markers indicate the number of re-ID papers.

Conference	Link	#Total	Person Re-ID
ICCV2021	click	34	24
CVPR2021	click	32	25
ECCV2020	Click	30	23
CVPR2020	Click	34	24
ICCV2019	Click	39	33
CVPR2019	Click	29	21
ECCV2018	Click	19	15
CVPR2018	Click	31	30
ICCV2017	Click	16	14
CVPR2017	Click	16	14



14. CVPR2023

- *Person re-identification*

1) *"Diverse Embedding Expansion Network and Low-Light Cross-Modality Benchmark for Visible-Infrared Person Re-Identification"* [paper]

2) *"PHA: Patch-Wise High-Frequency Augmentation for Transformer-Based Person Re-Identification"* [paper]

3) *"Shape-Erased Feature Learning for Visible-Infrared Person Re-Identification"* [paper]

4) *"TranSG: Transformer-Based Skeleton Graph Prototype Contrastive Learning With Structure-Trajectory Prompted Reconstruction for Person Re-Identification"* [paper]

5) *"PartMix: Regularization Strategy To Learn Part Discovery for Visible-Infrared Person Re-Identification"* [paper]

6) *"Event-Guided Person Re-Identification via Sparse-Dense Complementary Learning"* [paper]

7) *"Clothing-Change Feature Augmentation for Person Re-Identification"* [paper]

Awesome Person Re-identification (Person ReID), github

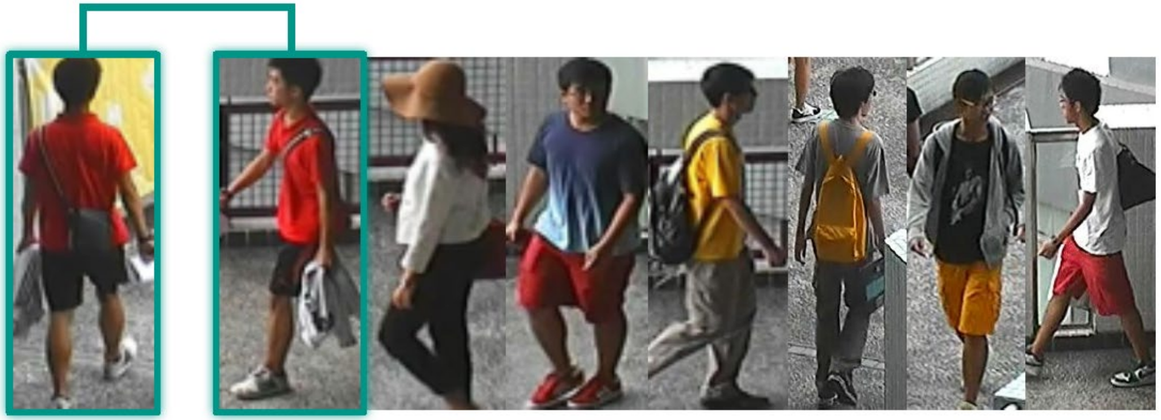
Zheng et al. Person Re-identification: Past, Present and Future. In arXiv 2016.
https://github.com/bismex/Awesome-person-re-identification.

SGVR Lab KAIST

Person Search

- Task to detect the person of interest from the entire image
- We need to detect for the target person from a gallery of whole scene images before doing a re-ID

(b) Person search: finding from whole scene images



(a) Person re-id: matching with manually cropped pedestrians

Credit: Joint detection and identification feature learning for Person Search (CVPR 20)

SGVR Lab KAIST

Datasets

- The dataset scale (both #image and #ID) has increased rapidly.
- The camera number is greatly increased to approximate the large-scale camera network in practical scenarios.

			Image datasets				
Dataset	Time	#ID	#image	#cam.	Label	Res.	Eval.
VIPeR	2007	632	1,264	2	hand	fixed	CMC
iLIDS	2009	119	476	2	hand	vary	CMC
GRID	2009	250	1,275	8	hand	vary	CMC
PRID2011	2011	200	1,134	2	hand	fixed	CMC
CUHK01	2012	971	3,884	2	hand	fixed	CMC
CUHK02	2013	1,816	7,264	10	hand	fixed	CMC
CUHK03	2014	1,467	13,164	2	both	vary	CMC
Market-1501	2015	1,501	32,668	6	both	fixed	C&M
DukeMTMC	2017	1,404	36,411	8	both	fixed	C&M
Airport	2017	9,651	39,902	6	auto	fixed	C&M
MSMT17	2018	4,101	126,441	15	auto	vary	C&M
			Video datasets				
Dataset	time	#ID	#track(#bbox)	#cam.	label	Res.	Eval
PRID-2011	2011	200	400 (40k)	2	hand	fixed	CMC
iLIDS-VID	2014	300	600 (44k)	2	hand	vary	CMC
MARS	2016	1261	20,715 (1M)	6	auto	fixed	C&M
Duke-Video	2018	1,812	4,832 (-)	8	auto	fixed	C&M
Duke-Tracklet	2018	1,788	12,647 (-)	8	auto	C&M	
LPW	2018	2,731	7,694(590K)	4	auto	fixed	C&M
LS-VID	2019	3,772	14,943 (3M)	15	auto	fixed	C&M

Ye et al. Deep Learning for Person Re-identification: A Survey and Outlook. In TPAMI 2021.

Evaluation Metrics

- mean Average Precision (mAP)
- Cumulative Matching Characteristics (CMC-*k*, Rank-*k* matching accuracy); the probability that a correct match appears in the top-*k* retrieved results.

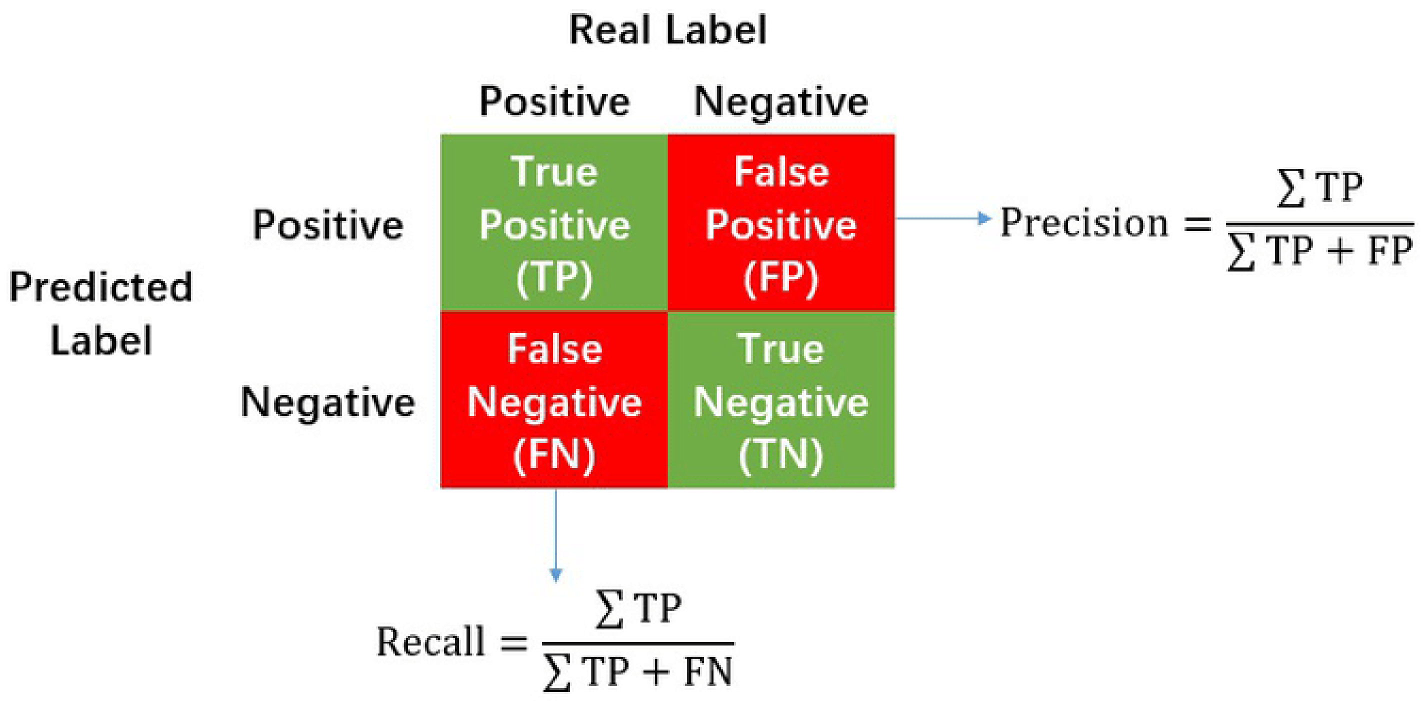


Image Ack; Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective, IEEE Access.

Evaluation Metrics

- mean Average Precision (mAP)
- Cumulative Matching Characteristics (CMC-*k*, Rank-*k* matching accuracy); the probability that a correct match appears in the top-*k* retrieved results.

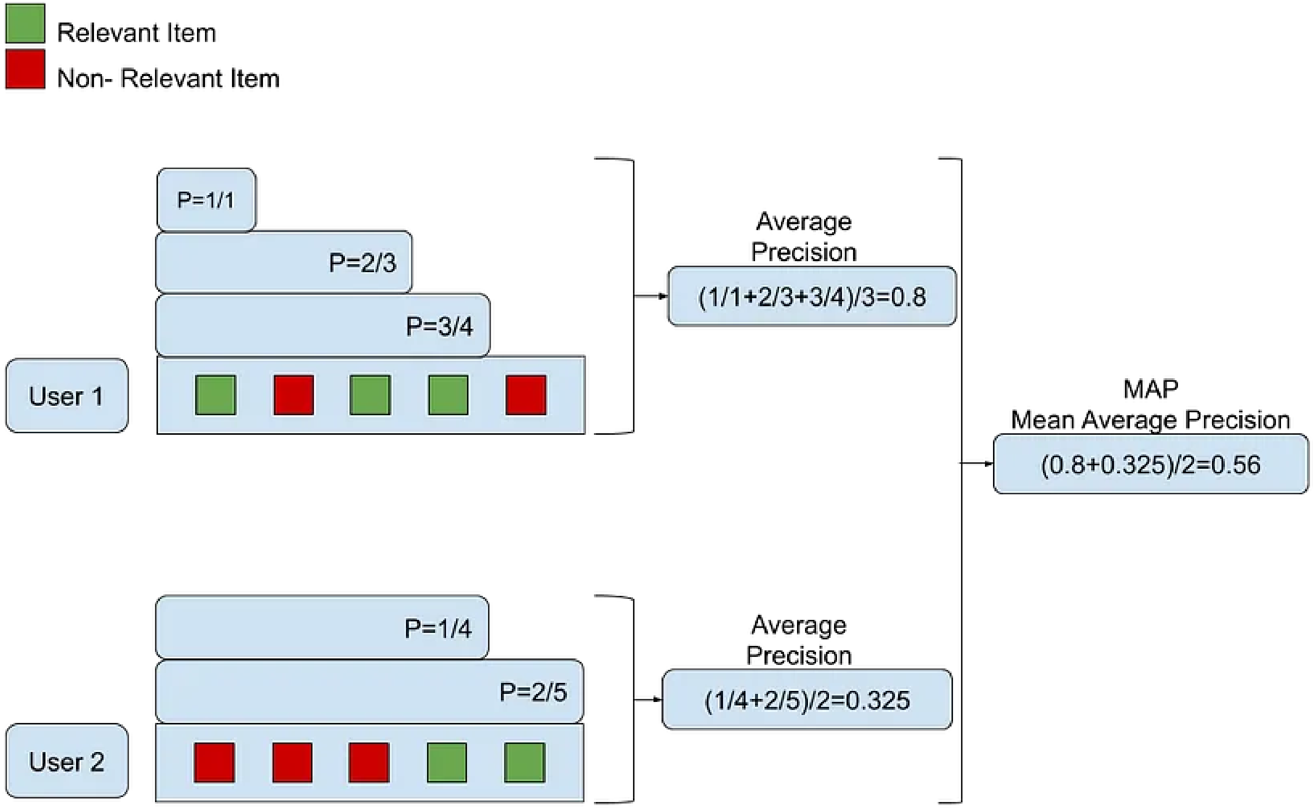


Image Ack; https://medium.com/swlh/rank-aware-recsys-evaluation-metrics-5191bba16832

Challenges in Person Re-ID

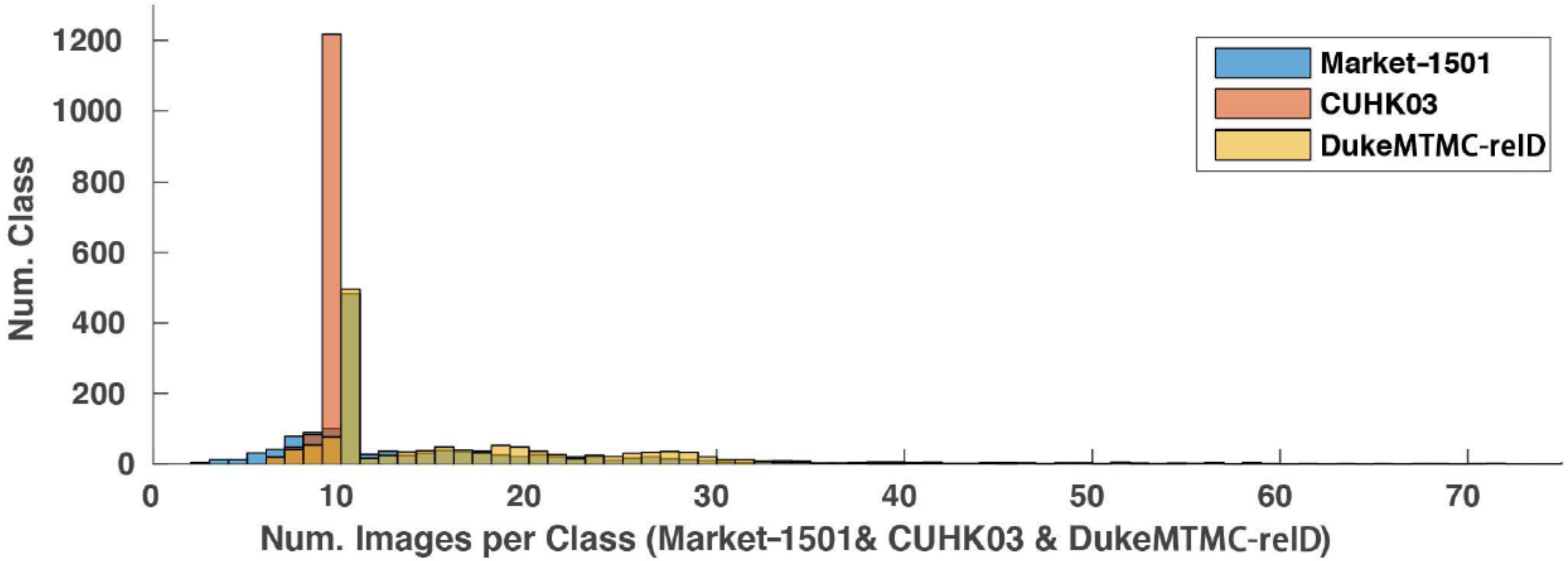
- Challenges by different camera views and time stamps.
  - Variance of viewpoints, illumination, pose, etc.
  - Occlusions.
  - Low resolutions.
- Large intra-variation & Small inter-variation

True match

False match

Challenges in Person Re-ID

- Long-tail problem.
  - In person re-ID, all datasets suffer from the insufficient training set.
  - Long-tail distribution training sets can yield unstable convergence and overfitting to head distributions.
  - MNIST 10 class/ 5000 per class, CIFAR 100 class/500 per class.

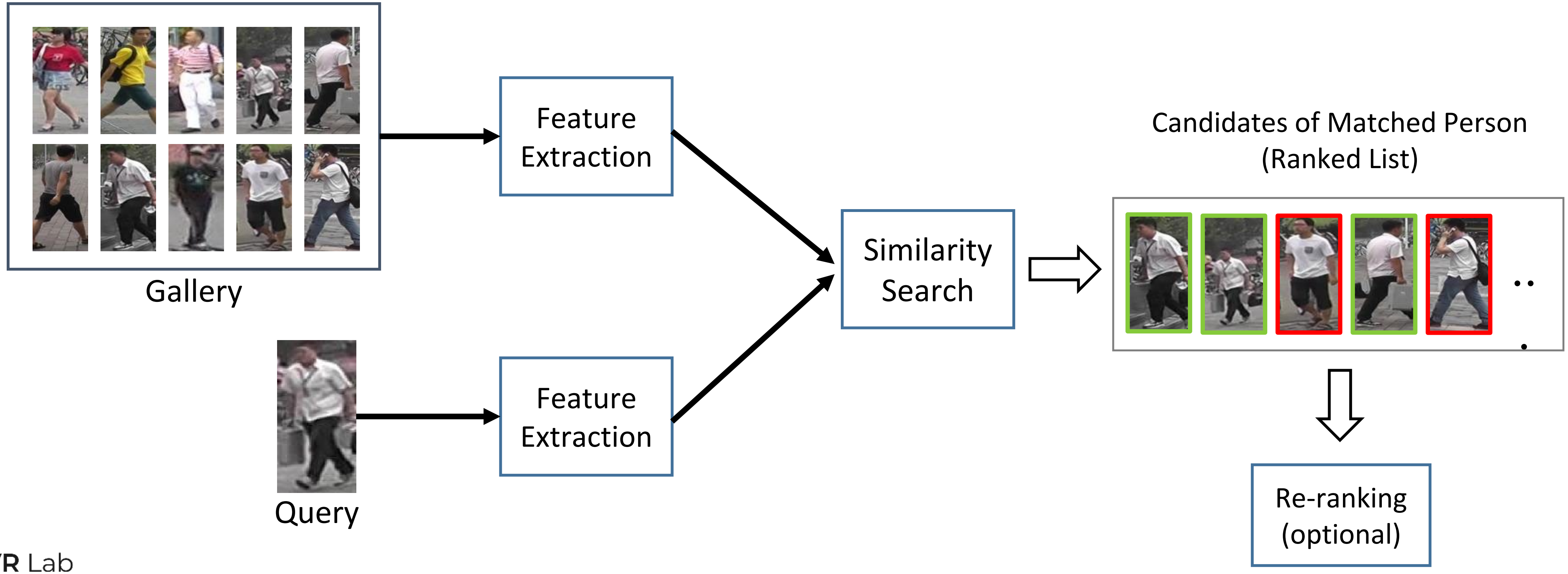


Zheng et al. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro. In ICCV 2017.

SGVR Lab KAIST

General Protocol of Person Re-ID

- Person re-identification pipeline.



SGVR Lab KAIST

Feature Representation Learning for Person Re-ID

- Most studies focus on **learning discriminative representation for person retrieval.**
- Recently, deep neural networks (DNN) have provided powerful descriptors.

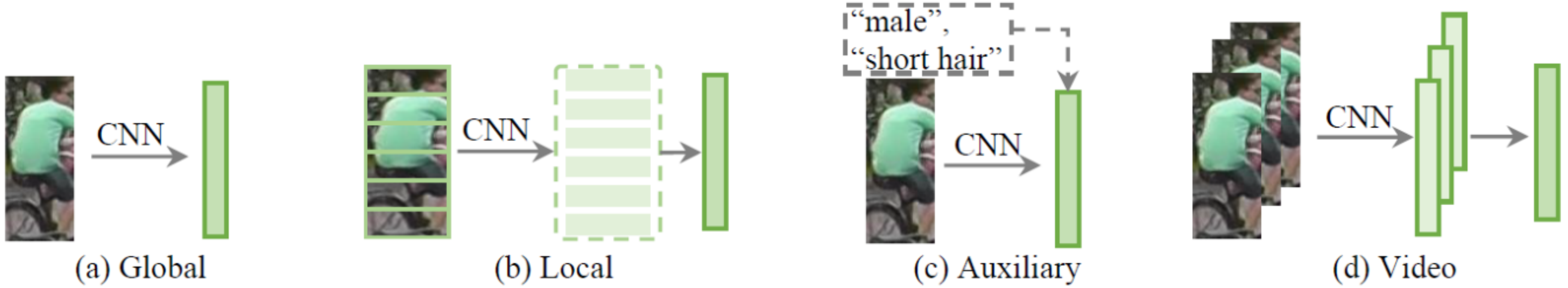
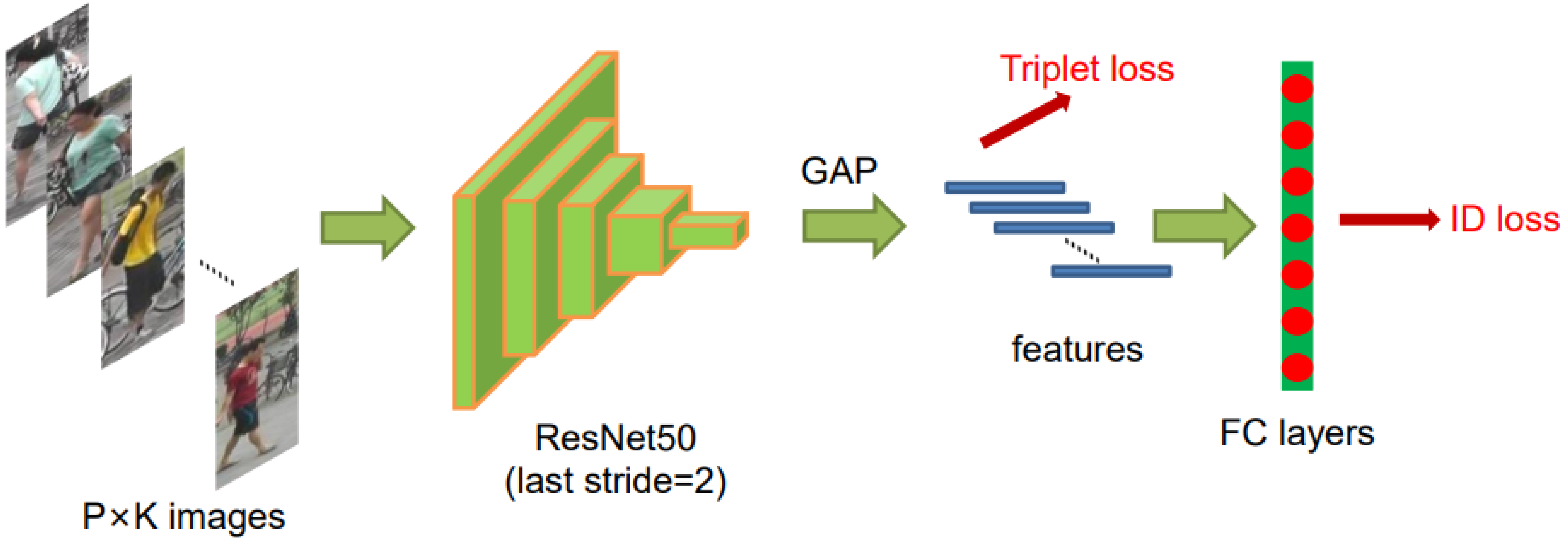


Fig. 2: Four different feature learning strategies. a) Global Feature, learning a global representation for each person image in § 2.1.1; b) Local Feature, learning part-aggregated local features in § 2.1.2; c) Auxiliary Feature, learning the feature representation using auxiliary information, *e.g.*, attributes [62], [63] in § 2.1.3 and d) Video Feature , learning the video representation using multiple image frames and temporal information [64], [65] in § 2.1.4.

Ye et al. Deep Learning for Person Re-identification: A Survey and Outlook. In TPAMI 2021.

Standard Approaches

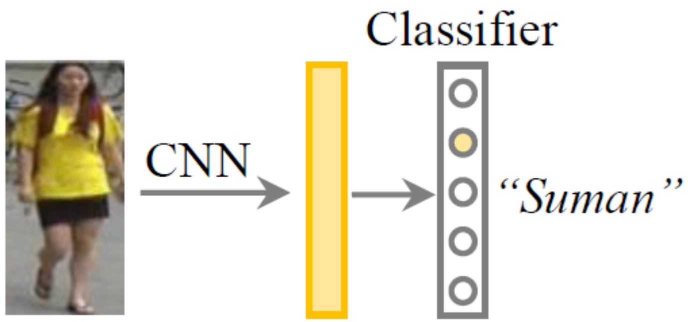
- Recent approaches **utilize both identification and triplet loss.**



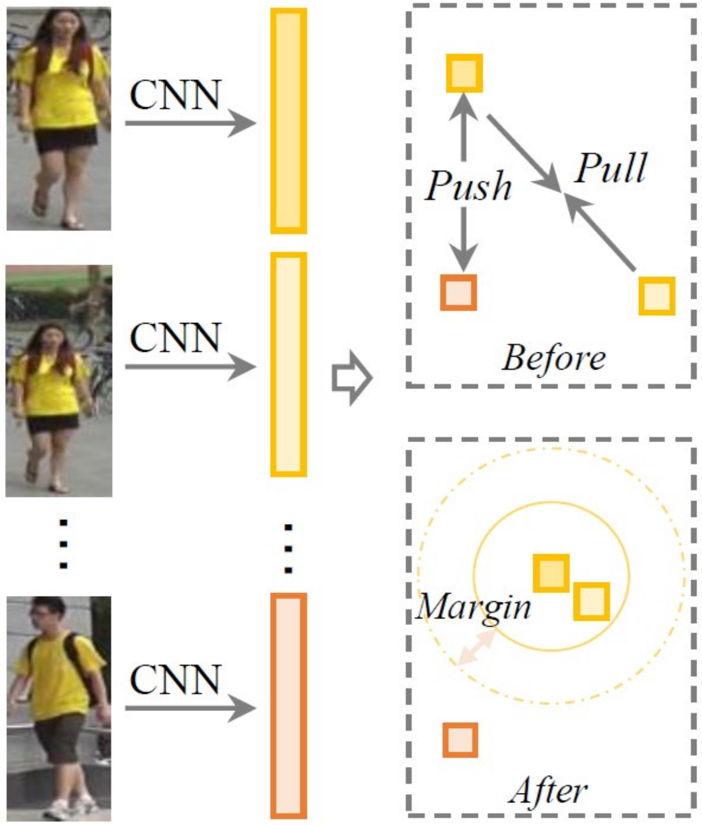
Luo et al. A Strong Baseline and Batch Normalization Neck for Deep Person Re-identification. In Transactions on Multimedia 2020.

Standard Approaches

- Recent approaches **utilize both identification and triplet loss.**



Identification loss



Triplet loss

Ye et al. Deep Learning for Person Re-identification: A Survey and Outlook. In TPAMI 2021.

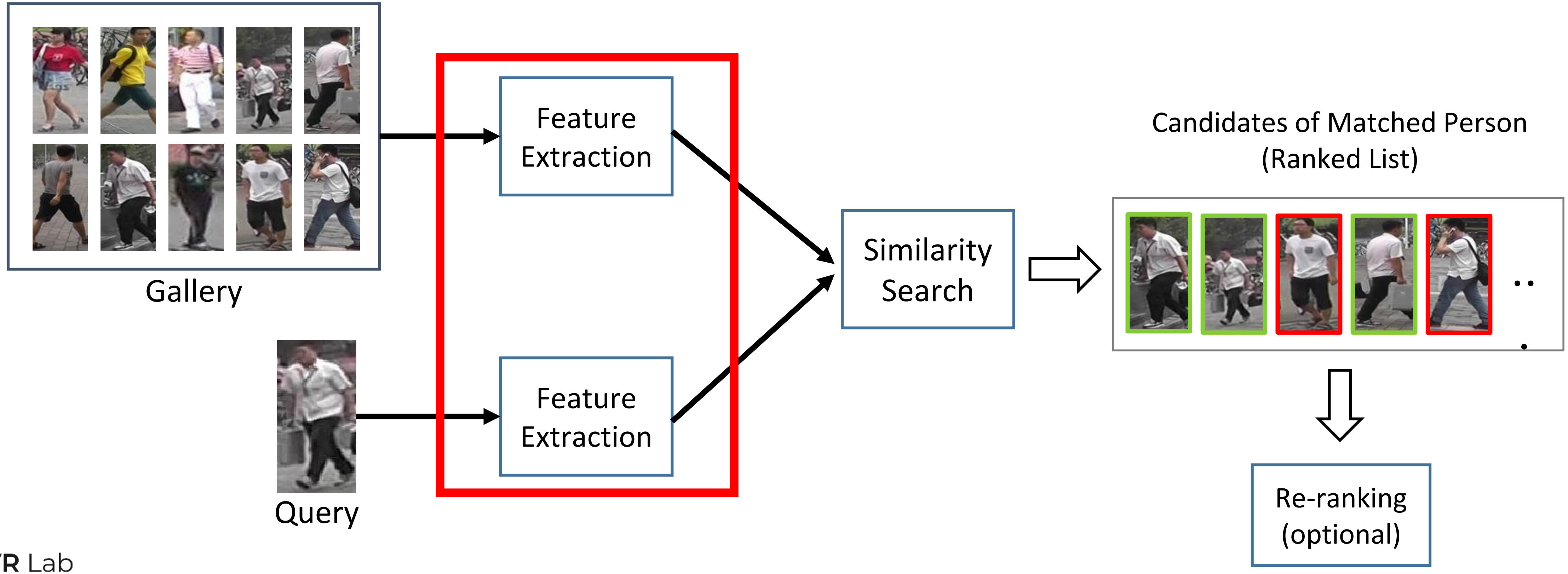
Unsupervised Person Re-identification

Recent techniques for unsupervised approaches

SGVR Lab KAIST

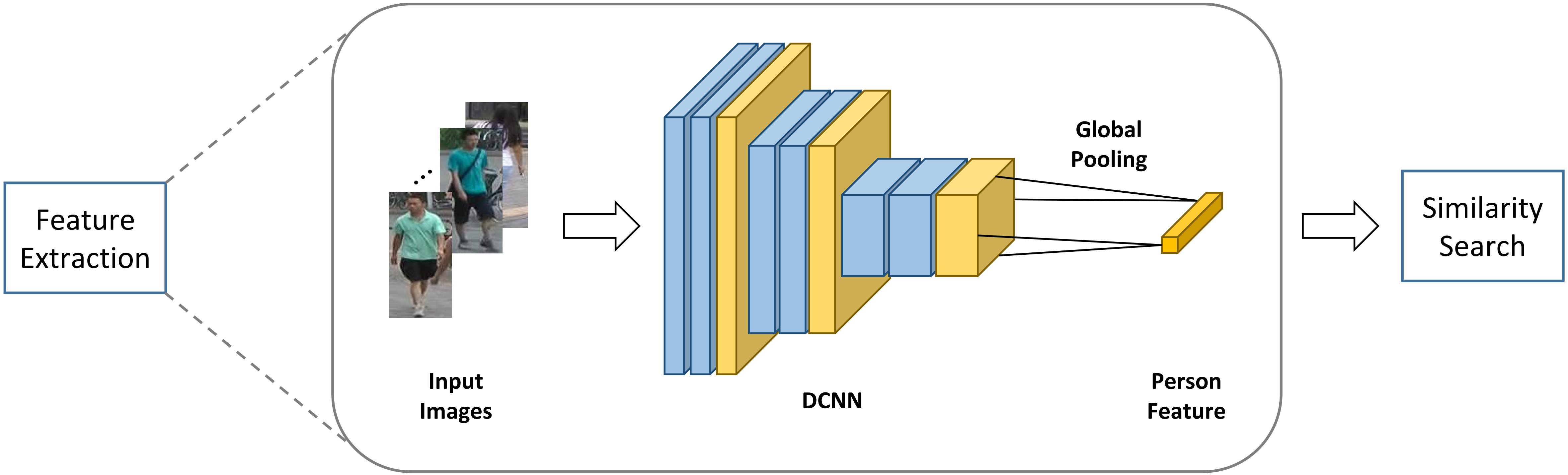
General Protocol of Person Re-ID

- Person re-identification pipeline.



General Protocol of Person Re-ID

- Deep convolutional neural network (DCNN) brings impressive improvements in person re-ID fields.

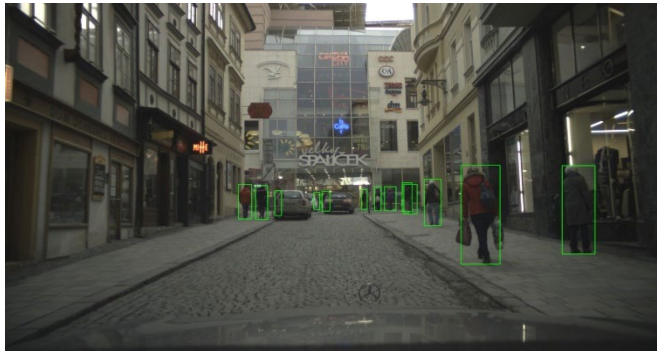


SGVR Lab KAIST

Problems in DCNN

- Require many **training data with labels.**
- Challenges in identity annotation.
  - illumination changes.
  - Low resolution.
  - **Occlusions.**

Camera view



Person Detection

Identity Annotation

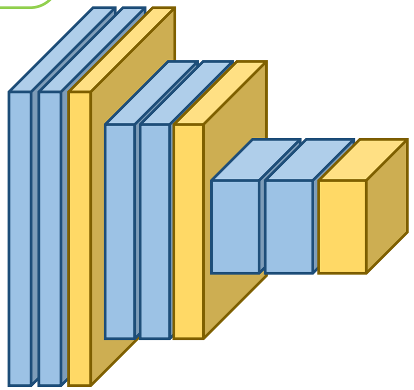
Problems in DCNN

- The real-world scenario of person re-ID is an open set problem.
- New people (= new class) will appear from the camera views.

Someone is detected!
Who is he/she?

Sorry. We don't know.
It's the first time we see him/her.

Surveillance system



Person re-ID system

Problem Setting

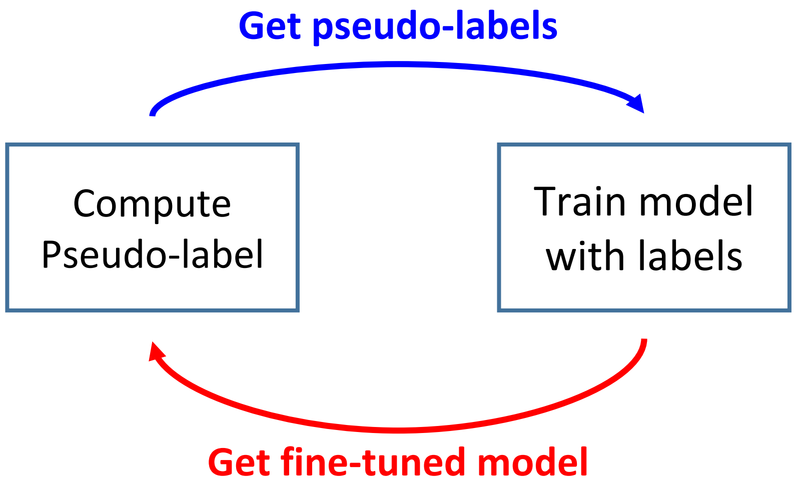
- Goal: Learn discriminative features for person retrieval **without ID labels.**

- Protocol: Training on target domain w/o labels → Testing on target domain.

- Challenges: Poor pseudo-supervision from unlabeled data.

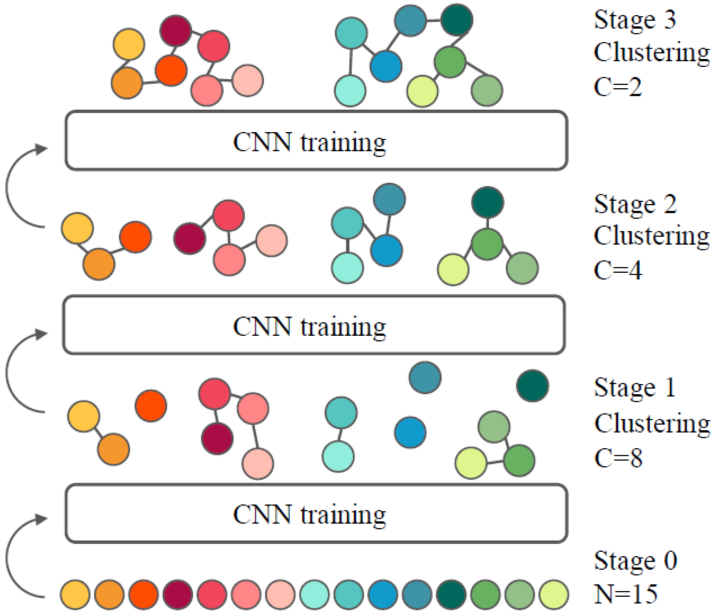
Pseudo Label-based Approaches

- Most recent studies **utilize pseudo-labels to train a re-ID model.**
  - K-nearest neighbor search; regard k-NN as the same class.
  - Clustering; regard each cluster as a class
- Clustering-based approaches have shown state-of-the-art results.

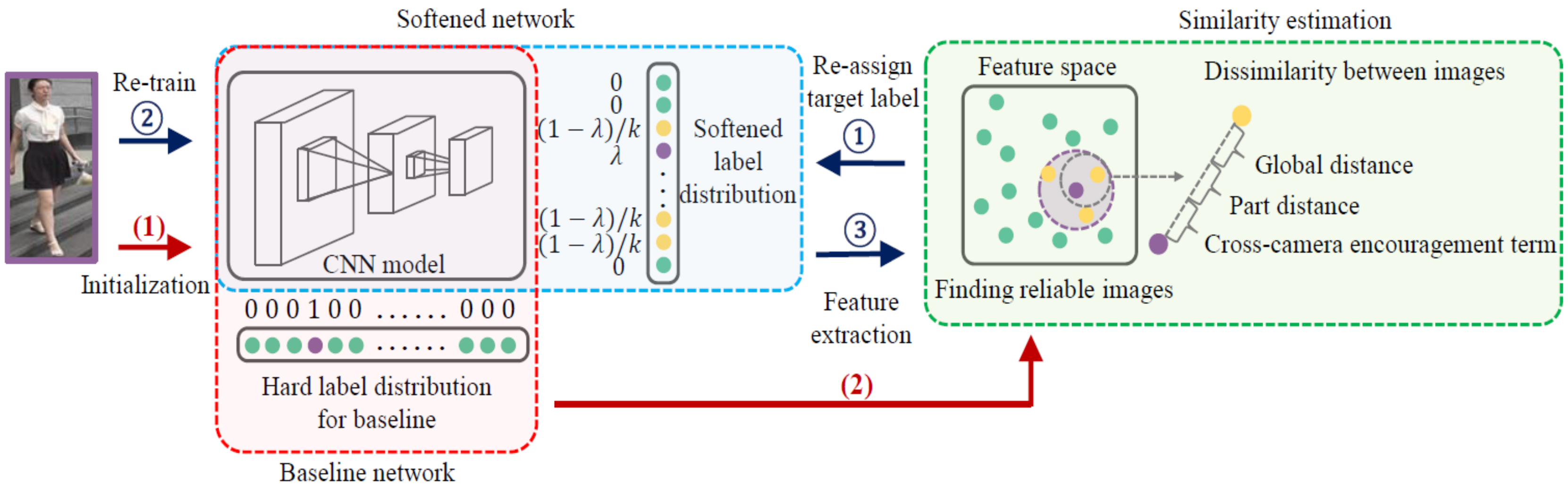


Pseudo Label-based Approaches

- In early studies of this field focus on **how to obtain pseudo-labels.**
- Nowadays, most methods utilize **DBSCAN** clustering with **re-ranked distances** (performed by k-NN neighboring, etc.).



Bottom-Up Clustering (AAAI 19)
Different clustering



Softened Similarity Learning (CVPR 20)
Different distance measure

SGVR Lab
KAIST

Lin et al. A Bottom-Up Clustering Approach to Unsupervised Person Re-Identification. In AAAI 2019.
Lin et al. Unsupervised Person Re-identification via Softened Similarity Learning. In CVPR 2020.

Pseudo Label-based Approaches

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
  - **Red; core point (when minPts = 4) that are closed to other nearby points**
  - **Yellow; not core points, but are reachable from A (belong to the same cluster)**
  - **Blue; neither a core point nor directly-reachable from A (not in the cluster)**

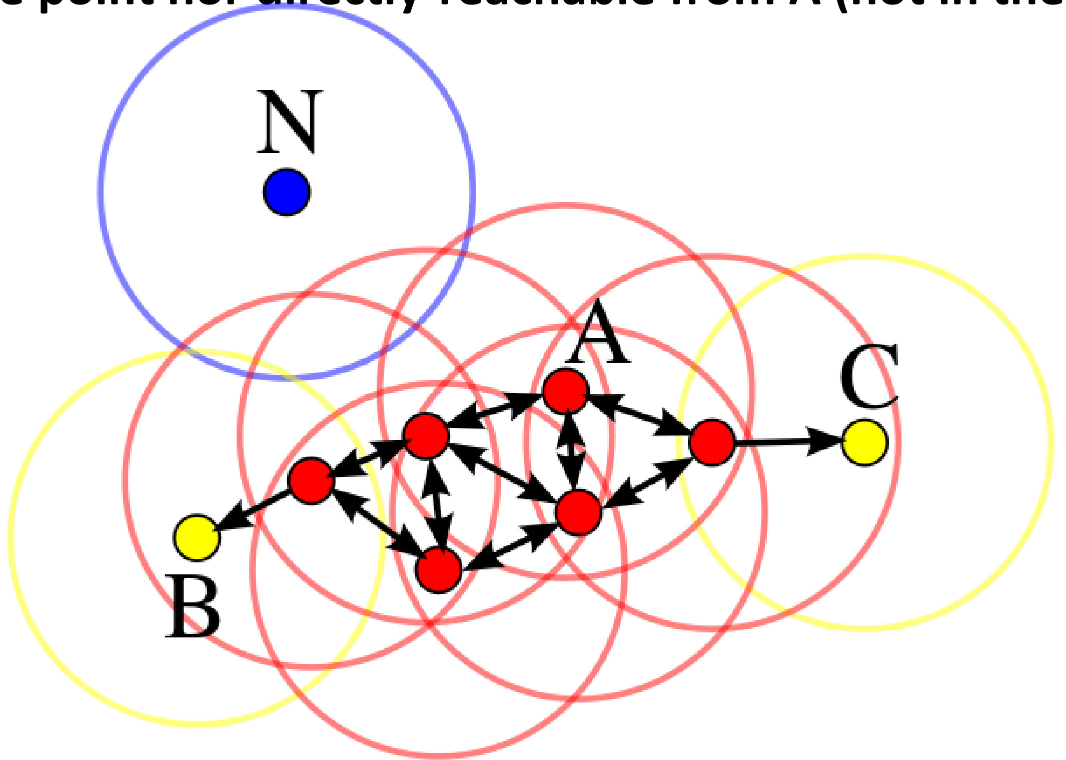
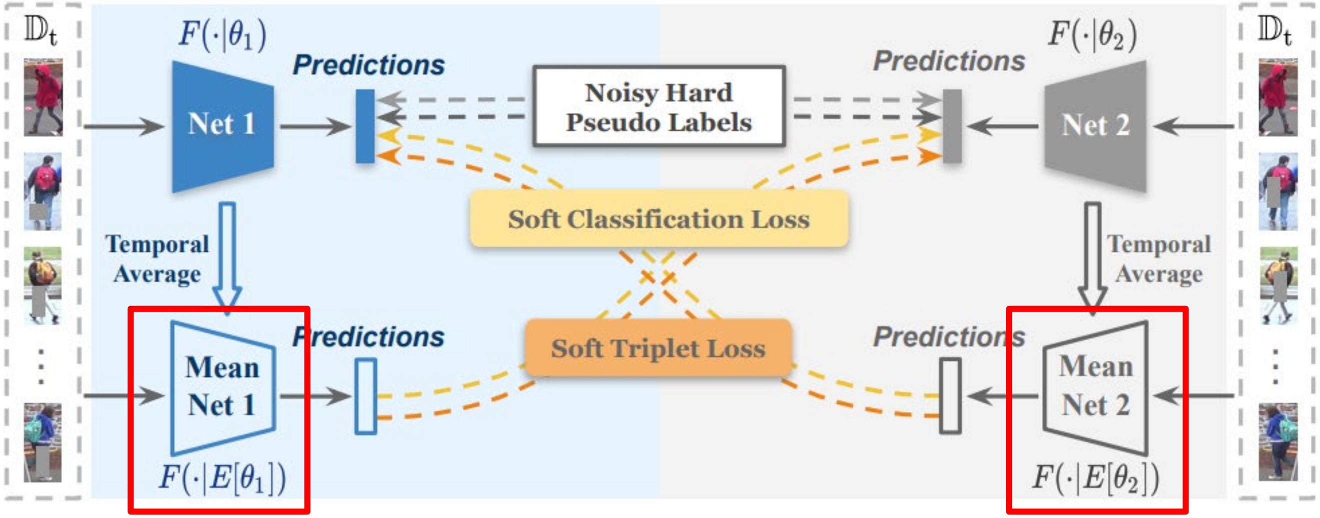


Image Ack; Wikipedia

Pseudo Label Refinement

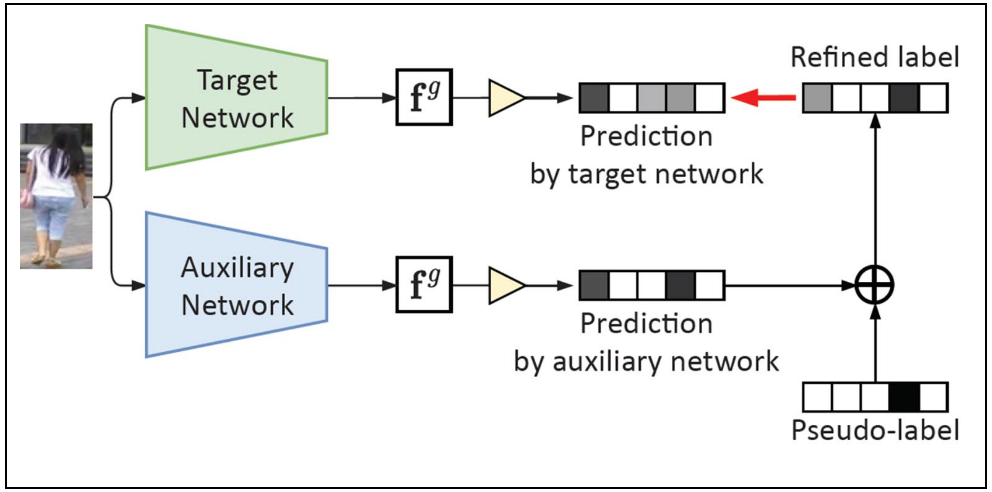
- Key idea: Re-ID performance ∝ Quality of pseudo-labels.
- There are inevitable noised in pseudo-labels (noisy label problem), and some studies utilize predictions of an auxiliary network to refine labels (to avoid bias due to noise)



Auxiliary Network

MMT (ICLR 20)

Auxiliary Network



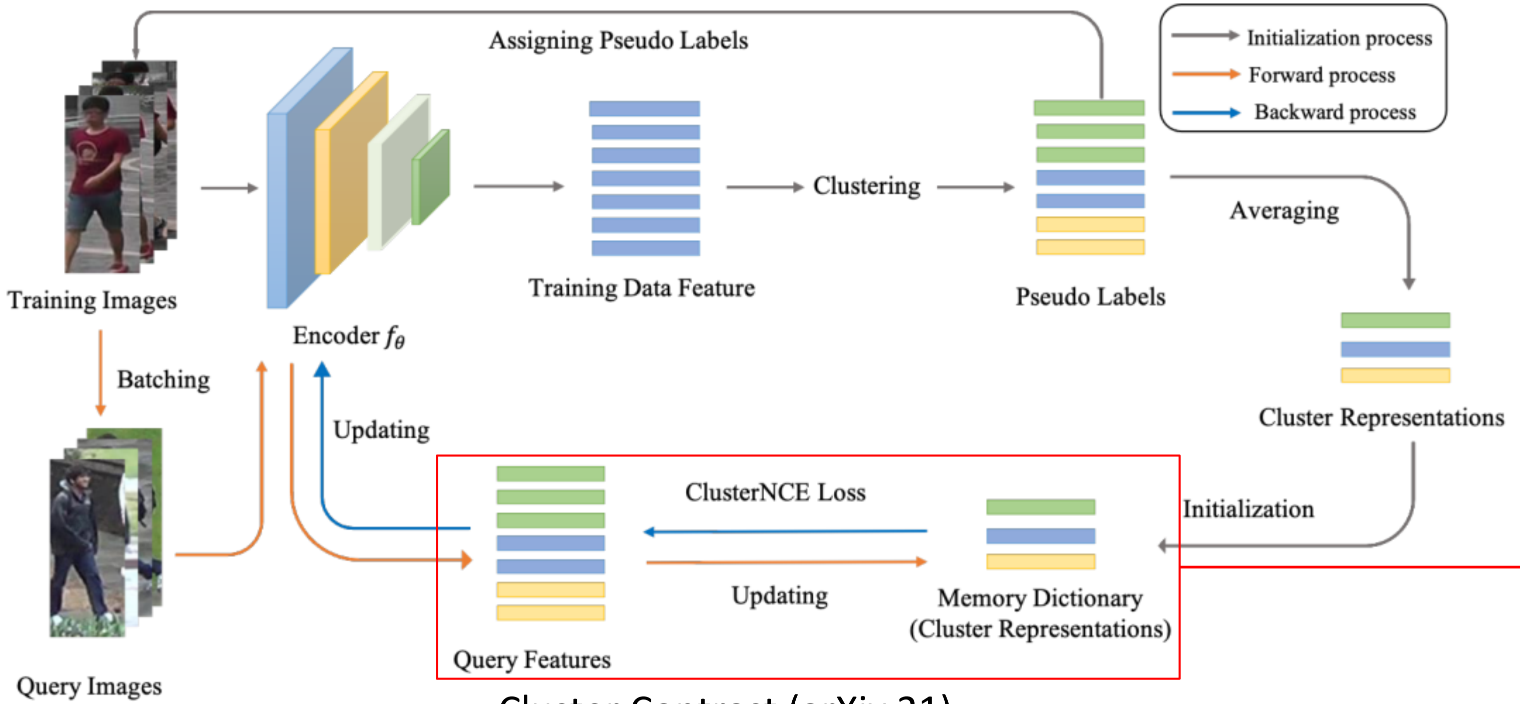
Label Refinement Process

SGVR Lab KAIST

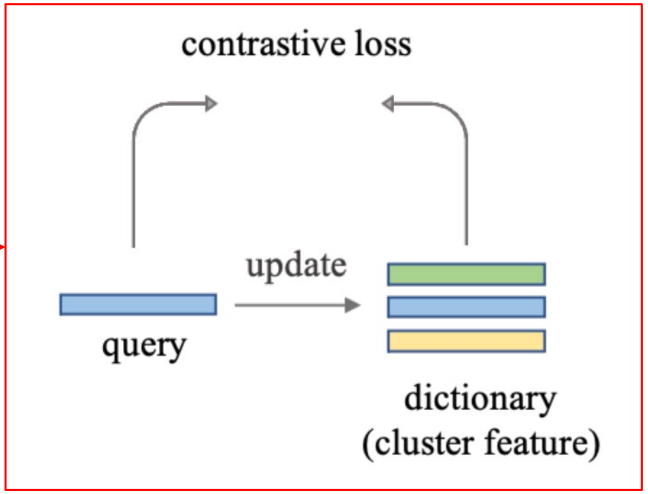
Ge et al. Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification. In ICLR 2020.

Cluster-based Contrastive Learning

- Key idea: Utilize clustering results for contrastive learning which is demonstrated its effectiveness in various unsupervised (self-supervised) tasks.
- Apply a contrastive learning in cluster-level



Cluster Contrast (arXiv 21)



Apply contrastive learning with cluster centroid features: Make q belong to its cluster closer, while pushing from other clusters

SGVR Lab KAIST

Dai et al. Cluster Contrast for Unsupervised Person Re-Identification. arXiv preprint arXiv:2103.11568, 2021.

SimCLR

[Chen et al, Asimple framework for contrastive learning of visual representations, ICML'20]

Maximizing the agreement of representations under data transformation, using a contrastive loss in the latent/feature space.

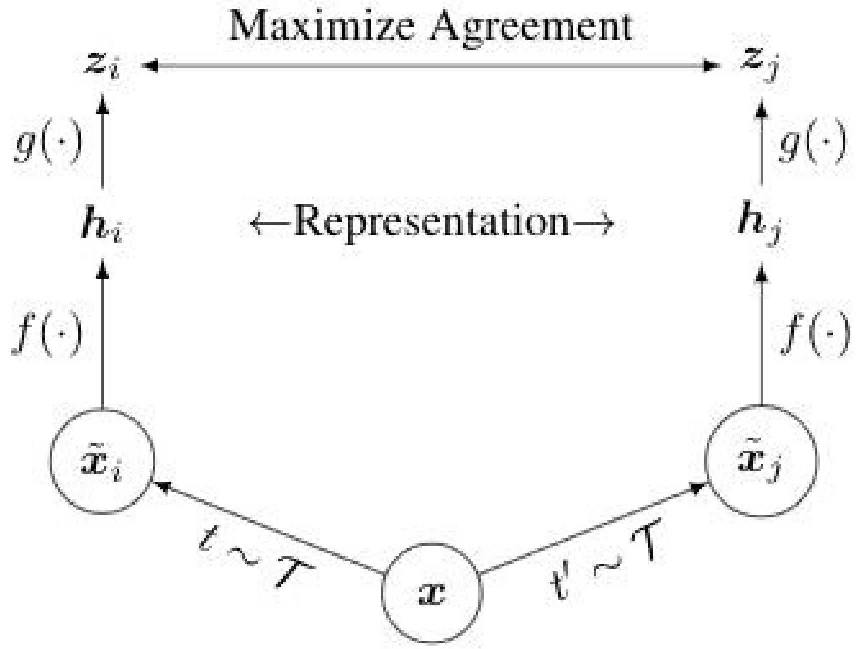


Figure 2. A framework for contrastive representation learning. Two separate stochastic data augmentations $t, t' \sim \mathcal{T}$ are applied to each example to obtain two correlated views. A base encoder network $f(\cdot)$ with a projection head $g(\cdot)$ is trained to maximize agreement in *latent representations* via a contrastive loss.

Google Research

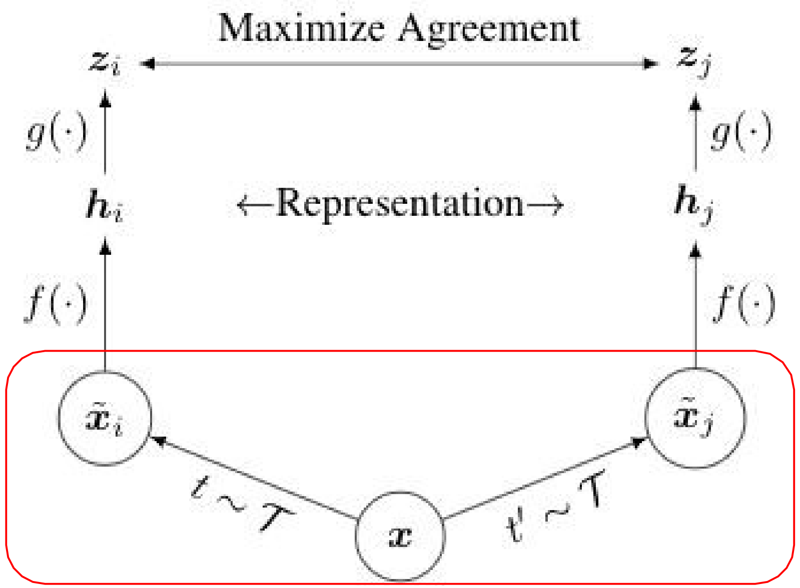
Semi-supervised learning

SimCLR as an example: strong semi-supervised learners, outperforms AlexNet with 100X fewer labels.

Method	Architecture	Label fraction 1%	10%
		Top 5	
Methods using other label-propagation:			
Pseudo-label	ResNet50	51.6	82.4
VAT+Entropy Min.	ResNet50	47.0	83.4
UDA (w. RandAug)	ResNet50	-	88.5
FixMatch (w. RandAug)	ResNet50	-	89.1
S4L (Rot+VAT+En. M.)	ResNet50 (4×)	-	91.2
Methods using representation learning only:			
InstDisc	ResNet50	39.2	77.4
BigBiGAN	RevNet-50 (4×)	55.2	78.8
PIRL	ResNet-50	57.2	83.8
CPC v2	ResNet-161(∗)	77.9	91.2
Ours	ResNet-50	75.5	87.8
Ours	ResNet-50 (2×)	83.0	91.2
Ours	ResNet-50 (4×)	**85.8**	**92.6**

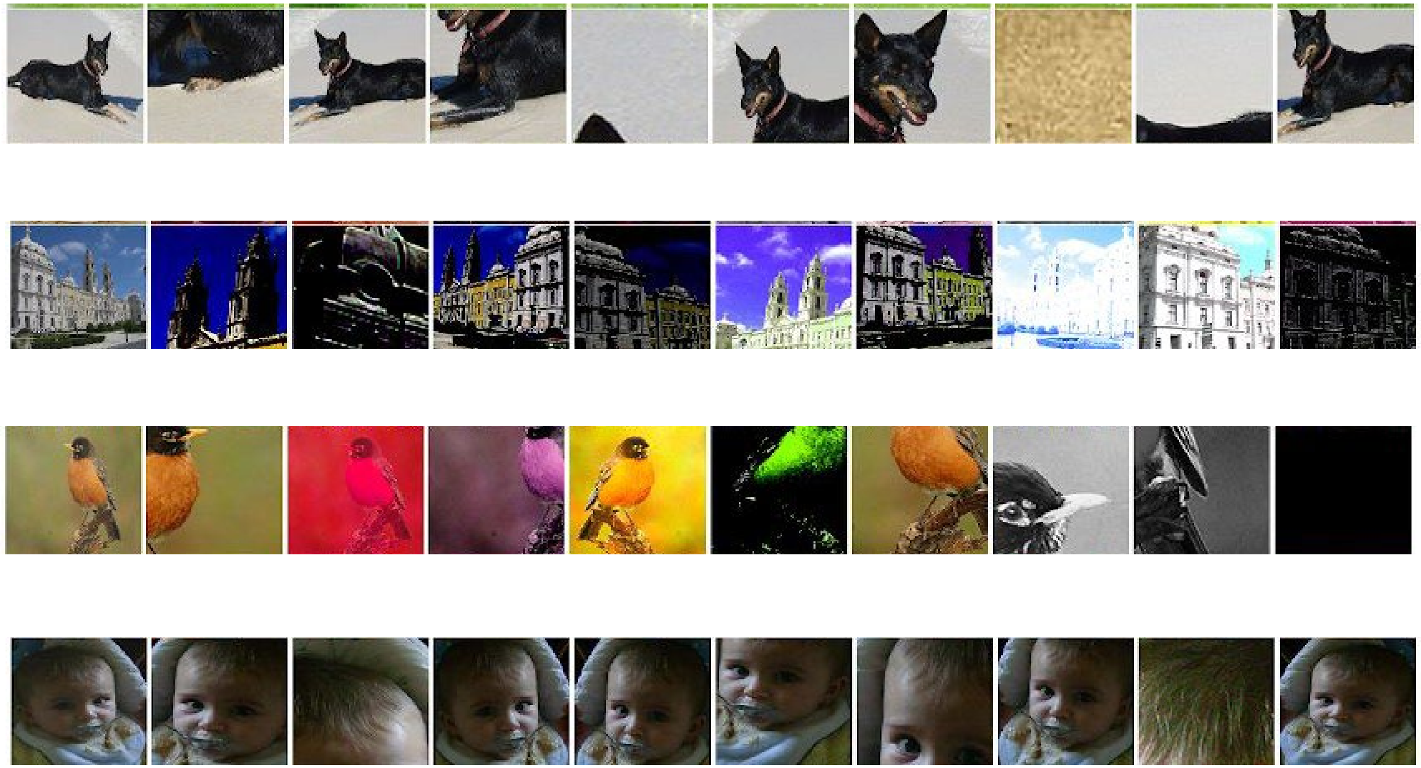
Table 7. ImageNet accuracy of models trained with few labels.

SimCLR component: data augmentation

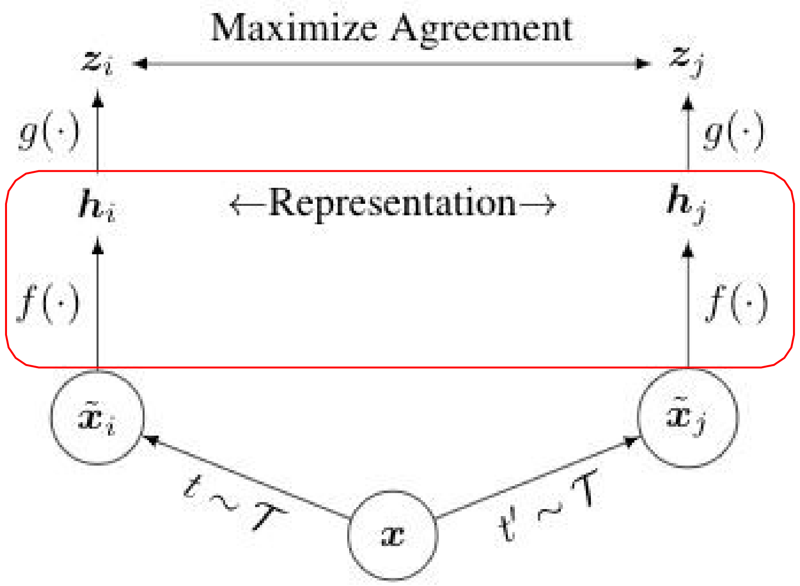


We use random crop and color distortion for augmentation.

Examples of augmentation applied to the left most images:

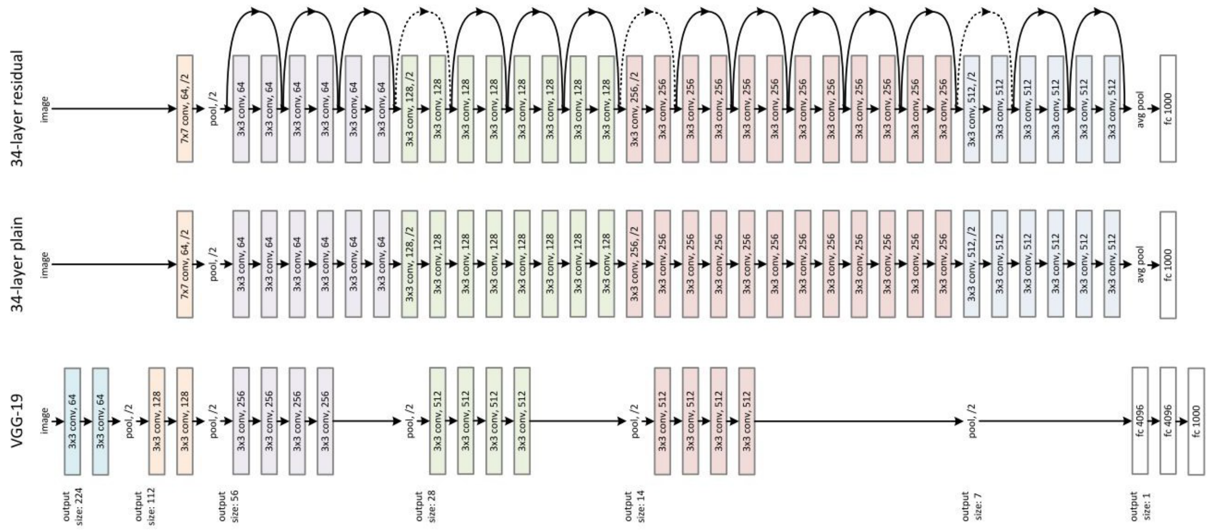


SimCLR component: encoder



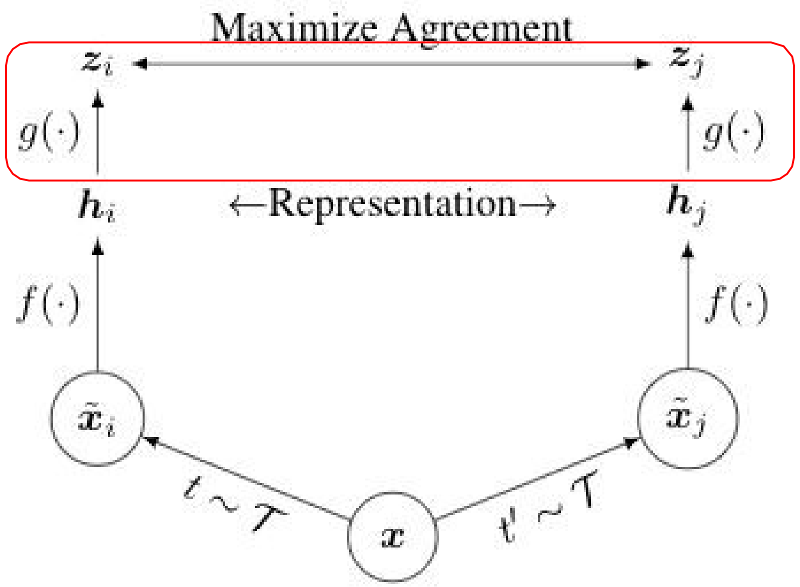
f(x) is the base network that computes internal representation.

We can use (unconstrained) ResNet in this work. However, it can be other networks.



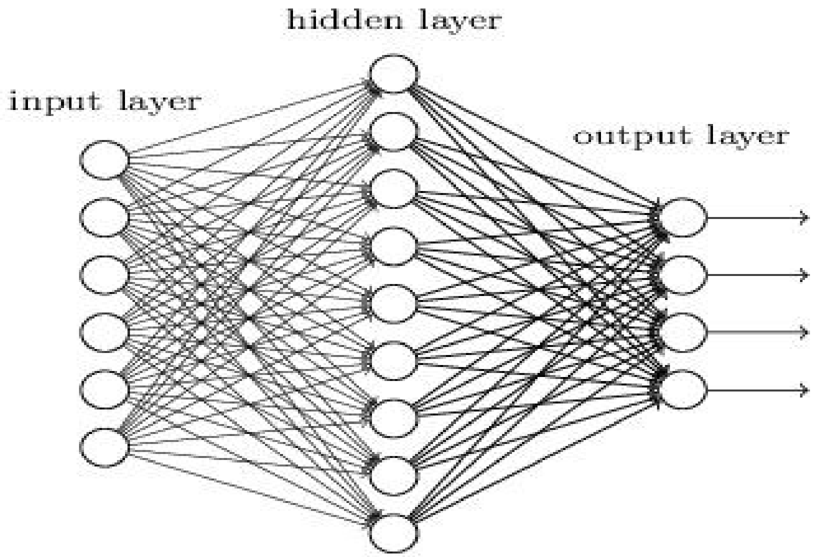
Google Research

SimCLR component: projection head

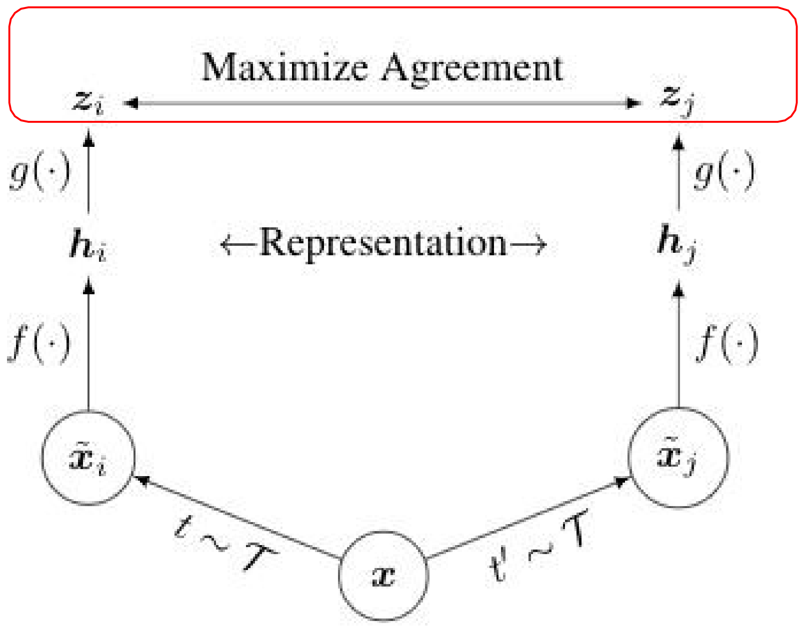


g(h) is a projection network that project representation to a latent space.

We use a MLP (with non-linearity).



SimCLR component: contrastive loss



Maximize agreement using a contrastive task:

Given {x_k} where two different examples x_i and x_j are a positive pair, identify x_j in {x_k}_{k!=i} for x_i.

Original image　crop 1　crop 2　contrastive image

Loss function:

$$\text{Let } \mathrm{sim}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^\top \boldsymbol{v} / \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

Part-based Pseudo Label Refinement for Unsupervised Person Re-identification

Yoonki Cho, Woo Jae Kim, Seunghoon Hong, Sung-Eui Yoon

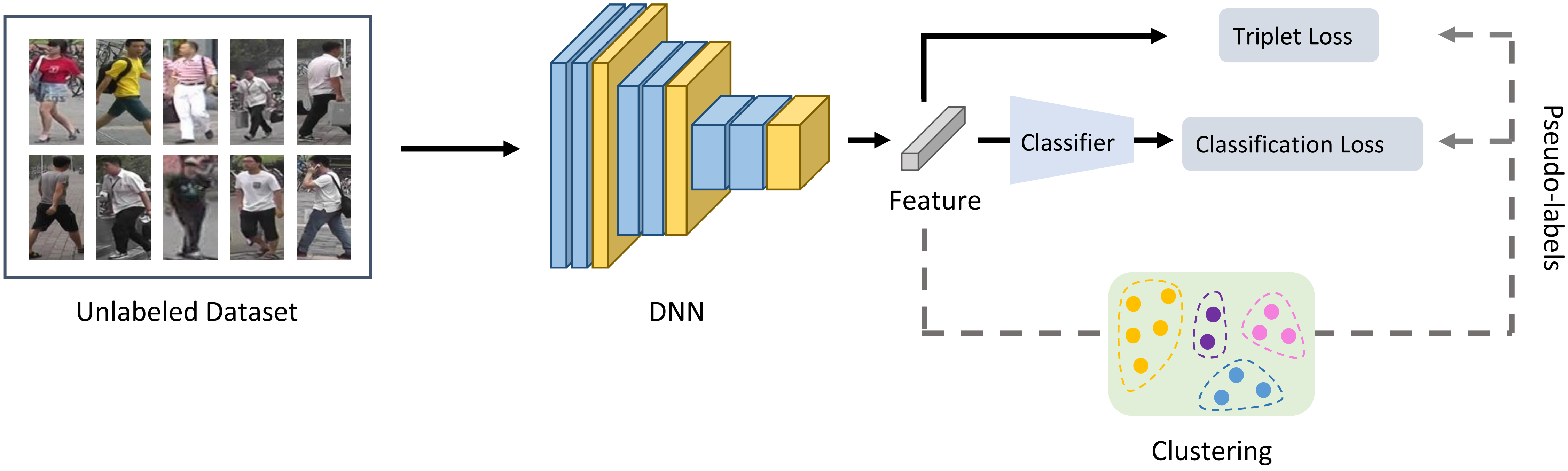
KAIST

CVPR 2022

SGVR Lab KAIST

Unsupervised Person Re-identification

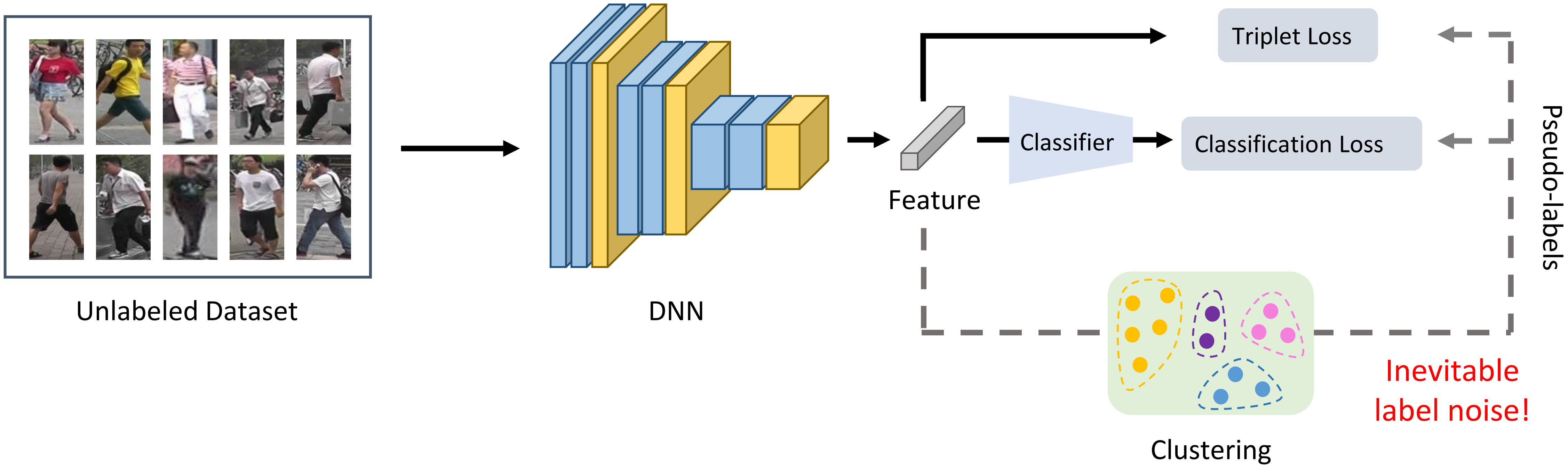
- Learn the discriminative features for person re-ID from unlabeled data



SGVR Lab KAIST

Unsupervised Person Re-identification

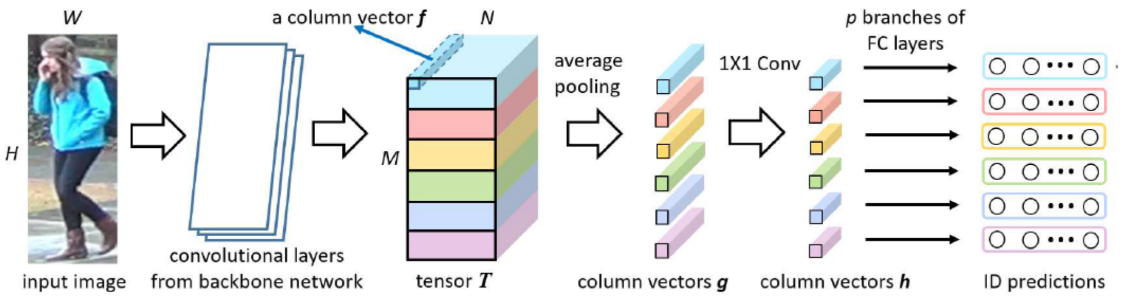
- Learn the discriminative features for person re-ID from unlabeled data



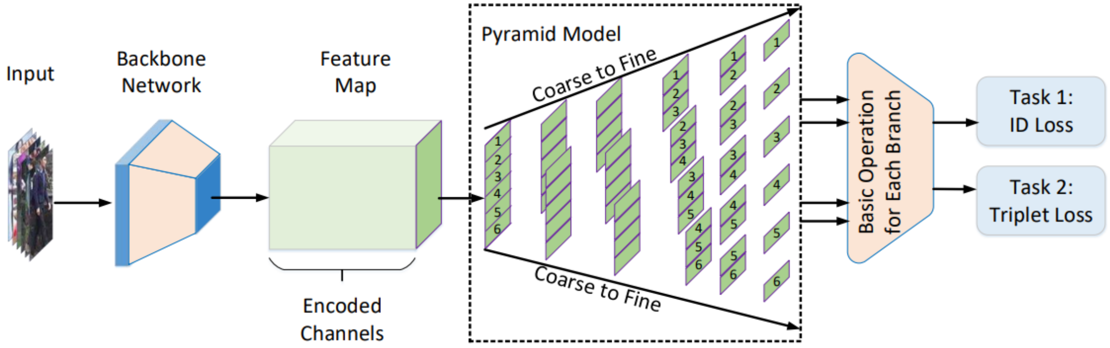
SGVR Lab KAIST

Motivation and Idea

- Existing works neglect fine-grained information essential to person re-ID



PCB [3]



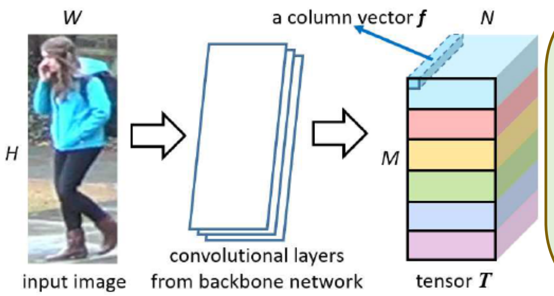
Pyramidal model [4]

SGVR Lab KAIST

[3] Sun et al. Beyond Part Models: Person Retrieval with Refined Part Pooling (and a strong convolutional baseline). In ECCV 2018.
[4] Zheng et al. Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training. In CVPR 2019.

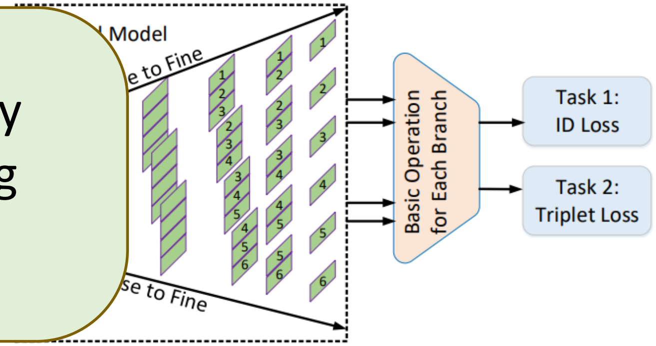
Motivation and Idea

- Existing works neglect fine-grained information essential to person re-ID



PCB [3]

Fine-grained information can provide not only useful cues for better representation learning but also **robustness against label noises**.



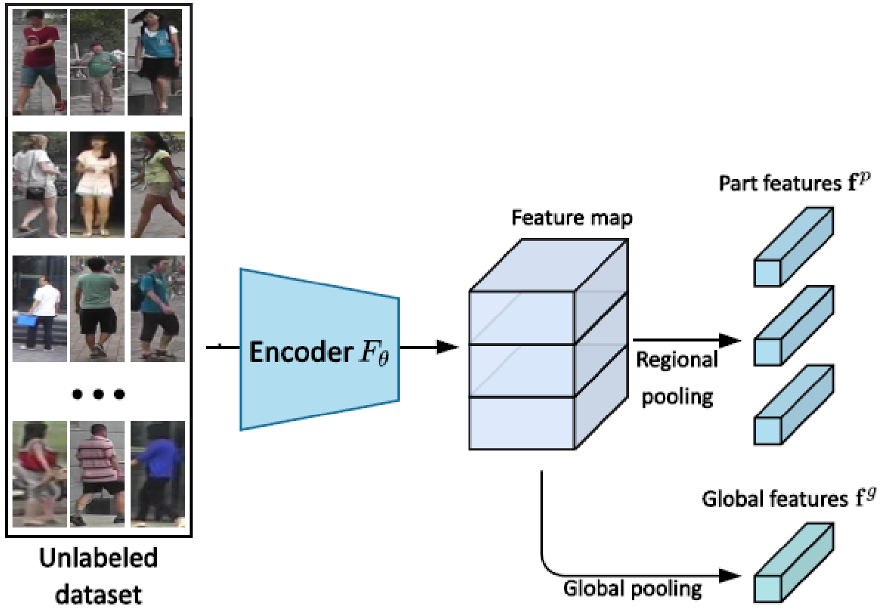
Pyramidal model [4]

[3] Sun et al. Beyond Part Models: Person Retrieval with Refined Part Pooling (and a strong convolutional baseline). In ECCV 2018.
[4] Zheng et al. Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training. In CVPR 2019.

SGVR Lab KAIST

Overview

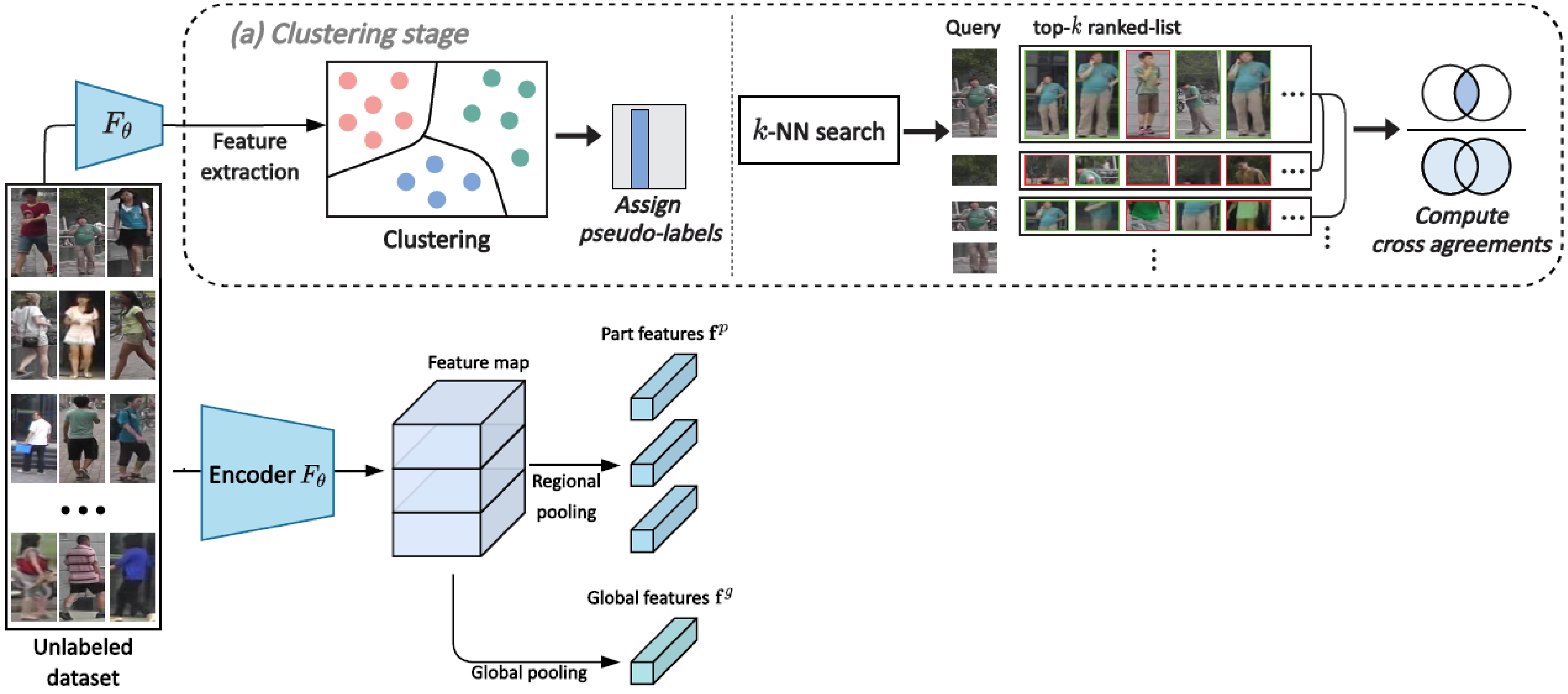
- **P**art-based **P**seudo **L**abel **R**efinement (**PPLR**) framework



SGVR Lab KAIST

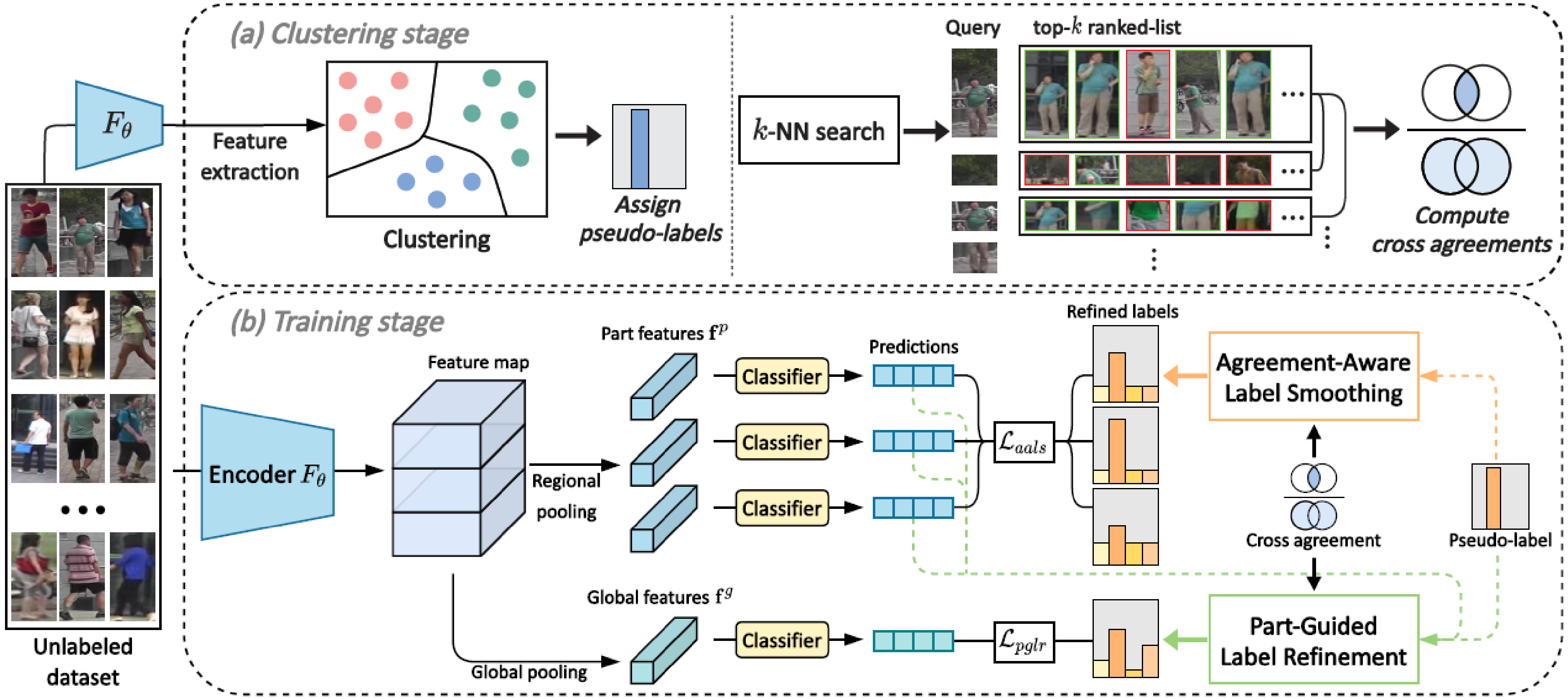
Overview

- **P**art-based **P**seudo **L**abel **R**efinement (**PPLR**) framework



Overview

- **P**art-based **P**seudo **L**abel **R**efinement (**PPLR**) framework



SGVR Lab KAIST

Cross Agreement Score

- Global and part features in the same image can capture very different semantic information

Examples of ID-166 of Market-1501

SGVR Lab KAIST

Cross Agreement Score

- Global and part features in the same image can capture very different semantic information

Examples of ID-166 of Market-1501

Using complementary relationship naïvely can result in unreliable information!

SGVR Lab KAIST

Cross Agreement Score

- Cross agreement score C_i between global feature space g and part feature space p_n for the image x_i
  - Jaccard similarity of nearest neighbors between global and part features

$$C_i(g, p_n) = \frac{|\underbrace{R_i(g,k)}_{k\text{-NN of global feature}} \cap \underbrace{R_i(p_n,k)}_{k\text{-NN of } n\text{-th part feature}}|}{|R_i(g,k) \cup R_i(p_n,k)|} \in [0,1]$$

SGVR Lab KAIST

Cross Agreement Score

- Cross agreement score C_i between global feature space g and part feature space p_n for the image x_i
  - Jaccard similarity of nearest neighbors between global and part features

$$C_i(g, p_n) = \frac{|R_i(g,k) \cap R_i(p_n,k)|}{|\underbrace{R_i(g,k)}_{k\text{-NN of global feature}} \cup \underbrace{R_i(p_n,k)}_{k\text{-NN of } n\text{-th part feature}}|} \in [0,1]$$

- $C_i(g, p_n) \uparrow$: g and p_n are highly correlated around the data i = reliable
- $C_i(g, p_n) \downarrow$: g and p_n are not correlated around the data i = unreliable

Agreement-Aware Label Smoothing (AALS)

- Smooths pseudo-labels according to a cross agreement score of each part

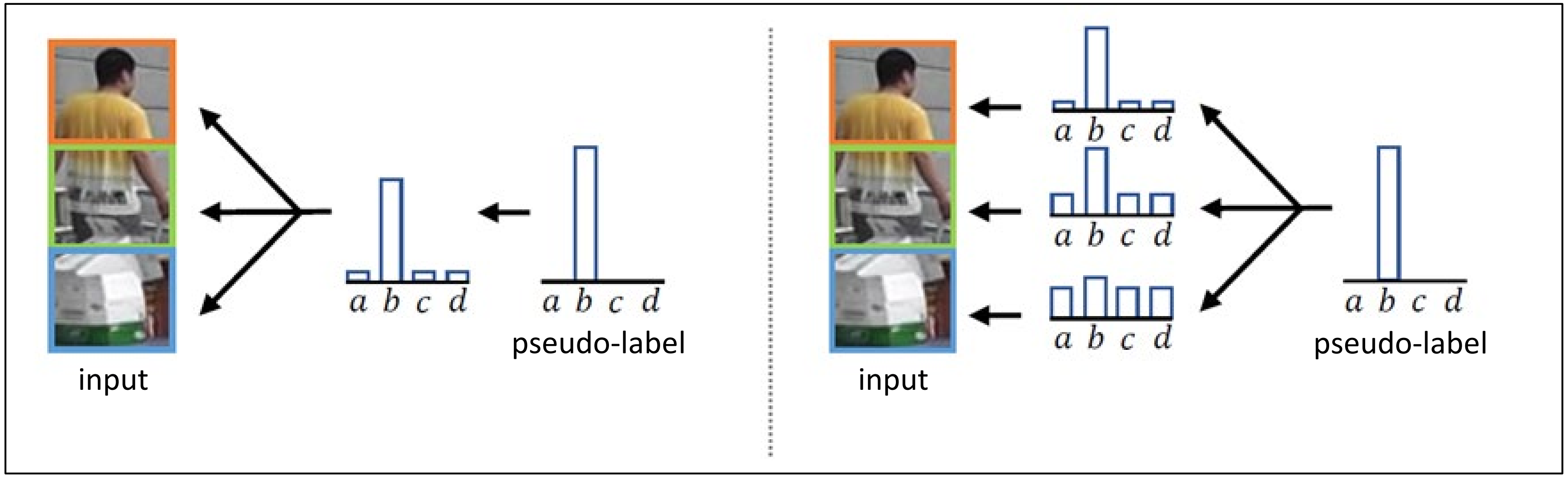
$$\tilde{y}_i^{p_n} = C_i(g, p_n) * \underbrace{y_i}_{\text{pseudo-label}} + \big(1 - C_i(g, p_n)\big) * \underbrace{u}_{\text{uniform distribution}}$$

- Cross agreement C_i ↑ : prediction should be close to pseudo-label
- Cross agreement C_i ↓ : prediction should be close to uniform distribution

SGVR Lab KAIST

Agreement-Aware Label Smoothing (AALS)

- Calibrates the predictions of part features leading to reliable part feature learning



Vanilla label smoothing

AALS

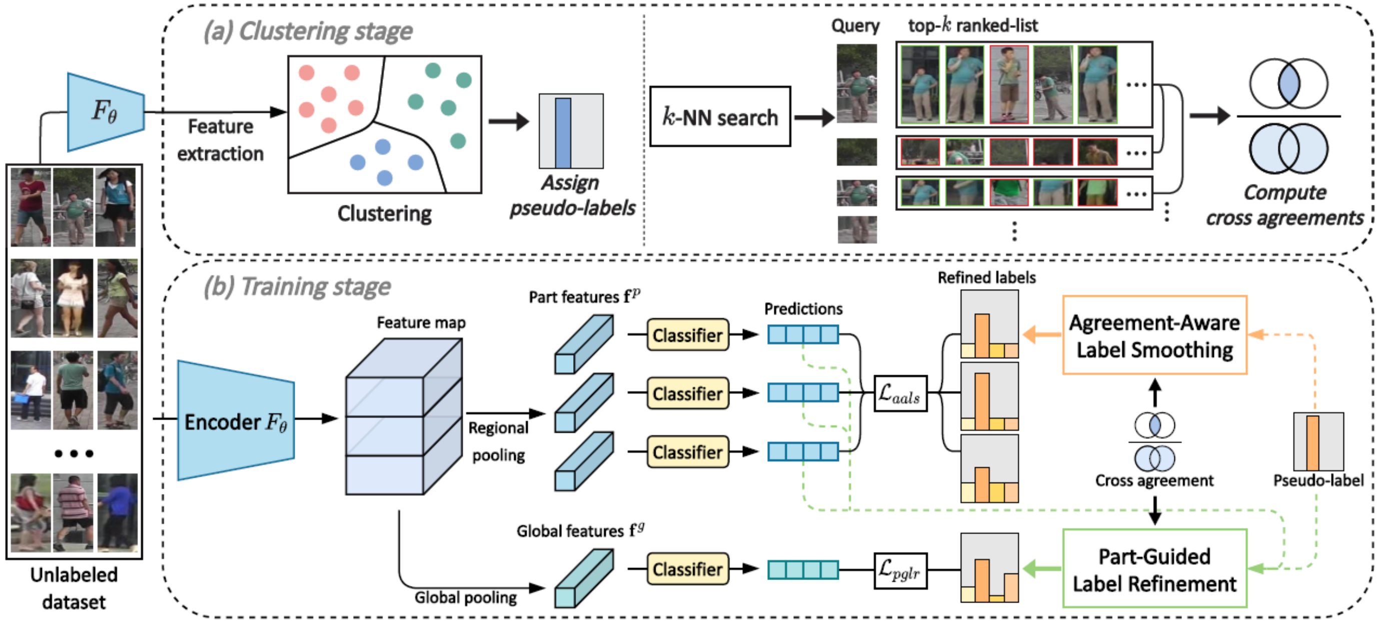
Part-Guided Label Refinement (PGLR)

- Refines pseudo-labels by aggregating predictions of part features with different weights depending on each cross agreement score

$$\tilde{y}_i^g = \beta y_i + (1-\beta)\sum_{n=1}^{N_p} \underbrace{w_i^{p_n}}_{\text{ensemble weight}} \underbrace{q_i^{p_n}}_{\text{prediction of } n\text{–th part feature}}, \quad \text{where} \quad w_i^{p_n} = \frac{\exp(C_i(g, p_n))}{\sum_k \exp(C_i(g, p_k))}$$

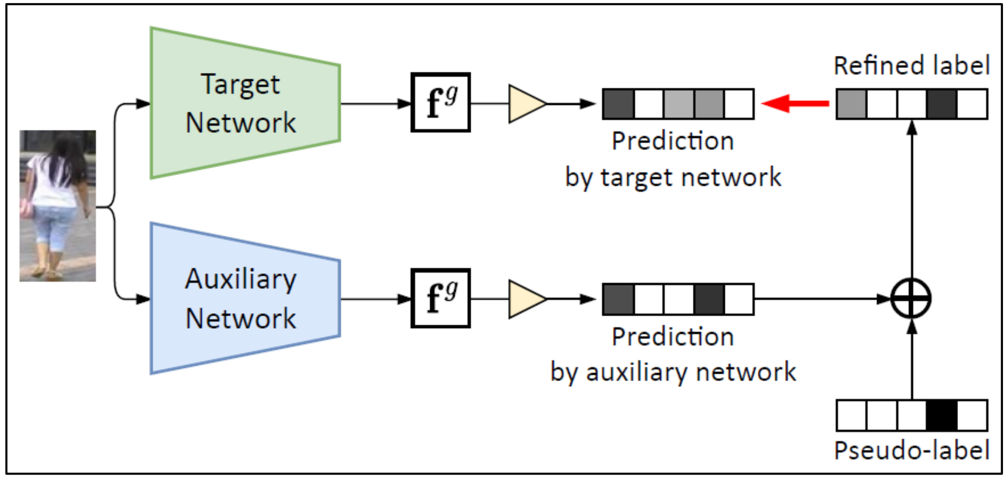
ensemble weight prediction of n–th part feature

- β: weighting parameter

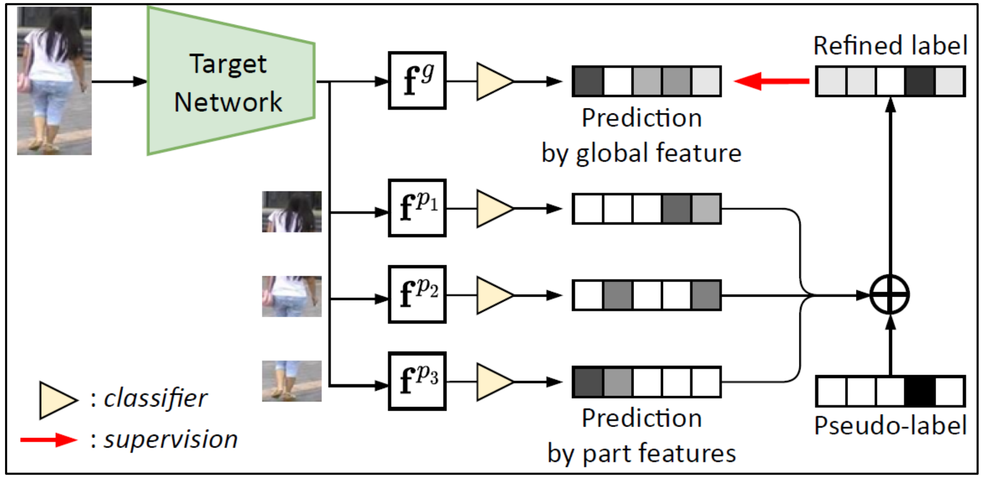


Part-Guided Label Refinement (PGLR)

- Global features learn from the ensembled part predictions with rich fine-grained information without additional teacher networks



Previous



PGLR

SGVR Lab KAIST

Experimental Results

- Ablation Study
  - Effectiveness of AALS and PGLR

Method		Market-1501		MSMT17	
AALS	PGLR	mAP	Rank-1	mAP	Rank-1
-	-	73.5	88.5	25.1	51.2

SGVR Lab KAIST

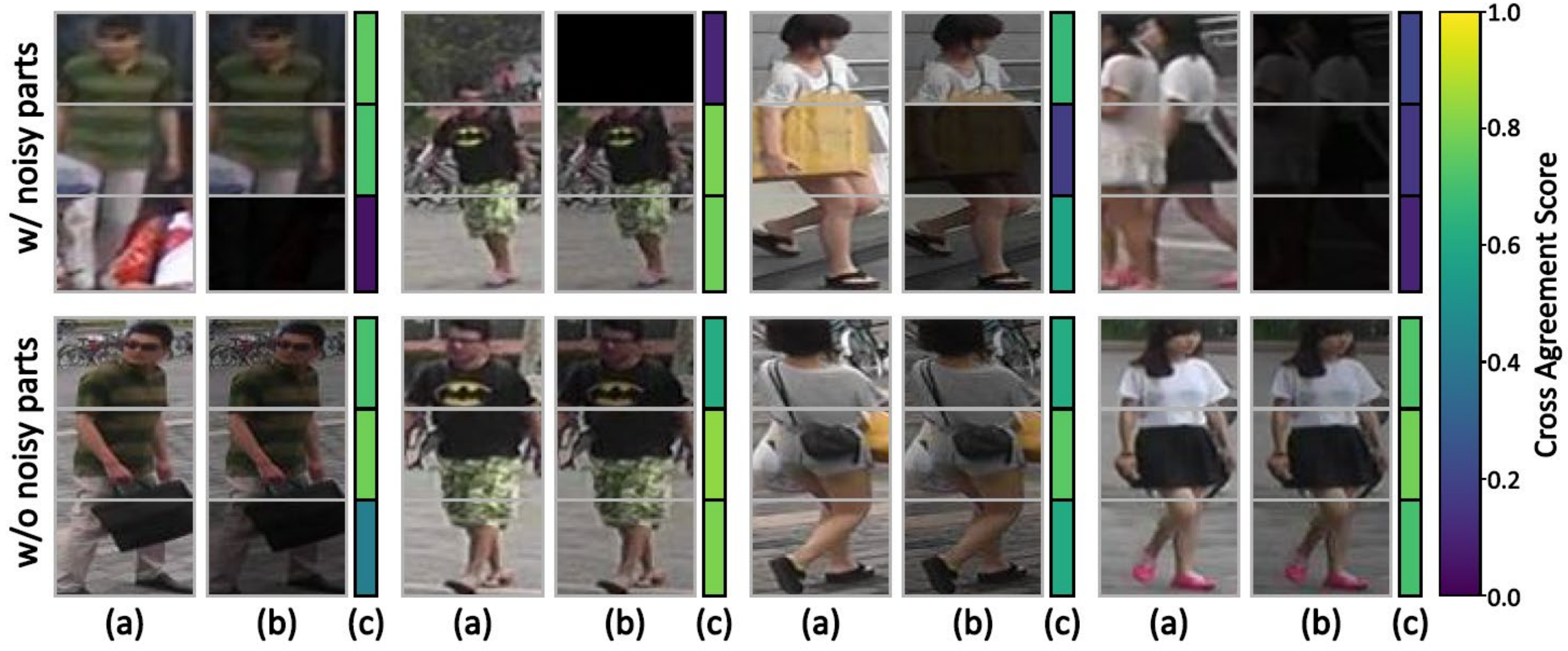
Experimental Results

- Comparison with State-of-the-Arts

Method	Market-1501		MSMT17	
	mAP	Rank-1	mAP	Rank-1
SpCL (NeurIPS 20)	73.1	88.1	19.1	42.3
GCL (CVPR 21)	66.8	87.3	21.3	45.7
IICS (CVPR 21)	72.9	89.5	26.9	56.4
RLCC (CVPR 21)	77.7	90.8	27.9	31.4
ICE (ICCV 21)	79.5	92.0	29.8	59.0
PPLR	**81.5**	**92.8**	**31.4**	**61.1**

Experimental Results

- Analysis of Cross Agreement Score
  - (a) Original image
  - (b) Soft masked image by the cross agreement score
  - (c) Color jet bar of the cross agreement score

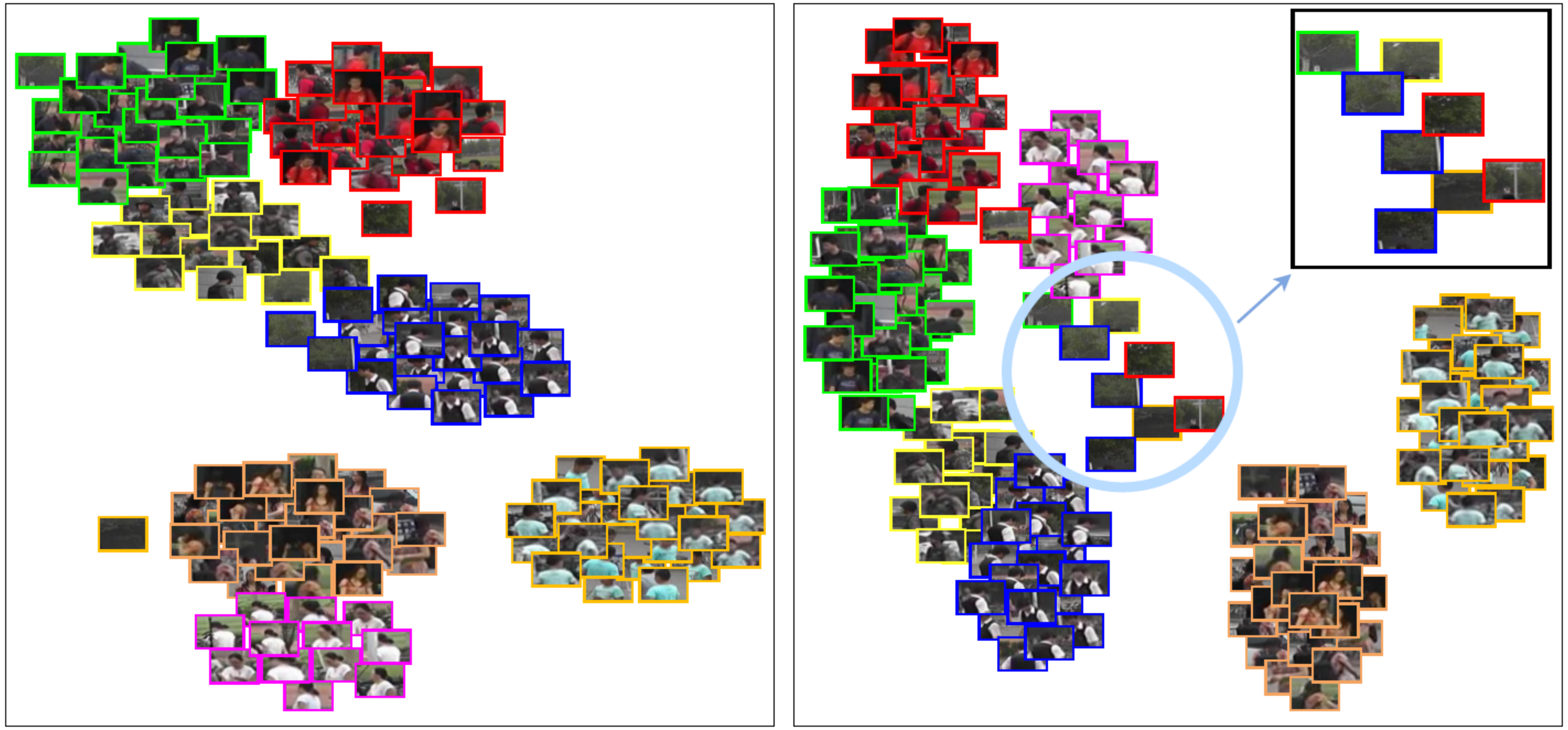


SGVR Lab KAIST

Experimental Results

- Effect of Agreement-Aware Label Smoothing
  - t-SNE visualization of the topmost part feature space

Meaningless parts are not overfitted to meaningful clusters



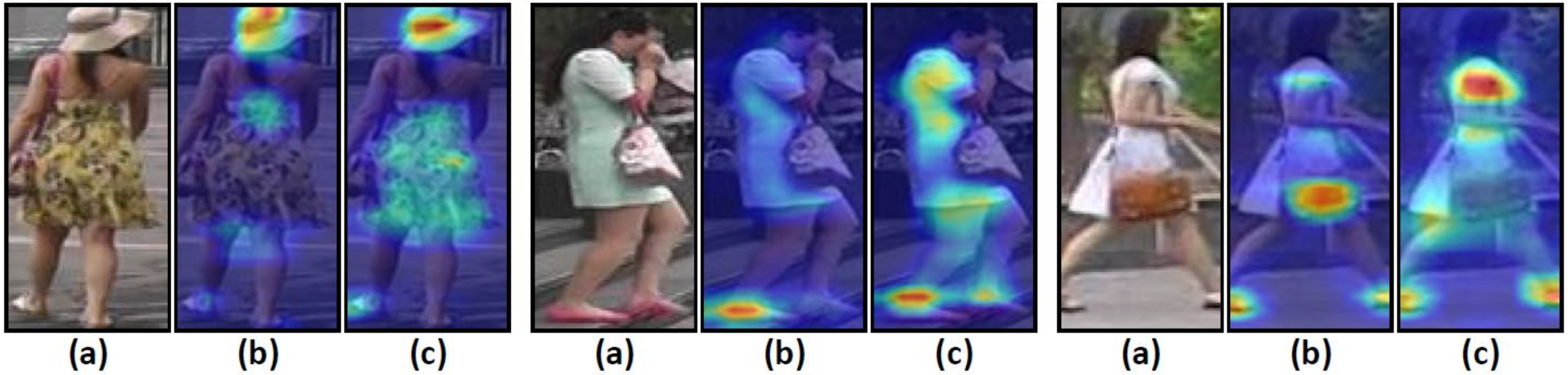
w/o AALS

w/ AALS

SGVR Lab KAIST

Experimental Results

- Analysis of Part-Guided Label Refinement
  - (a) Original image
  - (b) Without PGLR
  - (c) With PGLR



SGVR Lab KAIST

Summary

- Person Re-identification
- Unsupervised Approaches
- Part-based Pseudo Label Refinement for Unsupervised Person Re-ID (CVPR 2022)

Project Page: https://sgvr.kaist.ac.kr/~yoonki/PPLR/

SGVR Lab KAIST