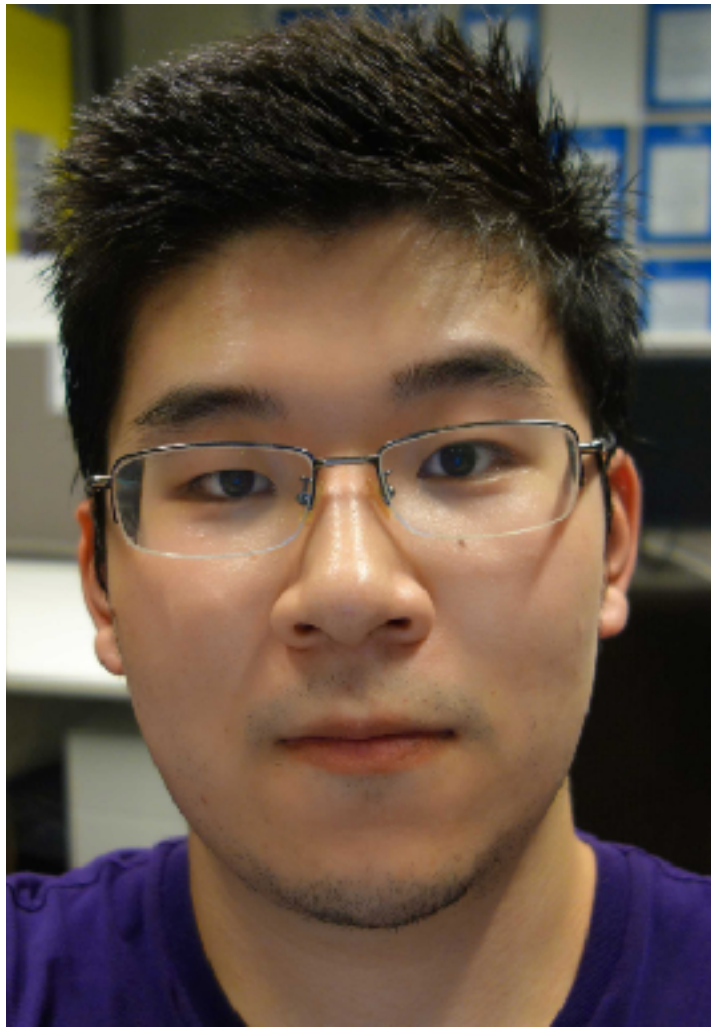
从行人重识别到无人机定位

郑哲东
http://zdzheng.xyz

About Me

Present

- **Final-year PhD student**
- Advised by Prof. Yi Yang and Dr. Liang Zheng
- Published 4 first-author conference papers and 5 first-author journal papers
- GoogleScholar 3000+ citations; Github 5000+ stars
- First-place winner in AICity Challenge CVPR 2020
- Looking for the post-doc position



Research Interests

Image Retrieval, Person Re-identification, Image-text Understanding

Image Generation, Domain Adaptation, Adversarial Samples

Zhedong.Zheng@student.uts.edu.au

【极市】郑哲东 -Deep-ReID：行人重识别的深度学习方法

7453播放 · 140弹幕 2017-09-25 22:29:40

（考古）2017年 19 期

-》

2020年 71 期

比较早的开源

2016 Matlab https://github.com/layumi/2016_person_re-ID

2017Python https://github.com/layumi/Person_reID_baseline_pytorch

Outlines

1. 行人重识别的一些实践 (行人动，camera不动)

2. 车辆重识别 CVPR2020 智慧城市比赛冠军 (车动，camera不动)

3. 无人机与重识别的机遇与挑战 ACM Multimedia2020 (camera动，建筑不动)

Outlines

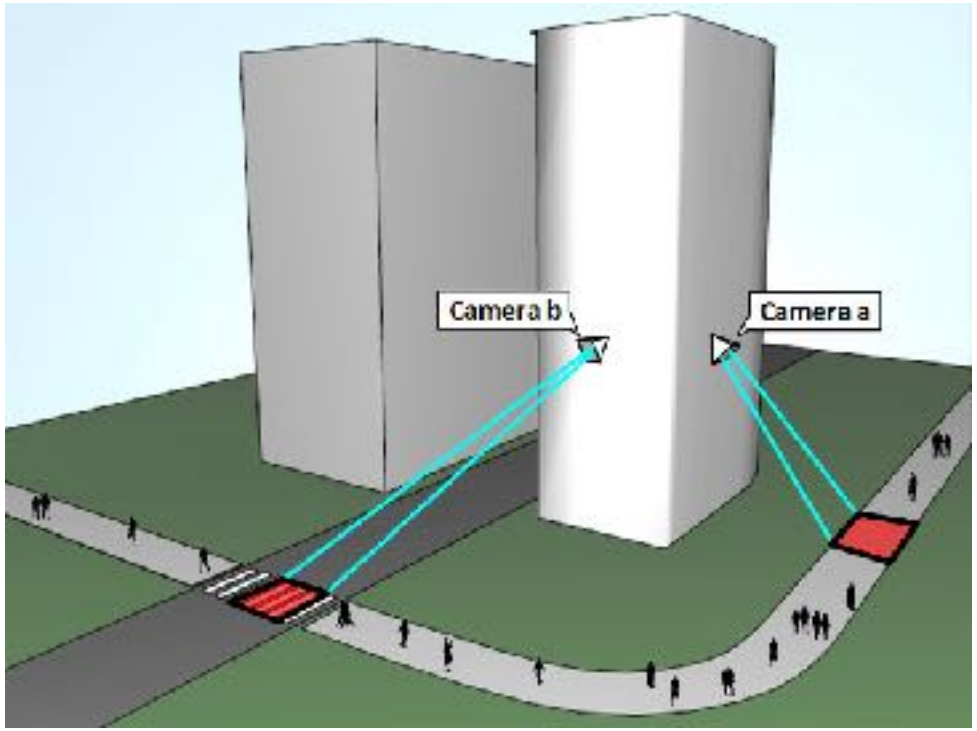
1. 行人重识别的一些实践 (行人动，camera不动)

2. 车辆重识别 CVPR2020 智慧城市比赛冠军

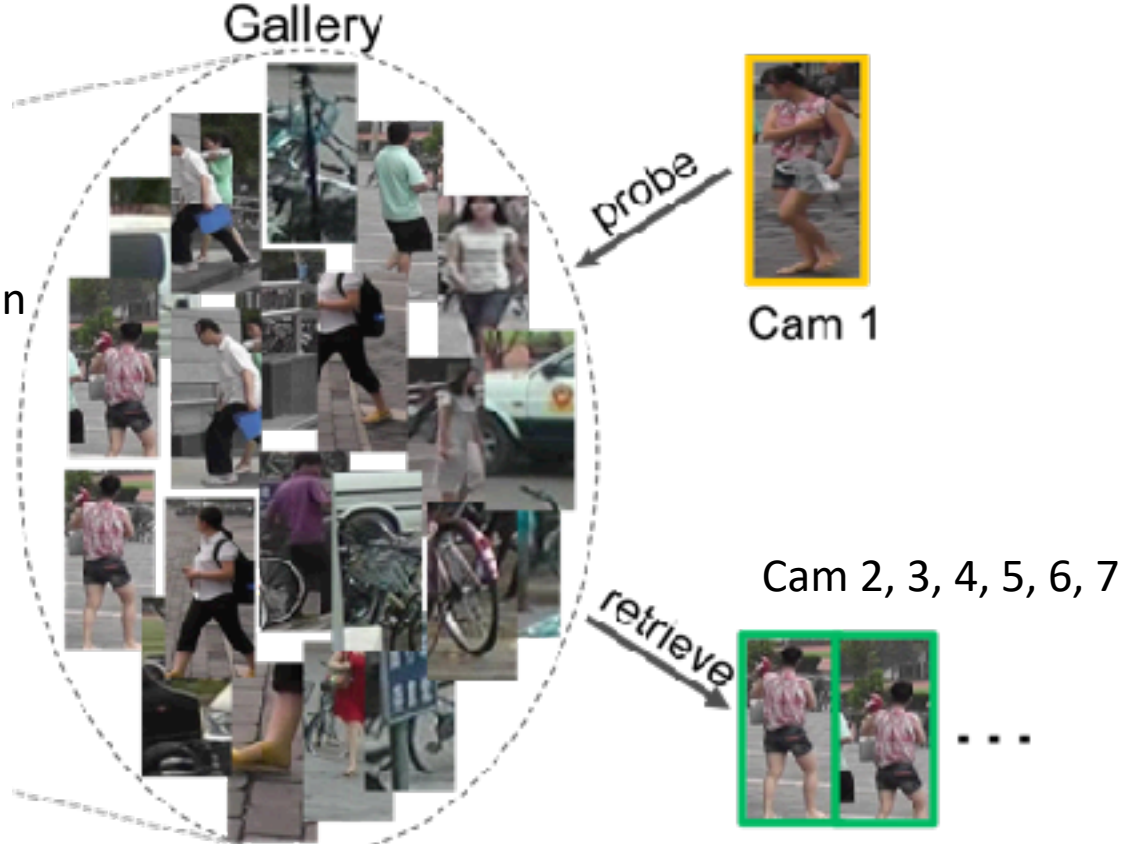
3. 无人机与重识别的机遇与挑战 ACM Multimedia2020

Background

- Person re-identification (re-id) is to find the person of interest from different camera views.
- The emergence of this task can be attributed to 1) the increasing demand of public safety and 2) the widespread large camera networks in public space.

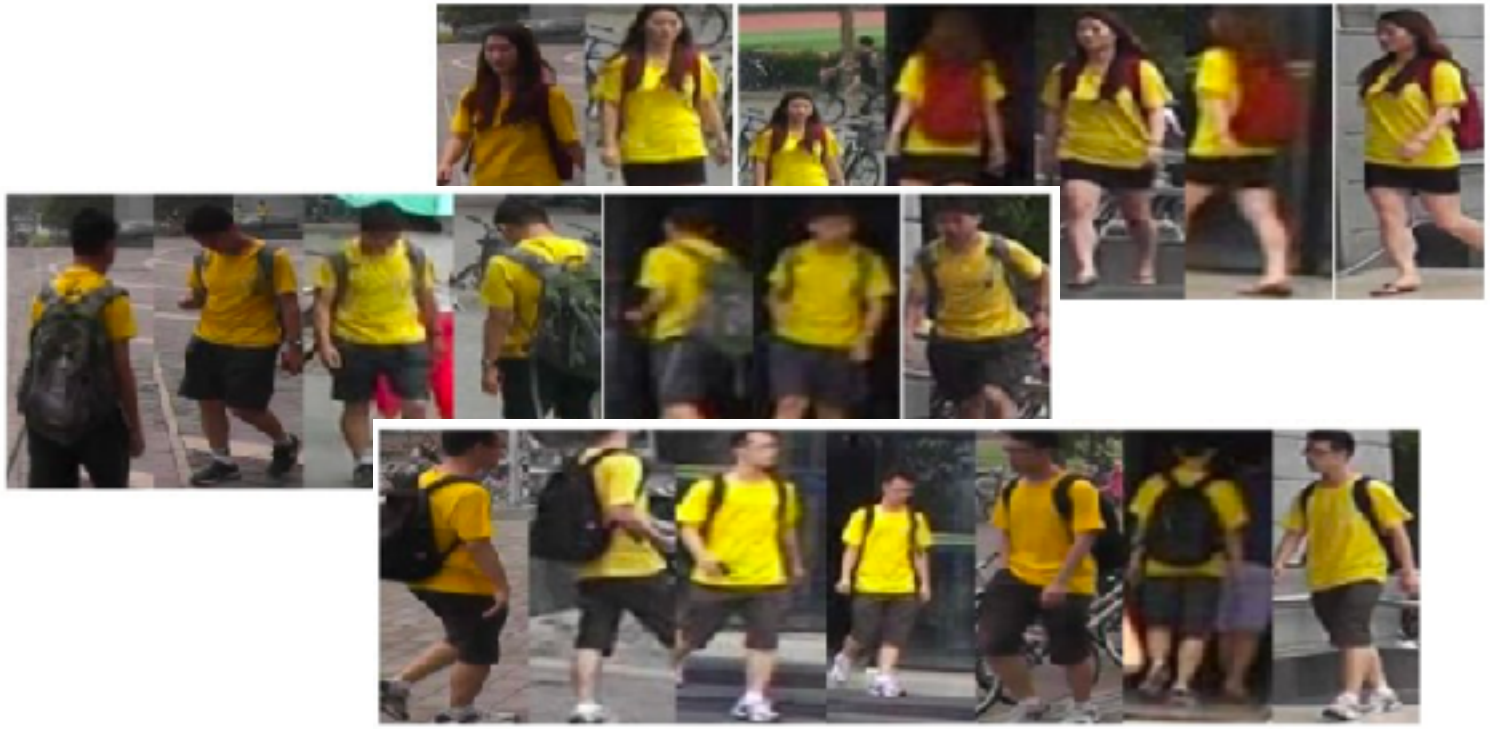


Person Detection



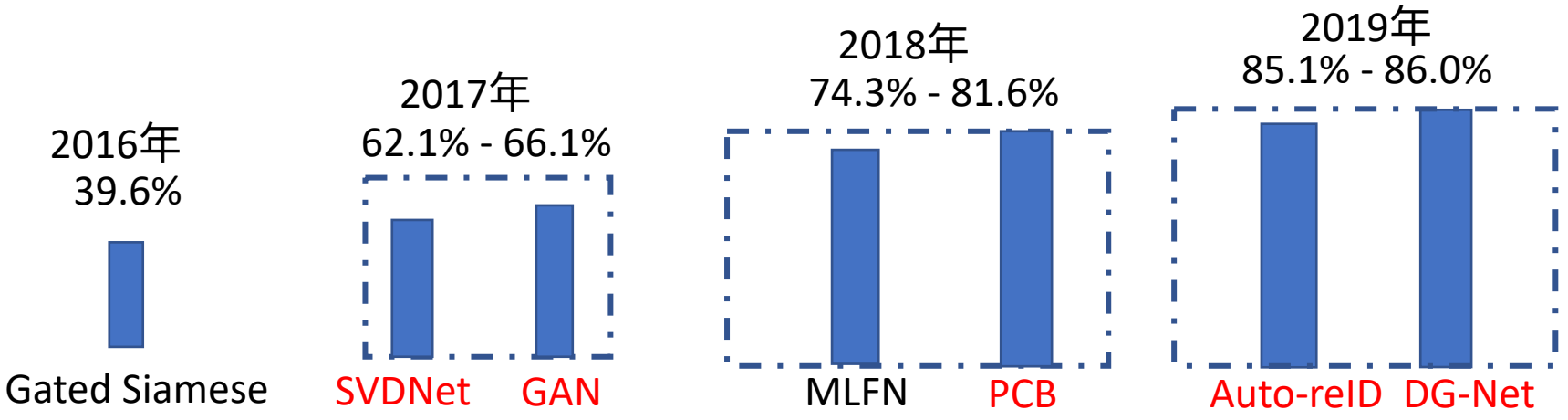
Background

- Person re-identification is challenging in the learning one discriminative and robust visual representation against the viewpoint changes.



All people love yellow shirt and short pants?

Evolution in State-of-the-art Performance



Market-1501, ResNet-50, mAP accuracy without re-ranking

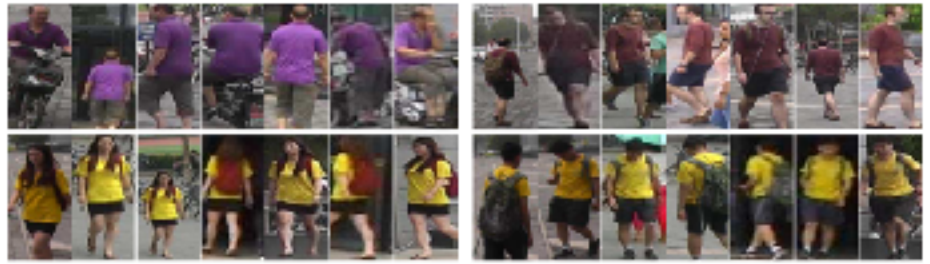
Methods	Market-1501		DukeMTMC-reID	
	Rank@1	mAP	Rank@1	mAP
Gated S-CNN [43]	65.9	39.6	-	-
Verif.-Identif. [56]	79.5	59.9	68.9	49.3
DCF [22]	80.3	57.5	-	-
DLPAR [53]	81.0	63.4	-	-
SSM [2]	82.2	68.8	-	-
SVDNet [40]	82.3	62.1	76.7	56.8
GLAD [48]	89.9	73.9	-	-
HA-CNN [24]	91.2	75.7	80.5	63.8
MLFN [4]	90.0	74.3	81.0	62.8
Part-aligned [39]	91.7	79.6	84.4	69.3
PCB [41]	93.8	81.6	83.3	69.2
Mancs [44]	93.1	82.3	84.9	71.8
Deform-GAN [36]	80.6	51.3	-	-
LSRO [57]	84.0	66.1	67.7	47.1
Multi-pseudo [16]	85.8	67.5	76.8	58.6
PT [27]	87.7	68.9	78.5	56.9
PN-GAN [33]	89.4	72.6	73.6	53.2
FD-GAN [8]	90.5	77.7	80.0	64.5
Ours	**94.8**	**86.0**	**86.6**	**74.8**

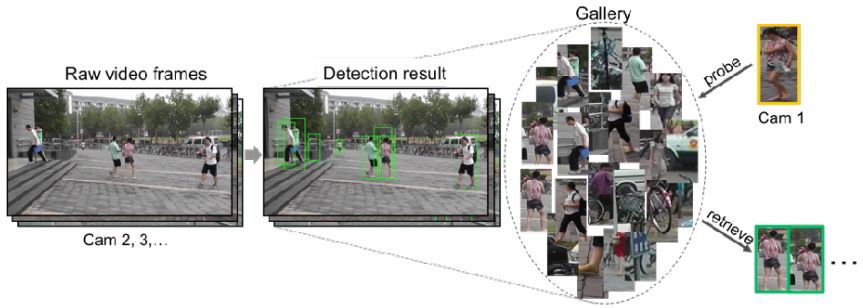
Table 4: Comparison with the state-of-the-art results on Market-1501 and DukeMTMC-reID. Group 1: the methods without using generated data. Group 2: the methods using separately generated images.

1. R. R. Varior, M. Haloi, and G. Wang. Gated Siamese convolutional neural network architecture for human reidentification. In ECCV, 2016.
2. Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In ICCV, 2017.
3. Y. Sun, L. Zheng, W. Deng, and S.Wang. SVDNet for pedestrian retrieval. In ICCV, 2017.
4. Suh et al., Part-Aligned Bilinear Representations for Person Re-identification, In ECCV 2018.
5. Sun et al., Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline). In ECCV 2018
6. Quan et al., Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification. In ICCV 2019.
7. Zheng et al. Joint discriminative and generative learning for person re-identification. In CVPR 2019

Datasets

- Image-based Datasets:
  - Market-1501 (ICCV 2015)
  - CUHK03 (CVPR 2014)
  - DukeMTMC-reID (ICCV 2017)
  - DG-Market (CVPR 2019)
  - MSMT-17 （CVPR 2019）
  - …
- Tracklet-based Datasets:
  - iLIDS (BMVC 2009)
  - MARS (ECCV 2016)
  - DukeMTMC-video (CVPR 2018)
  - …
- Scene-based Datasets:
  - PRW (CVPR 2017)
  - CUHK-SYSU (CVPR 2017)
  - …





https://github.com/NEU-Gou/awesome-reid-dataset

Large-scale Datasets are needed.

One Million Training Images for ImageNet

Dataset	CUHK01	VIPeR	PRID	CAVIAR	Market	DukeMTMC	MSMT-17
BBoxes	3,884	1,264	1,134	610	32,668	36,411	126,441
Identities	971	632	934	72	1,501	1,812	4,104
Cameras	10	2	2	2	6	8	15
Detector	hand	hand	hand	hand	DPM	hand	FasterRCNN
Scene	indoor	outdoor	outdoor	indoor	outdoor	outdoor	indoor/ outdoor

- Due to the annotation cost and privacy concerns, large-scale datasets are not easy to obtain.
- Although recent re-id datasets contain more images, they are still far from the real-world application.

Current Challenges

1. Limited Training Data 数据
2. Effectiveness 性能
3. Efficiency 效率
4. Domain Gap 实用
5. Unconstrained Environment (e.g., Occlusion) 实用

Potential Solutions

1. Synthetic Pedestrian Images
2. Parts/Losses
3. Auto-ML / Pruning
4. Domain Adaptation
5. Alignment / 3D Model

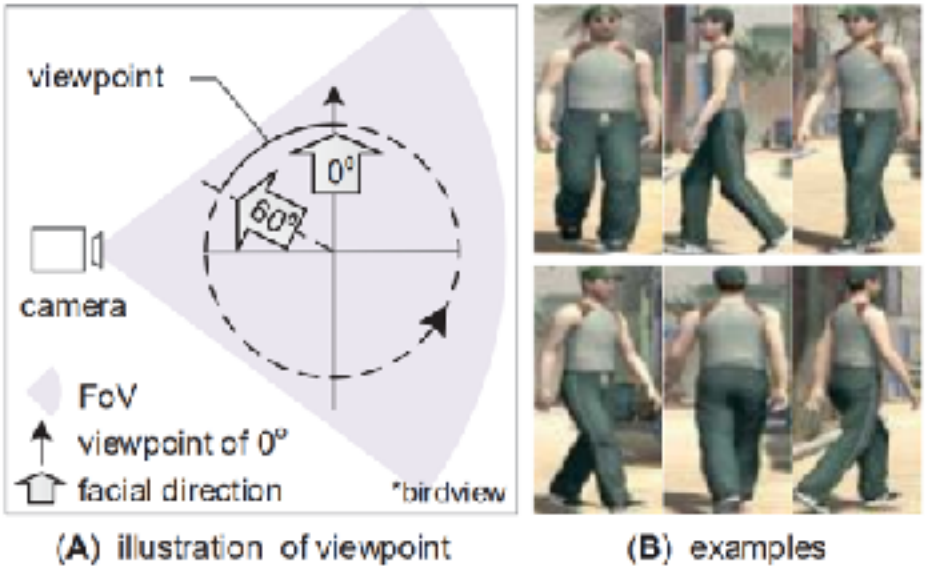
Limited Training Data

3D Game Engine

- 最直接的方法，我们可以通过3D游戏引擎来生成不同视角下的人体数据。但是，缺点是和实际的数据分布有一些差异。相比实际非限制场景来说，相对简单。

SOMAset and SOMAnet
(Barbosa *et al.,* CVIU 2018)

PersonX (Sun *et al.,* CVPR 2019)



Sun X, Zheng L. Dissecting person re-identification from the viewpoint of viewpoint In CVPR 2019.

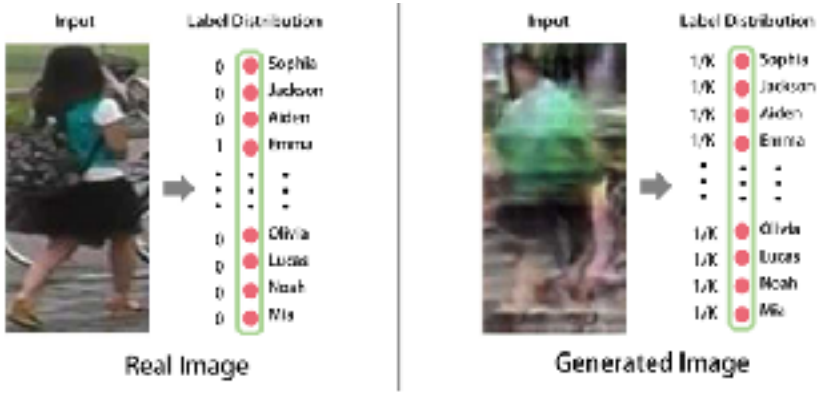
Generative Adversarial Network (GAN)

- 我们也可以使用GAN去模拟训练数据分布。借助GAN生成图像，我们缓解了数据集数据有限的问题。让模型看到更多样本，提升一定的鲁棒性。我们从ICCV17到CVPR19的两个工作，改进了生成质量，让模型可以端到端的生成和学习reID特征。

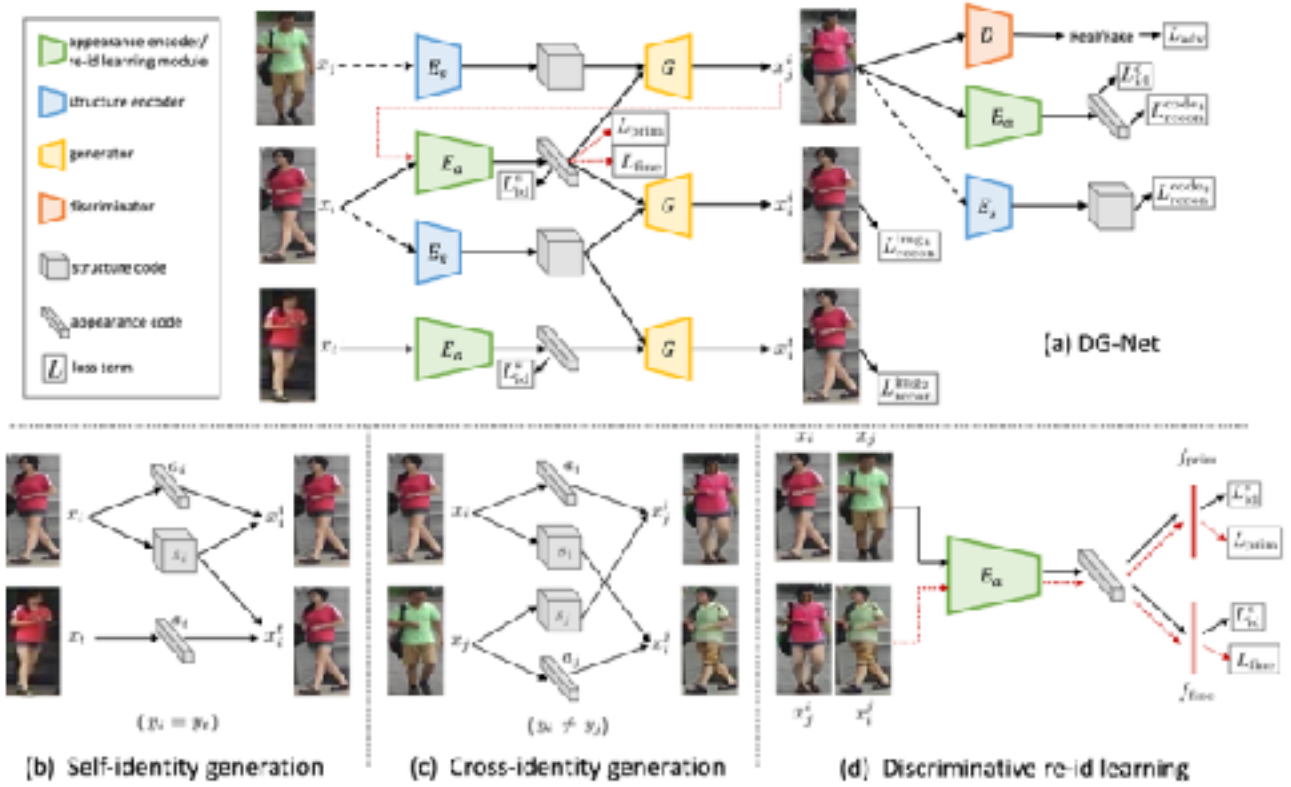
"What I cannot create, I do not understand."

—Richard Feynman

LSRO (Zheng *et al.,* ICCV 2017)



DGNet (Zheng *et al.,* CVPR 2019)

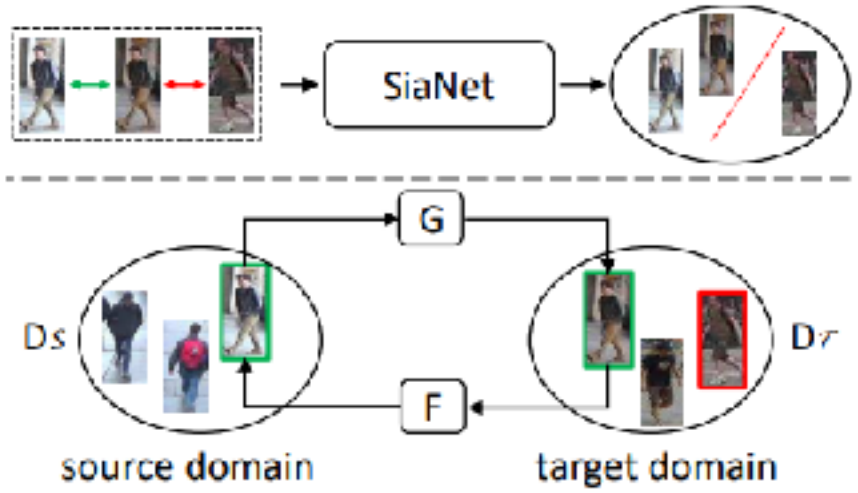


Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., & Kautz, J. (2019). Joint discriminative and generative learning for person re-identification. In CVPR 2019.

Generative Adversarial Network (GAN)

- 通过风格迁移，GAN也被广泛用于不同风格（光照）的行人图像生成。模型基于这些数据训练的，可以更容易迁移到不同应用场景。

SPGAN (Deng *et al.,* CVPR 2018)



PNGAN (Qian *et al., ECCV* 2018)
FDGAN (Ge *et al., NeurIPS 2018*)

PTGAN (Wei *et al.,* CVPR 2018)

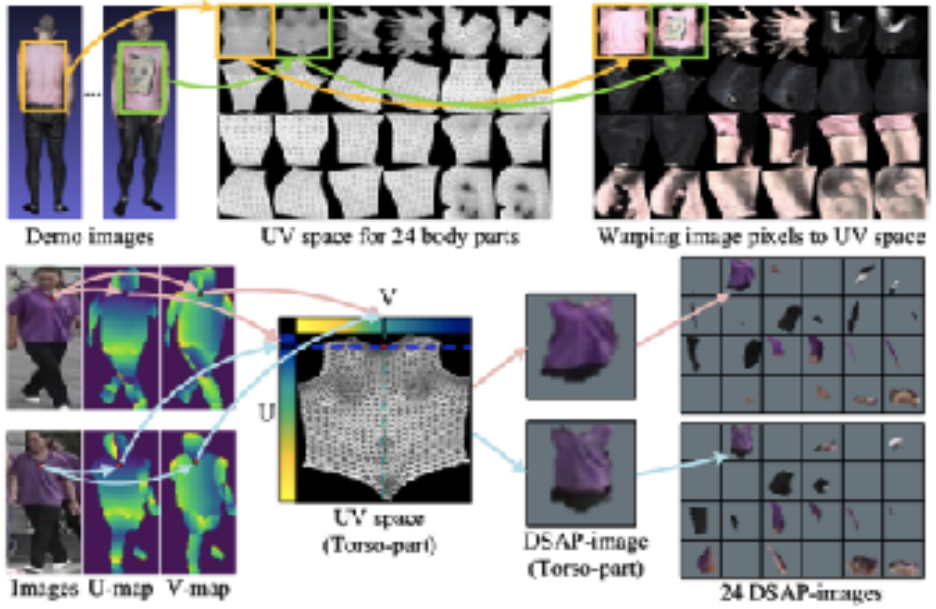
Effectiveness

Part Alignment in the Pixel Level

- 基于局部部件特征学习，最直接的方式就是在图像的像素层面先进行基于部件的像素对齐。

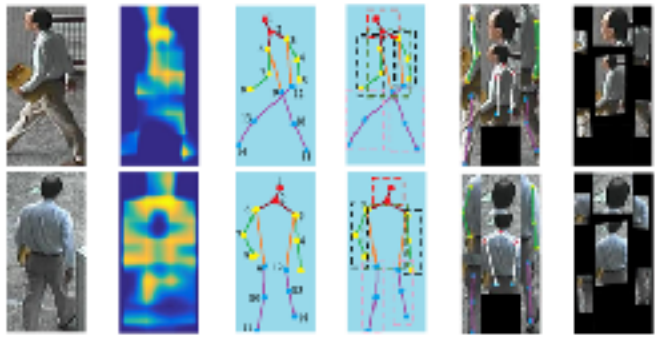
Dense semantic alignment

(Zhang *et al.*, CVPR2019)



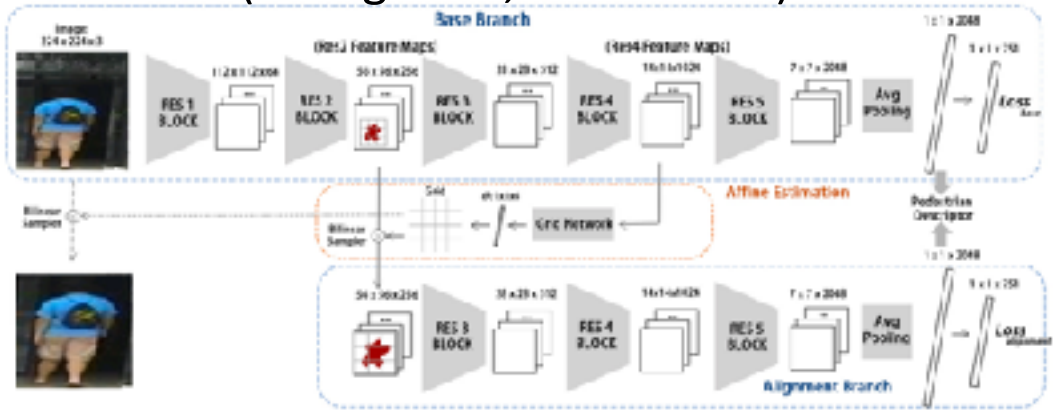
Pose-driven CNN

(Su *et al.*, CVPR2017)



Pedestrian alignment network

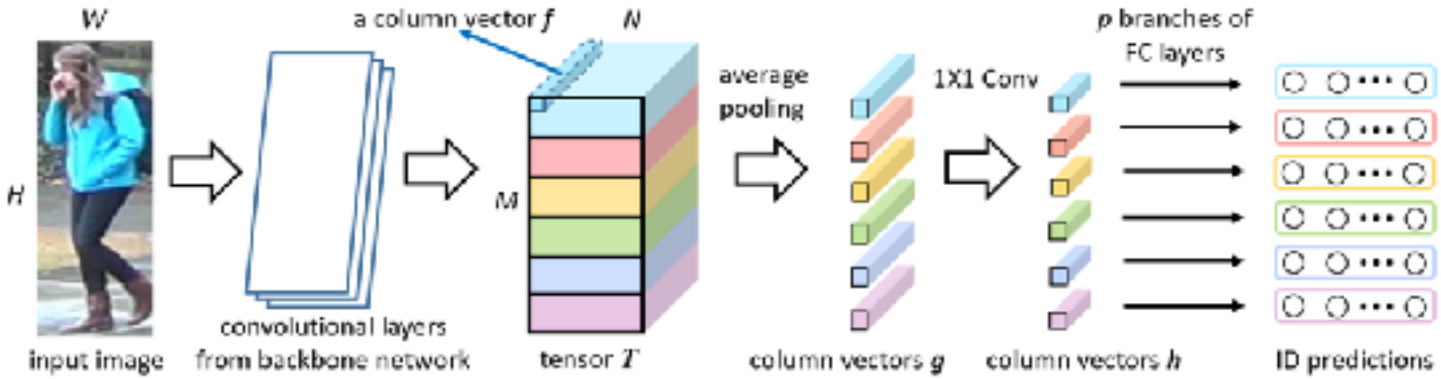
(Zheng *et al.*, TCSVT2018)



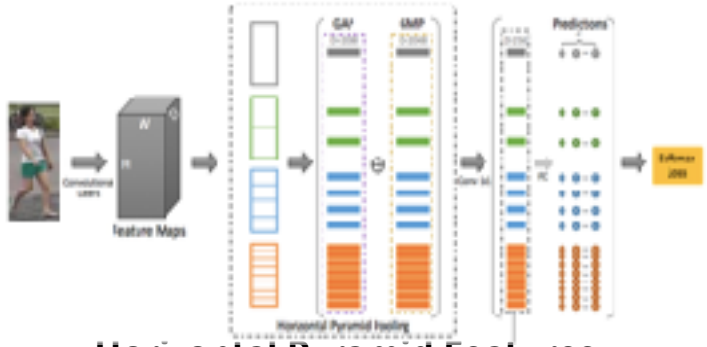
Part Alignment in the Feature Level

❑基于局部部件特征学习，以及特征对齐，我们在ECCV上达到了当时最佳重识别效果，形成了**解释性更强**特征表示的方法。

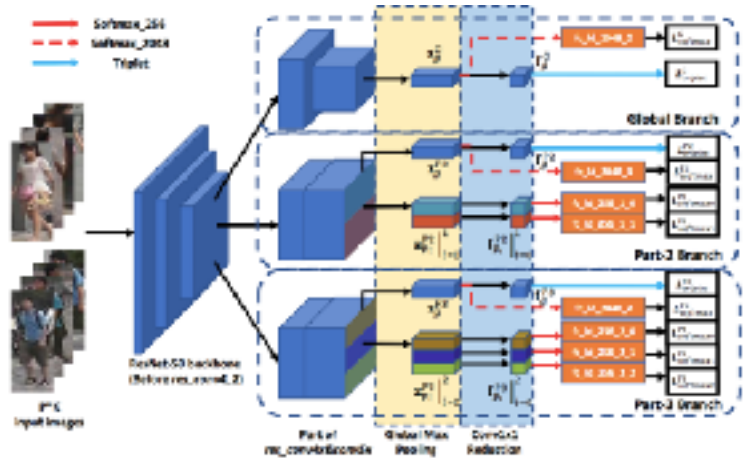
PCB: Part-based Convolutional Baseline (Sun *et al.,* ECCV2018)



Stripe partition
Simple and effective
93.8% R-1 on Market-1501



Horizontal Pyramid Features
(Yang *et al.,* AAAI 2019)



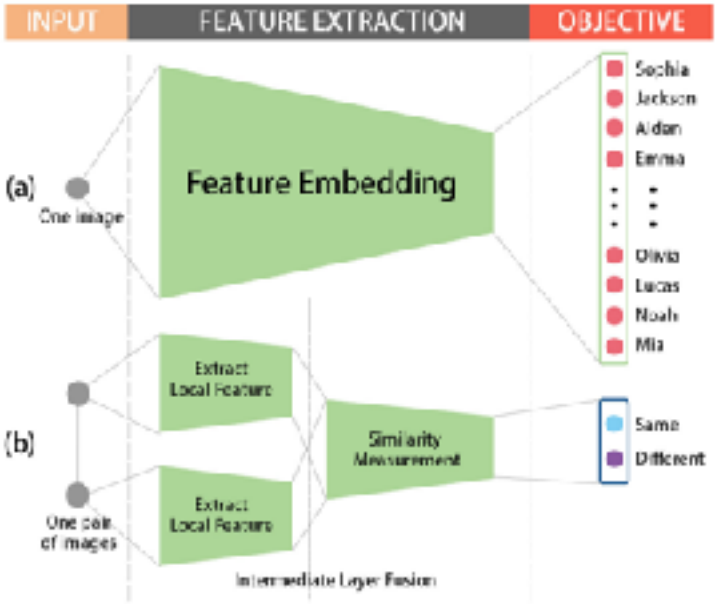
Multiple Granularities
(Wang *et al.,* ACM MM2019)

Yifan Sun, Liang Zheng, **Yi Yang**, Qi Tian, Shengjin Wang. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline). In *ECCV*, 2018.

Discriminative Losses

- 除了基于局部，也可采用不同的损失函数，从高维冗余特征中甄别行人身份，形成了**鉴别能力更强**特征表示。

Verification + Identification
(Zheng *et al.,* TOMM 2017)



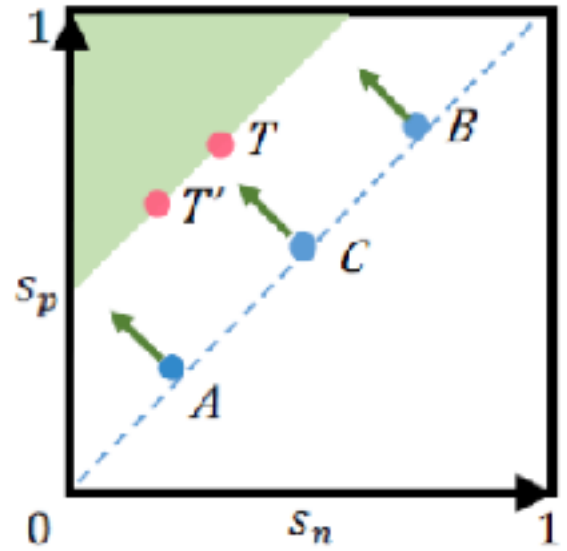
In Defence of Triplet
(Alexndar *et al.,* arXiv)

OIM Loss
(Xiao *et al.,* CVPR 2017)

Sphere ReID
(Fan *et al.,* arXiv)

Quadruplet Loss
(Chen *et al.,* CVPR 2017)

Circle Loss
(Sun *et al., CVPR 2020*)

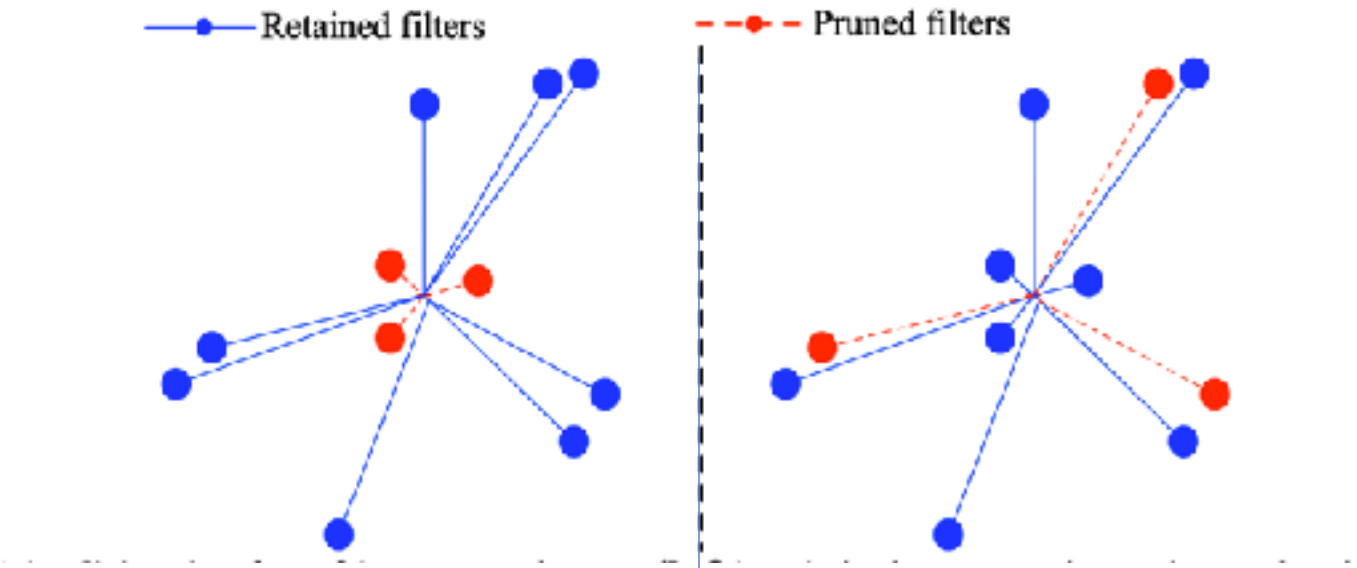


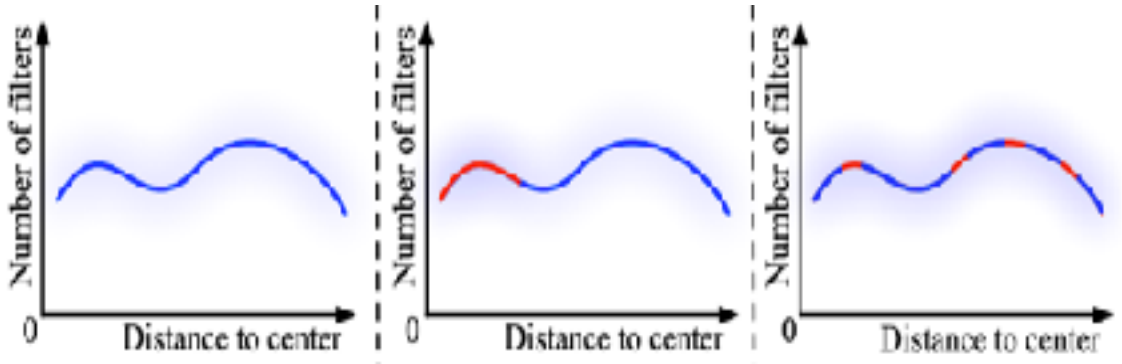
Zhedong Zheng, Liang Zheng, and Yi Yang. A discriminatively learned CNN embedding for person reidentification. *ACM TOMM*, 14(1):13, 2017.

Efficiency

Model Pruning for Image Retrieval

- 最直接的想法就是模型剪枝。今年我们提出了利用预训练网络的通道局部相关性的通道衰减“软”剪枝模型。针对最常用的ResNet-50主网络，可在减少88.85% FLOPs情况下，超越最新软剪枝方法8% mAP，原模型大小由27.1M，压缩为 6.97M。





传统方法　　我们

传统方法一般按照模长来删除filter

而我们根据局部聚类的冗余剪枝，可有效保持原始网络通道分布

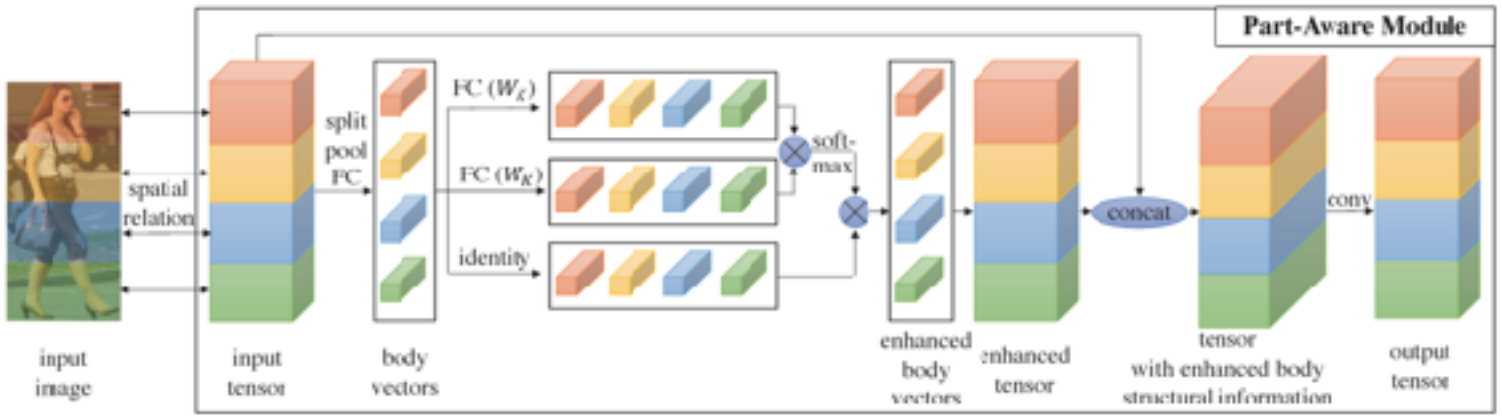
Wang, X., Zheng, Z., He, Y., Yan, F., Zeng, Z., & Yang, Y. (2020). Progressive local filter pruning for image retrieval acceleration. in arXiv.

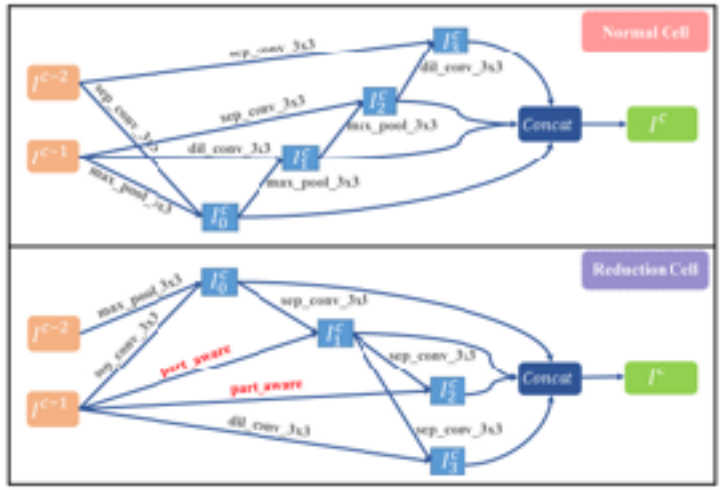
Auto-ML to Learn a Network for Re-id

- 使用自动机器学习方法自动化地设计针对行人再识别问题的神经网络结构，最终设计出比经典模型ResNet参数量少一半，运行开销也减少了**53%**，性能也达到最领先水平的行人再识别神经网络。

Auto Re-id

(Quan *et al.*, ICCV 2019)





ReID Search Space:

- *part-aware module*
- *3×3 max pooling*
- *3×3 average pooling*
- *3×3 depth-wise separable convolution*
- *3×3 dilated convolution*
- *zero operation, and identity mapping*

自动化设计行人再识别神经网络结构的可视化。如此复杂的结构很难被人为设计，其中设计的人体部位感知元件(part_aware)被自动化地安置在网络结构中。

Quan, R., Dong, X., Wu, Y., Zhu, L., & Yang, Y. (2019). Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification. In ICCV 2019.

Domain Gap

Domain Gap

❑实际场景中，往往由于摄像头角度、穿着、光照以及天气的不同，导致模型的准度下降，考验模型的泛化能力。

- 通常情况下，将一个已经训练好的模型直接应用到新场景、新数据集上，识别的准确率普遍较低。
- 右图展示了不同数据集里行人穿着、背景以及光照的差异。这些差异大大影响了模型的泛化能力。
- 如若对每一新场景都重新进行人工标记数据、训练模型，这将花费大量人力财力。

DukeMTMC-reID

Market-1501

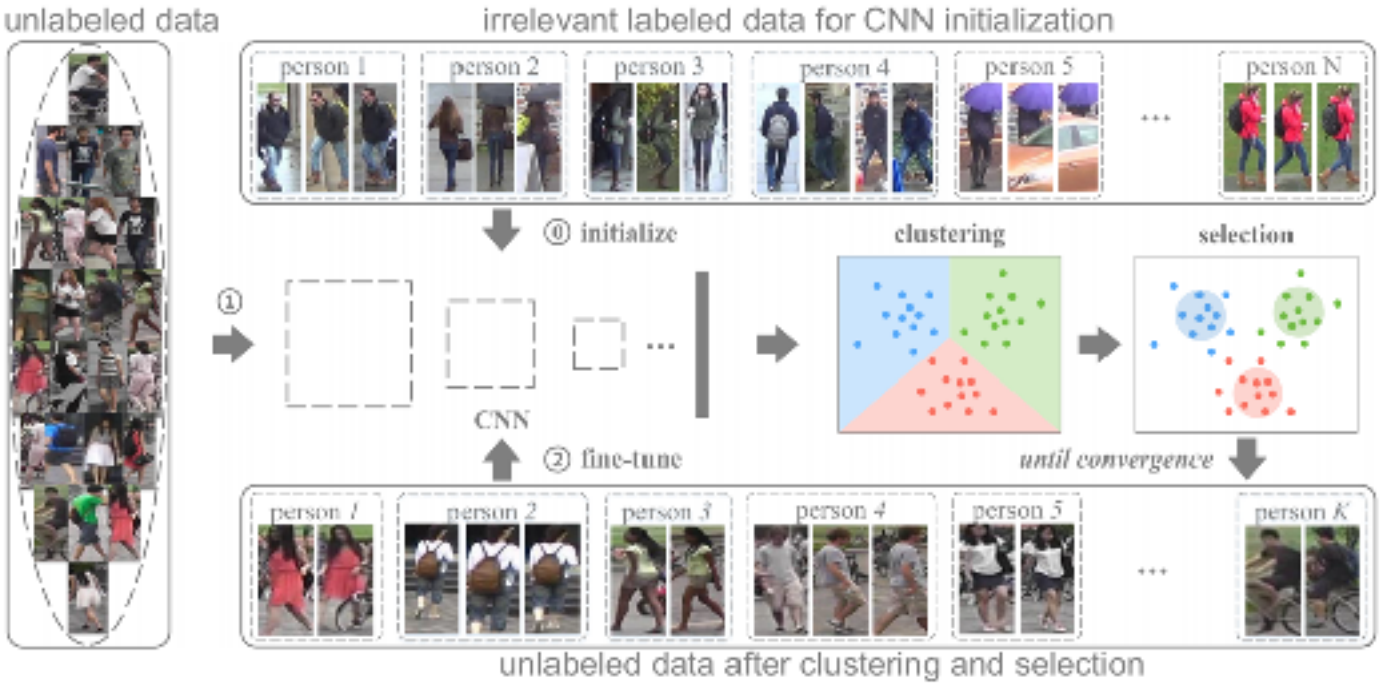
CUHK03

Domain Adaptation

- 对新场景的数据不进行人工标注，而是通过聚类的方式赋予新数据伪标签，最后对已有模型进行fine-tune。

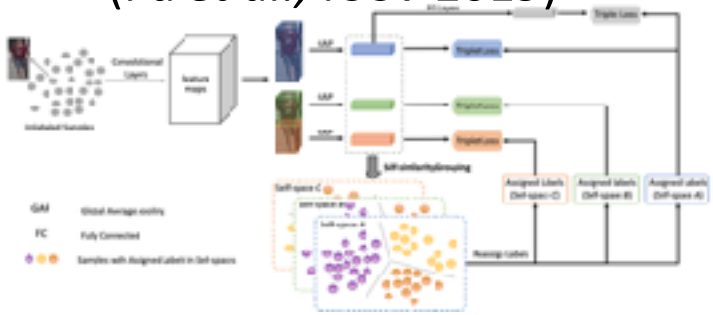
PUL

(Fan *et al.*, TOMM 2018)



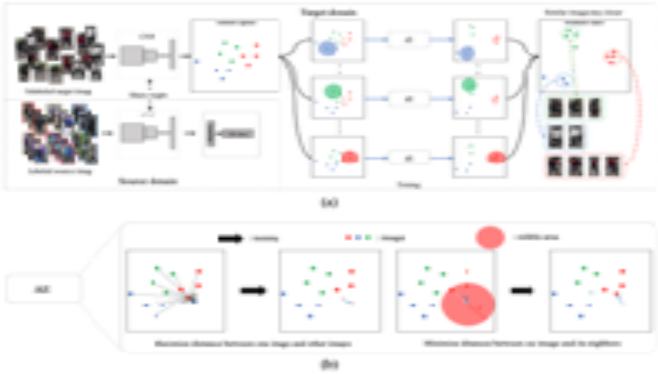
SSG

(Fu *et al.*, ICCV 2019)



Adaptive Exploration

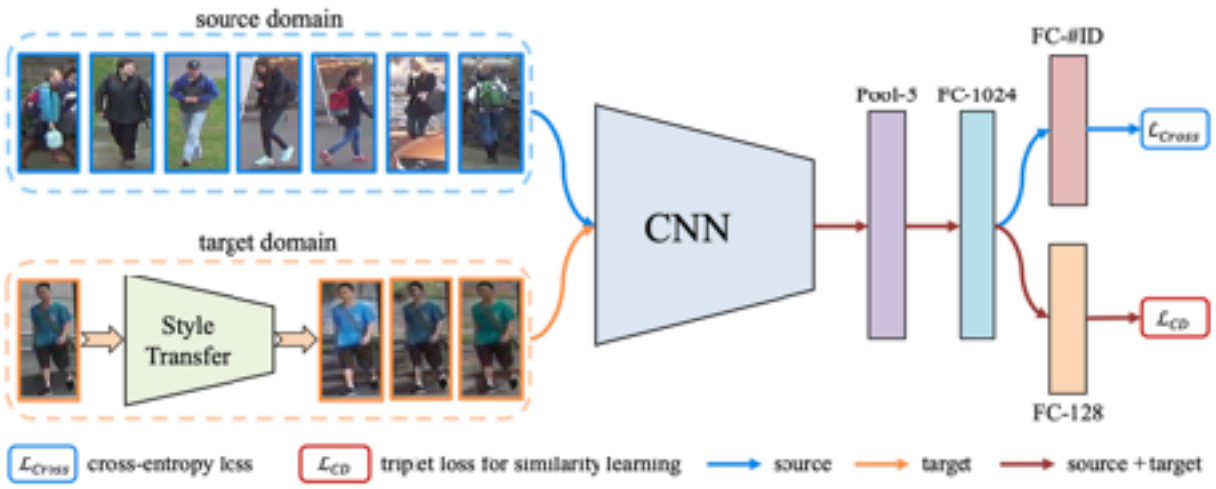
(Ding *et al.*, TOMM 2020)



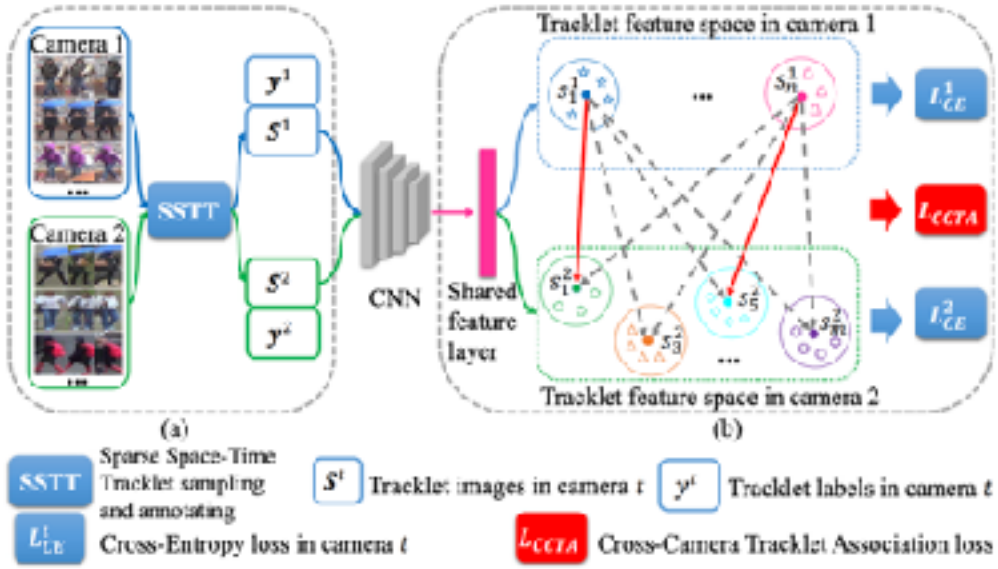
Domain Adaptation

❑利用先验知识和soft label 在目标域进一步挖掘信息。

HHL (Zhong *et al.*, CVPR 2018)



Tracklet Association
(Li *et al.*, ECCV 2018)



➢ **同构学习**：不同相机风格的同一样本组成正样本，学习相机不变性

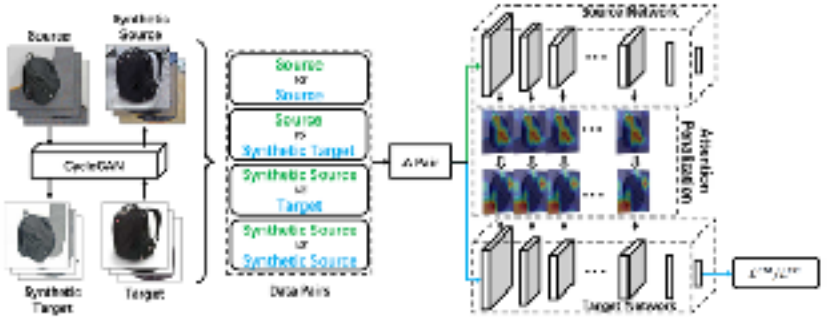
➢ **异构学习**：不同域的样本组成负样本对，学习域在特征空间中的潜在关系，提升域连通性

Generalizing A Person Retrieval Model Hetero- and Homogeneously. ECCV 2018

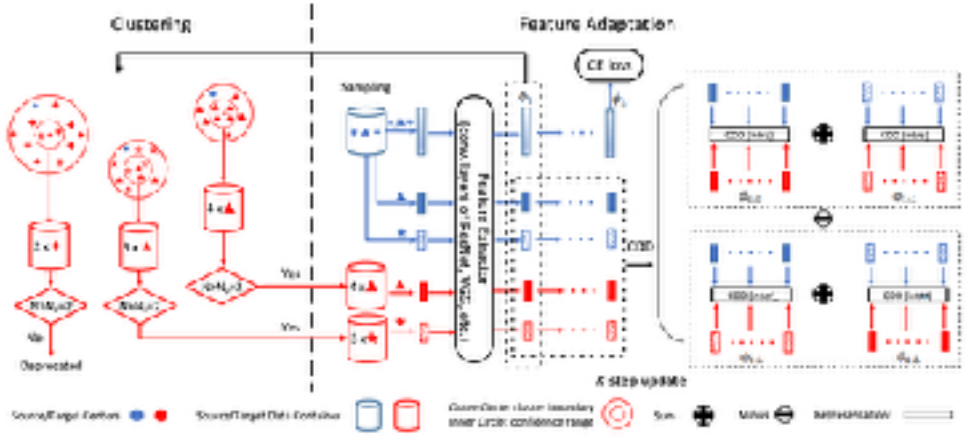
Domain Adaptation

- 我们在更通用的分类和语义分割问题上也取得了一些进展。

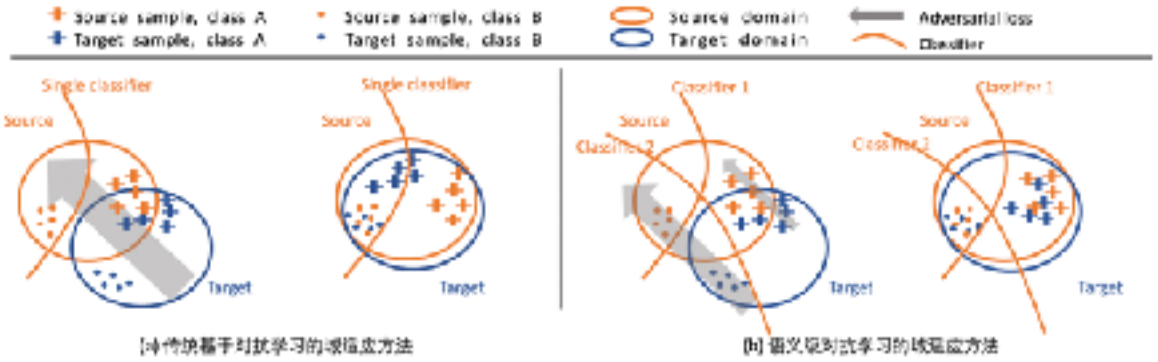
Deep Adversarial Attention Alignment
(Kang *et al.*, ECCV 2018)



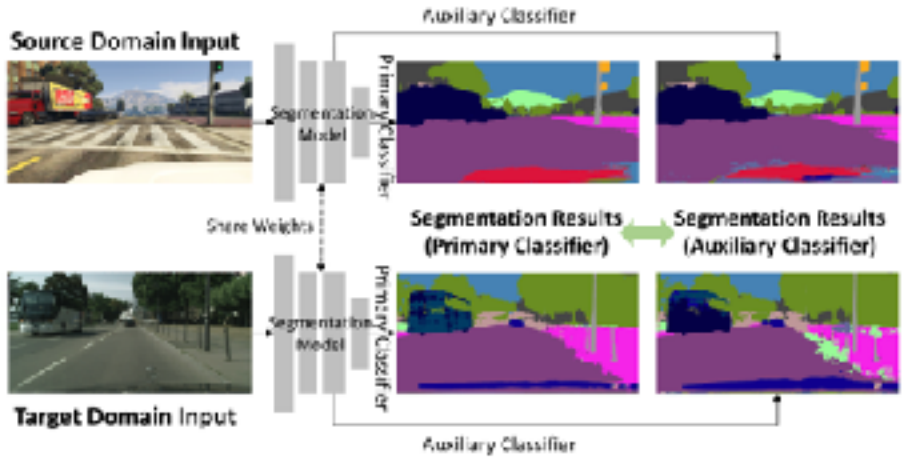
CAN (Kang *et al.*, CVPR 2019)



CLAN (Luo *et al.*, CVPR 2019)



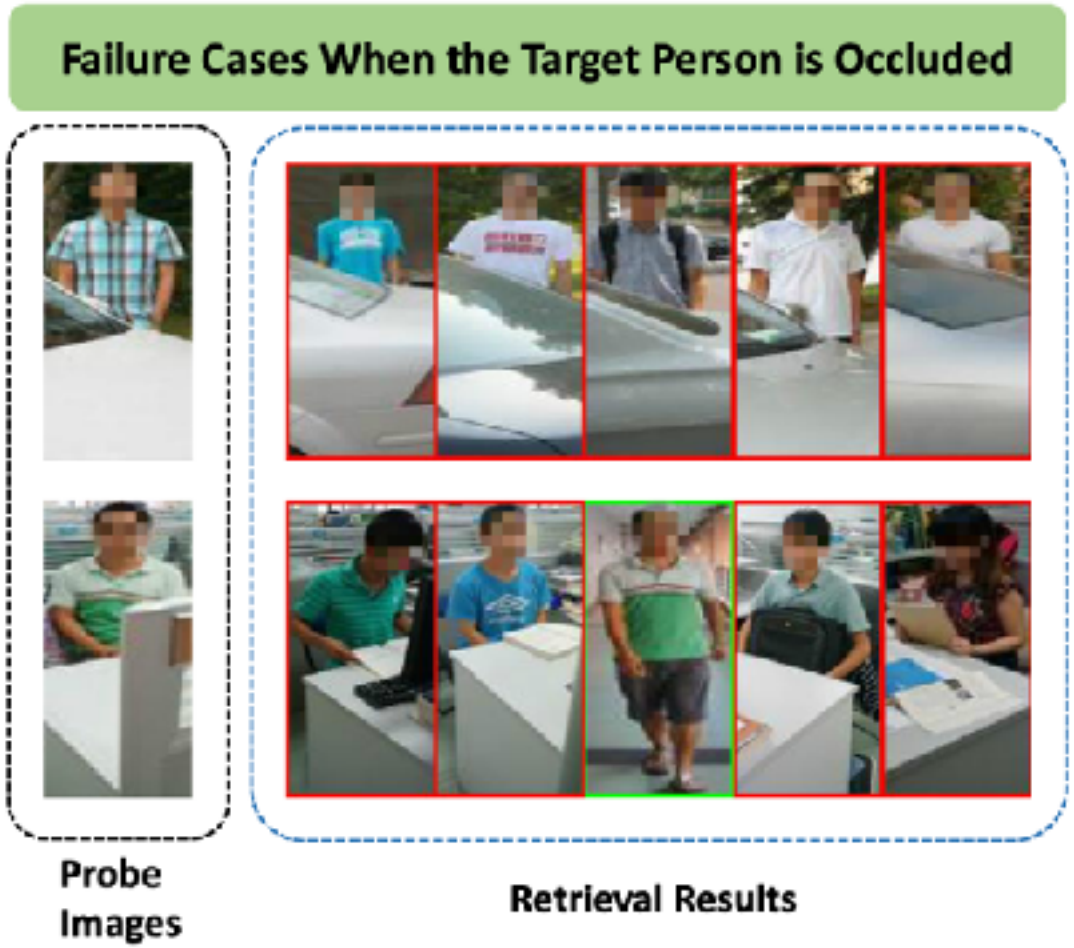
Memory Regularization (Zheng *et al.*, IJCAI 2020)



Unconstrained Environment

Occlusion

❑实际场景中，遮挡也常常无法避免，导致模型的准度下降，考验模型的泛化能力。

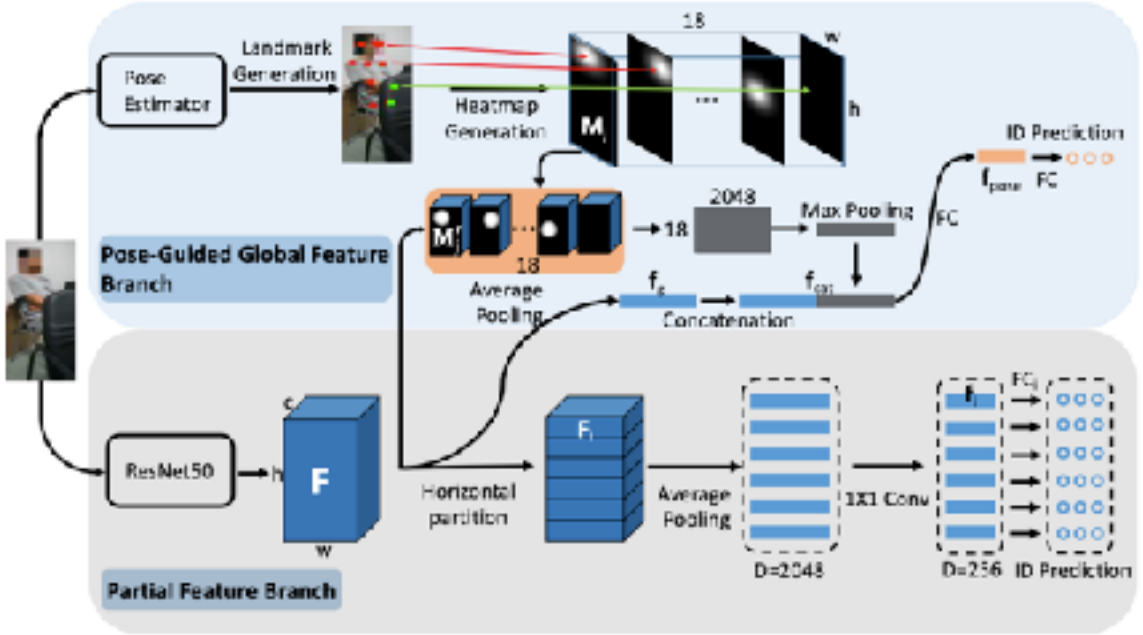


Leverage Person Structure

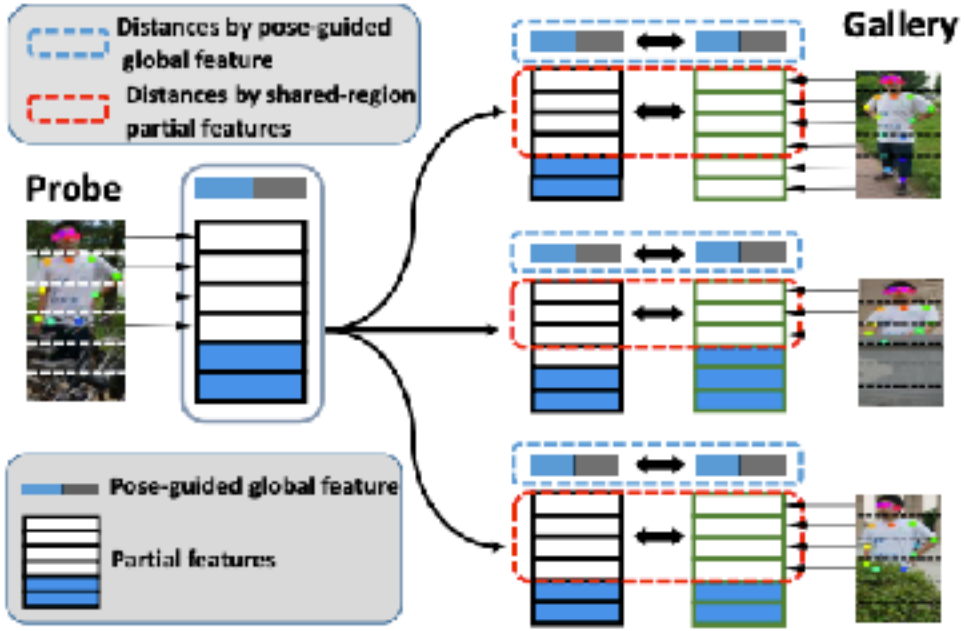
（1）行人重识别中的遮挡问题。解决思路：使用行人关键点检测，区分行人图片的遮挡部分和非遮挡部分。提取非遮挡部分的表征进行相似性匹配，忽略遮挡部分表征。

（2）构建大型遮挡行人识别数据库， Occluded-DukeMTMC。

训练：使用关键点生成的注意力图构建全局表征；
类似PCB，构建局部部件表征。



测试：使用关键点区分遮挡与非遮挡局部表征，
使用非遮挡表征进行相似性匹配，忽略遮挡部分。



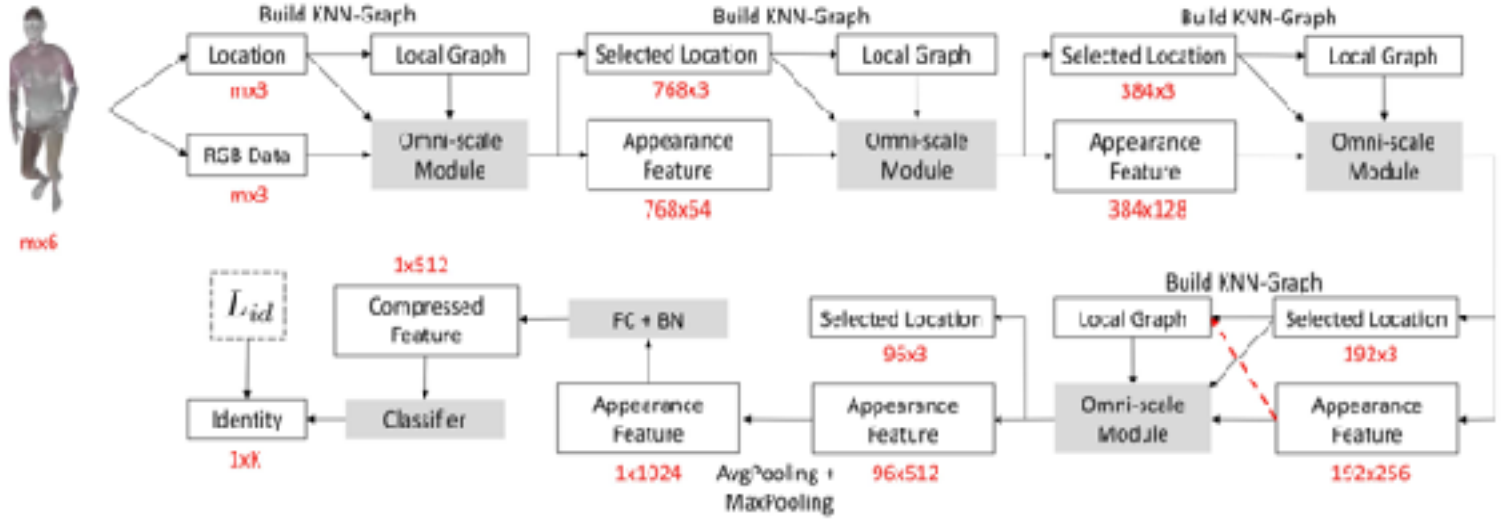
Miao J, Wu Y, Liu P, Ding Y, Yang Y . Pose-Guided Feature Alignment for Occluded Person Re-Identification. ICCV, 2019.

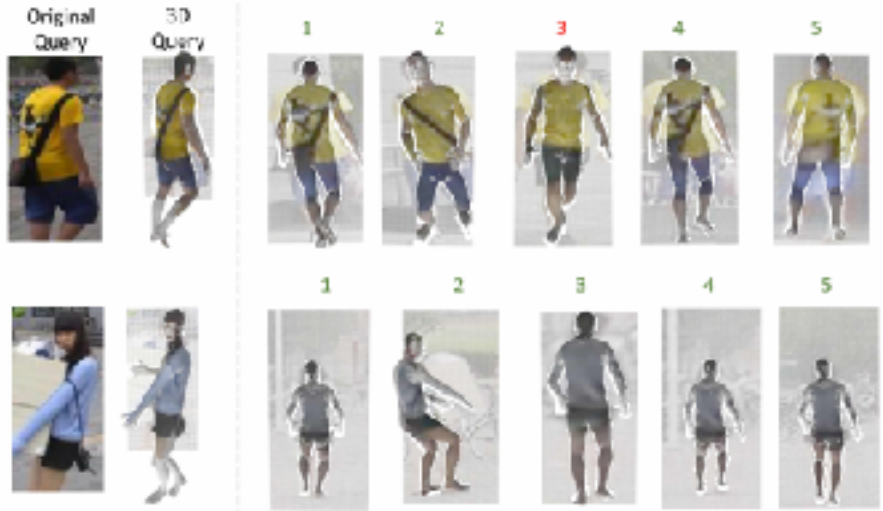
Person Re-identification in the 3D Space

- 人们生活在3D 的世界中。我们在2020年新的工作中，将人体映射到三维空间，然后再通过点云，融入几何结构(geometry structure)来学习人体表达，得到鲁棒特征，处理遮挡问题。同时，OG-Net只有9MB，推理速度也比常用的ResNet-50快。

OG-Net

(Zheng *et al.,* arXiv 2020)





Zhedong Zheng, Yi Yang . Person Re-identification in the 3D Space. ArXiv, 2020.

Multi-modal Inputs

- 最后，多模态的输入也可以提高模型在非限制场景下的鲁棒性，让模型更接近实用。比如通过语言描述来搜人。

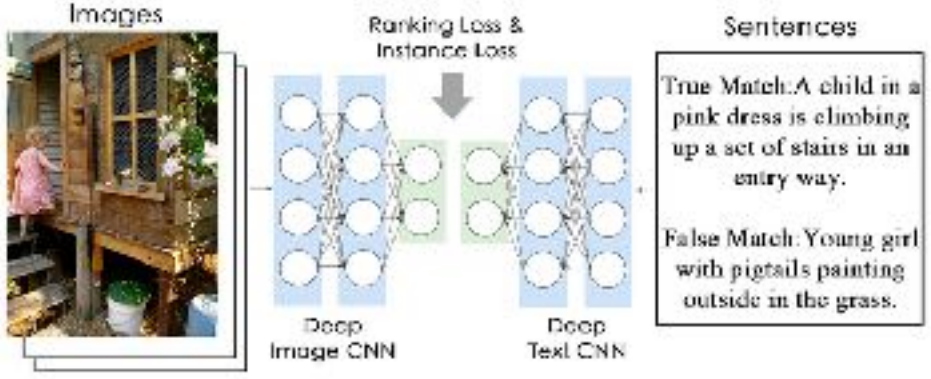
CUHK-PEDES (Li *et al.,* ICCV 2017)

Improving reid via Pedestrian Attribute
(Lin *et al.,* Pattern Recognition 2019)

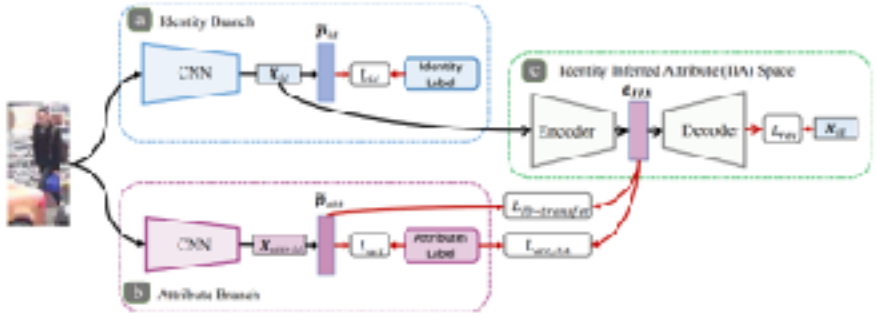
A *young man* wear a *yellow* shirt, *black* pants, and a *hat* without a *bag*

A *young woman* wear a *blue* shirt, *black* pants, and a *bag* without a *hat*

Instance Loss (Zheng *et al.,* TOMM 2020)



TJ-AIDL
(Wang *et al.,* CVPR 2018)



Future Research for Person Re-id

1. Millions of Distractor Images (Is the model robust?)
2. Efficient Training from Millions of Data (Which data is important? Long-tail)
3. Fast Domain Adaptation / Online Learning (Update model itself)
4. Unconstrained Environment (e.g., Occlusion and Illumination)

Outlines

1. 行人重识别的一些实践

2. 车辆重识别 CVPR2020 智慧城市比赛冠军 (车动，camera不动)

3. 无人机与重识别的机遇与挑战 ACM Multimedia2020

Going Beyond Real Data: A Robust Visual Representation for Vehicle Re-identification

AI City Challenge 2020

Zhedong Zheng[1,2*], Minyue Jiang[1*], Zhigang Wang[1],
Jian Wang[1], Zechen Bai[1], Xuanmeng Zhang[1,3],
Xin Yu[2], Xiao Tan[1], Yi Yang[2] , Shilei Wen[1] , Errui Ding[1]

[1] Baidu Inc.
[2] University of Technology Sydney
[3] Zhejiang University

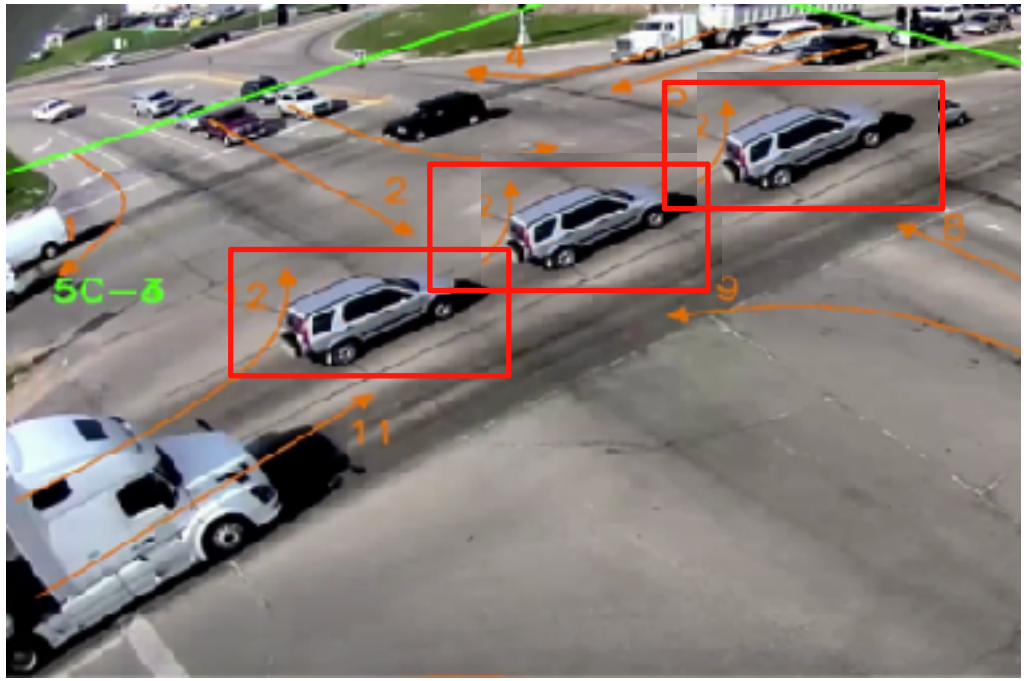
Baidu 百度 UTS

百度视觉技术部 COMPUTER VISION TECHNOLOGY

ZHEJIANG UNIVERSITY

What is Vehicle Re-identification?

Tracking



Re-identification

Main Challenge

?

What should we care about?

What should we care about?

- Limited Training Data.

We explore three different data generation approaches.

- Representation Learning.

We adopt two widely-used baseline.

- How to use Meta Data/ Attributes?

Meta-data also plays an important role in post processing.

What should we care about?

- **More Training Data**

We explore three different data generation approaches.

- A Strong Baseline

We adopt two widely-used baseline.

- Post-processing Methods

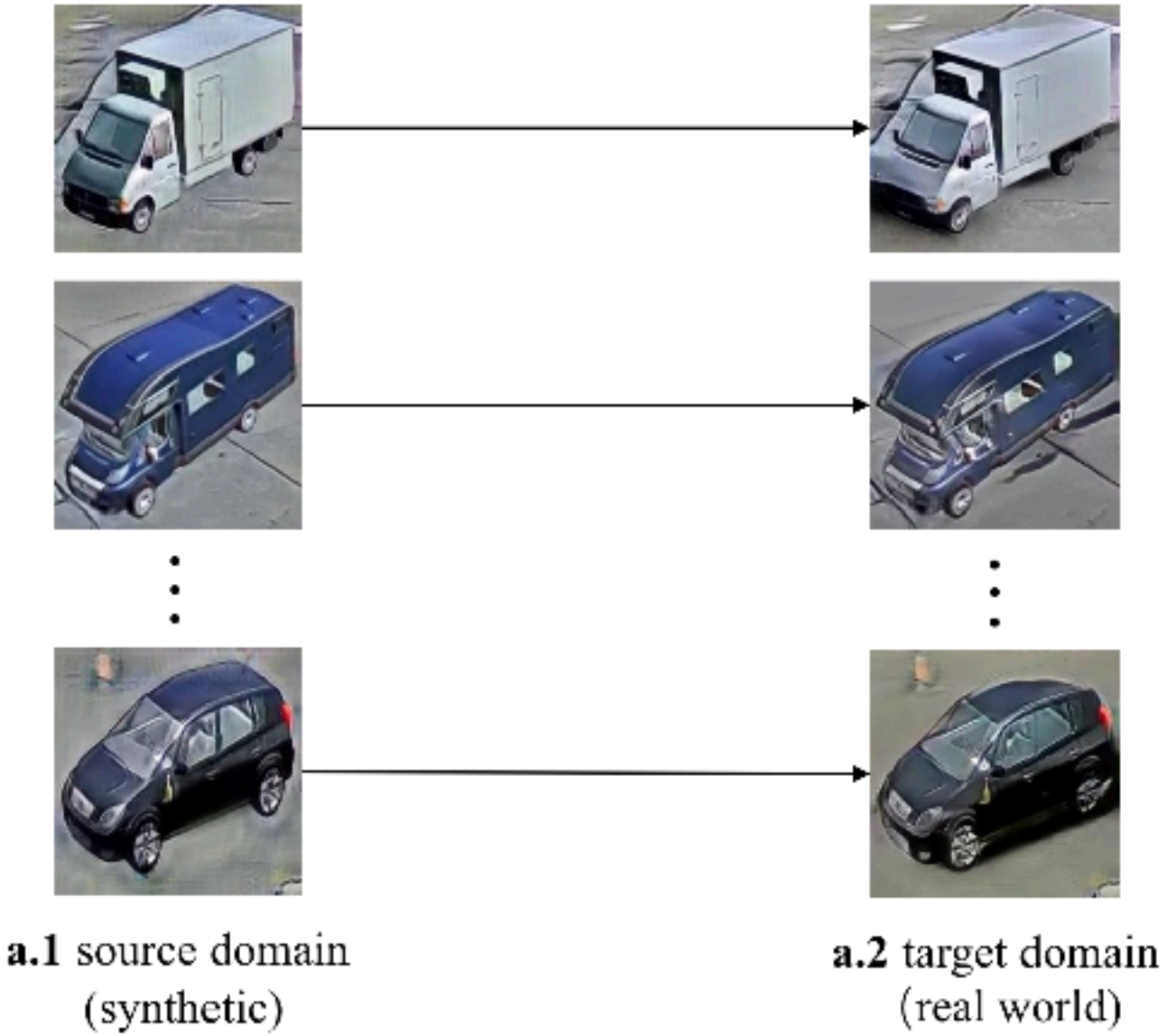
Meta-data also plays an important role.

Style Transfer

Real Data

Synthetic Data

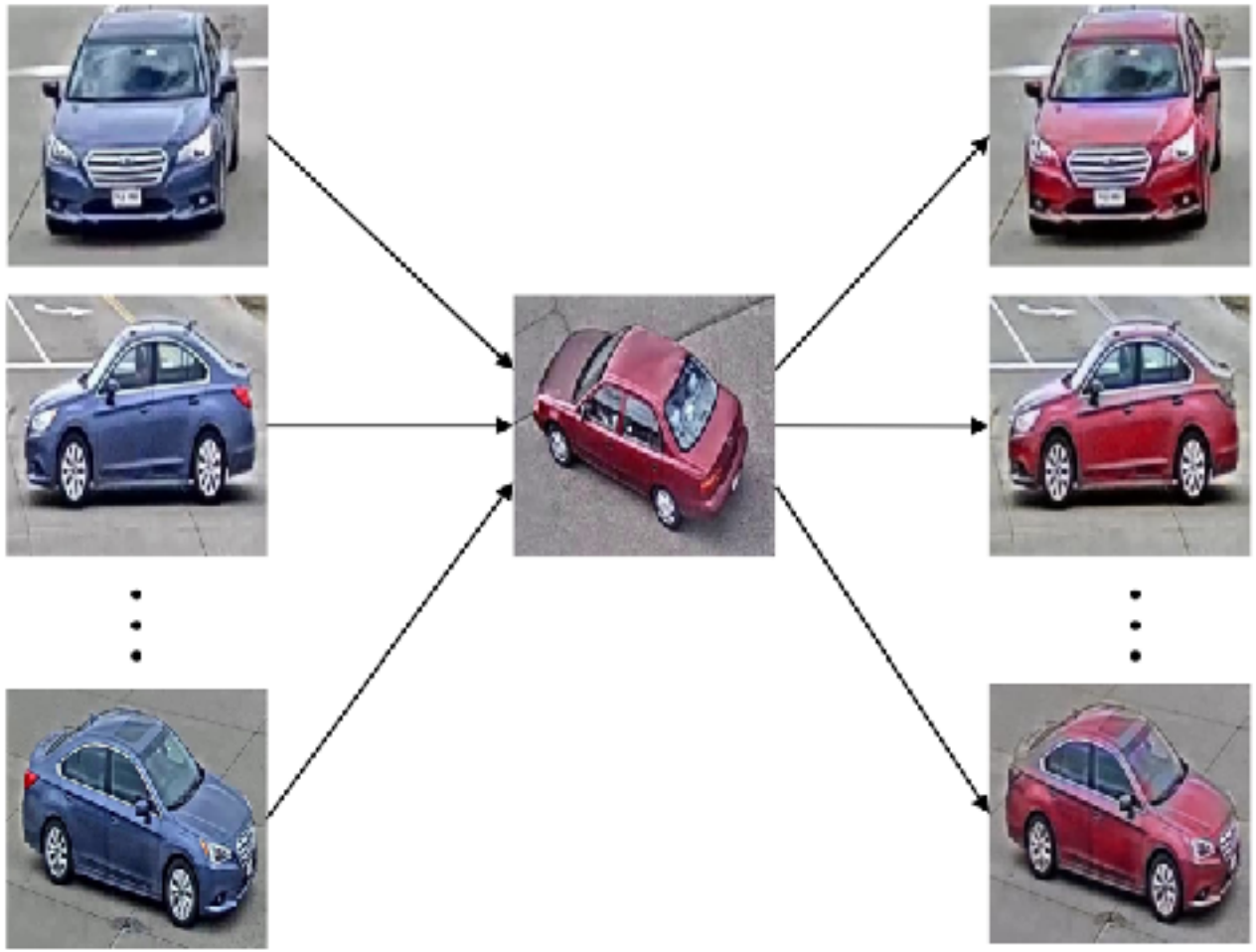
Style Transfer



a.1 source domain
(synthetic)

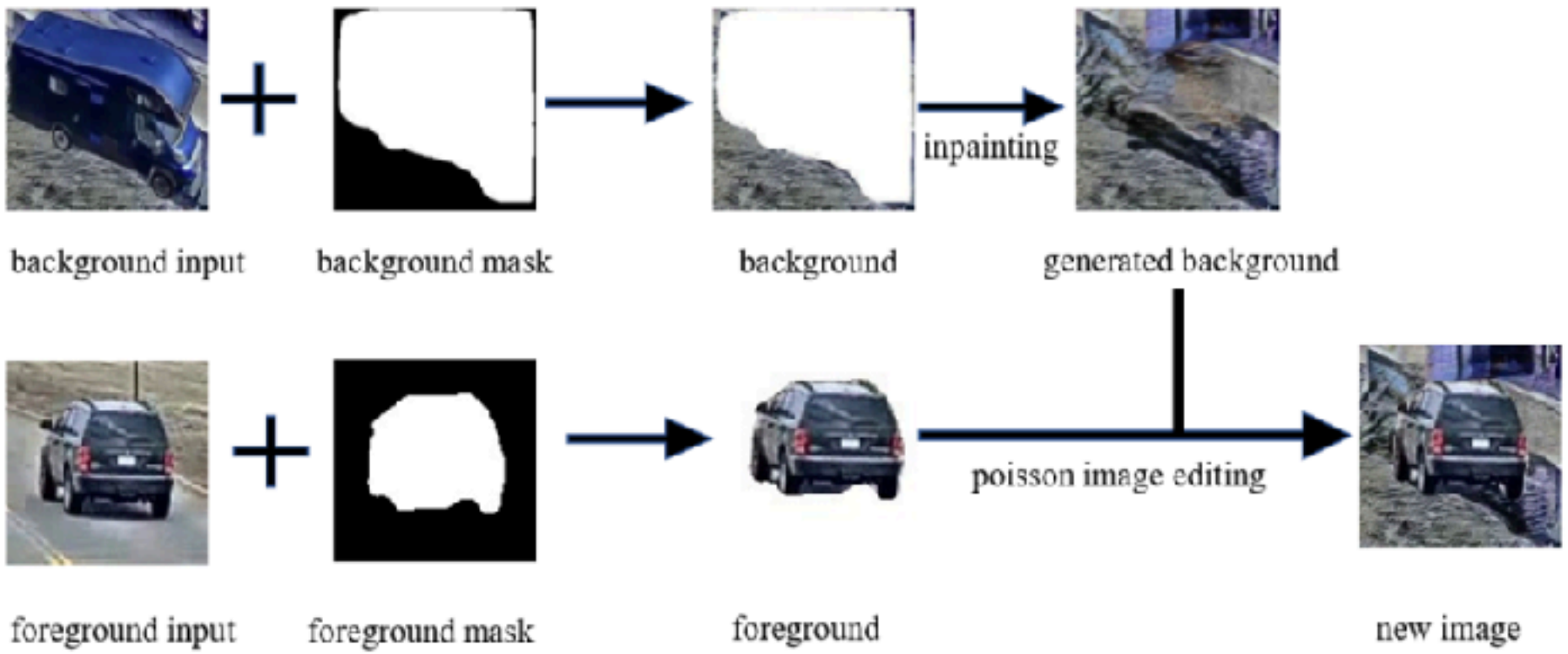
a.2 target domain
(real world)

Content Manipulation



b.1 source id **b.2** target appearance **b.3** new id

Copy & Paste


inpainting
background input
background mask
background
generated background
poisson image editing
foreground input
foreground mask
foreground
new image

What should we care about?

- More Training Data

We explore three different data generation approaches.

- **A Strong Baseline**

We adopt two widely-used baseline.

- Post-processing Methods

Meta-data also plays an important role.

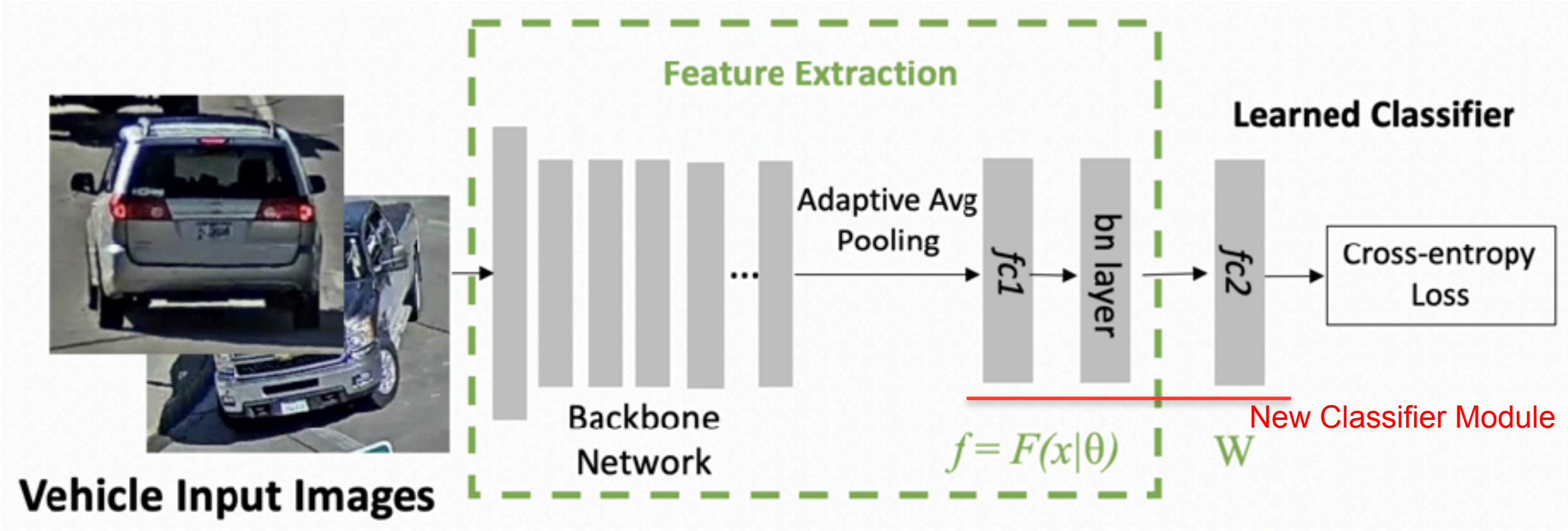
Cross-entropy Loss Definition

$$loss_{ce} = -\sum_{i=1}^{N} p_i log(\hat{p}_i),$$

Ranking Loss Definition

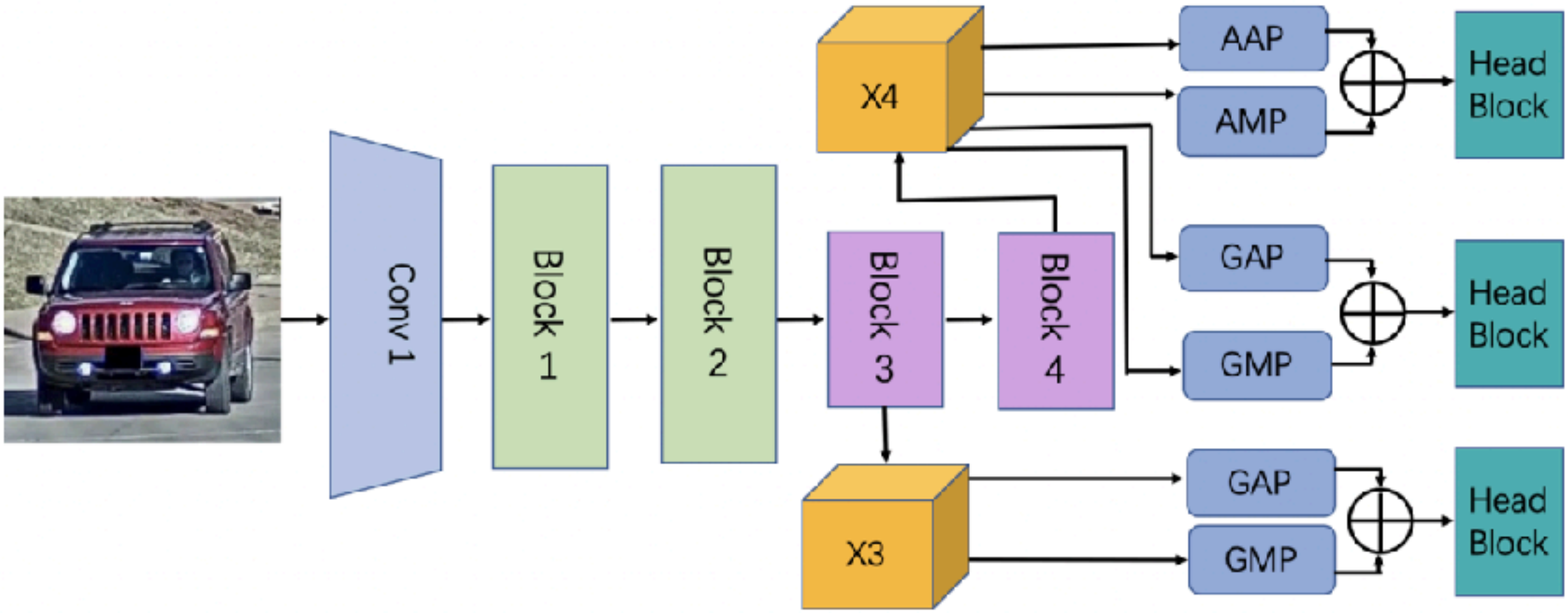
$$loss_{ranking} = [D_{ap} - D_{an} + m]_+,$$

Re-ID Baseline



- SVDNet for Pedestrian Retrieval
- In Defense of the Triplet Loss for Person Re-Identification

Re-ID Baseline



What should we care about?

- More Training Data

We explore three different data generation approaches.

- A Strong Baseline

We adopt two widely-used baseline.

- **Post-processing Methods**

Meta-data also plays an important role.

Camera Verification

Camera Verification

Experiment

AICity Competition

We achieve the 1st place in vehicle reID track of AICity Challenge, CVPR 2020.

Rank	Team ID	Team Name	Score
1	73	Baidu-UTS	0.8413
2	42	RuiYanAI	0.7810
3	39	DMT	0.7322
4	36	IOSB-VeRi	0.6839
5	30	BestImage	0.6684
6	44	BeBetter	0.6683
7	72	UMD_RC	0.6668
8	7	Ainnovation	0.6551
9	46	NMB	0.6206
10	81	Shahe	0.6191

- mAP Accuracy 高出第二名 6%

Code is available at https://github.com/layumi/AICIty-reID-2020

Ablation Study: With/without Synthetic Data

Table 2. Ablation Study. The Rank@1(%) and mAP (%) accuracy with / without synthetic training data.

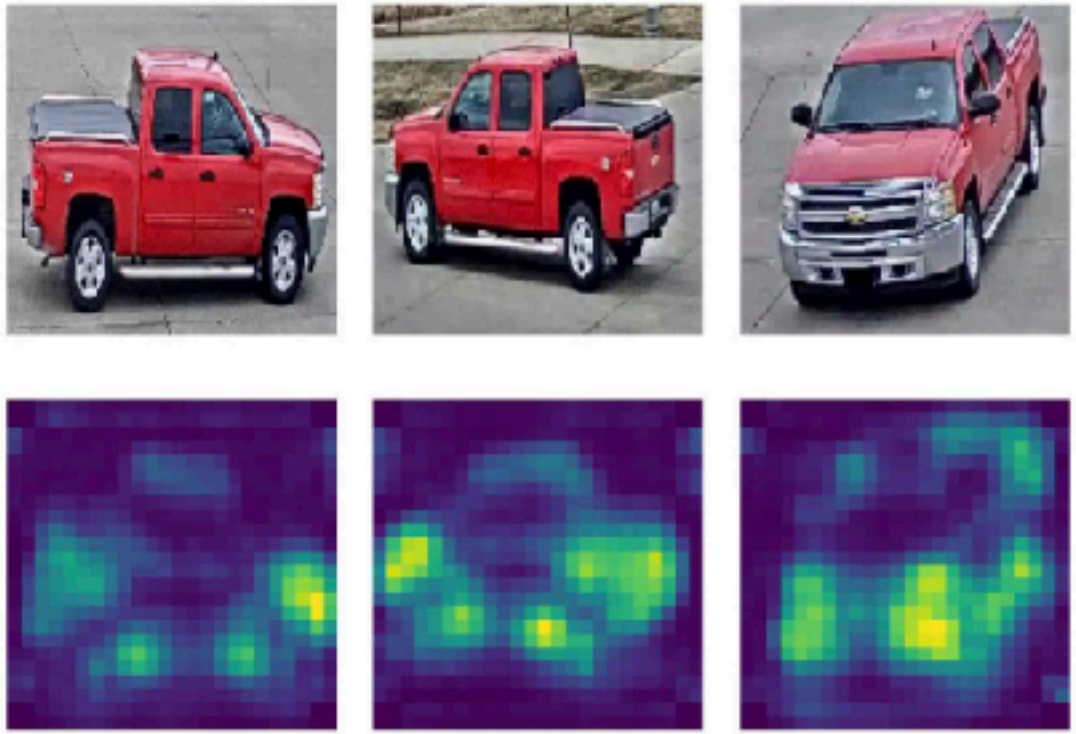
	Performance	
	Rank@1(%)	mAP(%)
without Synthetic Data	79.78	43.87
with Synthetic	80.86	46.90

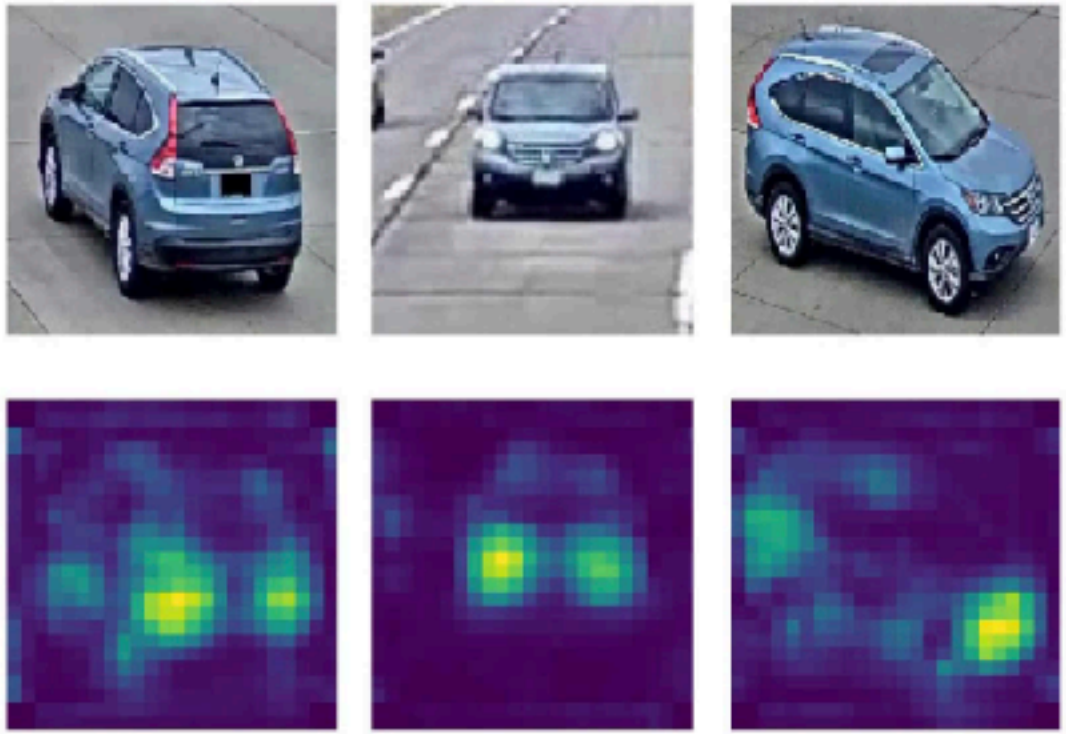
Ablation Study: Post-processing Methods

Table 3. Ablation Study. Effect of different post-processing techniques on the validation set.

Method	Performance					
with Alignment?		✓	✓	✓	✓	✓
Query Expansion?			✓	✓	✓	✓
Camera Verification?				✓	✓	✓
Group Distance?					✓	✓
Re-ranking?						✓
mAP (%)	46.90	47.66	49.06	50.07	51.58	61.26

Explainable





Future Works for Vehicle Re-id

Possible Approaches

1) Investigate the feasibility of high-fidelity **generated samples** for training. The generated samples could largely enrich the training set.
2) Mixture of **Unsupervised Learning/ Semi-supervised Learning**
3) **Domain Adaptation**

One last comment

Neural Networks are lazy

The models could easily overfit the datasets. Sometimes **adding prior knowledge and data augmentation** are important to obtain a robust re-id system.

Training Neural Networks sometime is tricky, and models will find the short way to overfit the objective. **Therefore, generated data with proper pseudo labels helps.**

The code is available at 

Outlines

1. 行人重识别的一些实践

2. 车辆重识别 CVPR2020 智慧城市比赛冠军

3. 无人机与重识别的机遇与挑战 ACM Multimedia2020 (camera动，建筑不动)

Drone is coming!

University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization

Zhedong Zheng, Yunchao Wei, Yi Yang
University of Technology Sydney

amazon.com.au

All drone

Hello, Sign in Account & Lists | Returns & Orders | Try Prime | 0 Cart

Hello Select your address | Best Sellers | Today's Deals | New Releases | Books | Gift Ideas | Electronics | Customer Service | Home | Computers | audible Free audiobook with trial

1-48 of over 60,000 results for "drone"

Sort by: Featured

Amazon Prime

- prime | Ships from Australia
- prime | International Shipping

Avg. Customer Review

- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up
- ★☆☆☆☆ & Up

Department

- Toys & Games
  - Hobby RC Drones & Multirotors
  - Toy Remote Control & Play Vehicles
- Electronics
  - Quadcopters & Accessories
  - Quadcopter Accessories
- Apps & Games
  - Game Apps
- See All 9 Departments

Brand

- DJI

Price and other details may vary based on size and colour

Amazon's Choice

DJI Mavic Mini – Drone FlyCam Quadcopter UAV with 2.7K Camera 3-Axis Gimbal GPS 30min Flight Time, less than 249g, Grey

★★★★½ ~ 773

$597.93

prime Get it by Wednesday, September

E Delivery by Amazon

e Buying Choices

6.00 (10 new offers)

REMOKING RC Drone with 720P FPV WI-FI HD Camera Live Video Racing Quadcopter Headless Mode 2.4GHz 360°flip 4 Channels Altitude Hold...

★★★★☆ ~ 32

Drone with 4K Camera Live Video,EACHINE E520 WIFI FPV Drone for Adults with 4K HD Wide Angle Camera 1200Mah Long Flight time...

★★★★☆ ~ 210

$169.40

prime Get it Tuesday, September 29 - Thursday, October 1

$13.53 shipping

Ages: 12 years and up

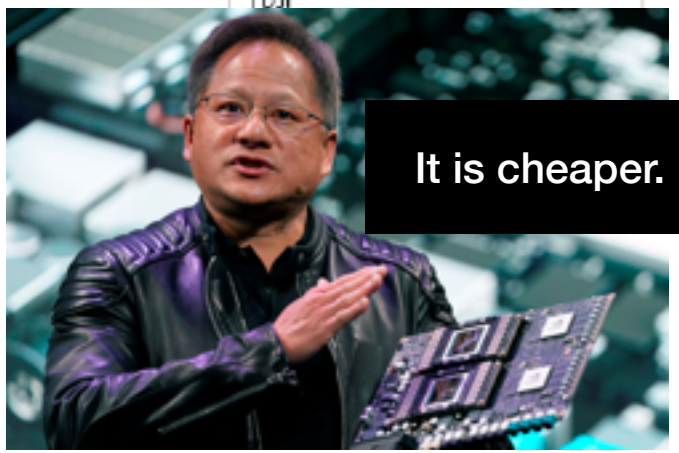
DJI Mavic 2 Pro – Drone Quadcopter UAV with Hasselblad Camera 3-Axis Gimbal HDR 4K Video Adjustable Aperture 20MP 1" CMOS Sensor, up...

★★★★½ ~ 81

$2,249.10

prime Get it by Wednesday, September 9

FREE Delivery by Amazon



Use Cases: What can Drones do? Why we study?

Drone is a new **aerial** platform.

- Accurate Delivery (e.g., send mask)
- Agriculture (e.g., pesticide)
- Event Detection (e.g. traffic jam)
- ….

- Task (Visual Gap)
- Dataset
- Baseline & Experiment

University-1652

- We consider one conventional task: cross-view Geo-localization.

Ground-view Images

Gap

Satellite-view Images (GPS tag)

Limited Roof

Whole Roof

We notice that the drone can be **a bridge.**

Ground-view ⟷ **Drone-view** ⟷ **Satellite-view (GPS tag)**

No dataset to verify it.

- Task
- Dataset (Missing in existing works)
- Baseline & Experiment

University-1652

- We collect the data from three platforms of 1652 buildings.
- More training images per class (instead of image pairs).
- More viewpoints -> More intra-class variants

Datasets	University-1652	CVUSA [34]	CVACT [16]
#training	701 × 71.64	35.5k × 2	35.5k × 2
Platform	Drone, Ground, Satellite	Ground, Satellite	Ground, Satellite
#imgs./location	54 + 16.64 + 1	1 + 1	1+1
Target	Building	User	User
GeoTag	✓	✓	✓
Evaluation	Recall@K & AP	Recall@K	Recall@K

- Me: I want to build one dataset.
- Supervisor: No! Too much cost.
- Me: We use free data from Internet.
- Supervisor: Do it!

Building names from Wikipedia

Building Names	
BibliothÃĺque Saint-Jean, University of Alberta	Clare Drake Arena
Foote Field	Myer Horowitz Theatre
National Institute for Nanotechnology	St Joseph's College, Edmonton
Stollery Children's Hospital	Universiade Pavilion, University of Alberta
University of Alberta Hospital	Alberta B. Farrington Softball Stadium
Decision Theater, University of Alberta	Gammage Memorial Auditorium
HarringtonäĂŞBirchett House	Industrial Arts Building
Irish Field	Louise Lincoln Kerr House and Studio
Matthews Hall, University of Alberta	Mona Plummer Aquatic Center
Old Main (Arizona State University)	Packard Stadium, University of Alberta
Security Building (Phoenix, Arizona)	Sun Devil Gym, University of Alberta
Sun Devil Stadium, University of Alberta	United States Post Office (Phoenix, Arizona)
Wells Fargo Arena (Tempe, Arizona)	Administration Building, University of Alberta
Wheeler Hall, University of Alberta	Marting Hall, University of Alberta
Malicky Center, University of Alberta	Burrell Memorial Observatory
Kleist Center for Art and Drama	Wilker Hall, University of Alberta
Kamm Hall, University of Alberta	Dietsch Hall, University of Alberta
Telfer Hall, University of Alberta	Ward Hall, University of Alberta
Thomas Center for Innovation and Growth (CIG)	Kulas Musical Arts Building, Baldwin Wallace University
Boesel Musical Arts Center, Baldwin Wallace University	Merner-Pfeiffer Hall, Baldwin Wallace University
Ritter Library, Baldwin Wallace University	Lindsay-Crossman Chapel, Baldwin Wallace University
Presidents House, Baldwin Wallace University	Student Activities Center (SAC), Baldwin Wallace University
Strosacker Hall (Union), Baldwin Wallace University	Bonds Hall, Baldwin Wallace University
Durst Welcome Center, Baldwin Wallace University	Lou Higgins Center, Baldwin Wallace University
Tressel Field @ Finnie Stadium, Baldwin Wallace University	Rutherford Library
Rudolph Ursprung Gymnasium, Baldwin Wallace University	Packard Athletic Center (formerly Bagley Hall), Baldwin Wallace University
Baldwin-Wallace College North Campus Historic District	Baldwin-Wallace College South Campus Historic District
Binghamton University Events Center, Binghamton University	Commonwealth Avenue, Boston University
Boston University Photonics Center, Boston University	Boston University School of Law, Boston University
Boston University Track and Tennis Center, Boston University	Boston University West Campus

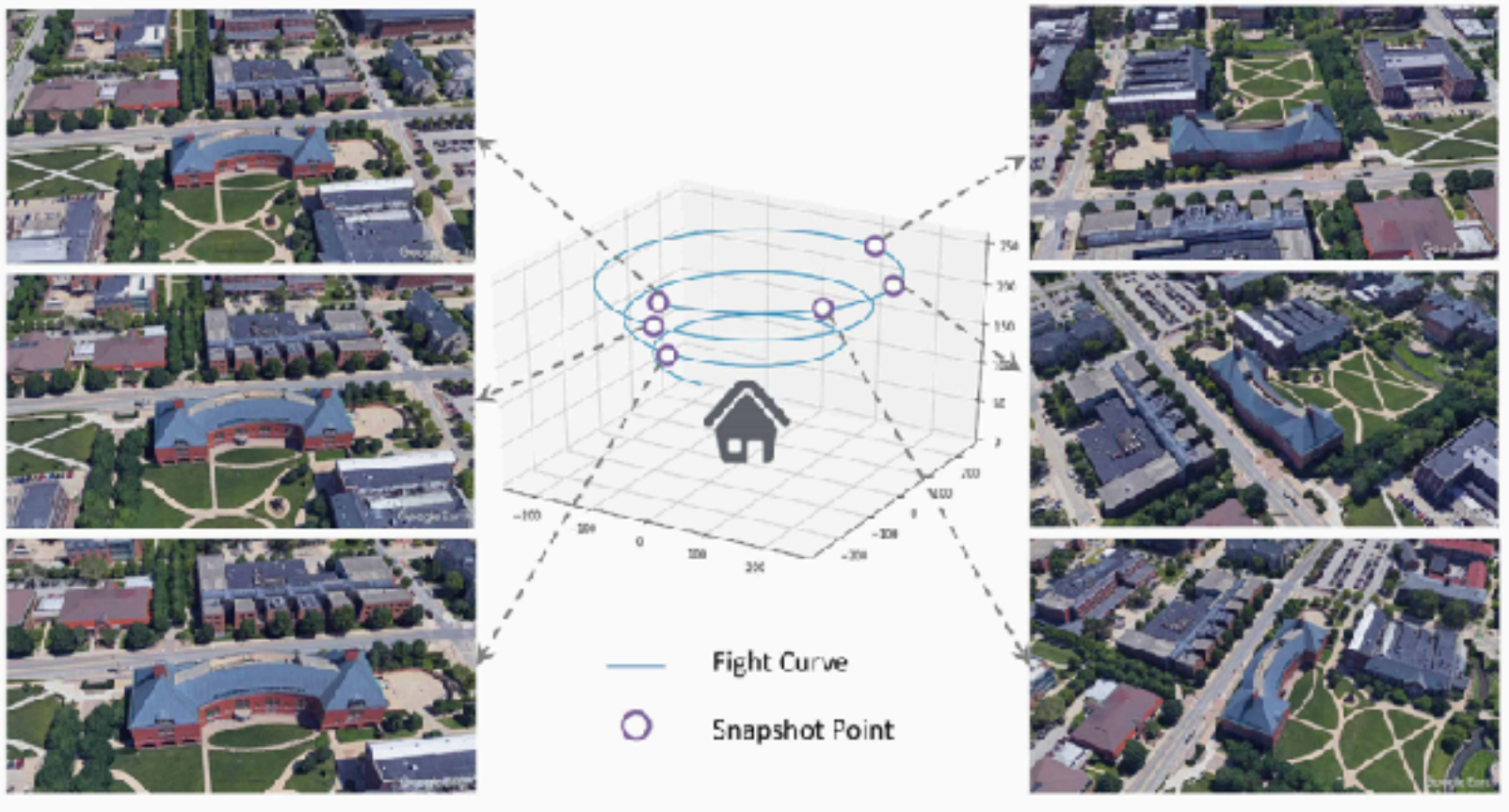
Get latitude/longitude from GoogleMap

Attributes	Value
name	Grainger Engineering Library
longitude	-88.22691719995214
latitude	40.11249969950067
altitude	18.565223428500 79

Attributes	Value
name	Foellinger Auditorium
longitude	-88.22728640012006
latitude	40.10594310015922
altitude	23.78598631063875

1. Drone-view Data

- Due to the privacy concerns and the cost, we deploy the simulated data via Google Earth. We write scripts to drive the engine as drone camera.




Google Earth


Google Earth


Google Earth


Google Earth

2. Ground-view Data from GoogleMap

3. Satellite-view Data from GoogleMap

4. Noisy Ground-view Data from GoogleImage

Building Names	
BibliothÃ¨que Saint-Jean, University of Alberta	Clare Drake Arena
Foote Field	Myer Horowitz Theatre
National Institute for Nanotechnology	St Joseph's College, Edmonton
Stollery Children's Hospital	Universiade Pavilion, University of Alberta
University of Alberta Hospital	Alberta B. Farrington Softball Stadium
Decision Theater, University of Alberta	Gammage Memorial Auditorium
HarringtonâĂŞBirchett House	Industrial Arts Building
Irish Field	Louise Lincoln Kerr House and Studio
Matthews Hall, University of Alberta	Mona Plummer Aquatic Center
Old Main (Arizona State University)	Packard Stadium, University of Alberta
Security Building (Phoenix, Arizona)	Sun Devil Gym, University of Alberta
Sun Devil Stadium, University of Alberta	United States Post Office (Phoenix, Arizona)
Wells Fargo Arena (Tempe, Arizona)	Administration Building, University of Alberta
Wheeler Hall, University of Alberta	Marting Hall, University of Alberta
Malicky Center, University of Alberta	Burrell Memorial Observatory
Kleist Center for Art and Drama	Wilker Hall, University of Alberta
Kamm Hall, University of Alberta	Dietsch Hall, University of Alberta
Telfer Hall, University of Alberta	Ward Hall, University of Alberta
Thomas Center for Innovation and Growth (CIG)	Kulas Musical Arts Building, Baldwin Wallace University
Boesel Musical Arts Center, Baldwin Wallace University	Merner-Pfeiffer Hall, Baldwin Wallace University
Ritter Library, Baldwin Wallace University	Lindsay-Crossman Chapel, Baldwin Wallace University
Presidents House, Baldwin Wallace University	Student Activities Center (SAC), Baldwin Wallace University
Strosacker Hall (Union), Baldwin Wallace University	Bonds Hall, Baldwin Wallace University
Durst Welcome Center, Baldwin Wallace University	Lou Higgins Center, Baldwin Wallace University
Tressel Field @ Finnie Stadium, Baldwin Wallace University	Rutherford Library
Rudolph Ursprung Gymnasium, Baldwin Wallace University	Packard Athletic Center (formerly Bagley Hall), Baldwin Wallace University
Baldwin-Wallace College North Campus Historic District	Baldwin-Wallace College South Campus Historic District
Binghamton University Events Center, Binghamton University	Commonwealth Avenue, Boston University
Boston University Photonics Center, Boston University	Boston University School of Law, Boston University
Boston University Track and Tennis Center, Boston University	Boston University West Campus

- We search the building name and download images from GoogleImage
- We then remove the indoor images and duplicate images.

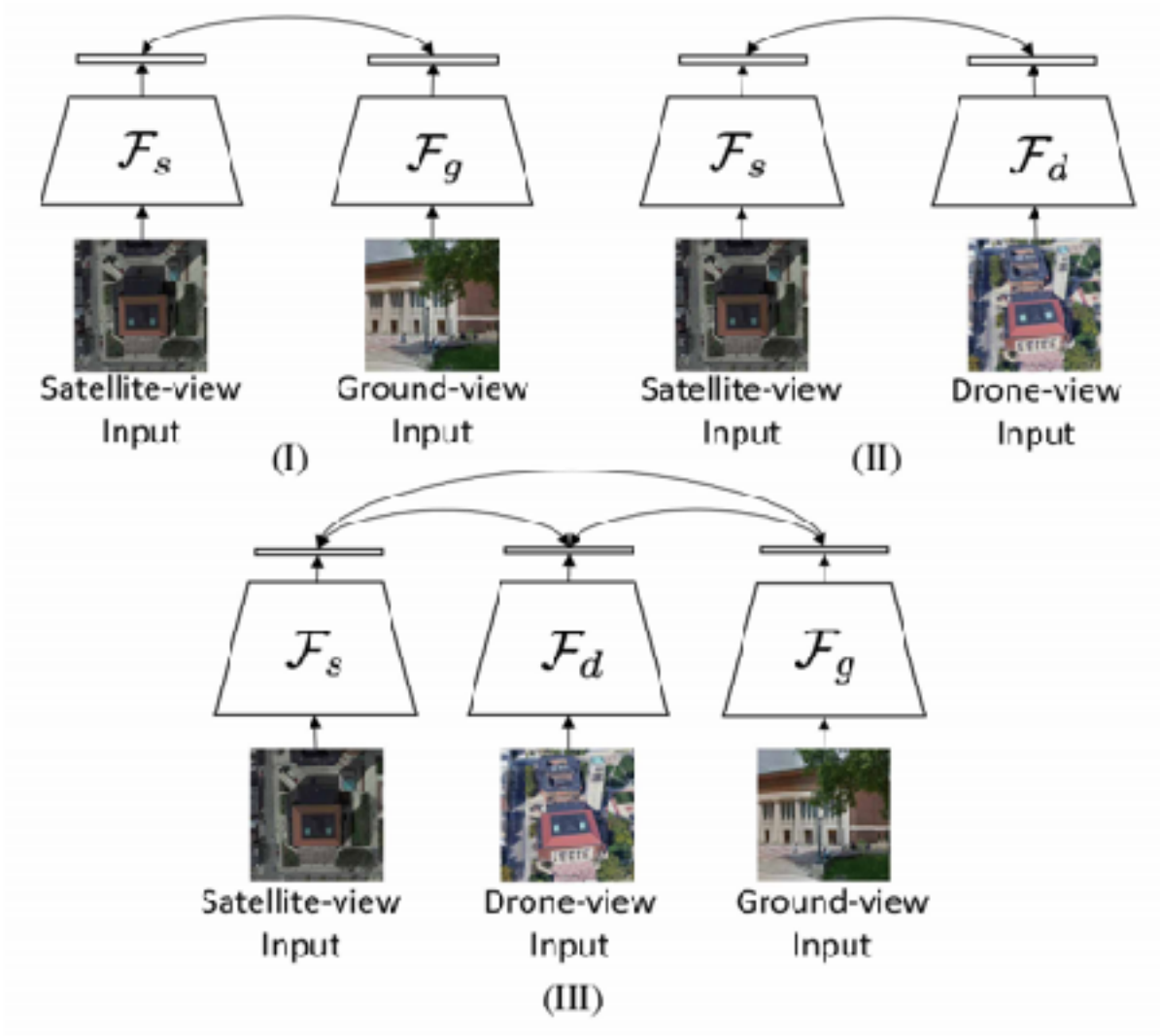
Krause J, Sapp B, Howard A, et al. The unreasonable effectiveness of noisy data for fine-grained recognition[C]//European Conference on Computer Vision. Springer, Cham, 2016: 301-320.

- Task
- Dataset (Now we have data.)
- Baseline & Experiment

Baseline

Flexible and Strong Baseline

- Objective: Instance Loss (Share Classifier)
- Structure: Generally, the backbone
- network do not share low-level patterns



New data ->

add one branch!

Zheng Z, Zheng L, Garrett M, et al. Dual-Path Convolutional Image-Text Embeddings with Instance Loss[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2020, 16(2): 1-23.

Baseline

CVUSA

Methods	R@1	R@5	R@10	R@Top1%
Workman [31] ICCV 2015	-	-	-	34.40
Zhai [34] CVPR 2017	-	-	-	43.20
Vo [29] ECCV 2016	-	-	-	63.70
CVM-Net [16] CVPR 2018	18.80	44.42	57.47	91.54
Orientation [16]† CVPR 2019	27.15	54.66	67.54	**93.91**
Ours	**43.91**	**66.38**	**74.58**	91.78

Table 9: Comparison of results on the two-view dataset CVUSA [34]. †: The method utilizes extra orientation information as input.

Oxford and Paris

Method	Oxford	Paris	ROxf (M)	RPar (M)	ROxf (H)	RPar (H)
ImageNet	3.30	6.77	4.17	8.20	2.09	4.24
$\mathcal{F}_s$	9.24	13.74	5.83	13.79	2.08	6.40
$\mathcal{F}_g$	25.80	28.77	15.52	24.24	3.69	10.29

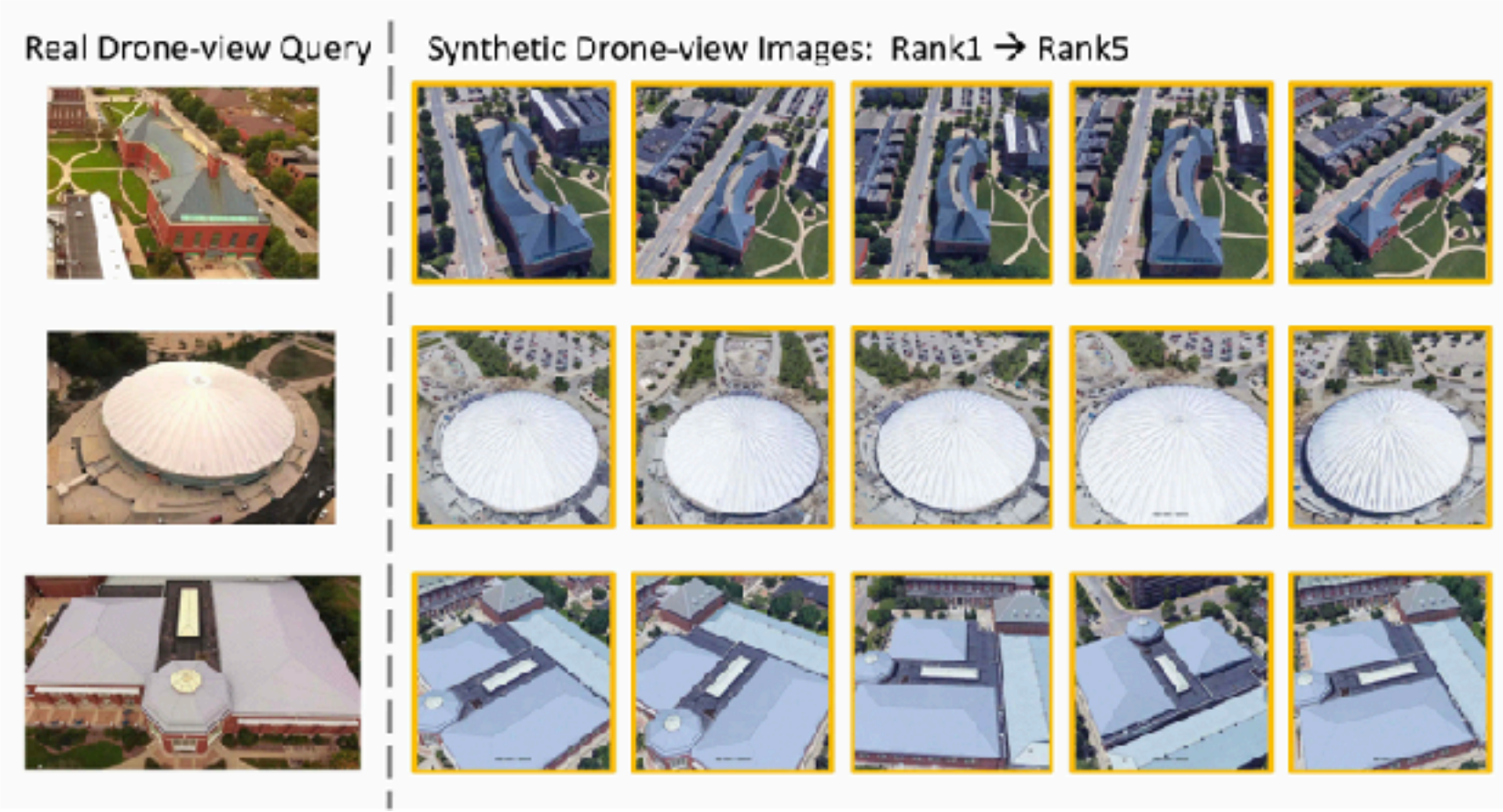
Table 10: Transfer learning from University-1652 to small-scale datasets. We show the AP (%) accuracy on Oxford [19], Paris [20], ROxford and RParis [21]. For ROxford and RParis, we report results in both medium (M) and hard (H) settings.

Ground-view query vs. drone-view query.

Query → Gallery	R@1	R@5	R@10	AP
Ground → Satellite	1.20	4.61	7.56	2.52
Drone → Satellite	58.49	78.67	85.23	63.13
*m*Ground → Satellite	1.71	6.56	10.98	3.33
*m*Drone → Satellite	69.33	86.73	91.16	73.14

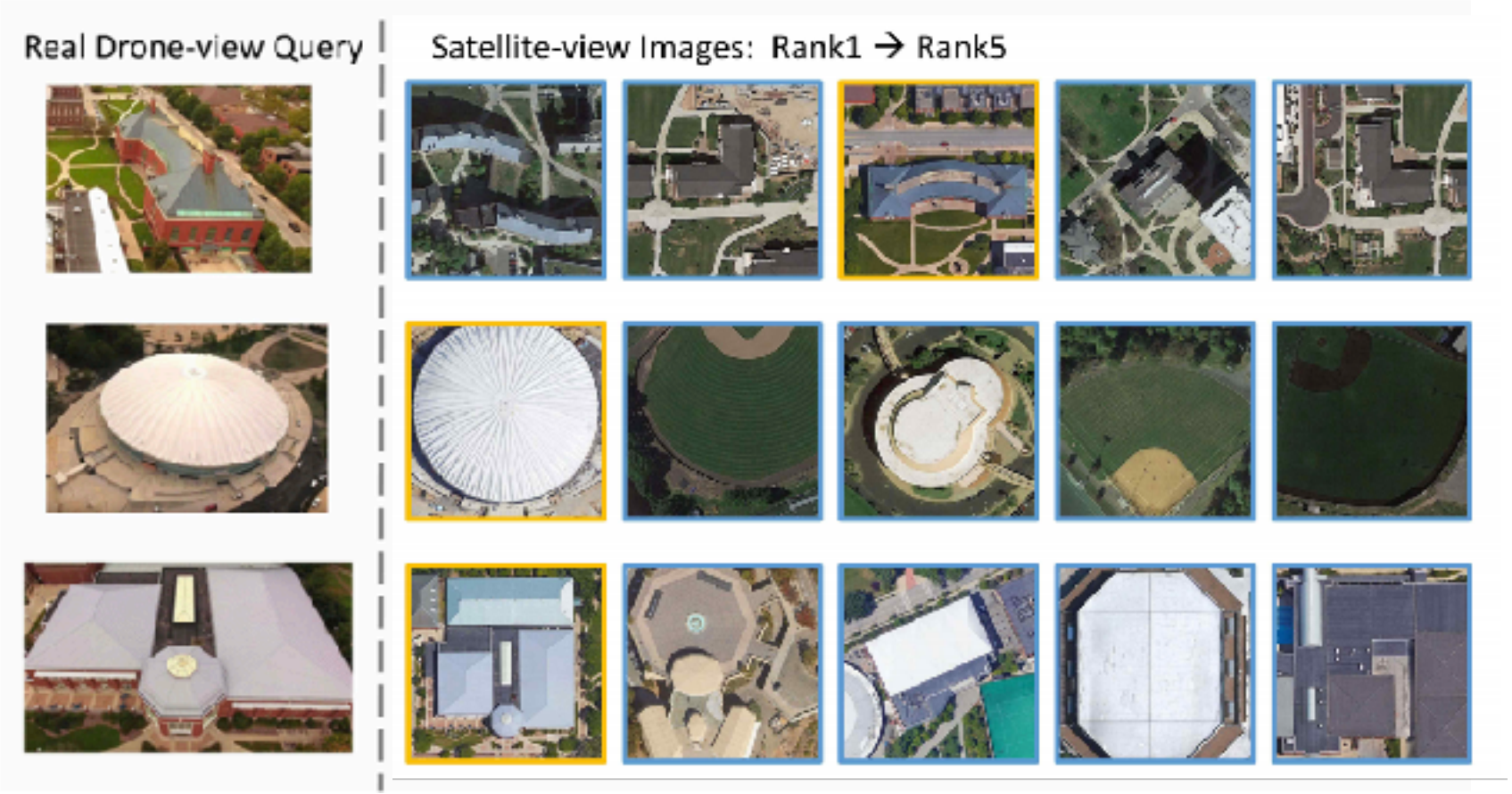
Table 4: Ground-view query vs. drone-view query. m denotes multiple-query setting. The result suggests that drone-view images are superior to ground-view images when retrieving satellite-view images.

Apply the model trained on University-1652 to real drone videos.



The model haven’t seen any data of UIUC.

Apply the model trained on University-1652 to real drone videos.



The model haven't seen any data of UIUC.

Scalability

Ablation Studies

Different Loss Functions

Loss	Drone → Satellite		Satellite → Drone	
	R@1	AP	R@1	AP
Contrastive Loss	52.39	57.44	63.91	52.24
Triplet Loss (margin=0.3)	55.18	59.97	63.62	53.85
Triplet Loss (margin=0.5)	53.58	58.60	64.48	53.15
Weighted Soft Margin Triplet Loss	53.21	58.03	65.62	54.47
Instance Loss	58.23	62.91	74.47	59.45

Table 5: Ablation study of different loss terms. To fairly compare the five loss terms, we trained the five models on satellite-view and drone-view data, and hold out the ground-view data. For contrastive loss, triplet loss and weighted soft margin triplet loss, we also apply the hard-negative sampling policy.

Whether Share Weights

Method	Drone → Satellite		Satellite → Drone	
	R@1	AP	R@1	AP
Not sharing weights	39.84	45.91	50.36	40.71
Sharing weights	58.49	63.31	71.18	58.74

Table 6: Ablation study. With/without sharing CNN weights on University-1652. The result suggests that sharing weights could help to regularize the CNN model.

Different Input Sizes

Image Size	Drone → Satellite		Satellite → Drone	
	R@1	AP	R@1	AP
256	58.49	63.31	71.18	58.74
384	62.99	67.69	75.75	62.09
512	59.69	64.80	73.18	59.40

Table 7: Ablation study of different input sizes on the University-1652 dataset.

Future Works - Boost Performance

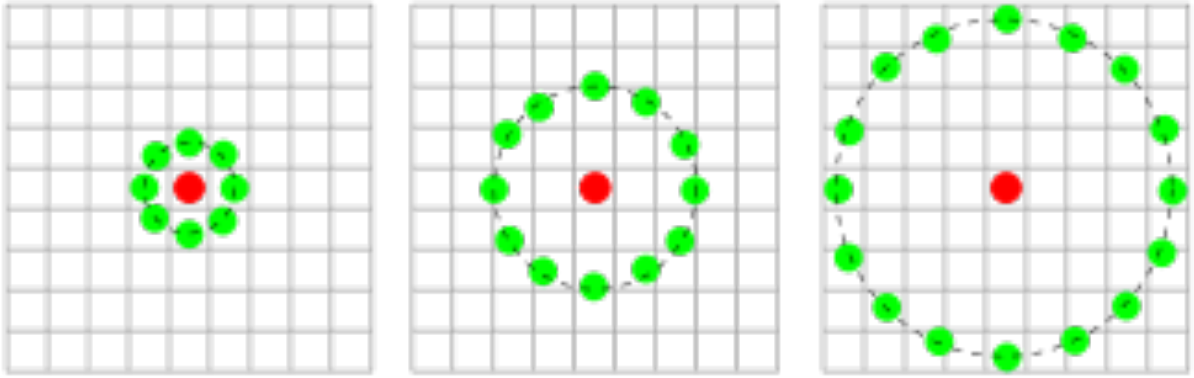
We run a leaderboard.
You are welcomed to push the state-of-the-art performance.

Awesome Geo-localization

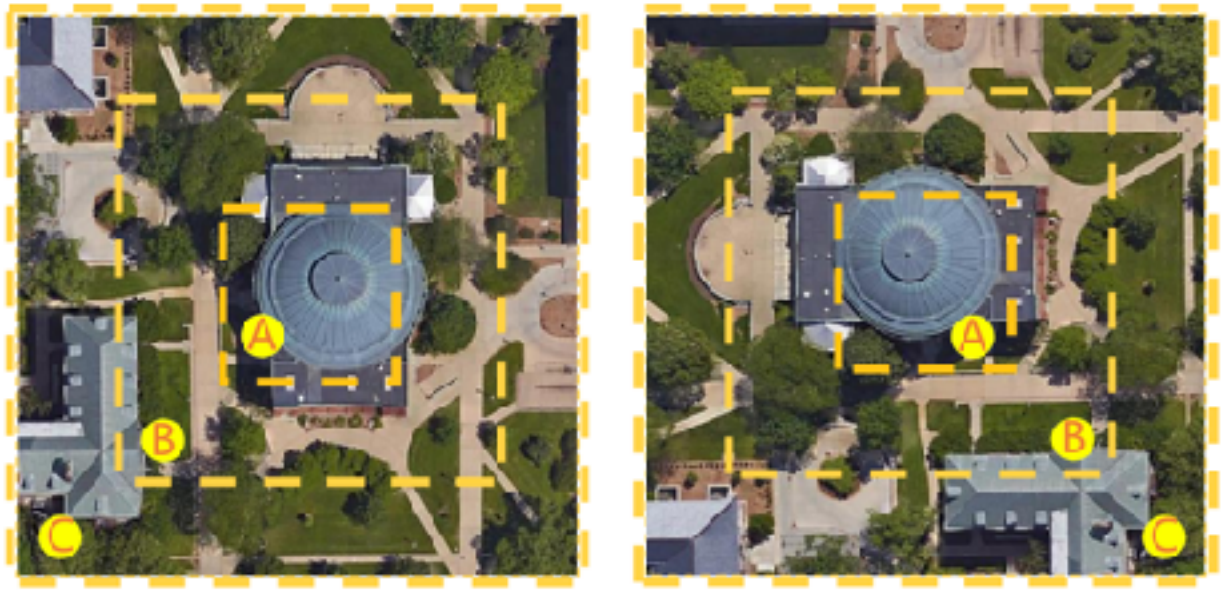
University-1652

Methods	R@1	AP	R@1	AP	Reference
Contrastive Loss	52.39	57.44	63.91	52.24	
Triplet Loss (margin=0.3)	55.18	59.97	63.62	53.85	
Triplet Loss (margin=0.5)	53.58	58.60	64.48	53.15	
Weighted Soft Margin Triplet Loss	53.21	58.03	65.62	54.47	
Instance Loss	58.23	62.91	74.47	59.45	

Future Works - LBP



Local binary patterns (LBP) descriptor

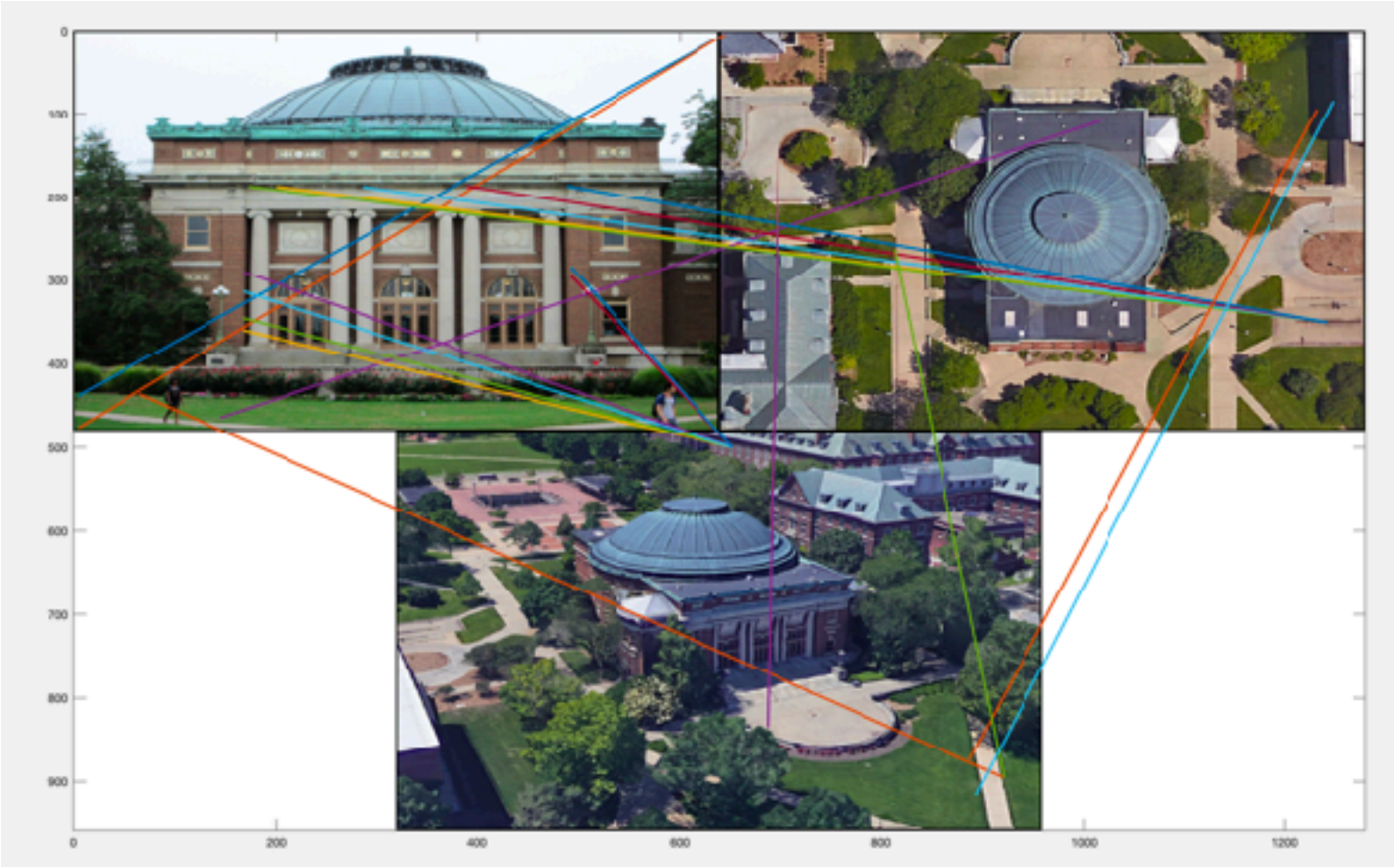


LPN

Tingyu W, Zhedong Z, Chenggang Y, and Yi, Y. Each Part Matters: Local Patterns Facilitate Cross-view Geo-localization. arXiv 2020.

Future Works - Keypoint Matching

SIFT does not work very well. Deeply-learned Methods are needed.



Thanks a lot!

University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization

Zhedong Zheng, Yunchao Wei, Yi Yang
University of Technology Sydney

Dataset & Code

Have been downloaded
By 300+ times.

Data License

- We carefully check the data license from Google. There are two main points.
- First, the data of Google Map and Google Earth could be used based on fair usage. We follow the guideline on this official website 3 .
- Second, several existing datasets have utilized the Google data. In practice, we adopt a similar policy of existing datasets 4, 5 to release the dataset based on the academic request.

3. https://www.google.com/permissions/geoguidelines/
4. http://www.ok.ctrl.titech.ac.jp/~torii/project/247/
5. http://mvrl.cs.uky.edu/datasets/cvusa/