Thoughts about Person Re-identification and Beyond

Liang Zheng
Australian National University
8-Jan-2019

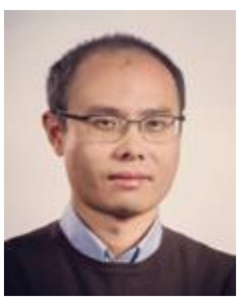
Collaborators

Xiaoxiao Sun
ANU

Yue Yao
ANU

Yunzhong Hou
ANU

Tom Gedeon
ANU



Xiaodong Yang
NVIDIA

Milind Naphade
NVIDIA

Zhongdao Wang
THU

Shengjin Wang
THU

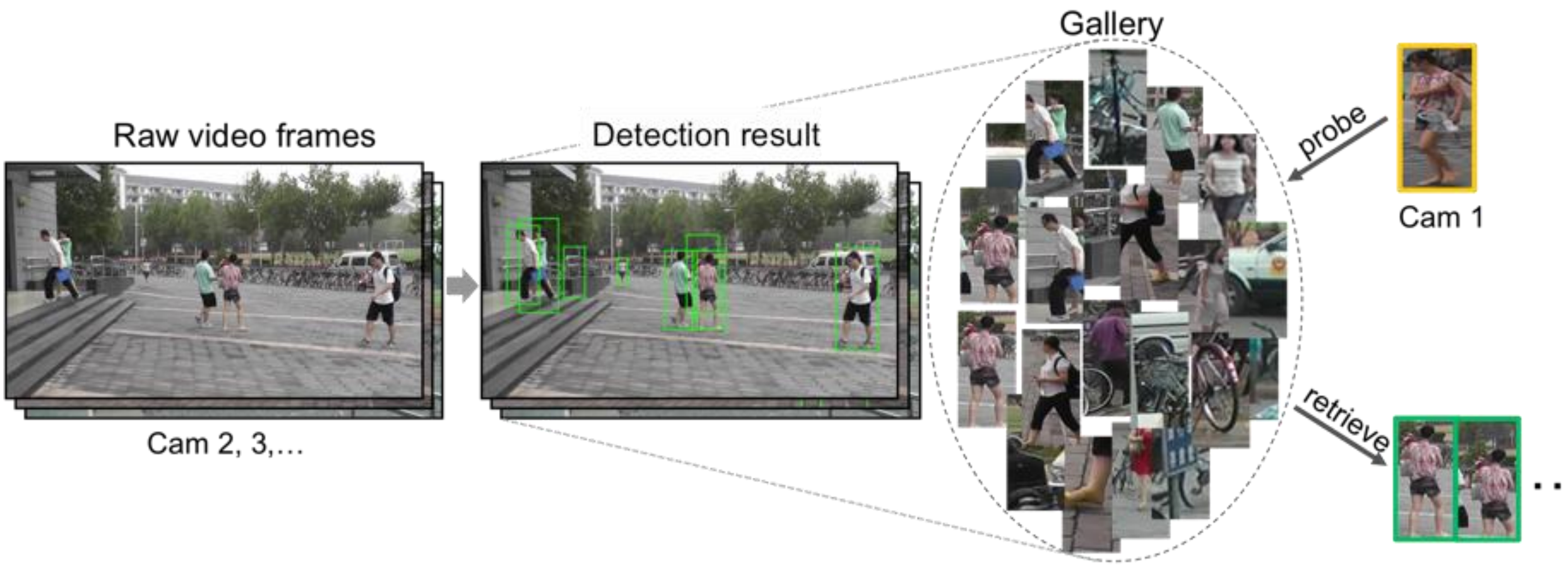
Outline

- Introduction
- Re-id vs multi-object tracking
- Data synthesis in object re-id
- Alice benchmark suite

Introduction

Person Detection

Person retrieval / re-identification



Person retrieval / re-identification

Outline

- Introduction
- Re-id vs multi-object tracking
- Data synthesis in object re-id
- Alice benchmark suite

Online Multi-Object Tracking (MOT)

1. Key Components in MOT:
   - Object Detection
   - Appearance feature model
   - Motion model
   - Association algorithm

Bottlenecks of the system for being real-time

2. Challenges in practical applications
   - Occlusions
   - **A real-time system !**

3. Our solution
   - Incorporating the detector and the appearance feature model into a **shared, one-stage network**.

Zhongdao Wang, Liang Zheng, Yixuan Liu, Shengjin Wang, Towards real-time multi-object tracking. Arxiv 2019.

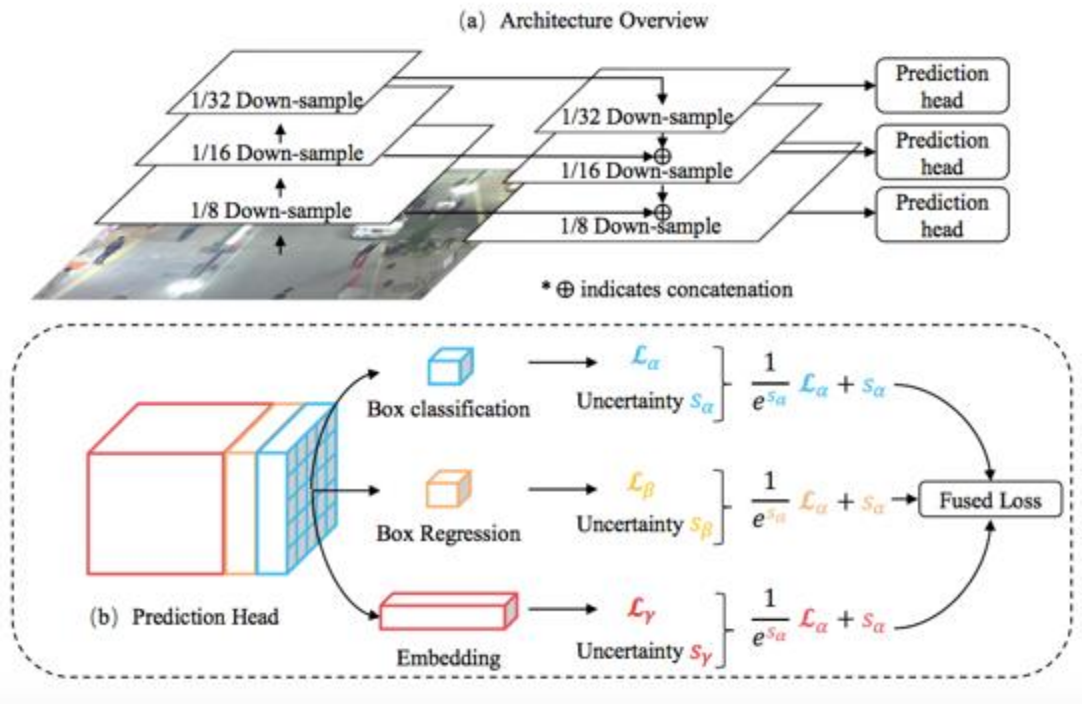
JDE: Joint Detection and appearance Embedding

1. Utilizing available training data (For multi-pedestrian tracking):
   a) Pedestrian detection datasets with box annotations. (Caltech, CityPersons, ETH)
   b) MOT/Person search datasets with box+identity annotations. (MOT16, PRW, CUHK-SYSU)

2. Architecture:
   FPN + Multi-task prediction head

3. Appearance embedding head:
   Classification with cross entropy loss

4. Loss fusion:
   Automatic loss balancing via modeling task-specific uncertainty

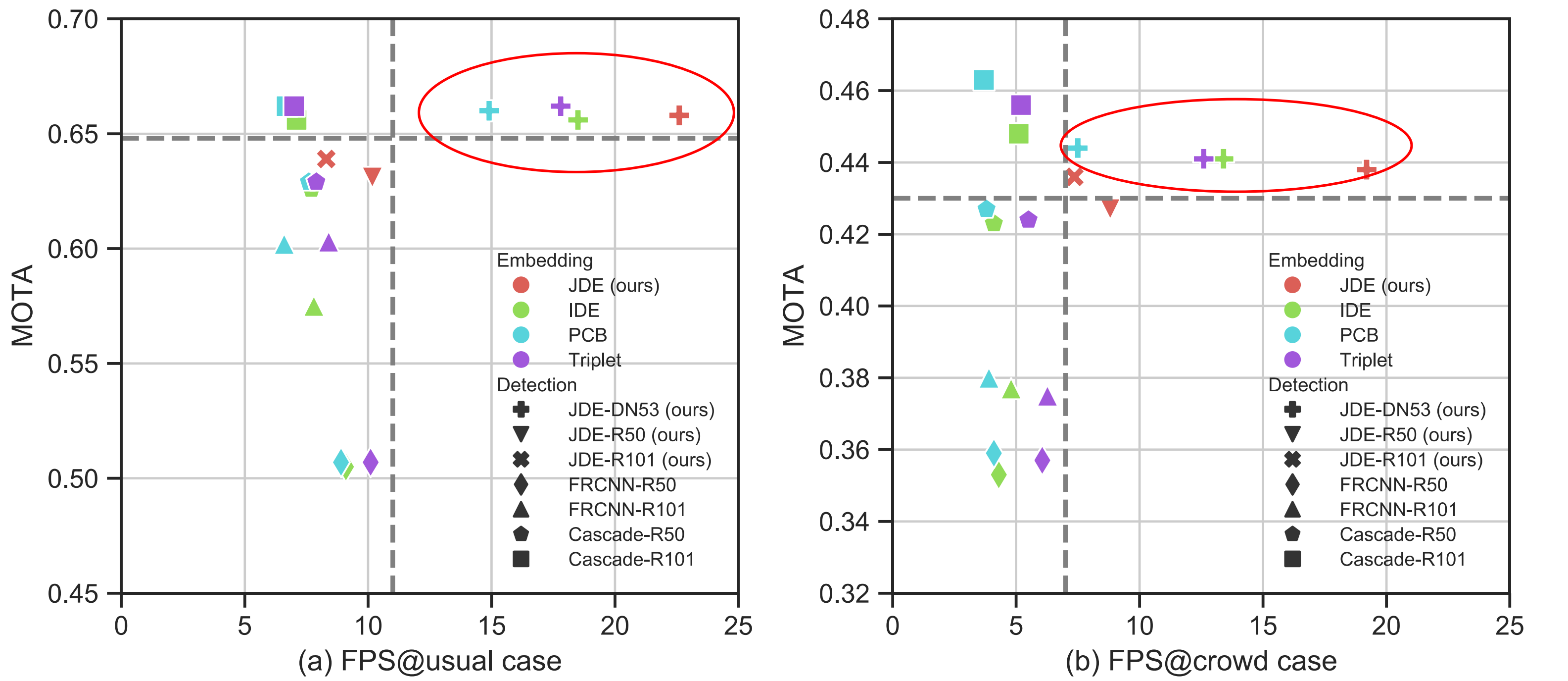


Zhongdao Wang, Liang Zheng, Yixuan Liu, Shengjin Wang, Towards real-time multi-object tracking. Arxiv 2019.

Result

- Good speed-accuracy trade-off

Joint training is mainly for speed consideration; accuracy might not be optimal.



Zhongdao Wang, Liang Zheng, Yixuan Liu, Shengjin Wang, Towards real-time multi-object tracking. Arxiv 2019.

Result

- Good speed-accuracy trade-off
- Near real-time
- Competitive accuracy on MOT-16 (MOTA)

Method	Det	Emb	#box	#id	MOTA	IDF1	MT	ML	IDs	FPSD	FPSA	FPS
DeepSORT_2	FRCNN	WRN	429K	1.2k	61.4	62.2	32.8	18.2	781	<15*	17.4	<8.1
RAR16wVGG	FRCNN	Inception	429K	-	63.0	63.8	39.9	22.1	**482**	<15*	1.6	<1.5
TAP	FRCNN	MRCNN	429K	-	64.8	**73.5**	**40.6**	22.0	794	<15*	18.2	<8.2
CNNMTT	FRCNN	5-Layer	429K	0.2K	65.2	62.2	32.4	21.3	946	<15*	11.2	<6.4
POI	FRCNN	QAN	429K	16K	**66.1**	65.1	34.0	21.3	805	<15*	9.9	<6
JDE-864(ours)	JDE	-	270K	8.7K	62.1	56.9	34.4	**16.7**	1,608	**34.3**	81.0	**24.1**
JDE-1088(ours)	JDE	-	270K	8.7K	64.4	55.8	35.4	20.0	1,544	24.5	**81.5**	18.8

Zhongdao Wang, Liang Zheng, Yixuan Liu, Shengjin Wang, Towards real-time multi-object tracking. Arxiv 2019.

Multi-Target Multi-Camera Tracking

- Multi-Target Multi-Camera Tracking focuses on determine who is where at all times.

- Similarity estimation is a key component in MTMCT.
  - Re-ID features are often adopted for similarity estimation.

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Difference between tracking and re-ID

- Local vs. global difference between tracking and re-ID.
  - Re-ID systems (top row) usually search globally.

Re-ID features are highly robust to variances.

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

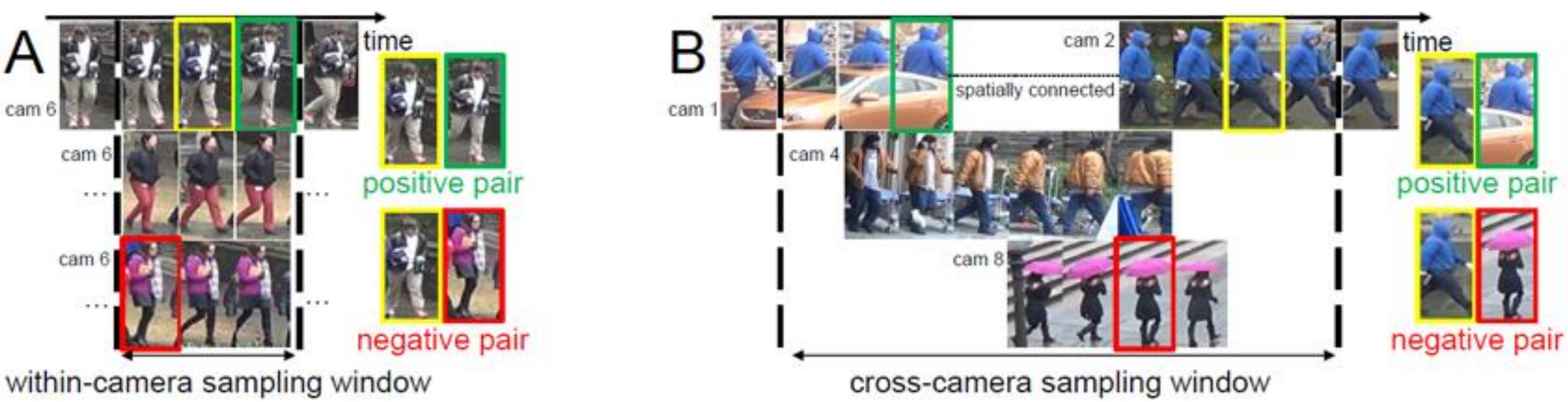
Difference between tracking and re-ID

- Local vs. global difference between tracking and re-ID.
  - Re-ID systems (top row) usually search globally.
  - Tracking systems usually search within local neighbors (neighboring frames/cameras).

Tracking features do not have to be that robust.
Directly using re-ID features leads to false positive matches.

Local metric for local matching

- Our idea: Local metric for local matching.
- A local metric for single camera tracking.
- A local metric for multi camera tracking.
- Select data pairs with temporal windows over single/multi camera.



Training data are locally sampled!

Result

- Tracking accuracy increases on multiple datasets.

Variant	CityFlow test set MCT results		
	IDF1	IDP	IDR

CityFlow dataset (vehicle tracking)

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Result

- Tracking accuracy increases on multiple datasets.

Variant	CityFlow test set MCT results		
	IDF1	IDP	IDR
Baseline	56.6	53.3	60.7

CityFlow dataset (vehicle tracking)

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Result

- Tracking accuracy increases on multiple datasets.

Variant	CityFlow test set MCT results		
	IDF1	IDP	IDR
Baseline	56.6	53.3	60.7
Global metric	57.1	54.4	60.7

CityFlow dataset (vehicle tracking)

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Result

- Tracking accuracy increases on multiple datasets.

Variant	CityFlow test set MCT results		
	IDF1	IDP	IDR
Baseline	56.6	53.3	60.7
Global metric	57.1	54.4	60.7
LAAM (intra/inter)	**63.0**	**60.7**	**66.0**

CityFlow dataset (vehicle tracking)

Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Result

- Tracking accuracy increases on multiple datasets.

Variant	Validation set IDF1 results					
	IDE		triplet		PCB	
	SCT	MCT	SCT	MCT	SCT	MCT
Baseline	86.4	81.4	86.2	80.9	85.8	80.6
Global metric	85.9	81.6	84.1	79.7	87.4	81.6
LAAM (intra/inter)	**87.9**	**83.8**	**87.9**	**84.5**	**87.7**	**82.9**

DukeMTMC dataset (pedestrian tracking)

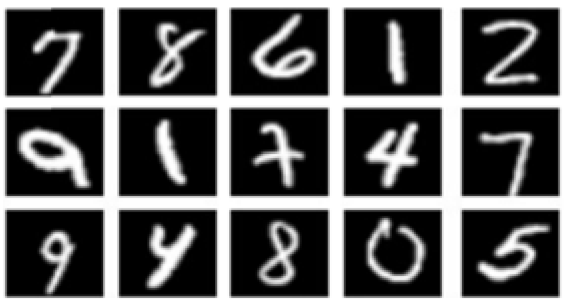
Yunzhong Hou, Liang Zheng, Zhongdao Wang, Shengjin Wang. Locality aware appearance metric for multi-target multi-camera tracking. Arxiv 2019.

Outline

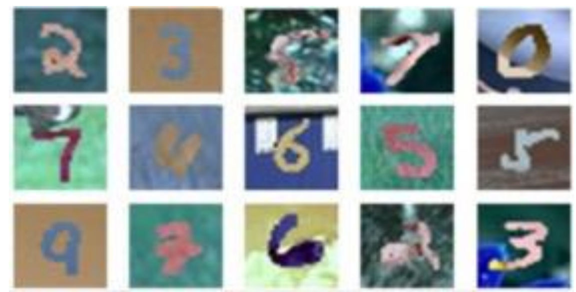
- Introduction
- Re-id vs multi-object tracking
- Data synthesis in object re-id
- Alice benchmark suite

Problem

- Domain shift
  - image classification



MNIST



MNIST-M

  - Crowd counting

GCC

ShanghaiTech

Existing domain adaptation methods

- Style level

Source image (GTA5) | Adapted source image (**Ours**) | Target image (CityScapes)

Hoffman et al. “CyCADA: Cycle-Consistent Adversarial Domain Adaptation.” ICML, 2017.

Our idea

	Training set	Testing set	model
Neural architecture search	fixed	fixed	To be searched
Content-level domain adaptation	To be searched	fixed	fixed

Content-level domain adaptation

idea

source

target

How to remedy domain gap?

Style/feature alignment

Content alignment

Content-level domain adaptation

idea

source

target

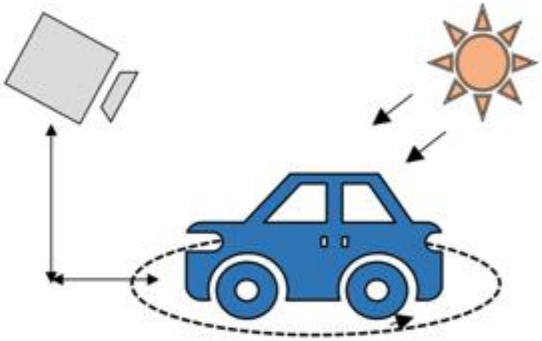
How to remedy domain gap?

Style/feature alignment

Content alignment

Content-level domain adaptation

- We collected the VehicleX Dataset
  - controllability and editability
  - 1,209 vehicles
  - ~350 types of vehicles
  - Platform: Unity
  - Editable attributes: lighting direction, lighting intensity, vehicle orientation, camera height, camera distance

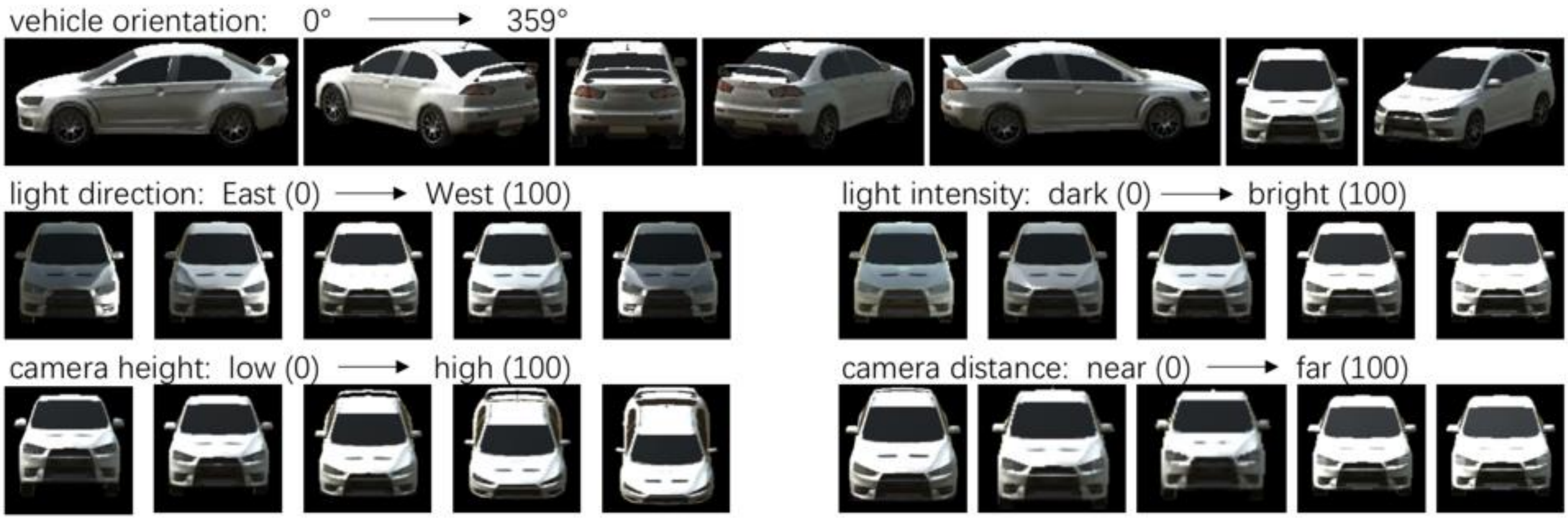


A Platform

B Vehicle identities

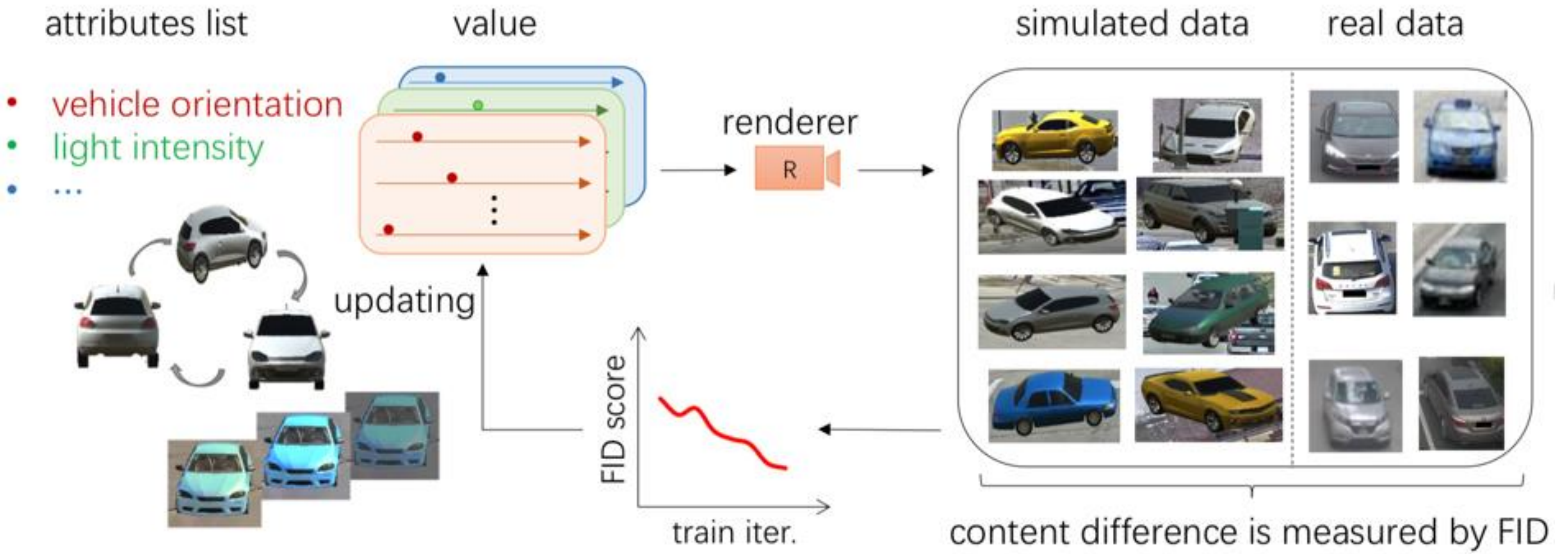
Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Editable Attributes



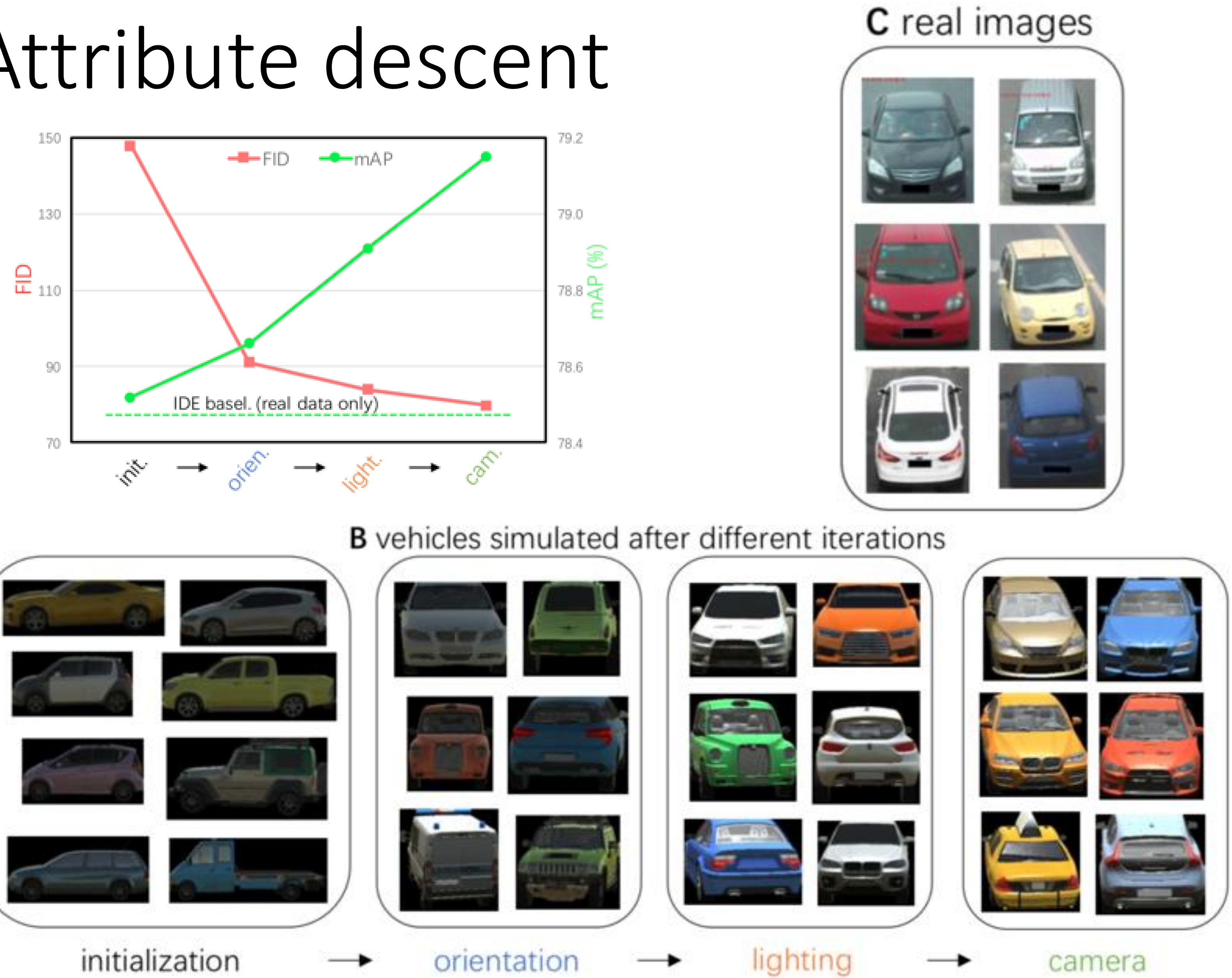
Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Overall method



Attribute modeling: Gaussian mixture models
Distribution difference measure: Fre ́chet Inception Distance (FID)

Attribute descent



We optimize the value of each attributes successively

For a given attribute, we search (brute-force) for its optimum value such that FID is minimized

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

Method	Data	Rank-1	Rank-20	mAP

We use rank-1, rank-20 and mAP as evaluation metrics

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

Method	Data	Rank-1	Rank-20	mAP
BA [18]	R	49.62	80.04	25.61

Existing methods

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

Method	Data	Rank-1	Rank-20	mAP
BA [18]	R	49.62	80.04	25.61
BS [18]	R	49.05	78.80	25.57
PAMTRI [32]	R+S	**59.7**	80.13	33.81

Existing methods

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

Method	Data	Rank-1	Rank-20	mAP
BA [18]	R	49.62	80.04	25.61
BS [18]	R	49.05	78.80	25.57
PAMTRI [32]	R+S	**59.7**	80.13	33.81
IDE(CE+Tri.)	R	56.75	72.24	30.21

Our baseline

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

Method	Data	Rank-1	Rank-20	mAP
BA [18]	R	49.62	80.04	25.61
BS [18]	R	49.05	78.80	25.57
PAMTRI [32]	R+S	**59.7**	80.13	33.81
IDE(CE+Tri.)	R	56.75	72.24	30.21
Random Attr.	R+S	54.09	78.04	32.03

We simulate data with random attributes.

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – training with real data + simulated data

- Method comparison on the CityFlow dataset

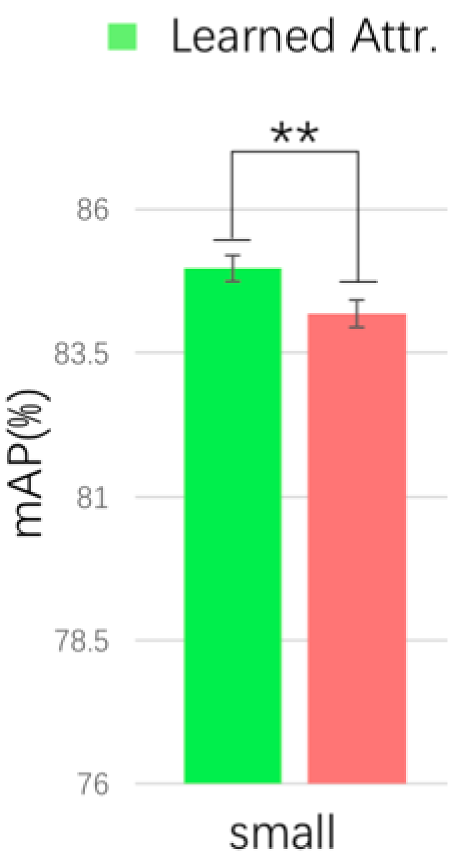
Method	Data	Rank-1	Rank-20	mAP
BA [18]	R	49.62	80.04	25.61
BS [18]	R	49.05	78.80	25.57
PAMTRI [32]	R+S	**59.7**	80.13	33.81
IDE(CE+Tri.)	R	56.75	72.24	30.21
Random Attr.	R+S	54.09	78.04	32.03
Learned Attr.	R+S	59.32	**80.42**	**34.63**

We simulate data with learned attributes.

Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, Tom Gedeon, Simulating Content Consistent Vehicle Datasets with Attribute Descent. Arxiv 2019.

Experiment – statistical significance

- Learned attribute vs. random attribute



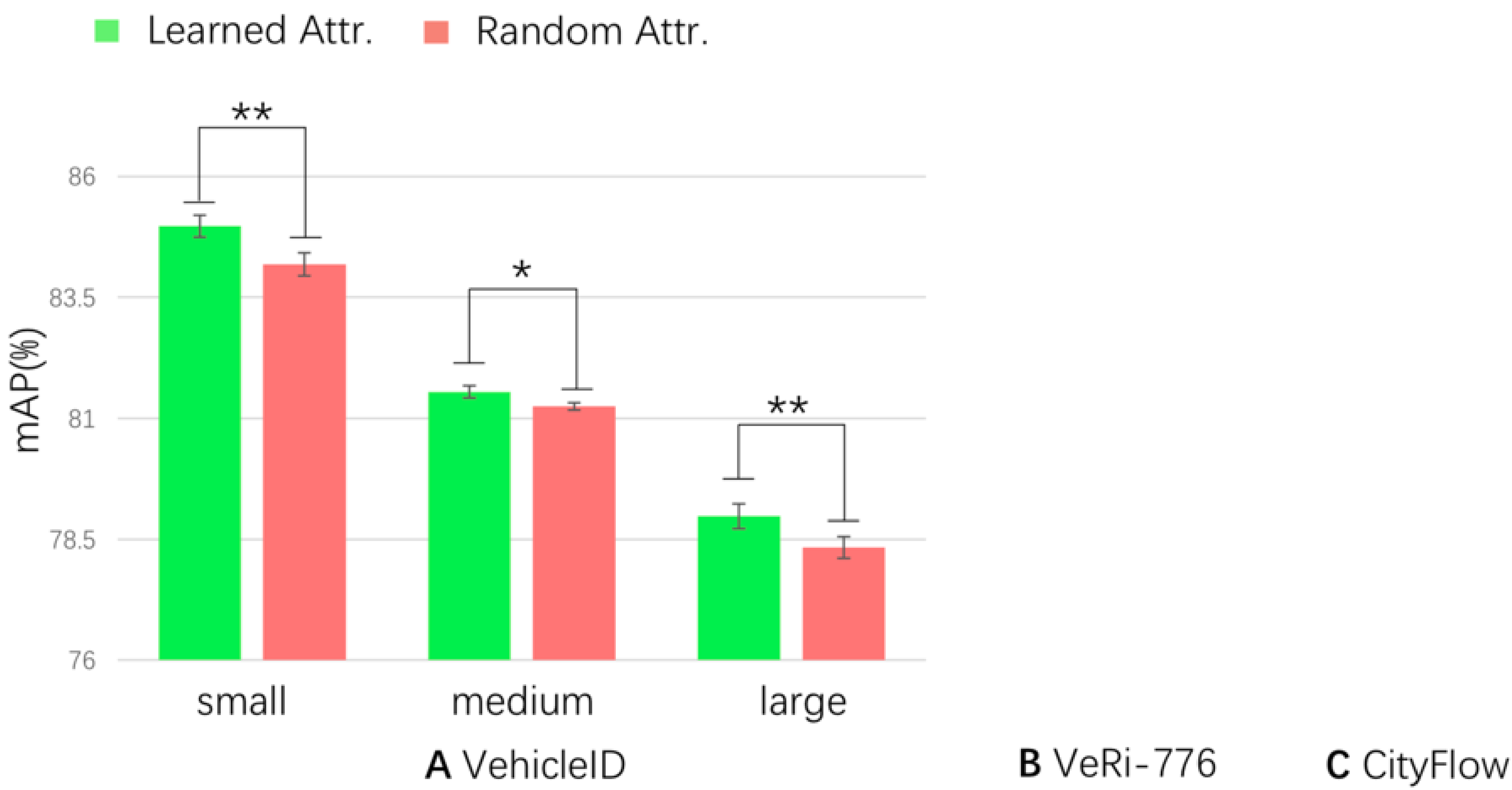
A VehicleID

B VeRi-776

C CityFlow

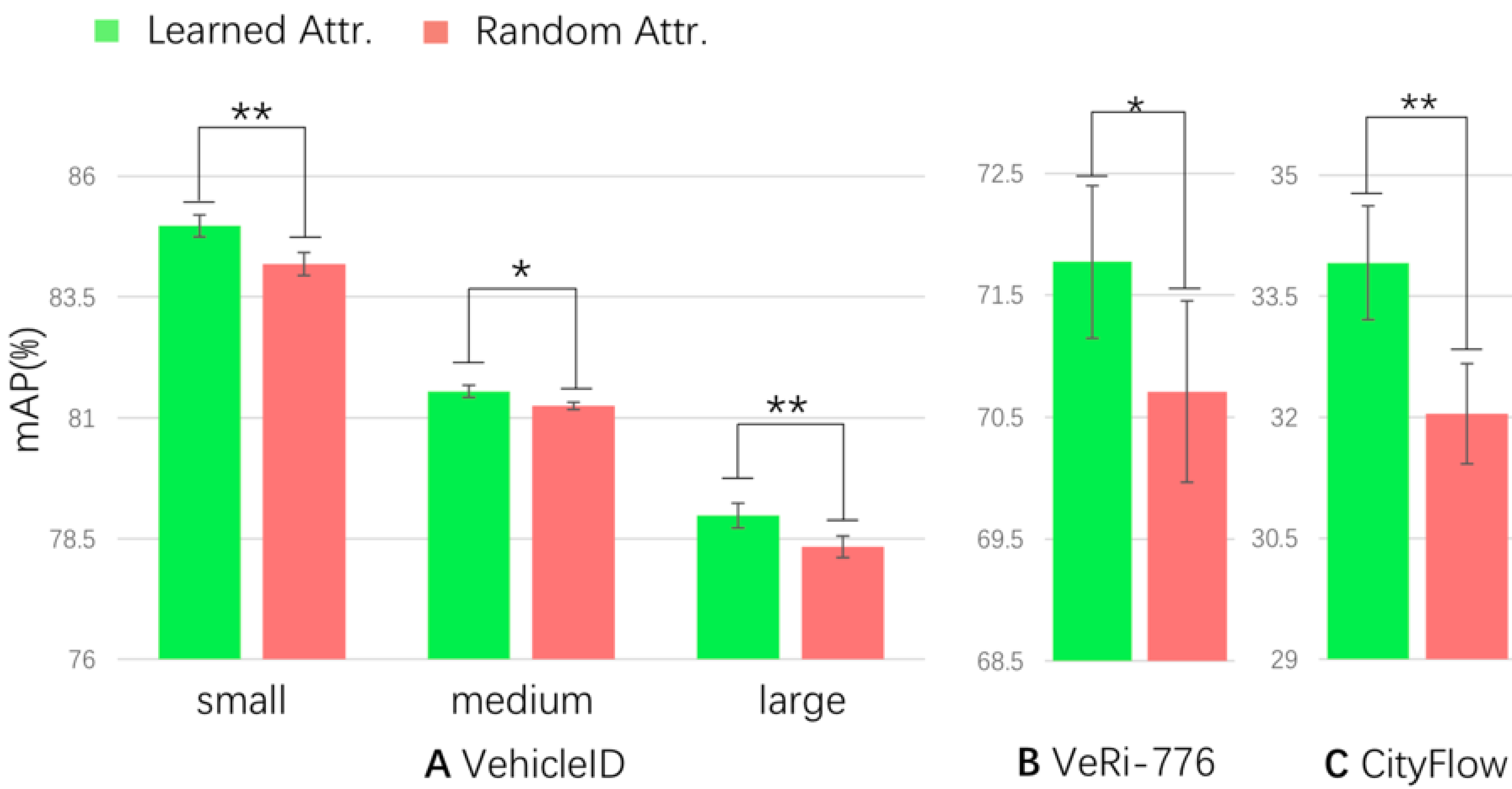
Experiment – statistical significance

- Learned attribute vs. random attribute



Experiment – statistical significance

- Learned attribute vs. random attribute



Outline

- Introduction
- Re-id vs multi-object tracking
- Data synthesis in object re-id
- Alice benchmark suite

Alice benchmark suite

Welcome to the Alice Synthetic 3D World

The Alice project is an online evaluation server of domain adaptation tasks in computer vision. In each task, the source domain consists of synthetic data, and the target domain can be either synthetic or real-world. A distinct feature of Alice is that the source domain is editable. We use 3D graphics engines like Unity to simulate the source data, such that the imagery can be precisely controlled through our APIs. In other words, the objective of Alice is to find a source training set to improve the recognition performance on a given target testing set.

News

- 23.12.2019: The very first task of Alice is online! It is a traditional task – domain adaptive pedestrian recognition. We use PersonX as the source domain and Market-1501 (unreleased labels) as the target domain. We are accepting leaderboard submissions!
- 06.12.2019: The Alice websit is online.

http://alice-challenge.site/

Alice benchmark suite

- Alice v0 is online, now accepting submissions

Xiaoxiao Sun, Liang Zheng, Dissecting person re-identification from the viewpoint of viewpoint. CVPR 2019.

- Task: style/feature domain adaptation
- Source: synthetic persons (PersonX, CVPR 2019)
- Target: real persons (AlicePerson, unreleased data from the Market-1501 data source)

Leader Board

TeamName	Rank-1	Rank-5	Rank-10	mAP	Rank
SPGAN (PCB)	0.28971	0.45277	0.538259	0.100823	1
Direct Transfer (PCB)	0.210554	0.347757	0.427441	0.078306	2
Direct Transfer (IDE)	0.170976	0.308179	0.38628	0.065966	3

Alice benchmark suite

- Future: content-level domain adaptation

Conclusion

- Re-id vs tracking
  - Feature sharing for efficiency considerations
  - Global (re-id) vs local (tracking)
- Content-level domain adaptation
  - Orthogonal to existing DA methods
  - Editable source domain
- Alice benchmark suite – content-level domain adaptation

Q & A

Thanks!