SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Playing with GANs in Person Retrieval

Liang Zheng

Singapore University of Technology and Design

April 20, 2018

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

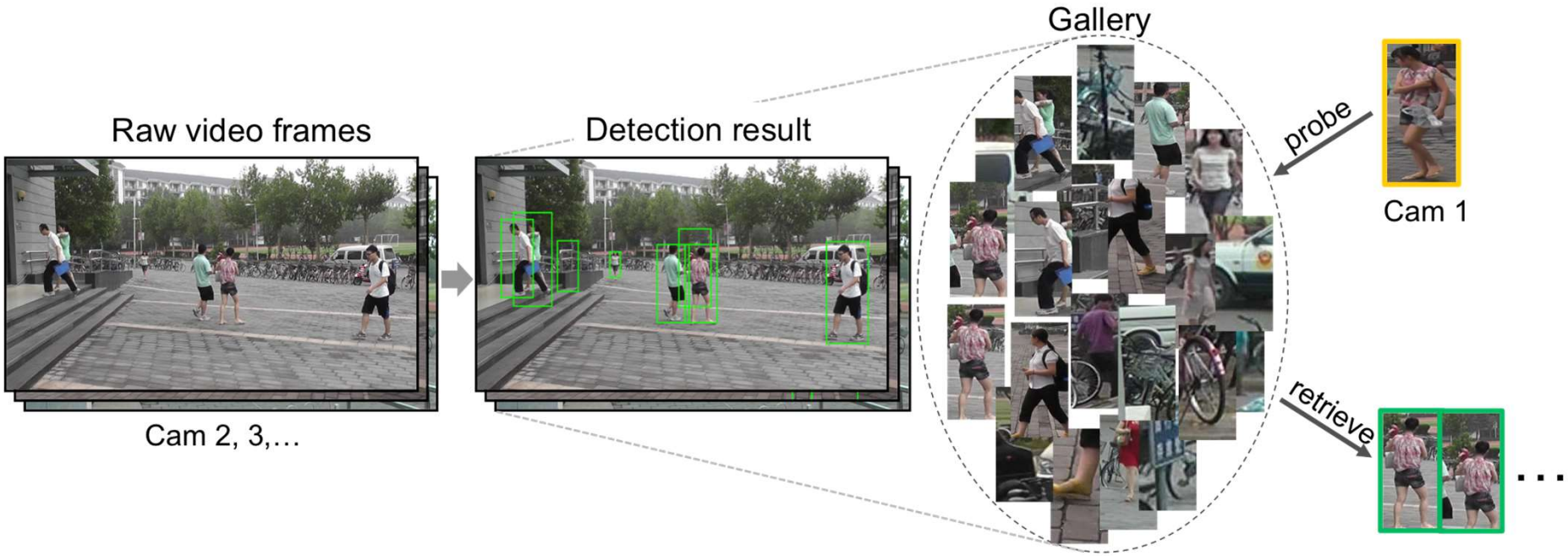
Outline

- Background
- Generating images? So what?
- Style transfer? Careful with errors!
- UDA, let's save the labels!
- Future directions

Introduction

Person Detection

Person retrieval / re-identification

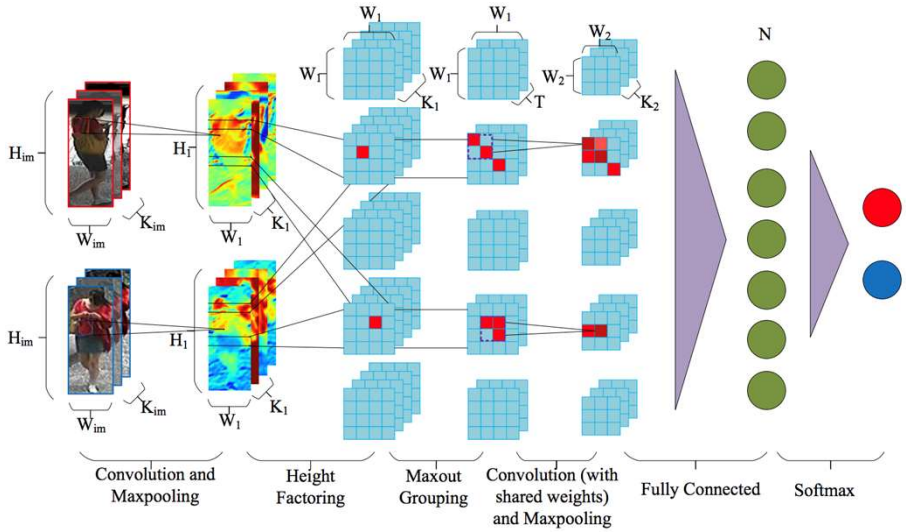


Person retrieval / re-identification

Background

- Before 2015, researchers addressed this task using an expensive matching procedure.

Zhao et al. CVPR 2013



Li et al. CVPR 2014

Two images are first compared using their local regions. The local similarities are aggregated into a global similarity.

-- very time-consuming under a large image database

Background

- View person re-identification as a special task of generic image retrieval (Person Re-identification Meets Image Search, Arxiv 2015)
- The importance of mean average precision (mAP)
- The importance of using a single descriptor
- The important of indexing
- Improve efficiency by 100x

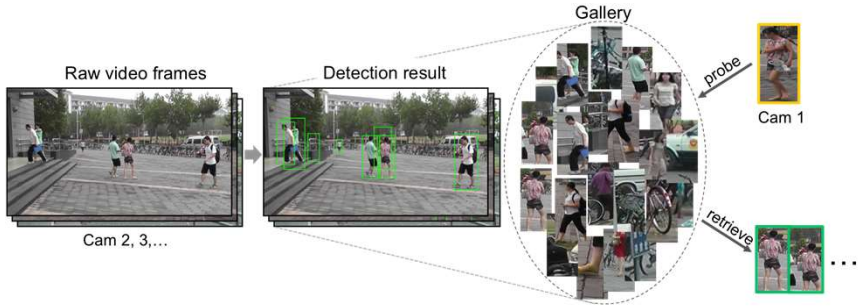
Stage	SDALF (CVPR 2011)	SDC (CVPR 2013)	**Ours**
Feature Extraction (s)	2.92	0.76	***0.62***
Search (s)	2644.80	437.97	***0.98***

Background

- Define five large-scale datasets / evaluation protocols
- Image-based:
- Market-1501 (ICCV 2015)
- CUHK03-NP (CVPR 2017)
- DukeMTMC (ICCV 2017)

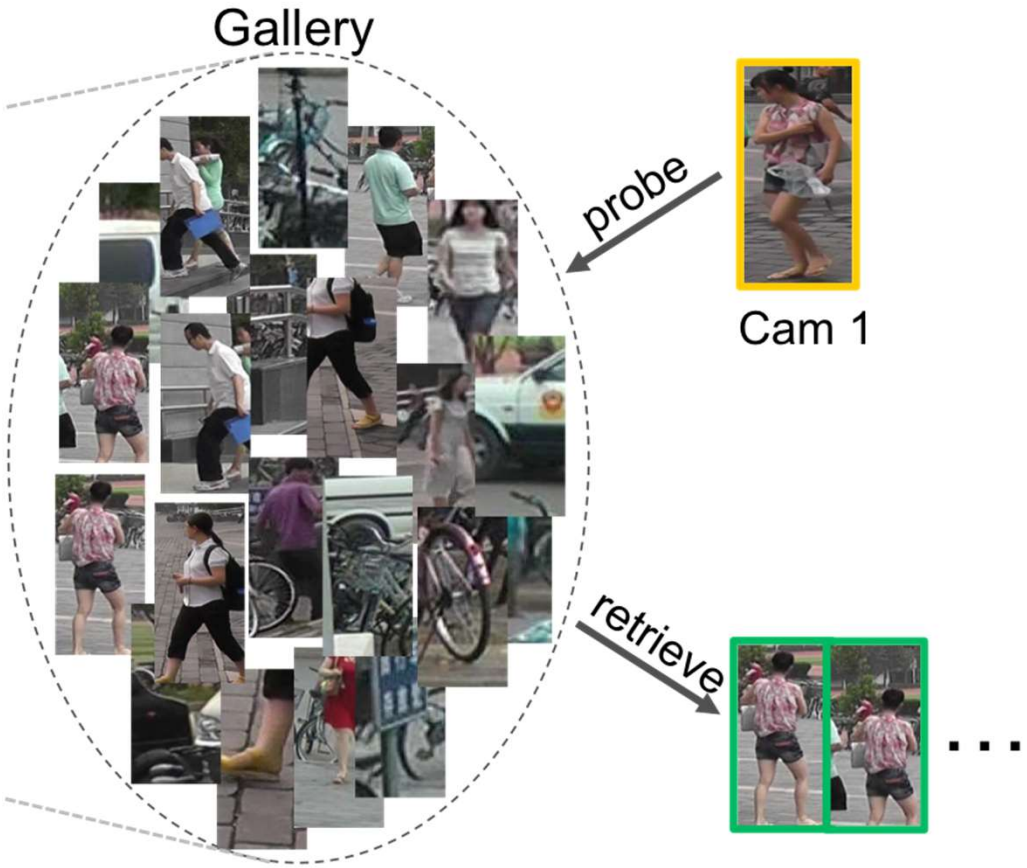
- Video-based:
- MARS (ECCV 2016)

- Detection + retrieval:
- PRW (CVPR 2017)



Background

- **Person re-identification is a retrieval problem**

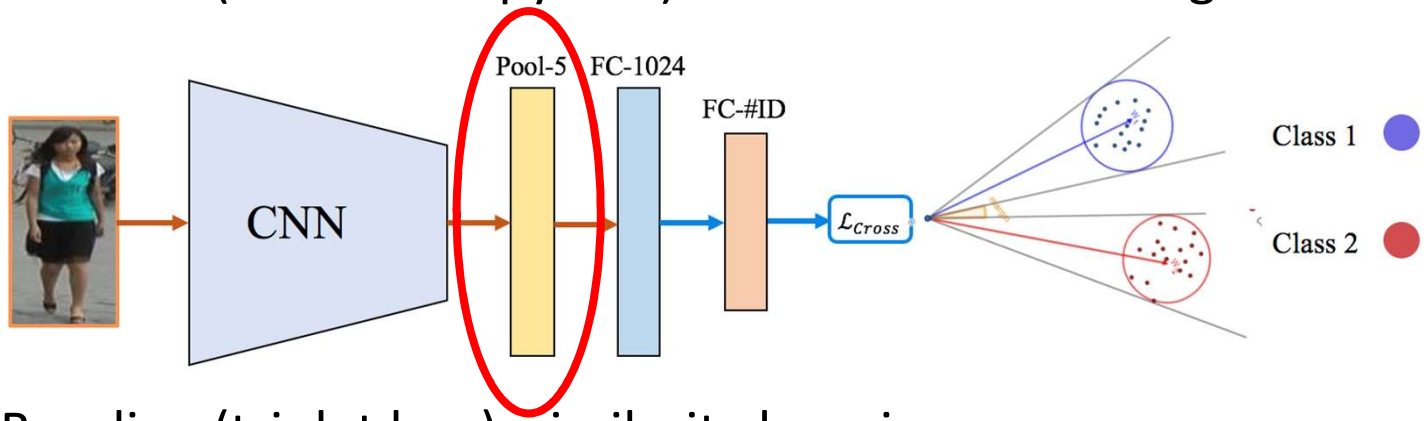


Training:
Representation learning

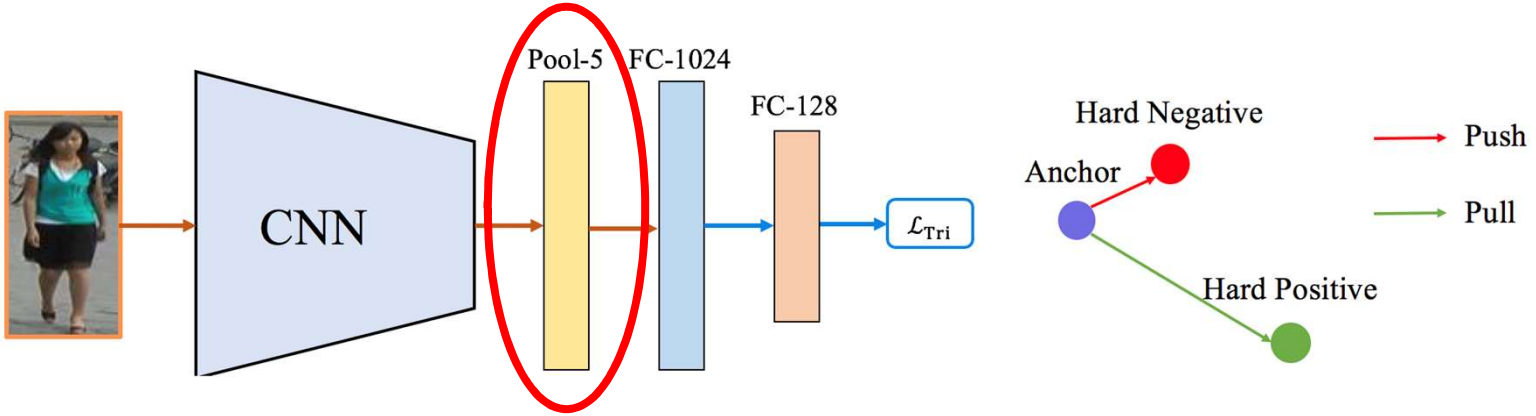
Testing:
- Feature extraction
- Similarity calculation

Background

- Baseline (cross-entropy loss): classification learning.



- Baseline (triplet loss): similarity learning.



- Testing: we extract the Pool-5 output as the visual representation for a person image

Background

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

On Market-1501 dataset, Resnet-50, rank-1 accuracy

Cross-entropy Loss

- 73.0%: [1], [2], [3] (2016.11)
- 83.1%: [4] (2017.7)
- 85.3%: [5] (2017.11)
- 86.7%: [5] (2017.11)

Triplet Loss

- 82.6 %: open-reID (2017.6)
- 84.9 %: [6] (2017.3)

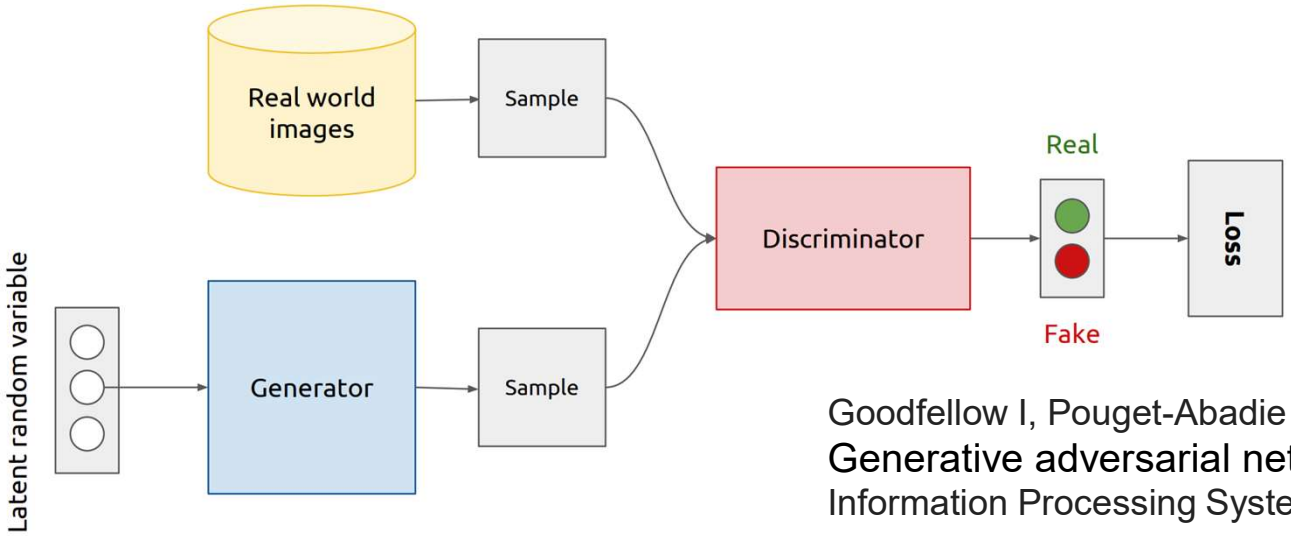
Cross-entropy Loss + Triplet Loss

- 90.1% (2017.12)

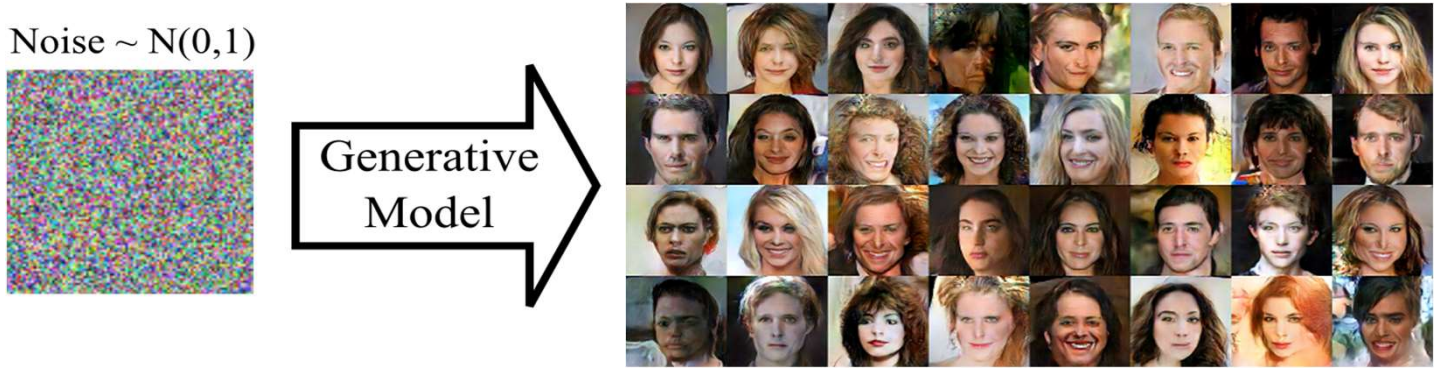
1. Zheng L, Yang Y, Hauptmann A G. Person re-identification: Past, present and future. In arXiv, 2016.
2. Zheng L, Huang Y, Lu H, et al. Pose invariant embedding for deep person re-identification. In arXiv, 2017.
3. Y. Sun, L. Zheng,W. Deng, and S.Wang. SVDNet for pedestrian retrieval. In ICCV, 2017.
4. Zhong Z, Zheng L, Kang G, et al. Random Erasing Data Augmentation. In arXiv preprint, 2017.
5. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. Beyond part models: Person retrieval with refined part pooling. In arXiv, 2017.
6. Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-Identification. In arXiv preprint arXiv, 2017.

Background

- **Generative adversarial networks**

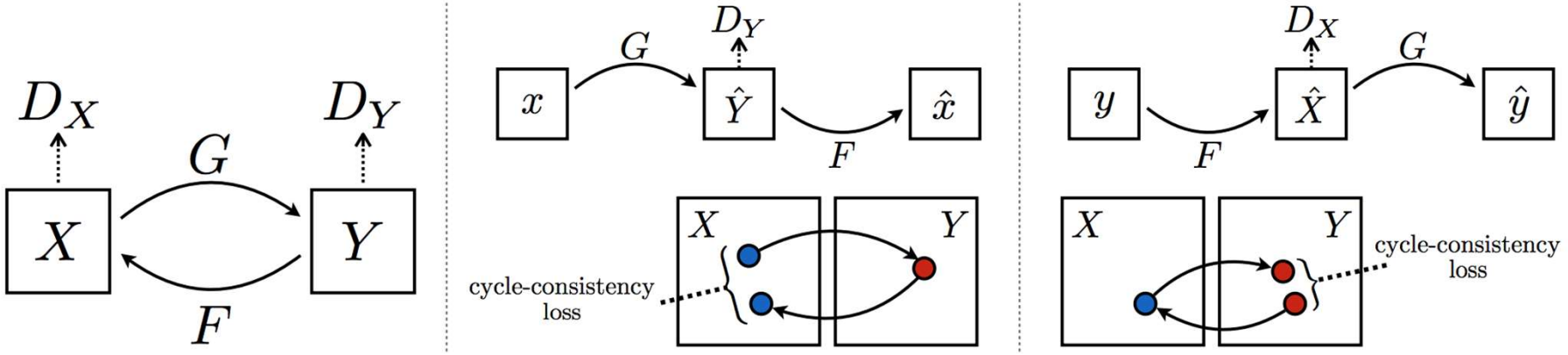


Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In Advances in Neural Information Processing Systems. 2014: 2672-2680.

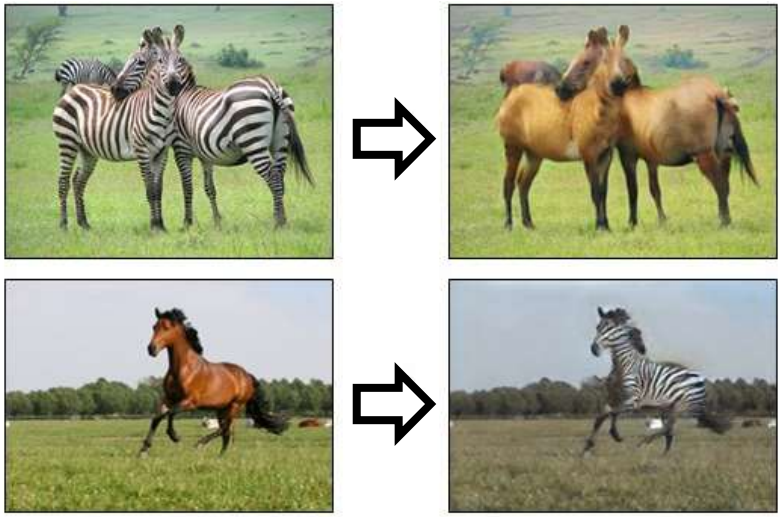


Background

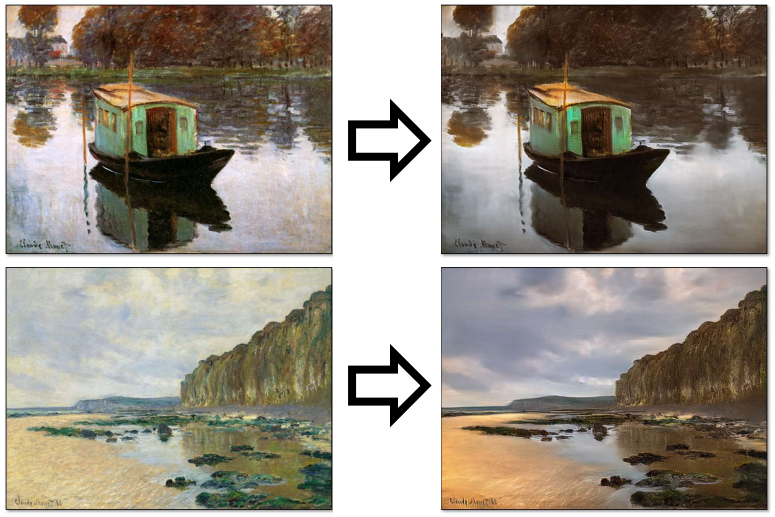
- **CycleGAN** (Zhu et al., ICCV 2017)



Zebra <-> horse



Painting -> photo



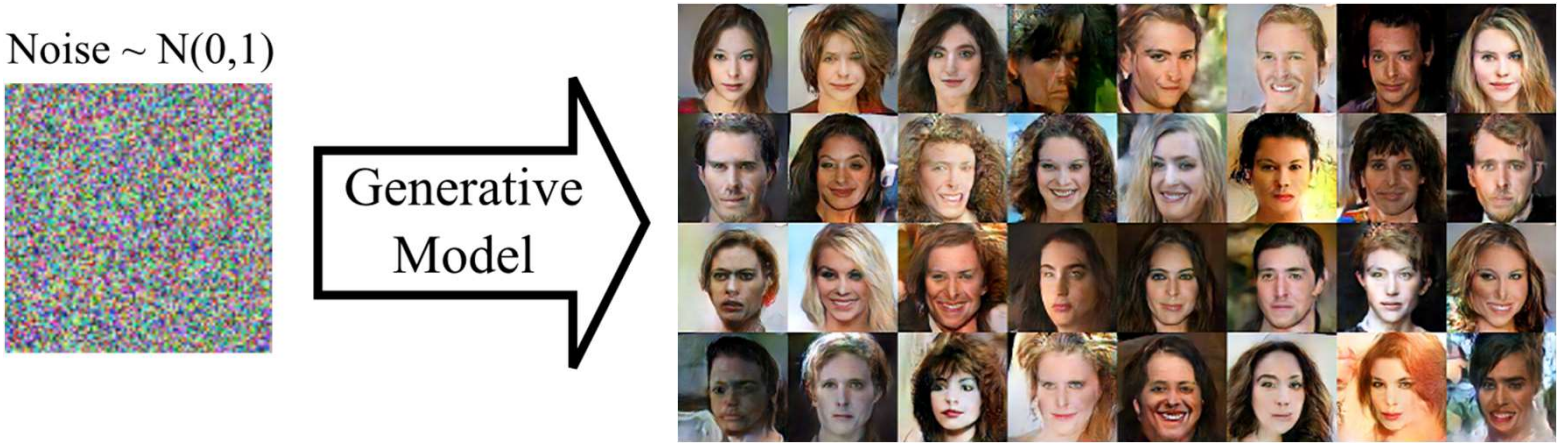
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline

- Background
- Generating images? So what?
- Style transfer? Careful with errors!
- UDA, let's save the labels!
- Future directions

Generating images? So what?

- In 2016, people were criticizing that the Generative Adversarial Network (GAN) might only be able to generate images.



Generating images? So what?

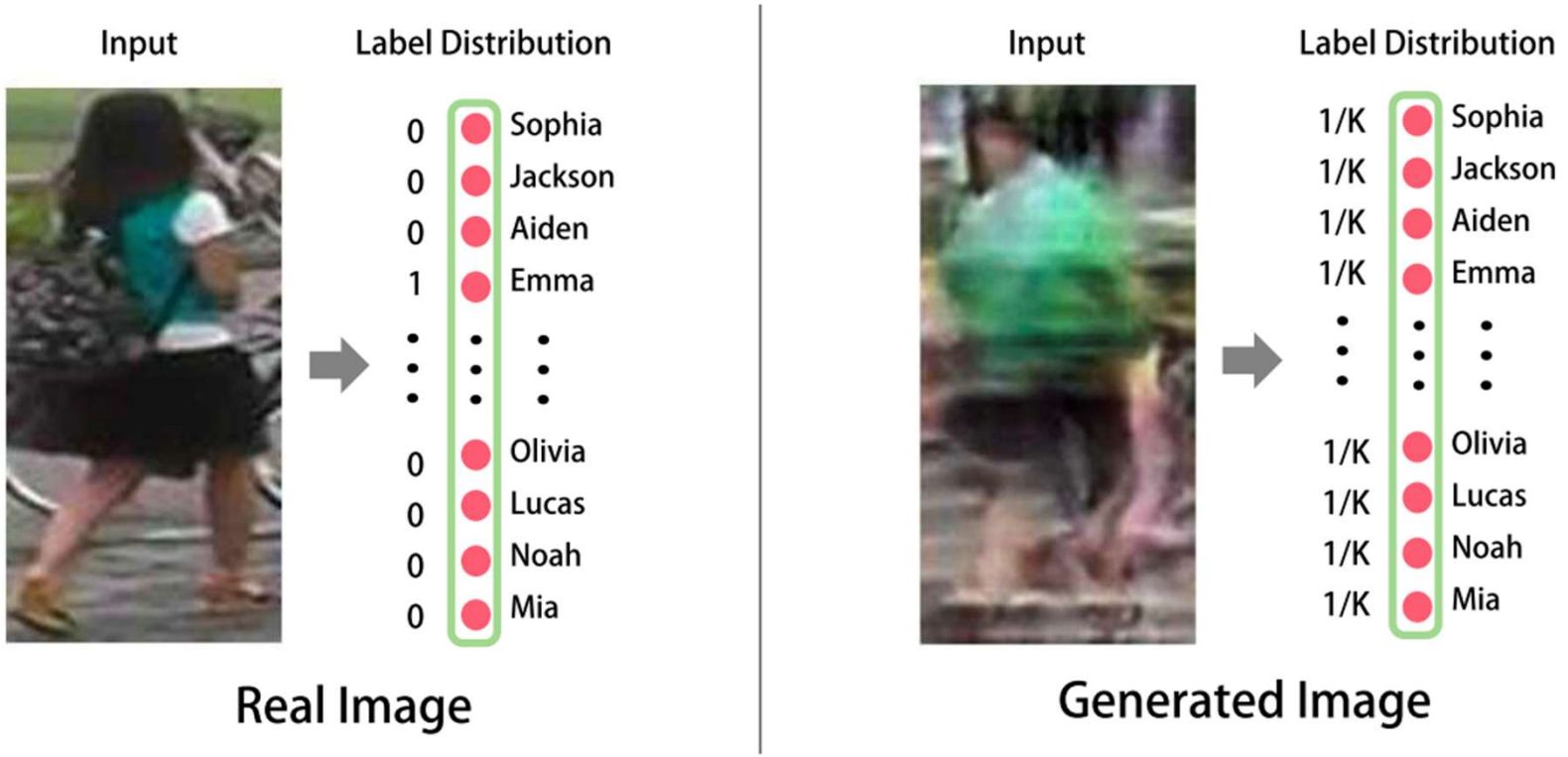
- One of the early success reports of GANs in supervised learning in vitro

Randomly generated persons

Randomly generated birds

Our Method

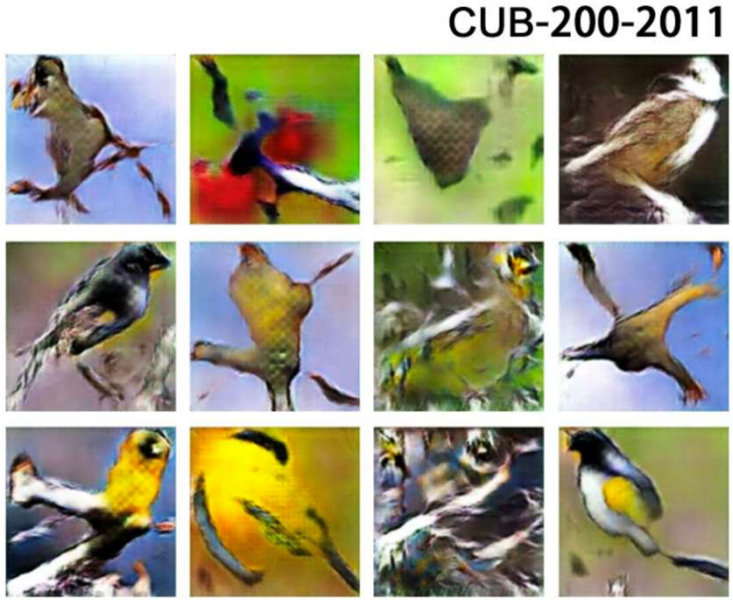
- Label Smooth Regularization for Outliers (LSRO)



Z. Zheng et al., Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*, ICCV 2017.

Results

We observe consistent improvement in person retrieval and fine-grained classification.



method	model	annotation	top-1
Zhang *et al.* [44]	AlexNet	2×part	76.7
Zhang *et al.* [44]	VGGNet	2×part	81.6
Liu *et al.* [19]	ResNet-50	attribute	82.9
Wang *et al.* [35]	3×VGGNet	×	83.0
Basel. [19]	ResNet-50	×	82.6
Basel.+LSRO	ResNet-50	×	83.2
Basel.+LSRO	2×ResNet-50	×	**84.4**

Table 6. We show the recognition accuracy (%) on CUB-200-2011. The proposed method has a 0.6% improvement over the competitive baseline. The two-model ensemble shows a competitive result.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

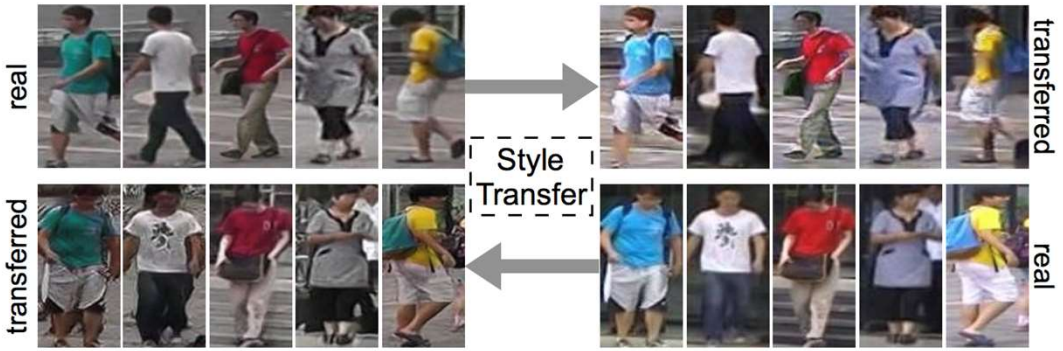
Outline

- Background
- Generating images? So what?
- Style transfer? Careful with errors!
- UDA, let's save the labels!
- Future directions

Camera style

- A new data augmentation method

(a) Example images under two cameras from Market-1501



(b) Examples of camera-aware style transfer between two cameras

Generated images:
labels are preserved

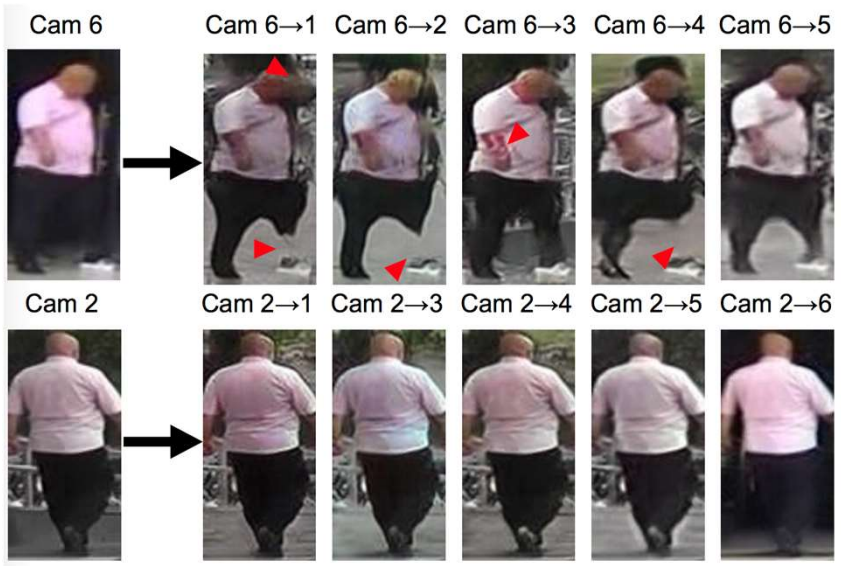
Training set:
generated images + original images

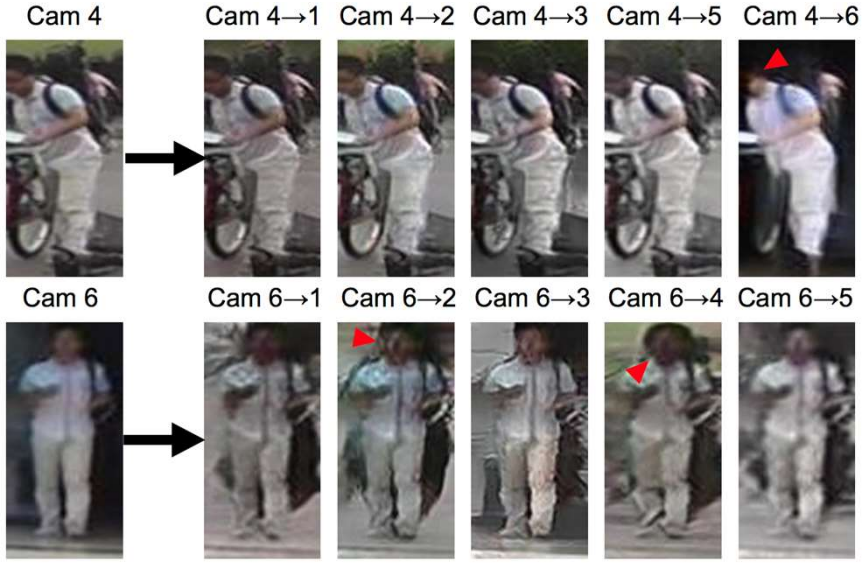
Zhong et al., Camera Style Adaptation for Person Re-identification. CVPR 2018.

Style transfer? Careful with errors!

- Errors occurs during image-image translation

Examples of style-transferred samples in Market-1501

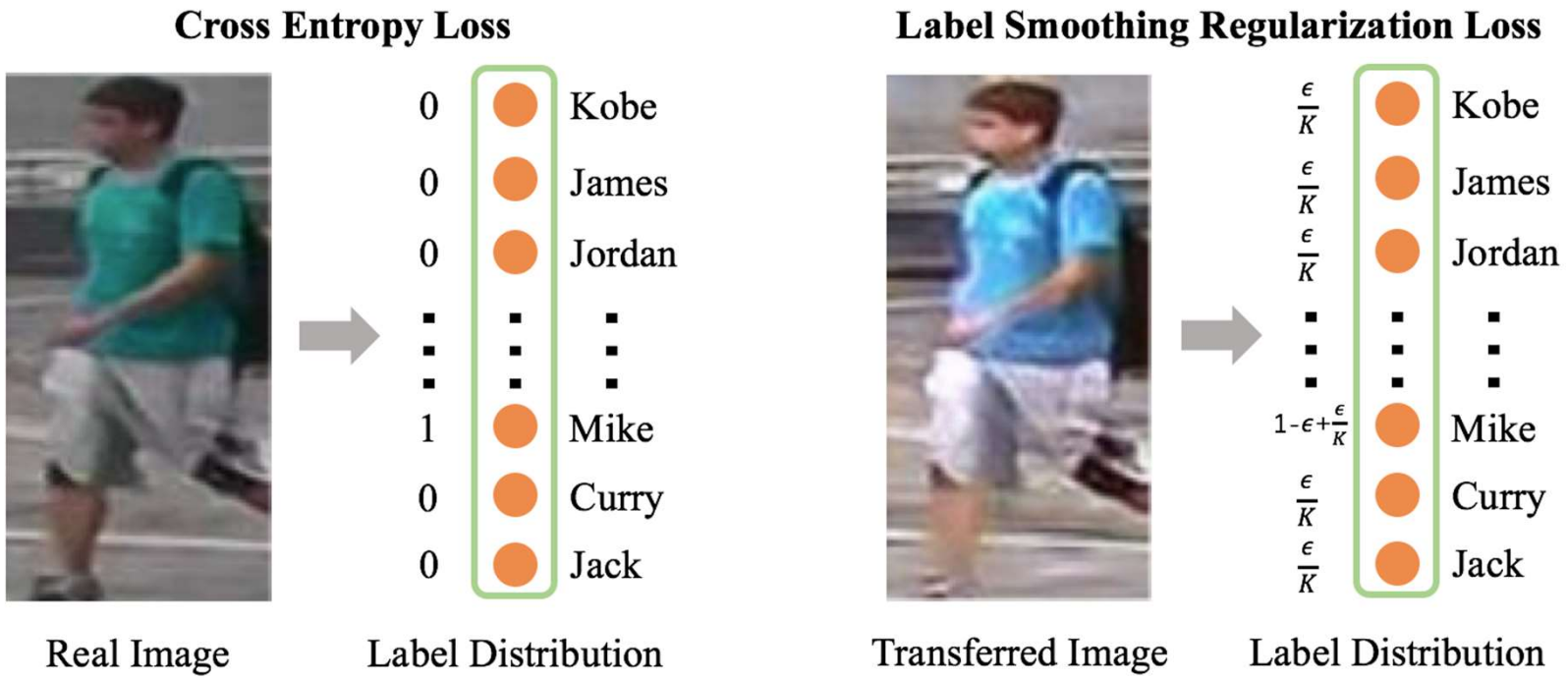




Zhong et al., Camera Style Adaptation for Person Re-identification. CVPR 2018.

Camera style data augmentation

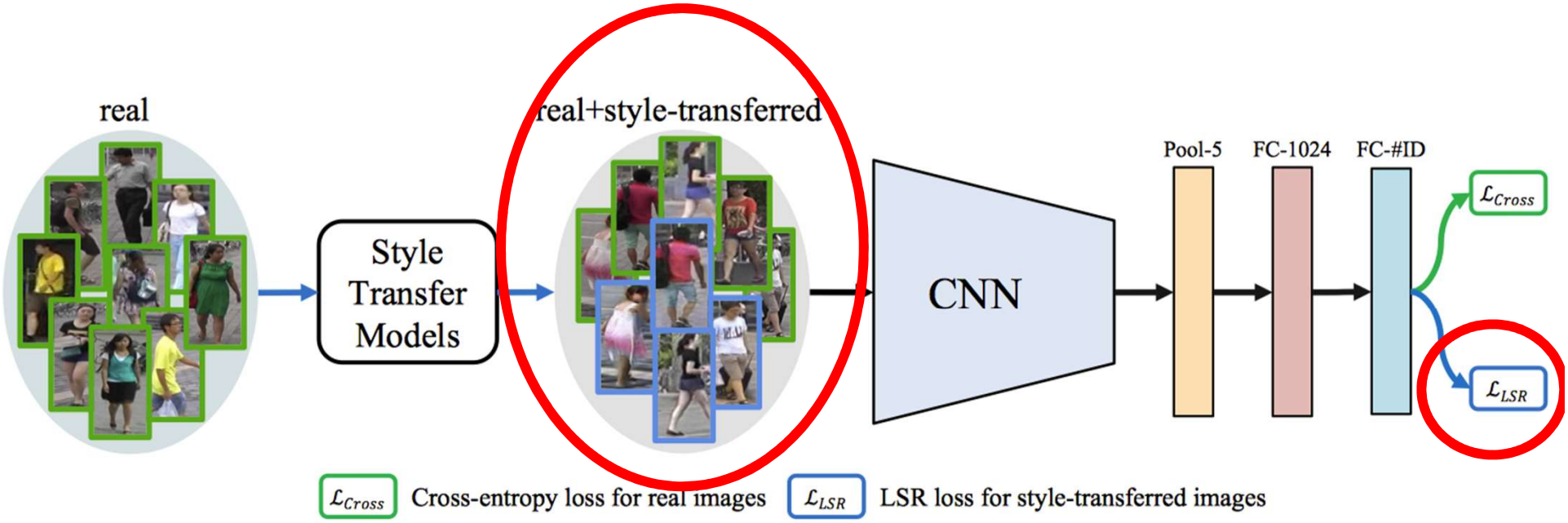
- Label smoothing regularization (LSR) on the style-transferred samples to softly distribute their labels



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

Camera style data augmentation

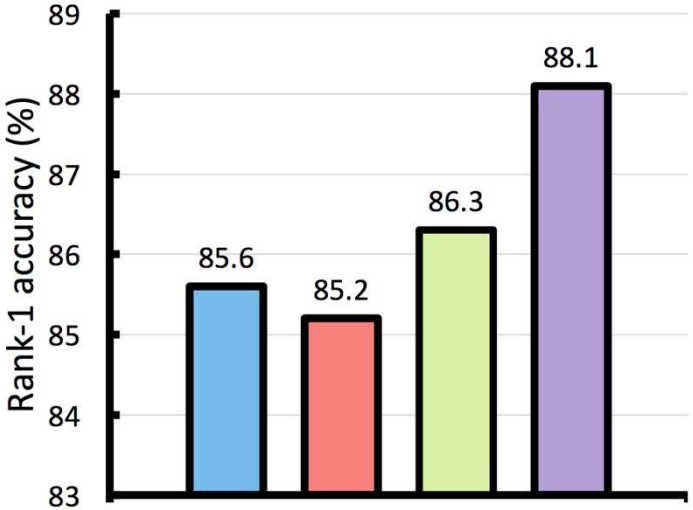
- Label smoothing regularization (LSR) on the style-transferred samples to softly distribute their labels



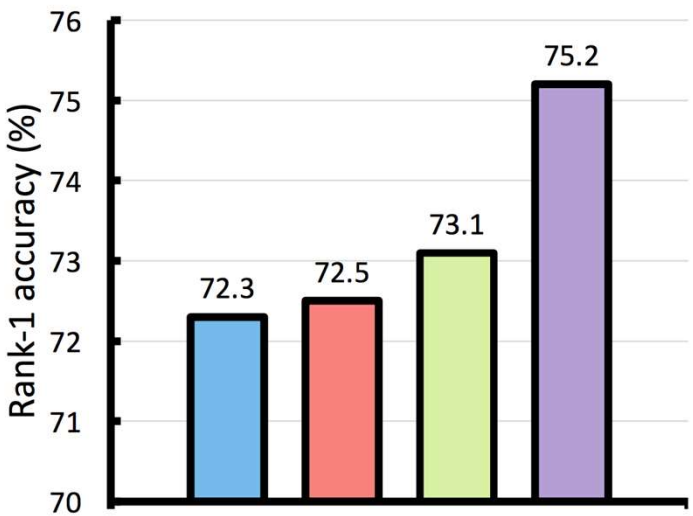
Zhong et al., Camera Style Adaptation for Person Re-identification. CVPR 2018.

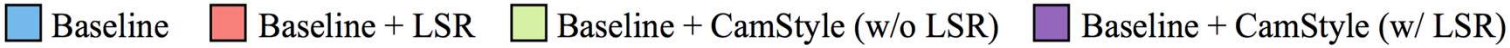
Results

Market-1501 dataset, 6 cameras



DukeMTMC-reID dataset, 8 cameras





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

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline

- Background
- Generating images? So what?
- Style transfer? Careful with errors!
- UDA, let's save the labels!
- Future directions

Unsupervised domain adaptation in person retrieval

- Problem: dataset bias

Unsupervised domain adaptation on the image level

Source domain

Target domain

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Pipeline

- Learning via translation

Step 1: source-target image translation

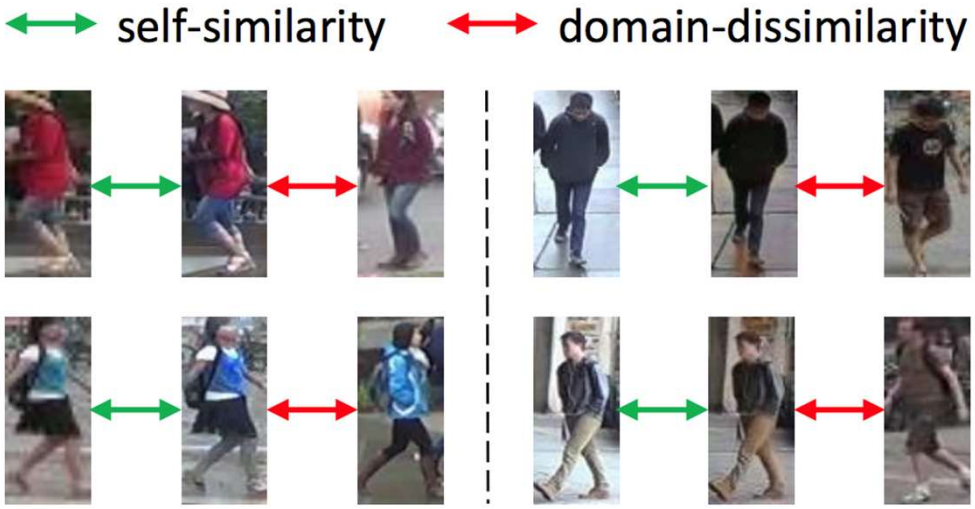
Step 2: feature learning

We focus on Step 1, i.e., improving image-image translation.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Let's save the labels!

- Identity labels should be preserved before and after translation

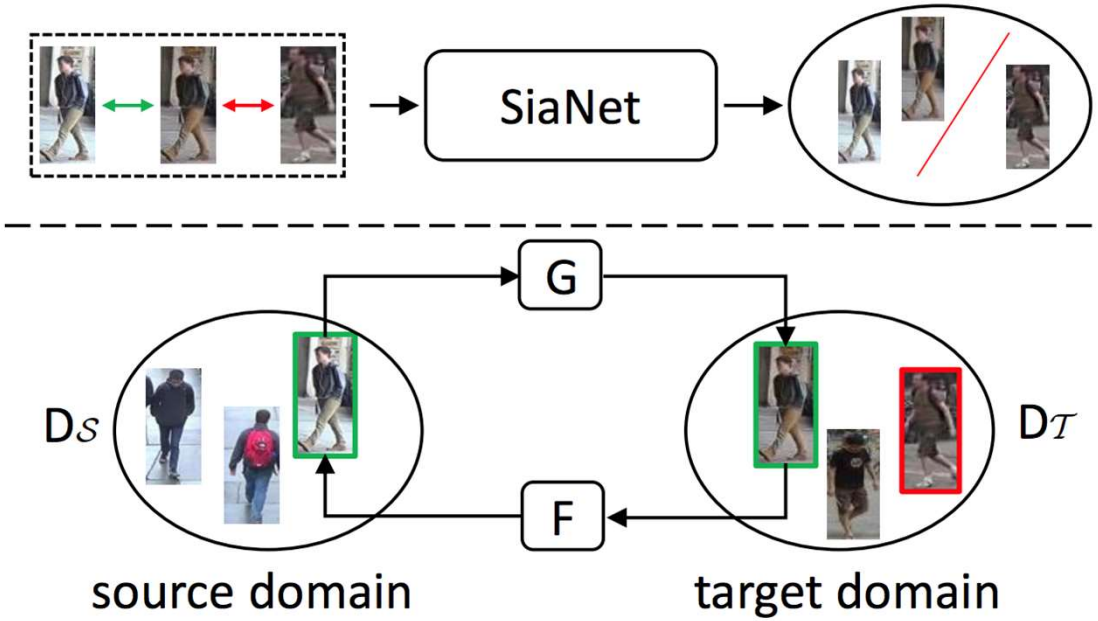


Self-similarity: a translated image, despite of its style changes, should contain the same underlying identity with its corresponding source image

Domain-dissimilarity: a translated image should be different from any image in the target dataset in terms of the underlying ID.

Let's save the labels!

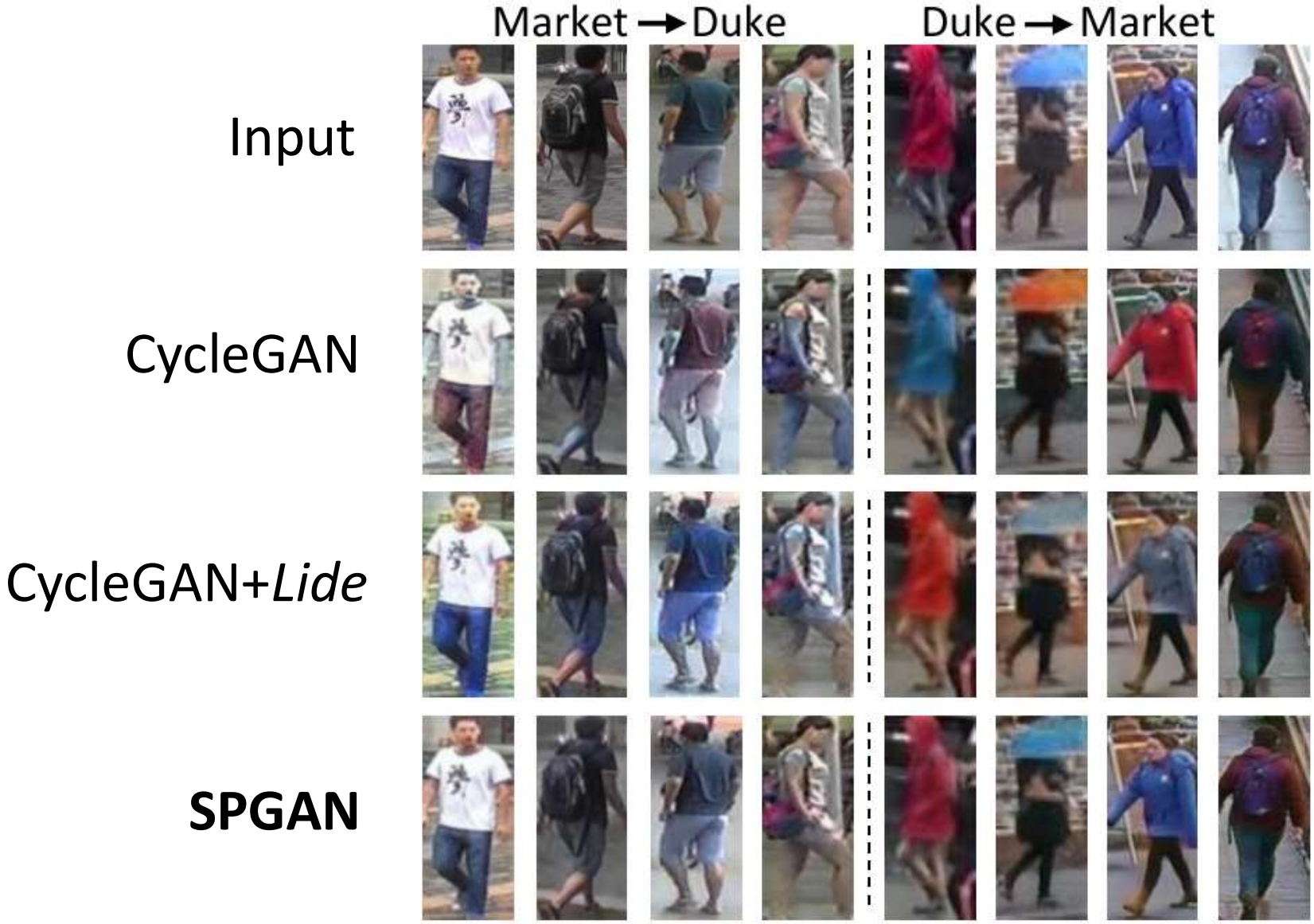
- Similarity Preserving cycle-consistent Generative Adversarial Network (SPGAN)
- SPGAN: Siamese Network (top) and CycleGAN (bottom)



SiaNet constrains the mapping functions towards identity preserving

CycleGAN maps a source image to the style of the target domain

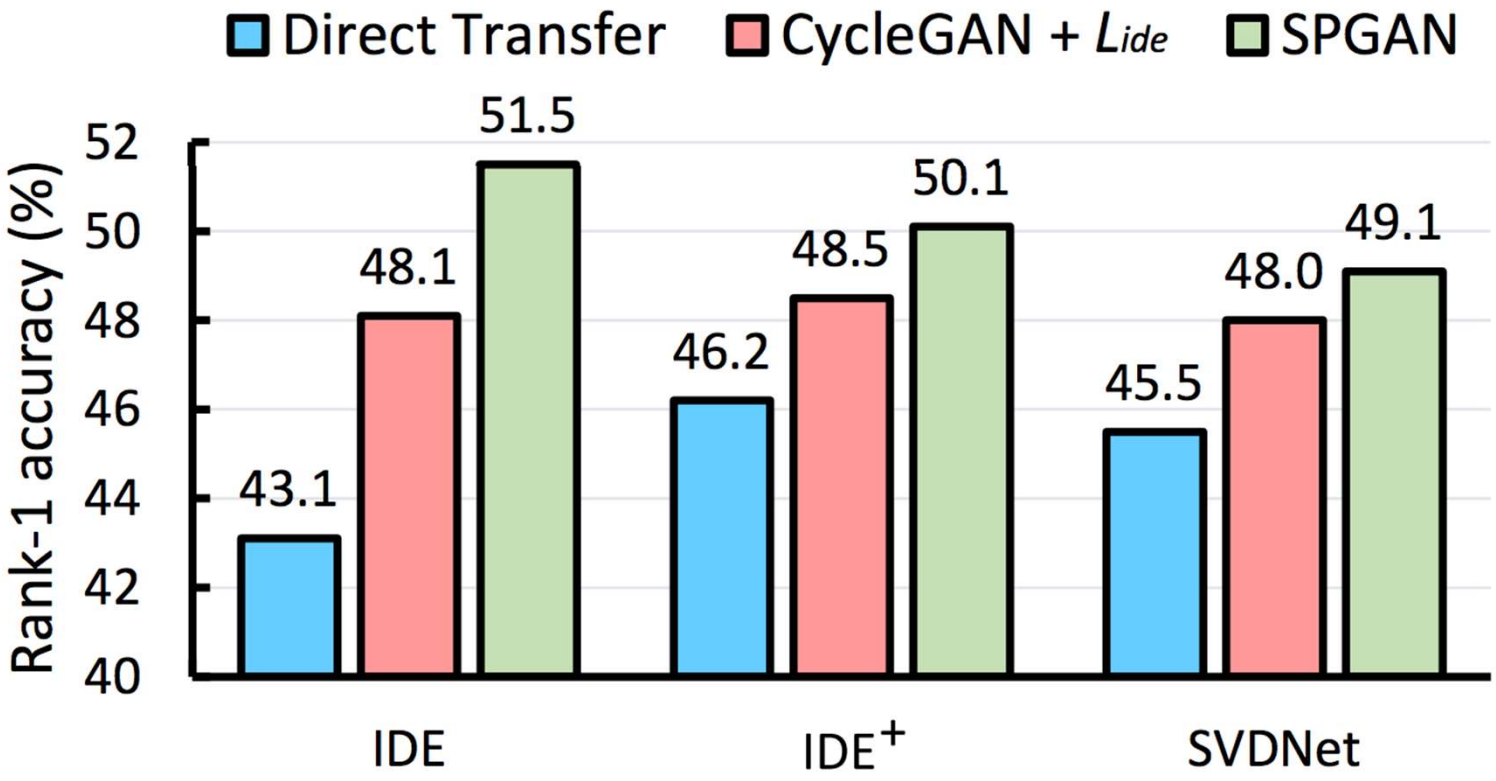
Image-image Translation Examples



Results

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

- Comparison with the baselines



Results

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

- Comparison with the state of the art

Methods	DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP
Bow [51]	17.1	28.8	34.9	8.3
LOMO [26]	12.3	21.3	26.6	4.8
UMDL [35]	18.5	31.4	37.6	7.3
PUL [6]*	30.0	43.4	48.5	16.4
Direct transfer	33.1	49.3	55.6	16.7
SPGAN	41.1	56.6	63.0	22.3
SPGAN+LMP	**46.9**	**62.6**	**68.5**	**26.4**

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline

- Background
- Generating images? So what?
- Style transfer? Careful with errors!
- UDA, let's save the labels!
- Future directions

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Future directions

- **Person retrieval under very large databases**

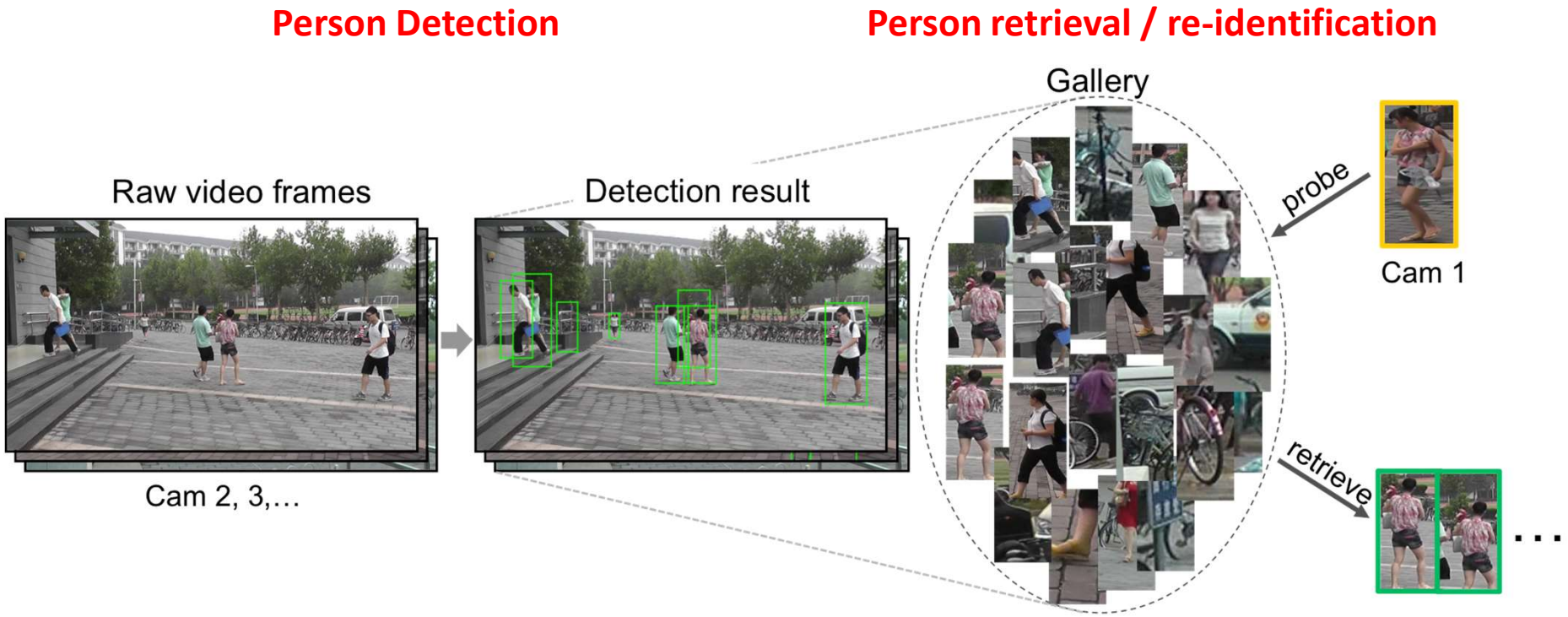
Gou et al. ICCVW, 2017

Performance drops significantly under much larger databases.

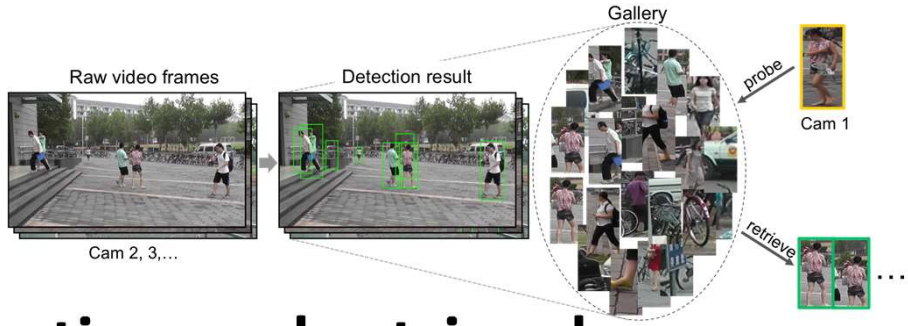
Design efficient systems that attend to pedestrian characteristics.

Future directions

- **Relationship between person detection and retrieval**



Key Applications



- **Relationship between detection and retrieval**
- How detector quality affects retrieval / recognition
- How to design a proper detector for the subsequent recognition tasks?

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Thank you!

Q&A