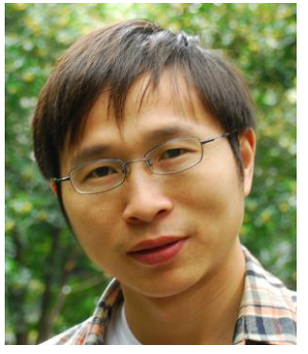
Person Re-Identification:
Recent Advances and Challenges

Shiliang Zhang
Peking University
Beijing, China



Jingdong Wang
Microsoft Research
Beijing, China

Qi Tian
University of Texas at
San Antonio, USA

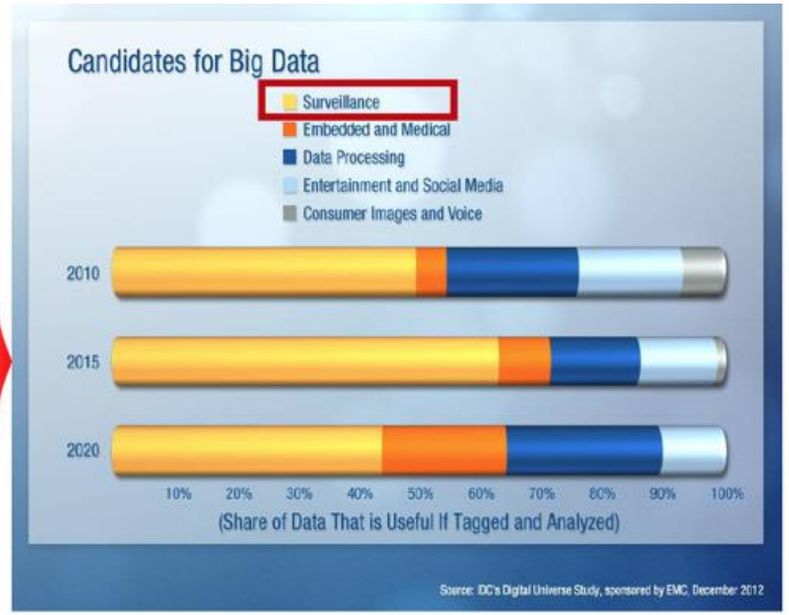
Wen Gao
Peking University
Beijing, China

Longhui Wei
Peking University
Beijing, China

Cameras are Everywhere in Smart City

Large camera networks

Fast growth: millions of cameras deployed



Surveillance ~50% of all big data

Complicated and rich: cars, persons, etc.

Data ≠ Information: Visual Mining Needed

Boston 2013

Paris 2015

Person Detection
Re-Identification
Tracking
.....

London 2017

Las Vegas 2017

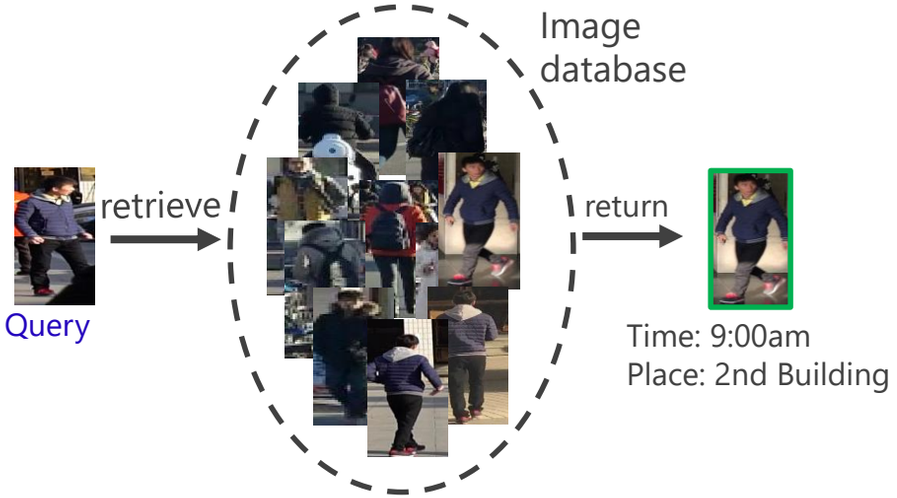
Person Re-Identification

- Associate the same individuals across multiple cameras
  = Multi-camera tracking

Person Re-Identification

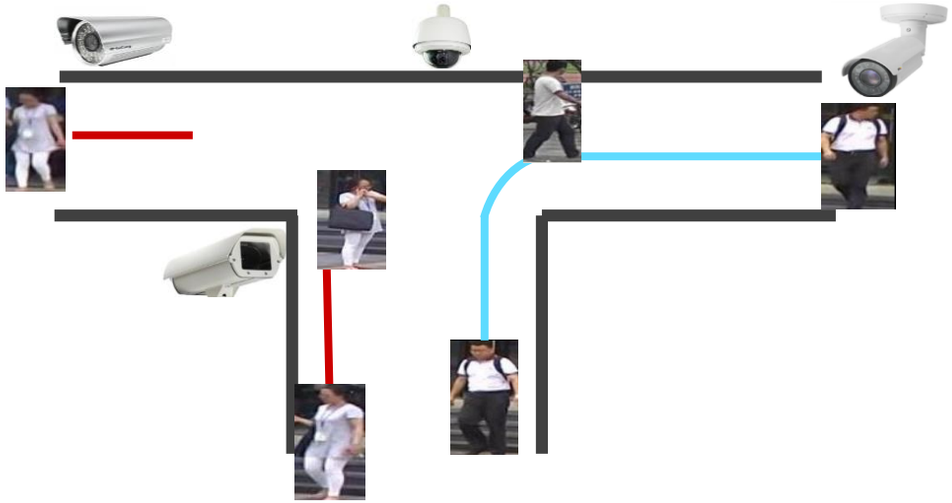
- Associate the tracks for a single camera
  - One person re-appears after a time
  - Target lost due to occlusion

Applications

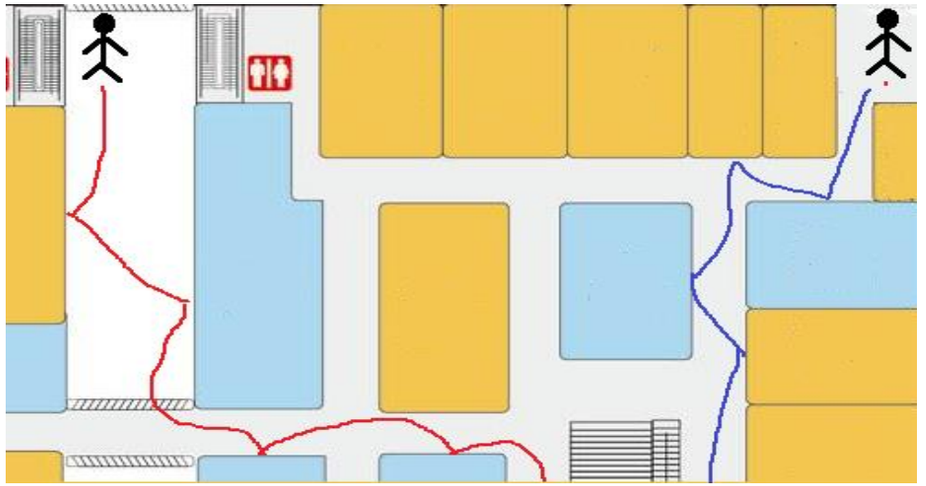


Person search

Event/action detection



Person counting by tracking



Behavior analysis

Solutions

- Biometric
  - Face
    - Frontal?
    - Enough resolution?

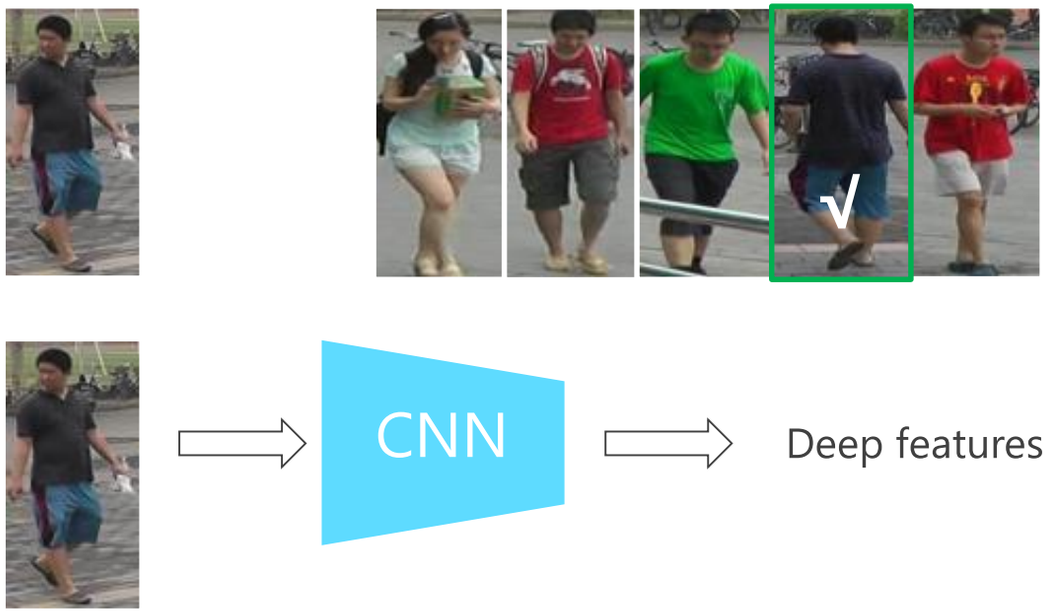
Solutions

- Biometric
  - Face
    - Frontal?
    - Enough resolution?

  - Gait
    - Silhouette extraction is not easy

Solutions

- Appearance
  - Clothing color, texture
  - Deep learning



Our Roadmap

- What we face
  - **Variable images:** Low quality, part misalignment, …

  Our work: Part-aligned representation

Low image quality

Detection error

- What we leverage
  - **Mid-level information:** Long sleeve, short hair, …

  Our work: Mid-level attributes

Occlusion

View& pose change

- What we contribute
  - Limited data scale/diversity

  Our work: Enlarge/diversify the datasets

Our Roadmap

- What we face
  - Variable images: Low quality, part misalignment, …

  Our work: Part-aligned representation

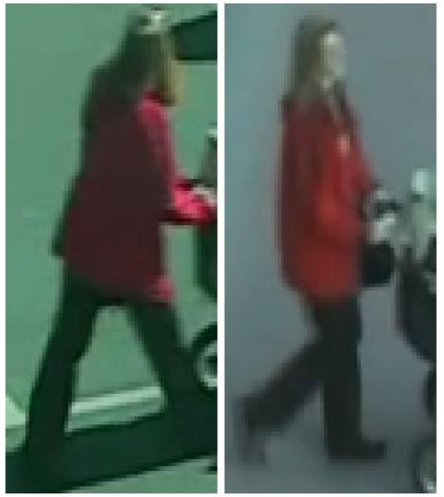
- What we leverage
  - Mid-level information: Long sleeve, short hair, …

  Our work: Mid-level attributes

- What we contribute
  - Limited data scale/diversity

  Our work: Enlarge/diversify the datasets

Upper body: long sleeve, black
Lower body: Jeans
Personal: female



Upper body: red
Lower body: black
Personal: female

Our Roadmap

- What we face
  - Variable images: Low quality, part misalignment, …

  Our work: Part-aligned representation

- What we leverage
  - Mid-level information: Long sleeve, short hair, …

  Our work: Mid-level attributes

- What we contribute
  - Limited data scale/diversity

  Our work: Enlarge/diversify the datasets

CUHK: 1000~2000 identities, short period, single weather

Our Roadmap

- What we face
  - Variable images: Low quality, part misalignment, …

  Our work: Part-aligned representation

- What we leverage
  - Mid-level information: Long sleeve, short hair, …

  Our work: Mid-level attributes

- What we contribute
  - Limited data scale/diversity

  Our work: Enlarge/diversify the datasets

Diverse time/weather/location

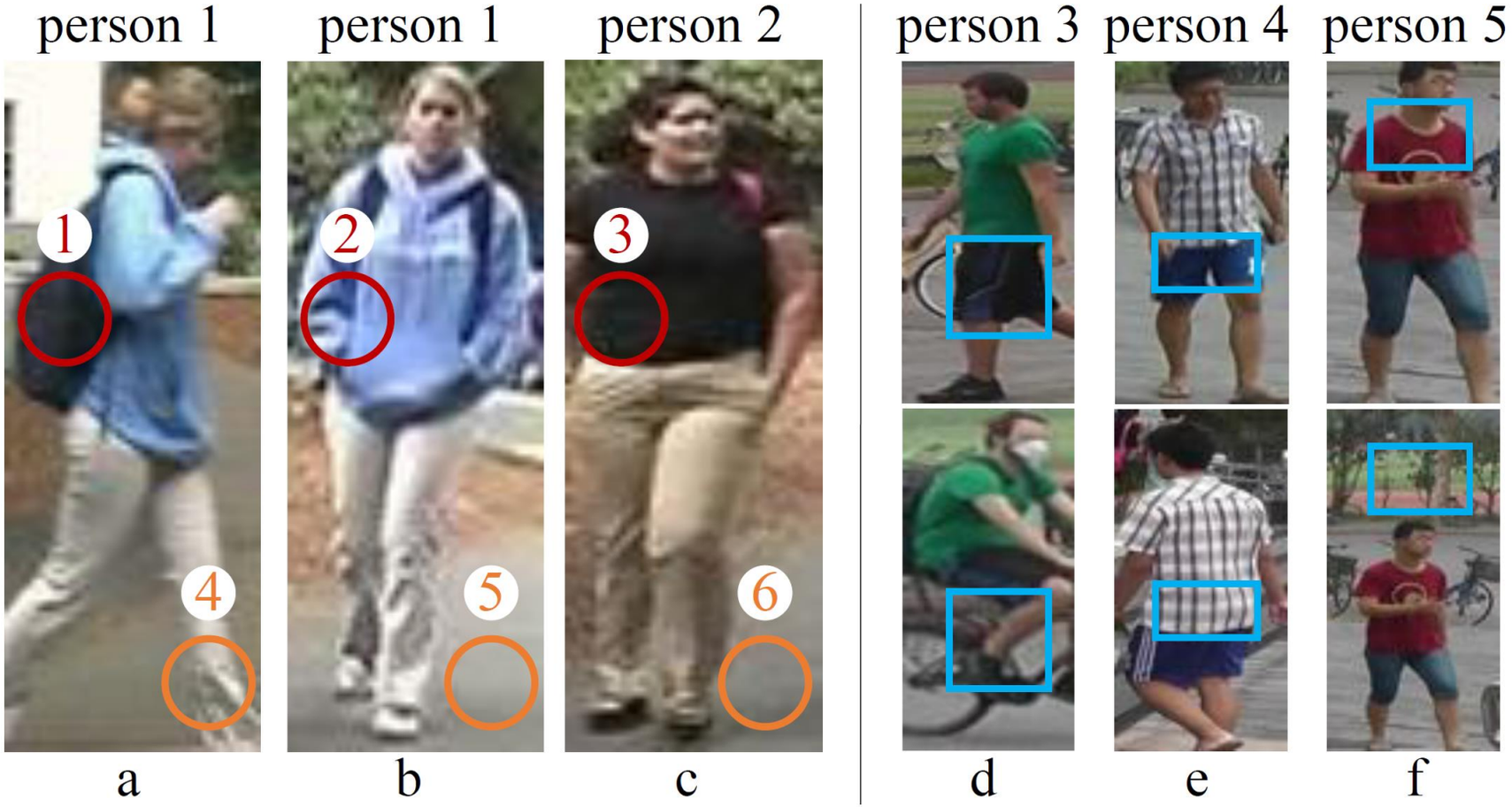
Scale: MegaPerson/Millions identities/boxes

Type: image/video/depth/infrared

Evaluation: training/validation/testing

……

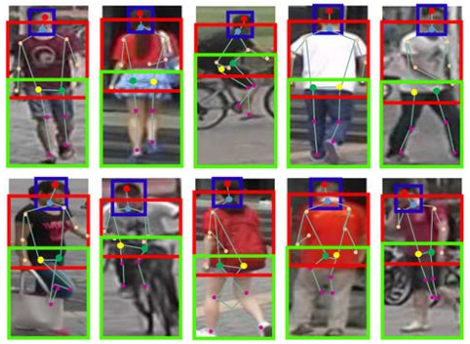
Session 1: Part-Aligned Representation Learning

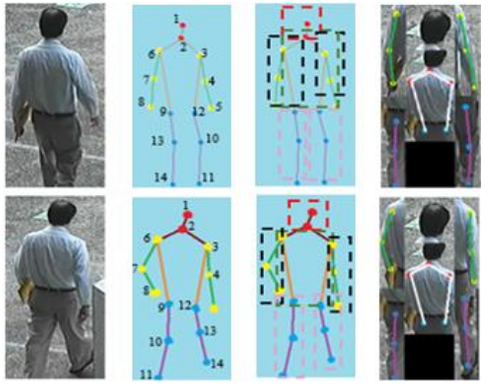


Part-Aligned Representation Learning

Attention, w/o pose detector
ICCV 2017

Coarse part boxes
from pose detector
ACMMM 2017



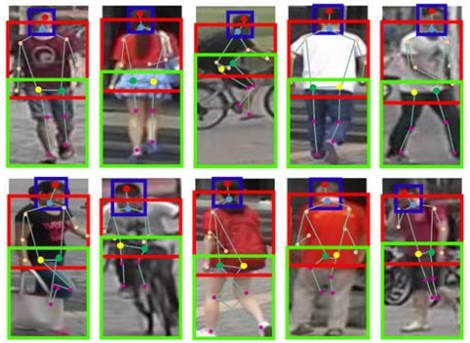


Normalized fine part boxes
from pose detector
ICCV 2017

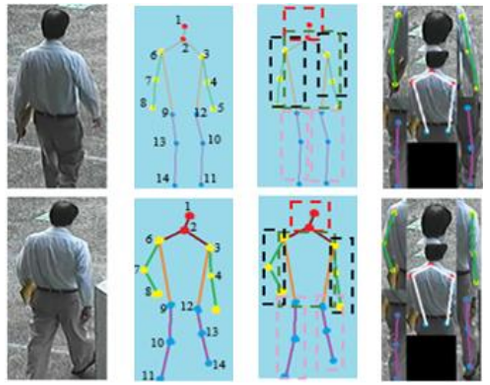
Retraining part maps
initialized from pose detector
Under review

Part-Aligned Representation Learning

Attention, w/o pose detector
ICCV 2017



Coarse part boxes
from pose detector
ACMMM 2017



Normalized fine part boxes
from pose detector
ICCV 2017

Retraining part maps
initialized from pose detector
Under review

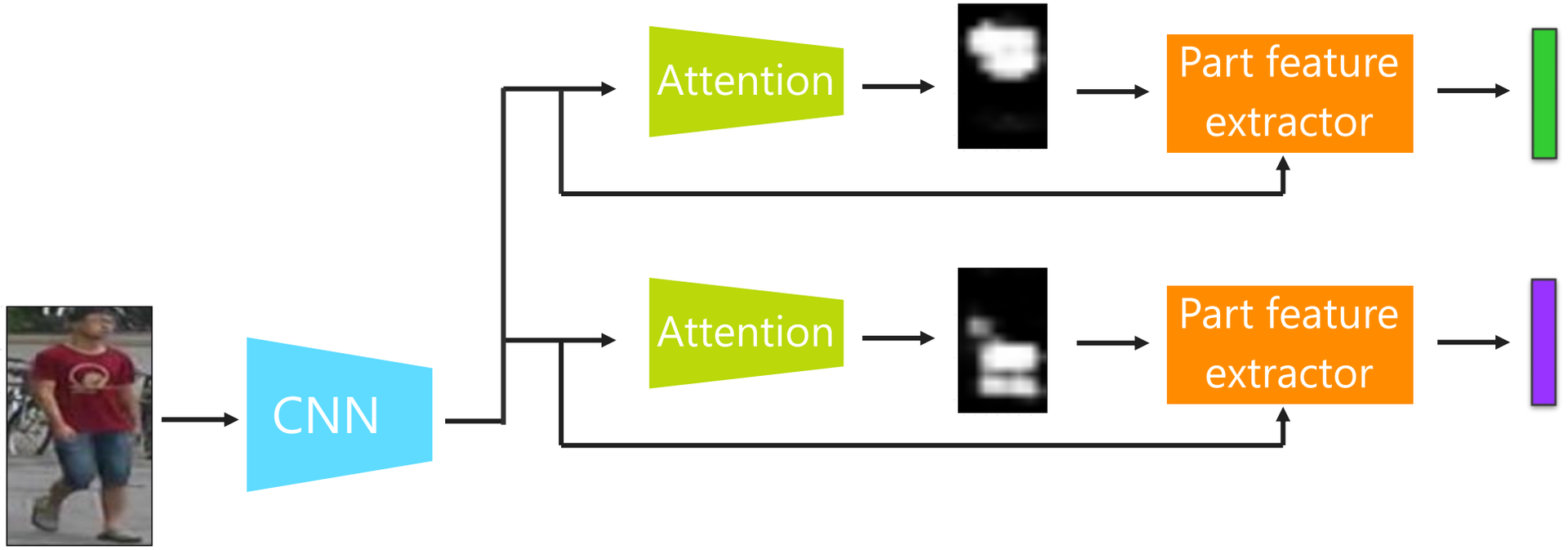
Deeply-Learned Part-Aligned Representations

- An end-to-end solution to
  - Learn the discriminative parts (attention maps) from scratch w/o part annotation
  - Directly for person matching

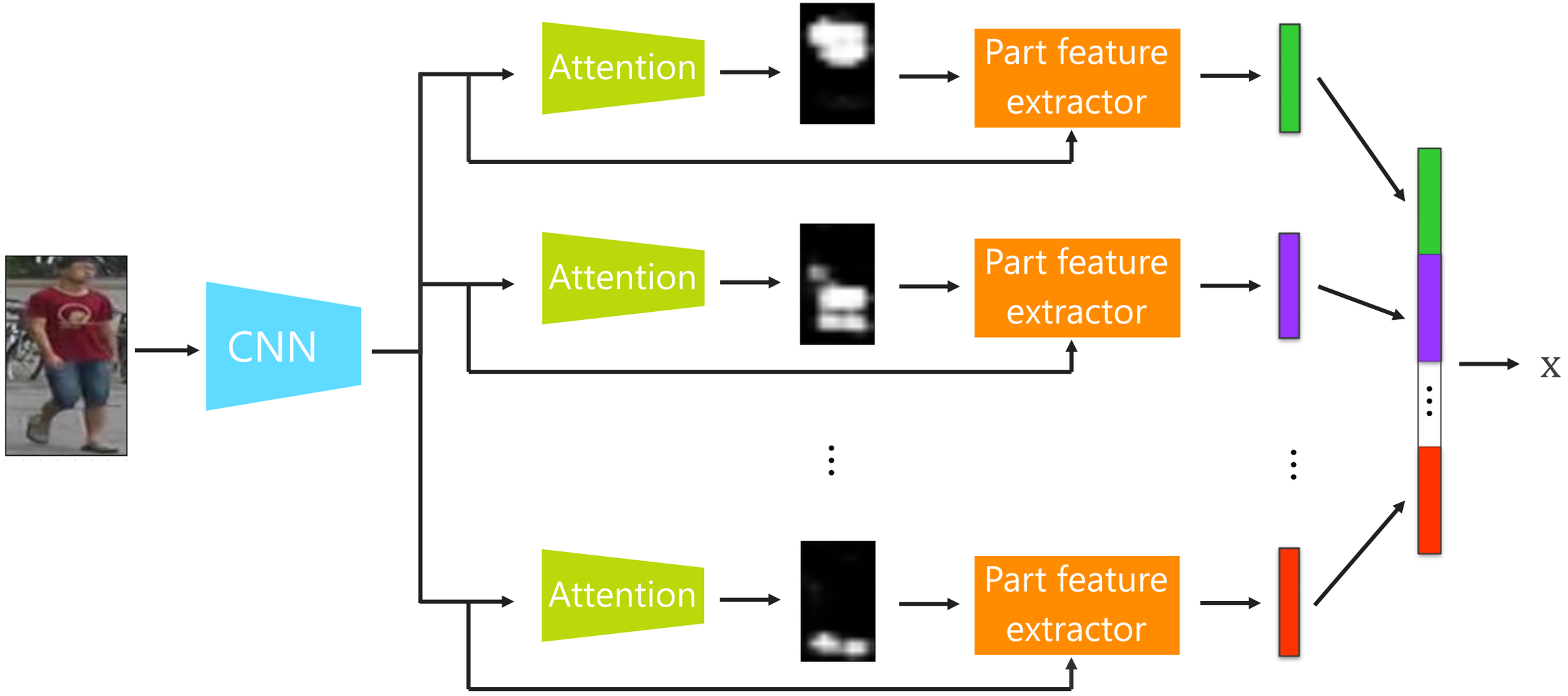
Liming Zhao, Xi Li, Yueting Zhuang, Jingdong Wang: Deeply-Learned Part-Aligned Representations for Person Re-identification. ICCV 2017: 3239-3248

Code: https://github.com/zlmzju/part_reid

Pipeline


Attention
Part feature extractor
Attention
Part feature extractor
CNN

Pipeline


CNN
Attention
Attention
Attention
Part feature extractor
Part feature extractor
Part feature extractor
x

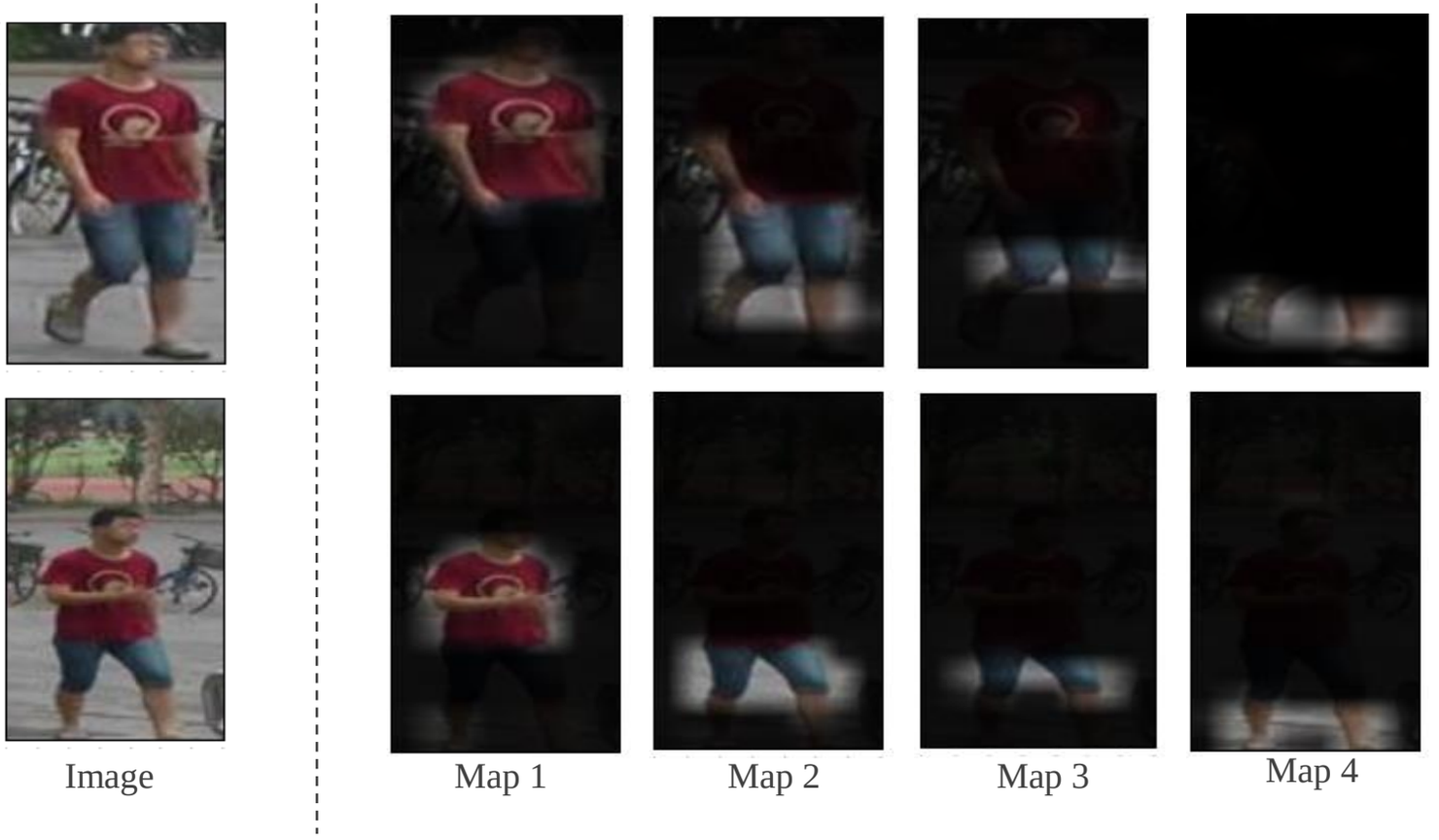
Optimization Loss

- Rank loss: Triplet hinge loss

$$l(\mathrm{x}, \mathrm{x}^{+}, \mathrm{x}^{-}) = [d(\mathrm{x}, \mathrm{x}^{+}) - d(\mathrm{x}, \mathrm{x}^{-}) + 1]_{+}$$

$$[y]_{+} = \begin{cases} y, if\ y > 0 \\ 0, if\ y \leq 0 \end{cases}$$

Parts are well aligned



The Effect of Attention

	rank-1	rank-5	rank-10
W/o attention	75.89	89.25	92.93
W/ attention	81.15	92.25	94.92
Gain	**5.26**	**3**	**1.99**

Experiments

- Datasets
  - Market-1501
  - CUHK03
  - CUHK01
- Evaluation metric
  - Cumulative matching characteristics (CMC) at Rank position
  - MAP

Experimental Results

- Market-1501
  - Training 750 identities; testing 751 identities
  - 3,368 queries; 15,913 galleries (2,798 distractors)

	rank-1	rank-5	rank-10	mAP
BoW (ICCV 2015)	35.84	52.40	60.33	14.75
PersonNet (Arxiv 2016)	37.21	-	-	18.57
Deep Attributes (ECCV 2016)	39.40	-	-	19.60
LOMO Feature(CVPR 2015)	43.79	-	-	22.22
WARCA Metric (ECCV 2016)	45.16	68.23	76.00	-
Bilinear CNN (Arxiv 2015)	45.58	67.00	76.00	26.11
Null Space (CVPR 2016)	55.43	-	-	29.87
Gated Siamese CNN (ECCV 2016)	65.88	-	-	39.55
Our Method	**81.15**	**92.25**	**94.92**	**63.69**

Experimental Results

- CUHK03
  - Training 1160 identities; testing 100 identities
  - Randomly split training and testing sets

	rank-1	rank-5	rank-10	rank-20
DeepReID (CVPR 2014)	20.65	51.32	68.74	83.06
LOMO Feature (CVPR 2015)	52.20	82.23	92.14	96.25
Improved Deep (CVPR 2015)	54.74	86.42	93.88	98.10
Null Space (CVPR 2016)	58.90	85.60	92.45	96.30
Bilinear CNN (Arxiv 2015)	63.87	91.43	95.85	98.56
PersonNet (Arxiv 2016)	64.80	89.40	94.92	98.20
Gated Siamese CNN (ECCV 2016)	68.10	88.10	94.60	-
WARCA Metric (ECCV 2016)	78.38	94.50	97.52	99.11
Our Method	**85.43**	**97.57**	**99.36**	**99.86**

Experimental Results

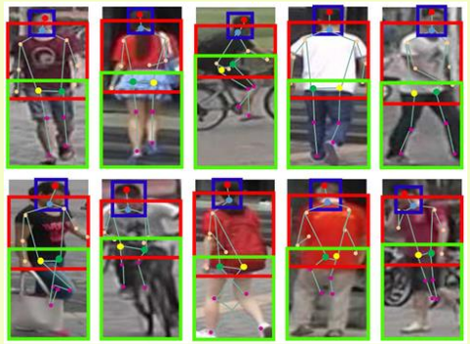
- CUHK01
  - The CUHK01 data set has 971 identities
  - 2 images per person in each view/camera
  - 486 test IDs; 485 for training

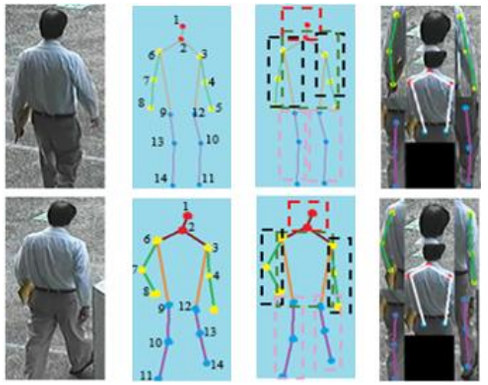
	rank-1	rank-5	rank-10	rank-20
Salience Matching (ICCV 2013)	28.45	45.85	55.67	67.95
Improved Deep (CVPR2015)	47.53	71.60	80.25	87.45
LOMO Feature (CVPR 2015)	63.21	83.89	90.04	94.16
Null Space (CVPR 2016)	64.98	84.96	89.92	94.36
WARCA Metric (ECCV 2016)	65.64	85.34	90.48	95.04
Our Method	**71.40**	**89.80**	**94.52**	**97.31**

Part-Aligned Representation Learning

Attention, w/o pose detector
ICCV 2017

Coarse part boxes
from pose detector
ACMMM 2017



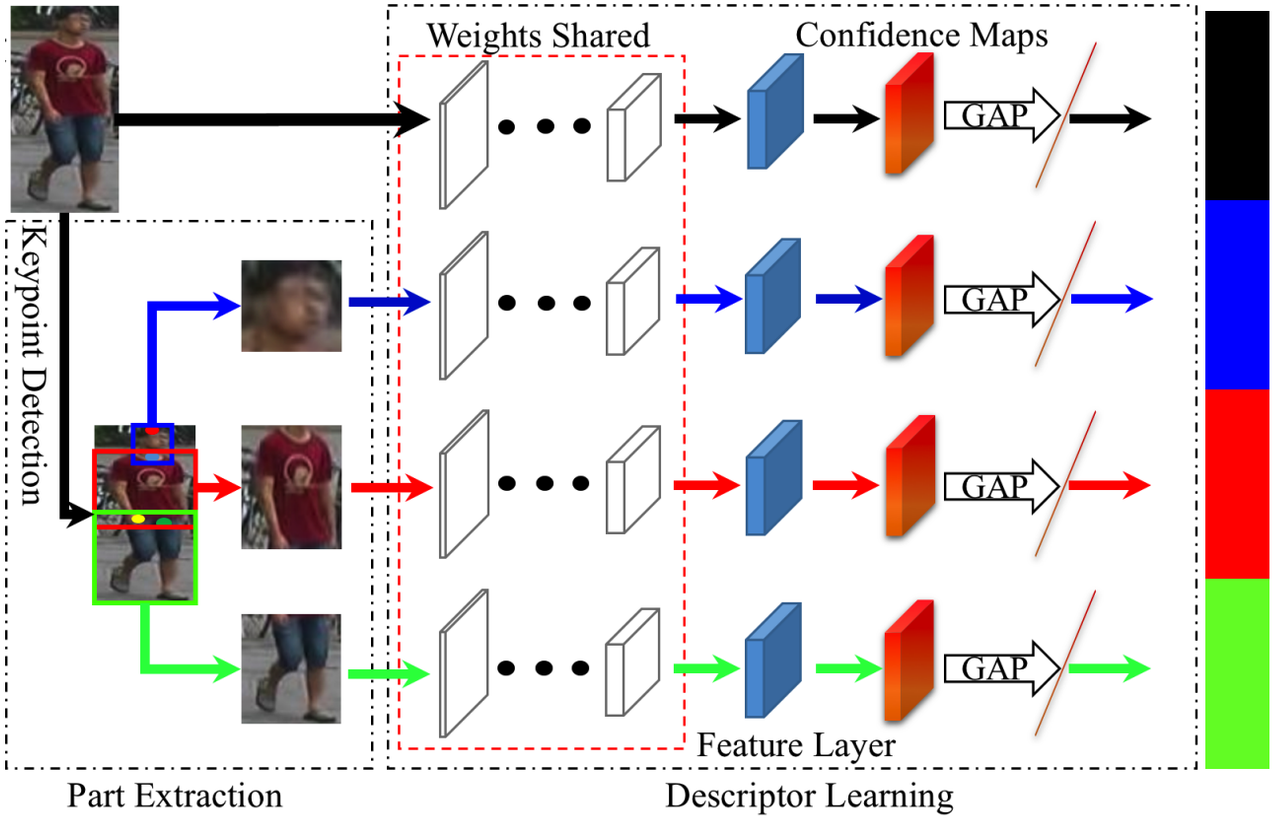


Normalized fine part boxes
from pose detector
ICCV 2017

Retraining part maps
initialized from pose detector
Under review

GLAD: Global-Local Alignment Descriptor

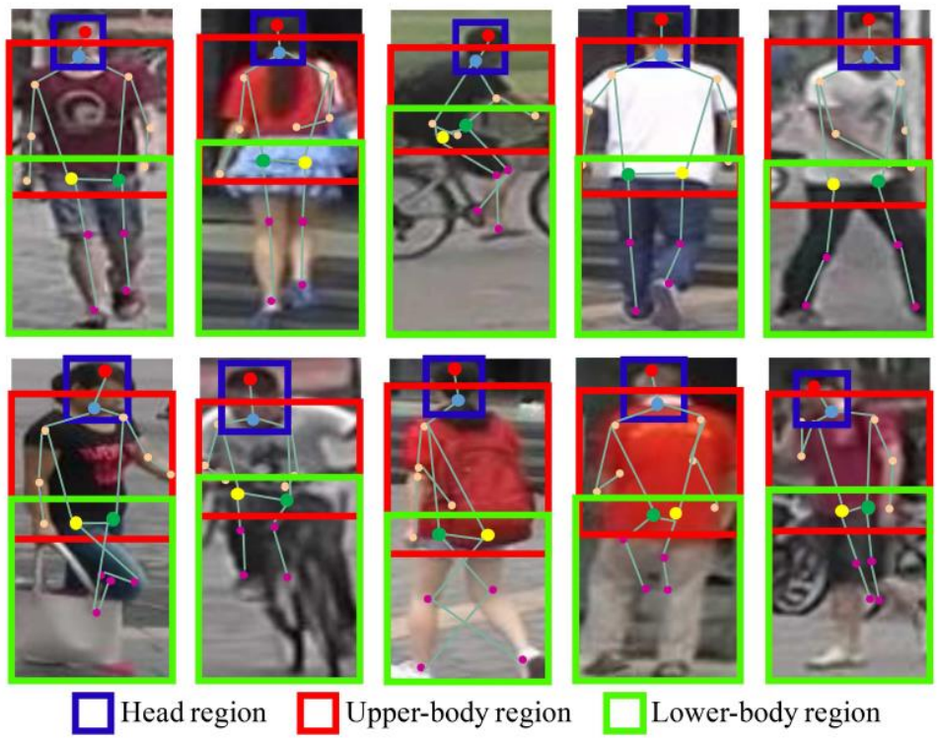
- Extract from features from coarsely partitioned parts
- Concatenate three part features and the global feature



Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, Qi Tian: GLAD: Global-Local-Alignment Descriptor for Pedestrian Retrieval. ACM Multimedia 2017: 420-428

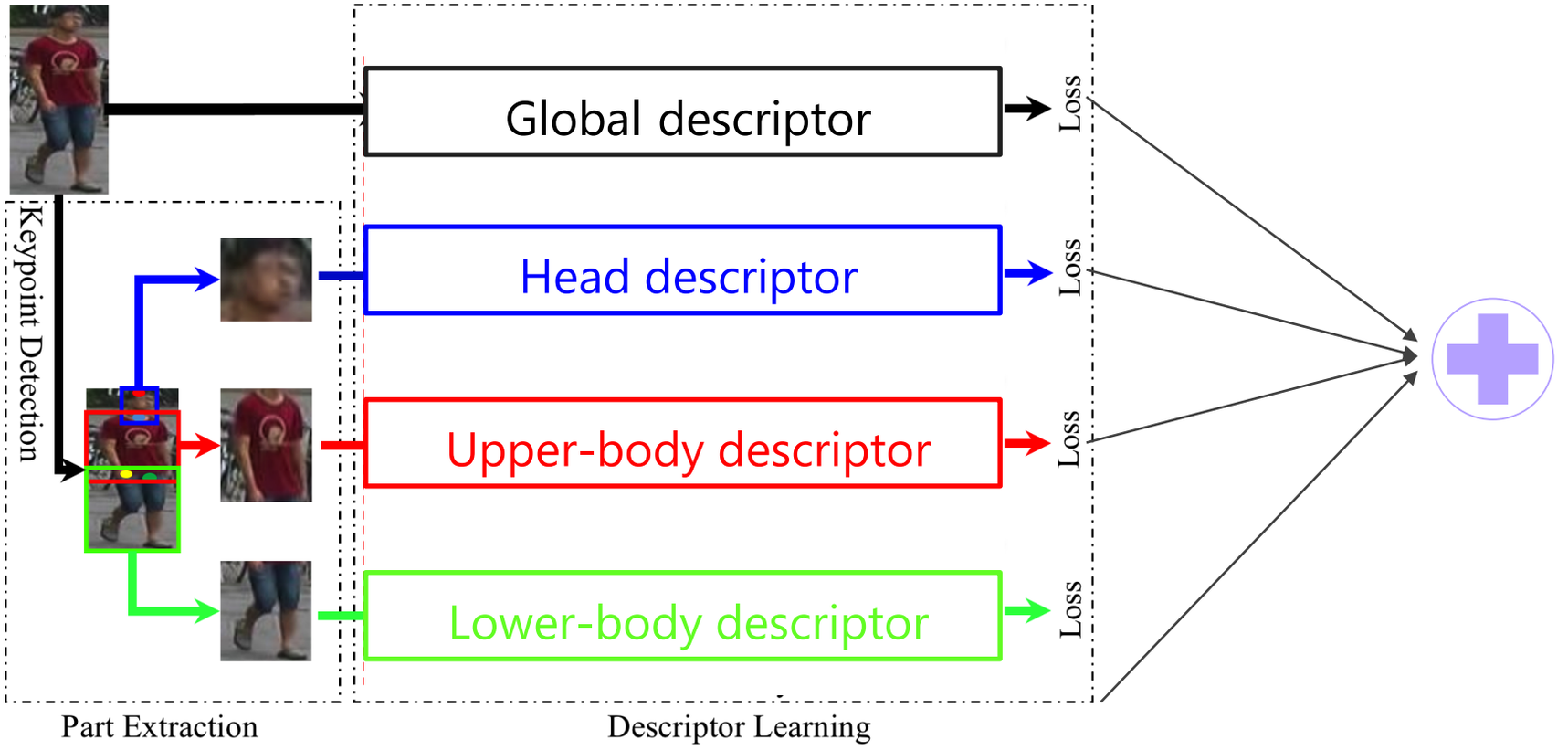
Coarse Part Partition

- Coarsely divide a pedestrian image into head, upper-body, and lower-body
  - Head region: based on upper-head point and neck point.
  - The upper and lower body regions: based on the neck point, left-hip and right-hip points, and bounding box bottom



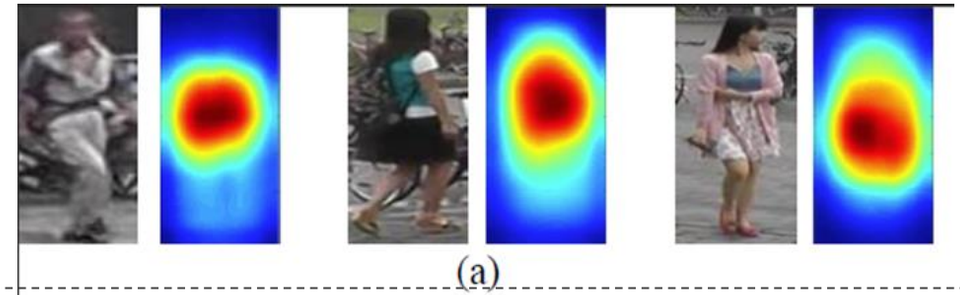
Loss

- Optimize the classification loss on each part feature/global feature



The Activation Maps of the Global Branch

w/o part



w/ part

The Effect of Part Boxes and Weight Sharing

Table 1: Comparison of different feature fusion and training strategies on Market1501. Baseline denotes the descriptor generated by our modified GoogLeNet [28] on the original image.

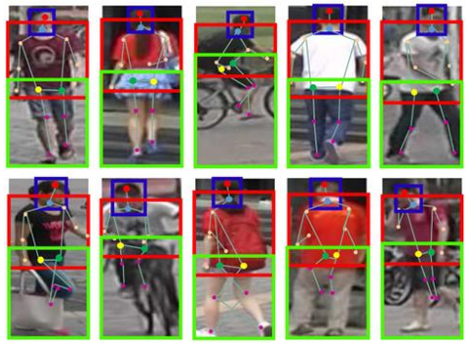
Training Strategy	Descriptor	mAP	Rank-1
-	Baseline	60.3	80.7
WO/S	Global	60.3	80.7
	Upper+Lower body	53.8	79.8
	Head+Upper+Lower body	49.6	77.3
	Head+Upper+Lower body (W)	55.7	81.0
	GLAD	**71.0**	**87.9**
W/S	Global	66.1	84.6
	Upper+Lower body	60.9	84.2
	Head+Upper+Lower body	55.6	81.8
	Head+Upper+Lower body (W)	62.8	85.5
	GLAD	**73.9**	**89.9**

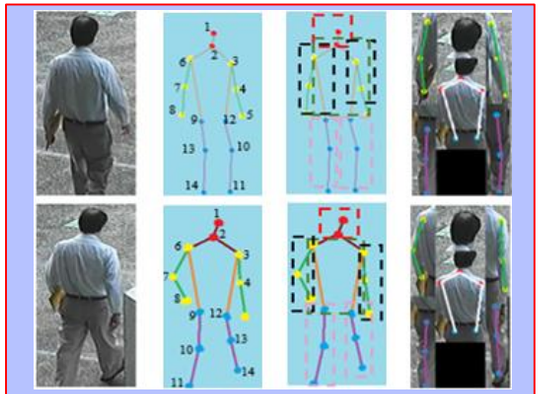

Probe
Probe
Probe

Part-Aligned Representation Learning

Attention, w/o pose detector
ICCV 2017

Coarse part boxes
from pose detector
ACMMM 2017



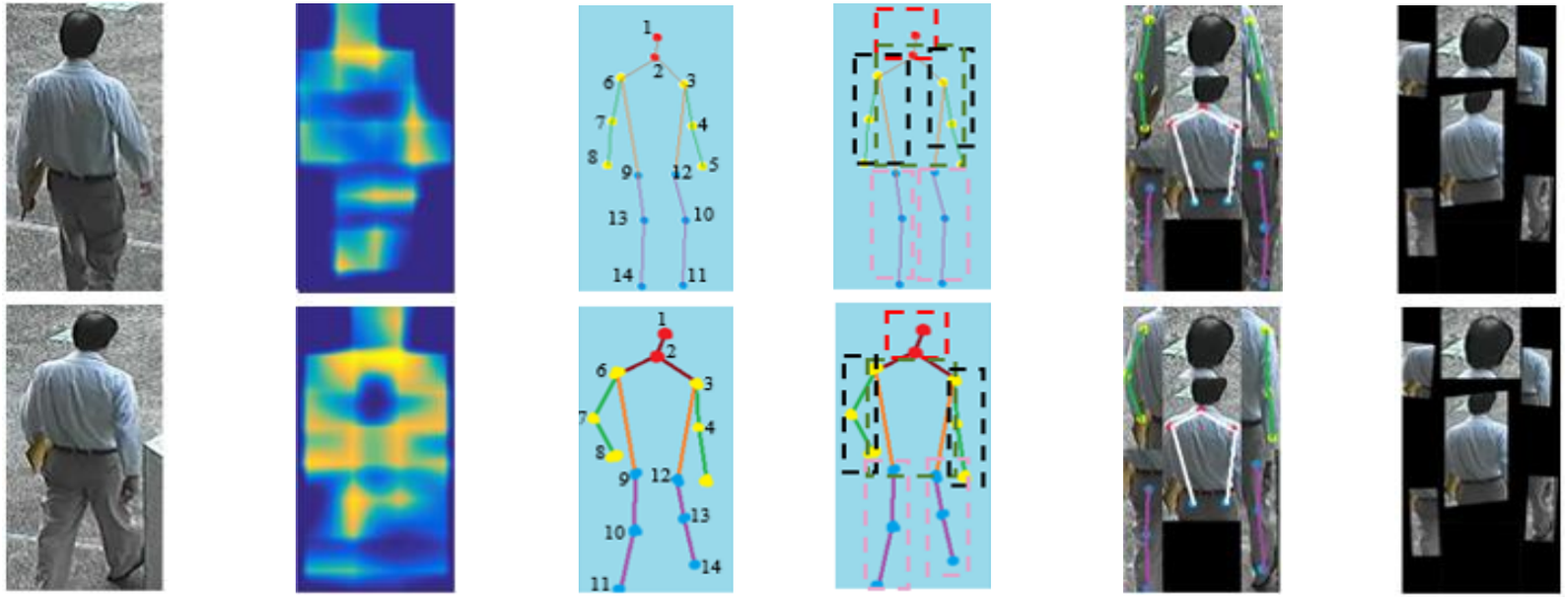


Normalized fine part boxes
from pose detector
ICCV 2017

Retraining part maps
initialized from pose detector
Under review

Pose Driven Deep Convolutional Model

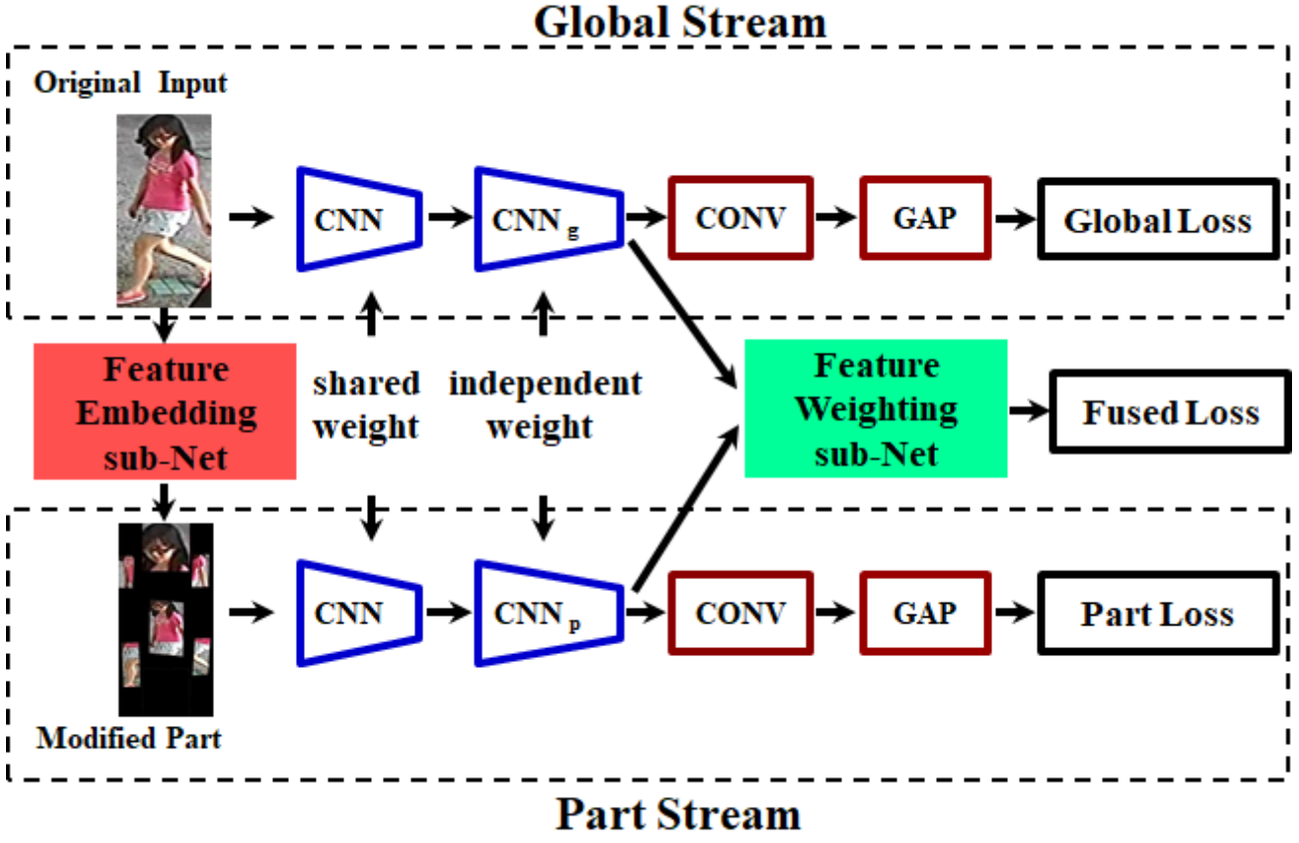
- Extract body parts in the form of boxes
- Normalize them into a fixed position



Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, Qi Tian: Pose-Driven Deep Convolutional Model for Person Re-identification. ICCV 2017: 3980-3989

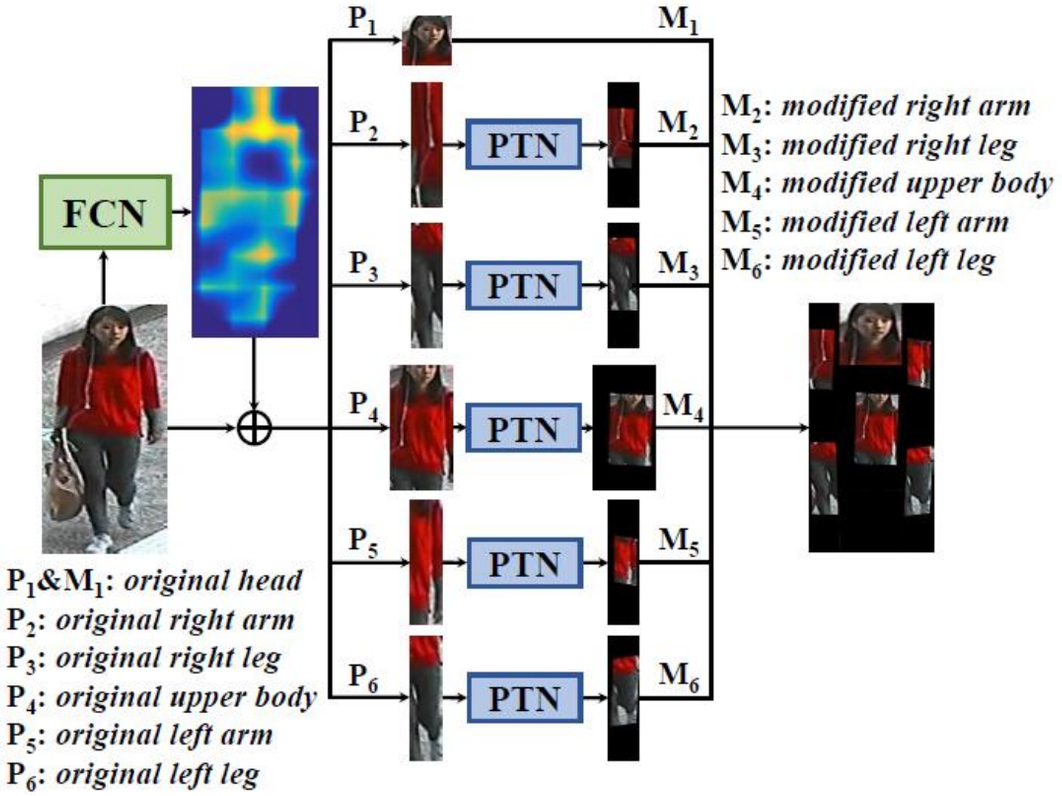
Pipeline

- Pose-driven Deep Convolutional model (PDC) has two branches for global feature extraction and part feature extraction.
- Part feature is extracted from normalized part image to achieve pose invariance.

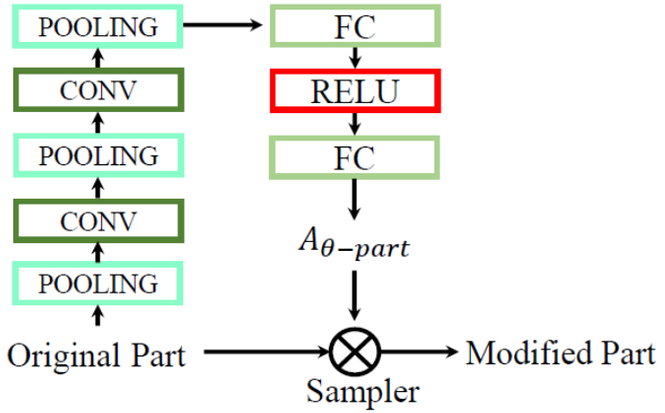


Feature Embedding sub-Net

- Part Transform Network (PTN) normalizes the orientation and scale of each part



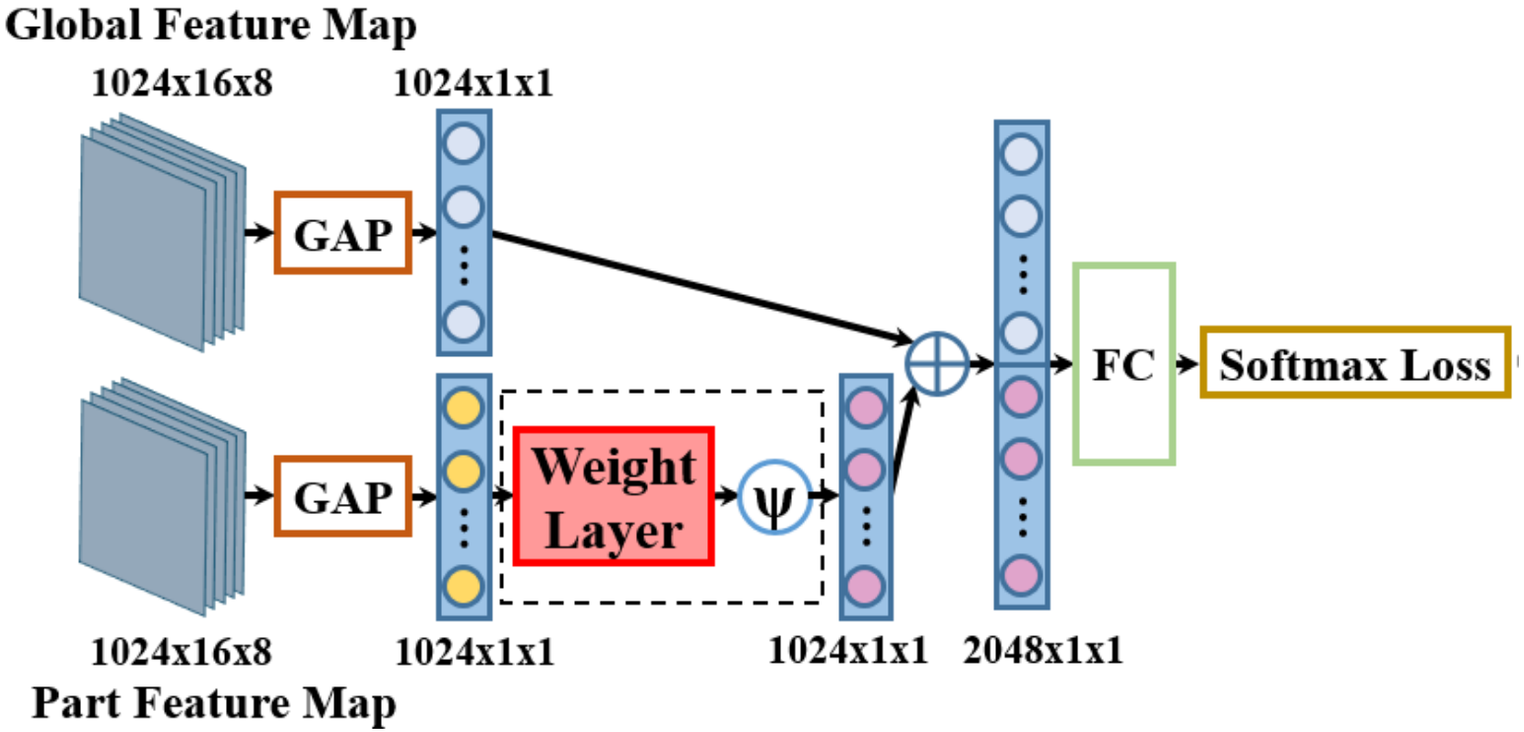
Part Transform Network:



$$\begin{pmatrix} x^s \\ y^s \end{pmatrix} = A_\theta \begin{pmatrix} x^t \\ y^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix} \begin{pmatrix} x^t \\ y^t \\ 1 \end{pmatrix}$$

Feature Weighting sub-Net

- Feature Weighting sub-Net (FWN) adaptively predicts weights for global and part feature fusion



The Effect of PDC

Table 2. The results on the *CUHK 03*, *Market 1501* and *VIPeR* datasets by five variants of our approach and the complete PDC.

dataset	CUHK03		Market1501		VIPeR
	labeled	detected			
method	rank1	rank1	mAP	rank1	rank1
Global Only	79.83	71.89	52.84	76.22	37.97
Part Only	53.73	47.29	31.74	55.67	22.78
Global+Part	85.07	76.33	62.20	81.74	48.42
Global+Part+FEN	87.15	77.57	62.58	83.05	50.32
Global+Part+FWN	86.41	77.62	62.58	82.69	50.00
PDC	**88.70**	**78.29**	**63.41**	**84.14**	**51.27**

Market-1501

Table 3: Comparison on Market1501 in single query mode.

Methods	mAP	Rank-1
BoW+Kissme [39]	20.8	44.4
WARCA [15]	-	45.2
LOMO+XQDA [18]	22.2	43.8
Null Space [37]	35.7	61.0
SCSP [4]	26.4	51.9
PersonNet [33]	26.4	37.2
Gated Siamese [28]	39.6	65.9
LSTM Siamese [29]	35.3	61.6
DLCNN [40]	59.9	79.5
PIE [38]	56.0	79.3
Baseline	60.3	80.7
GLAD	**73.9**	**89.9**

CUHK03-Detected

Methods	rank1	rank5	rank10	rank20
MLAPG [28]	51.15	83.55	92.05	96.90
LOMO + XQDA [27]	46.25	78.90	88.55	94.25
LDNS [52]	54.70	84.75	94.80	95.20
IDLA [1]	44.96	76.01	84.37	93.15
SI+CI [44]	52.17	84.30	92.30	95.00
LSTM S-CNN [43]	57.30	80.10	88.30	-
Gate S-CNN [42]	61.80	80.90	88.30	-
EDM [39]	52.09	82.87	91.78	97.17
PIE [55]	67.10	92.20	96.60	98.10
PDC	**78.29**	**94.83**	**97.15**	**98.43**

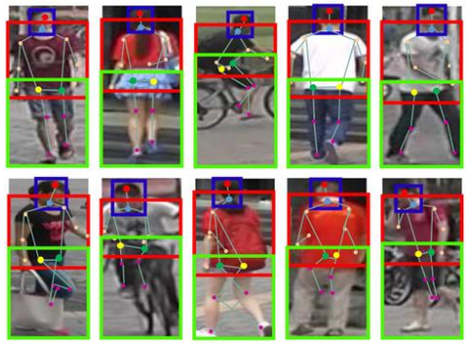
CUHK03-Labeled

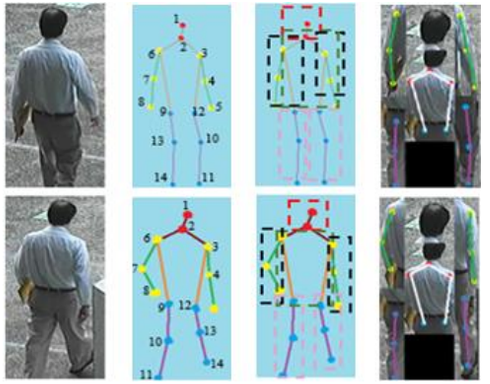
Methods	rank1	rank5	rank10	rank20
MLAPG [28]	57.96	87.09	94.74	96.90
LOMO + XQDA [27]	52.20	82.23	94.14	96.25
WARCA [22]	78.40	94.60	-	-
LDNS [52]	62.55	90.05	94.80	98.10
IDLA [1]	54.74	86.50	93.88	98.10
PersonNet [47]	64.80	89.40	94.90	98.20
DGDropout [48]	72.58	91.59	95.21	97.72
EDM [39]	61.32	88.90	96.44	99.94
Spindle [16]	88.50	97.80	98.60	99.20
PDC	**88.70**	**98.61**	**99.24**	**99.67**

Part-Aligned Representation Learning

Attention, w/o pose detector
ICCV 2017

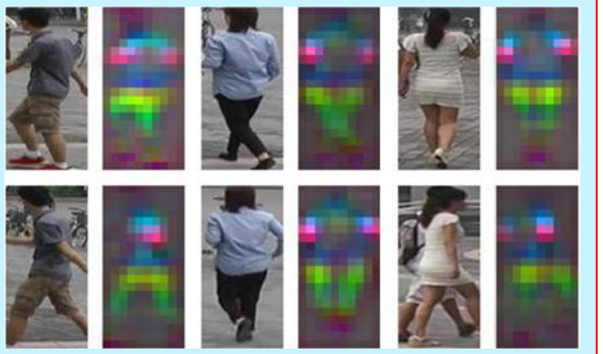
Coarse part boxes
from pose detector
ACMMM 2017



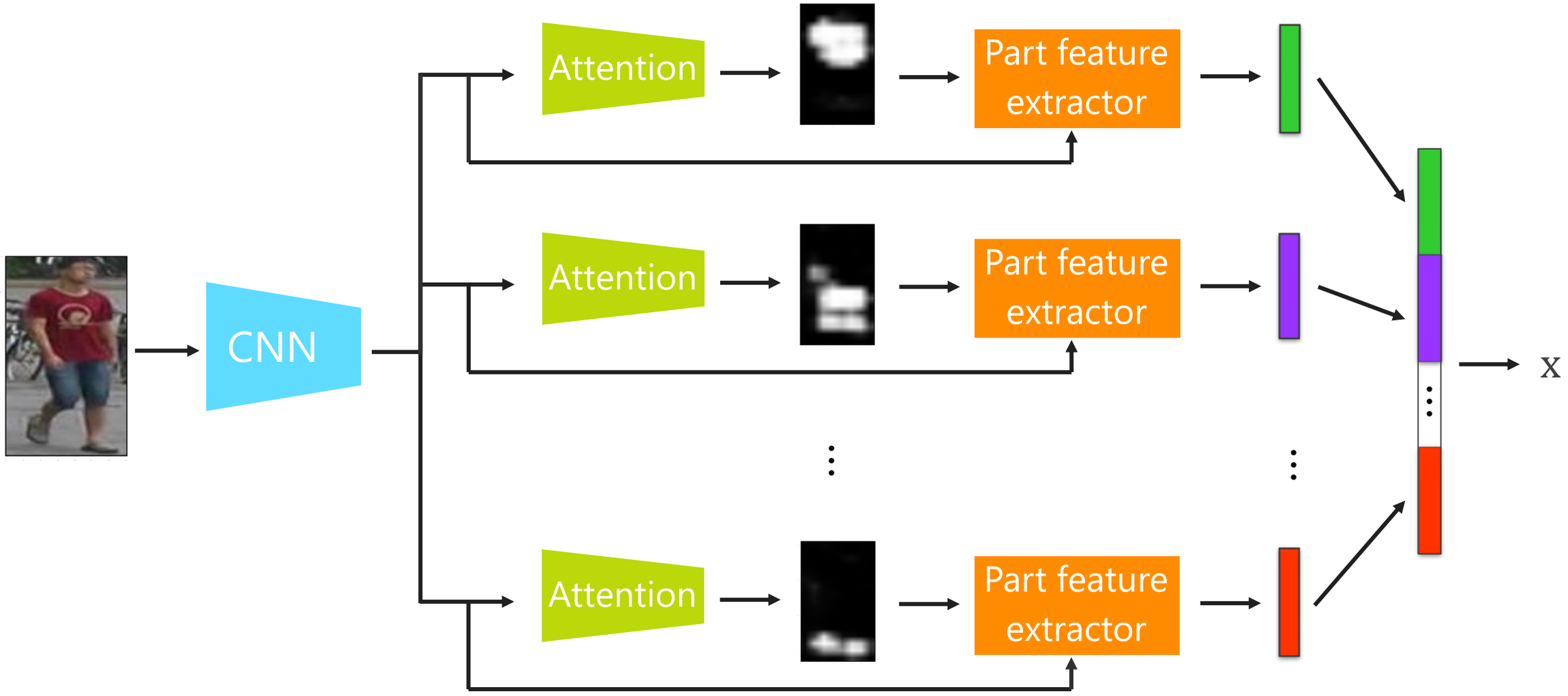


Normalized fine part boxes
from pose detector
ICCV 2017

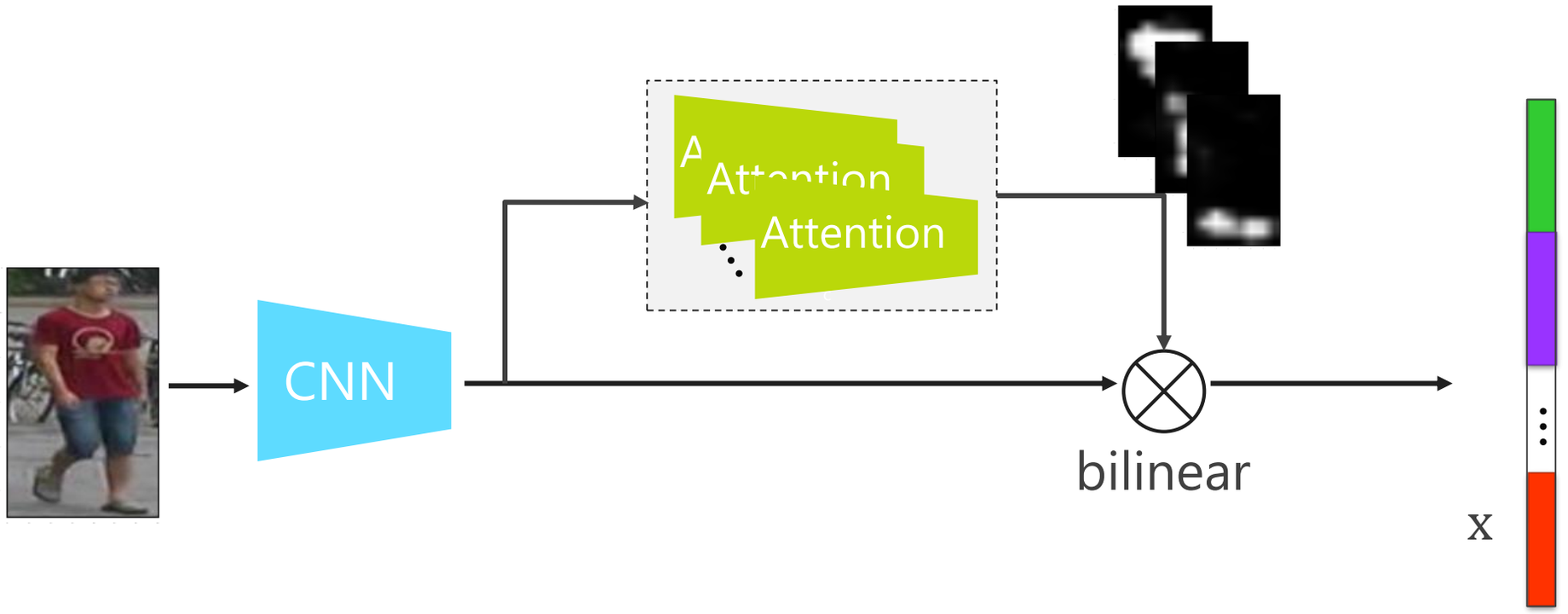
Retraining part maps
initialized from pose detector
Under review



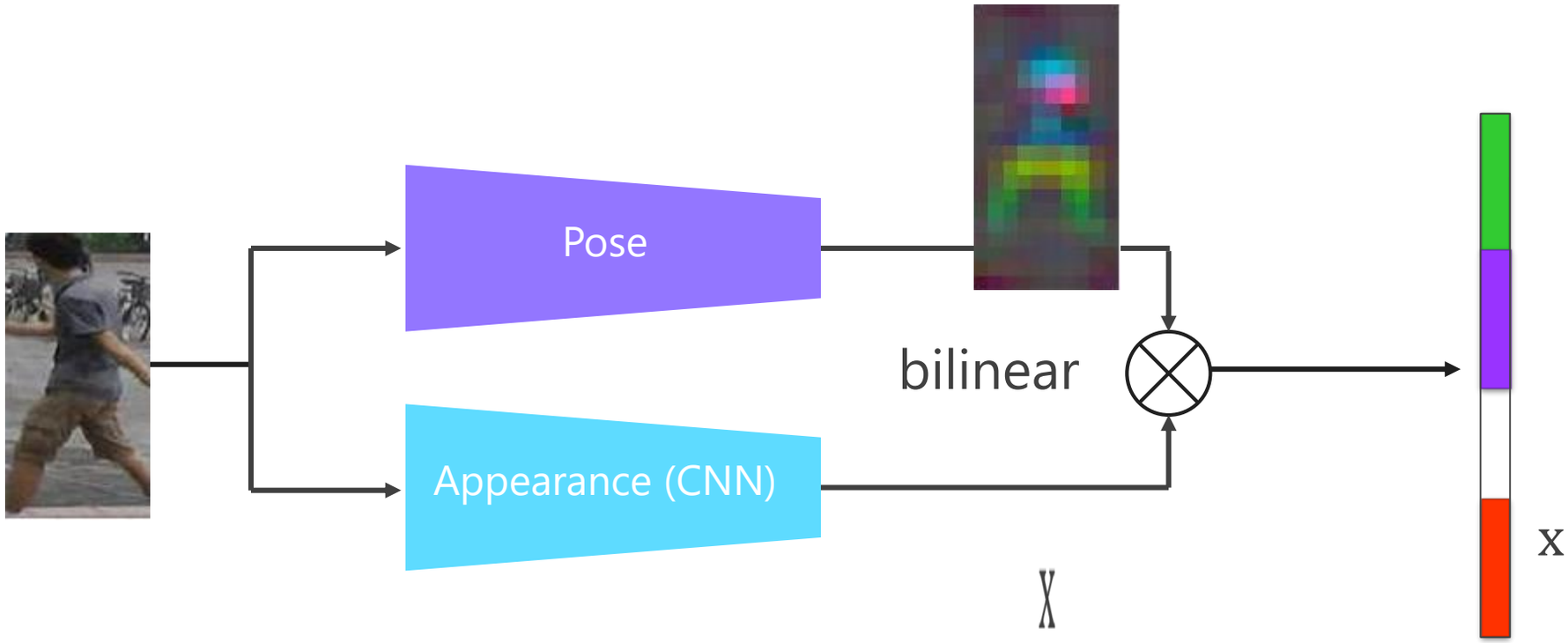
Attention–based Representations as Bilinear Mapping



Attention–based Representations as Bilinear Mapping



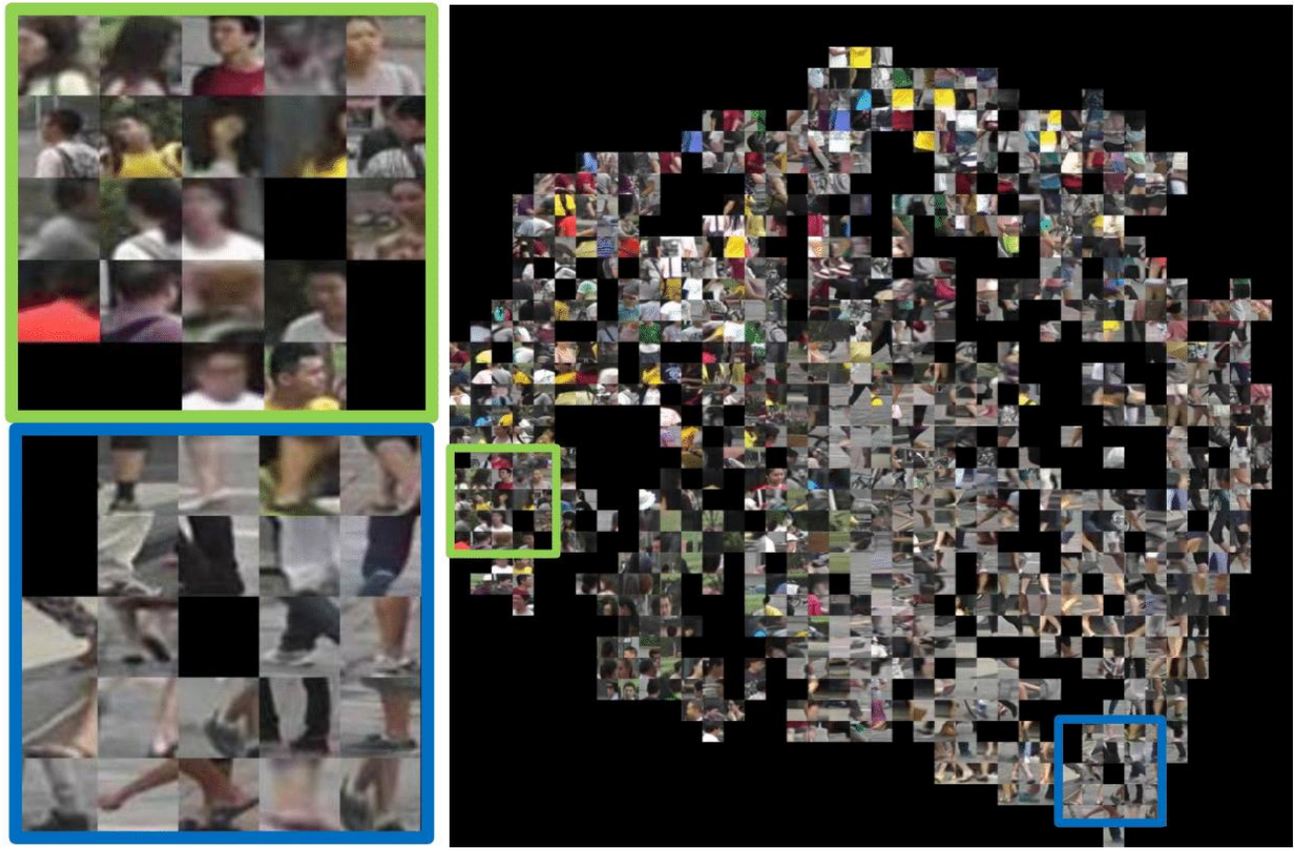
Part-Aligned Bilinear Representations



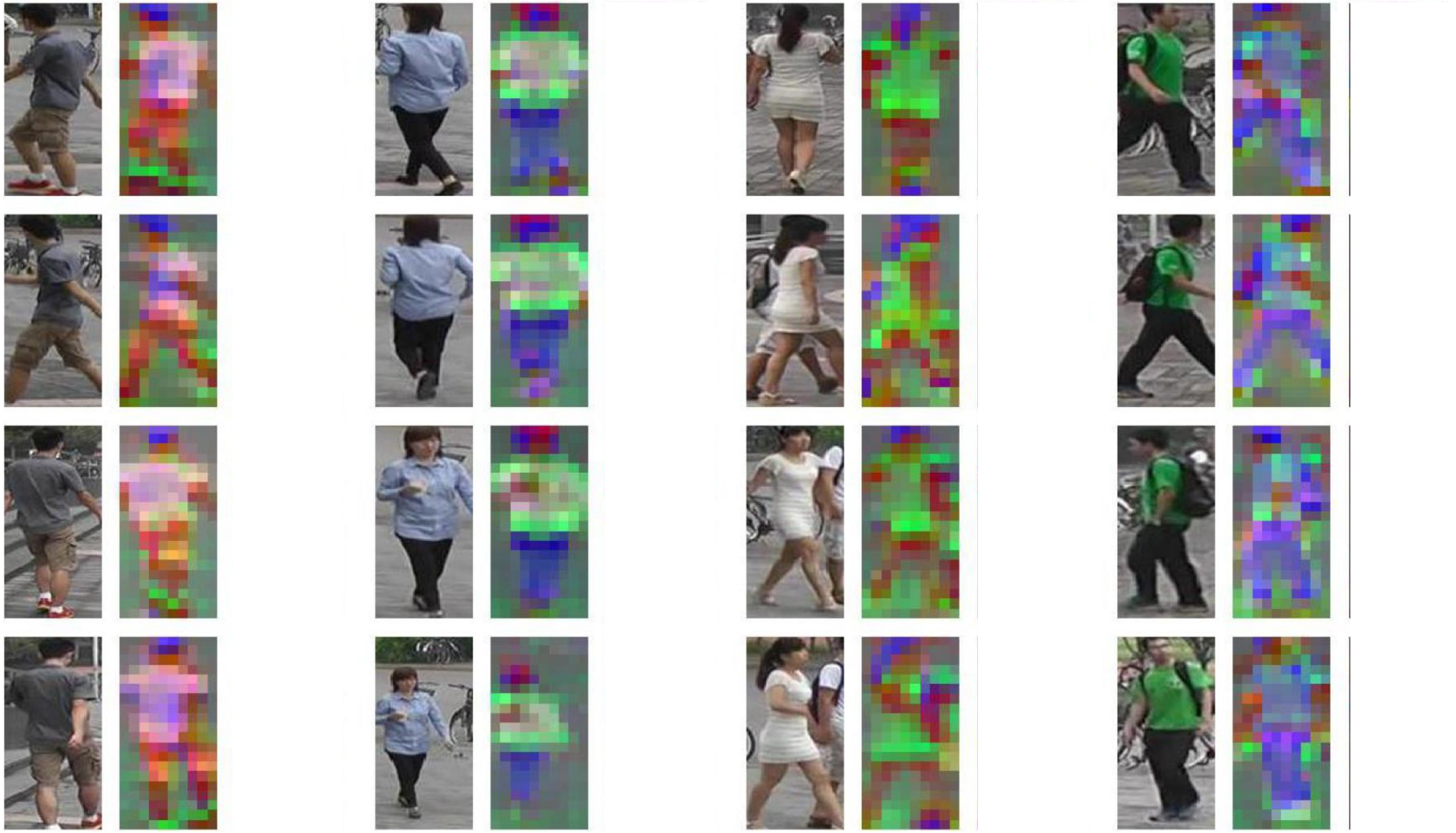
Pose estimator *pretrained on COCO*, and re-trained *only with the re-id loss*

Appearance Descriptors Clustered by Colors

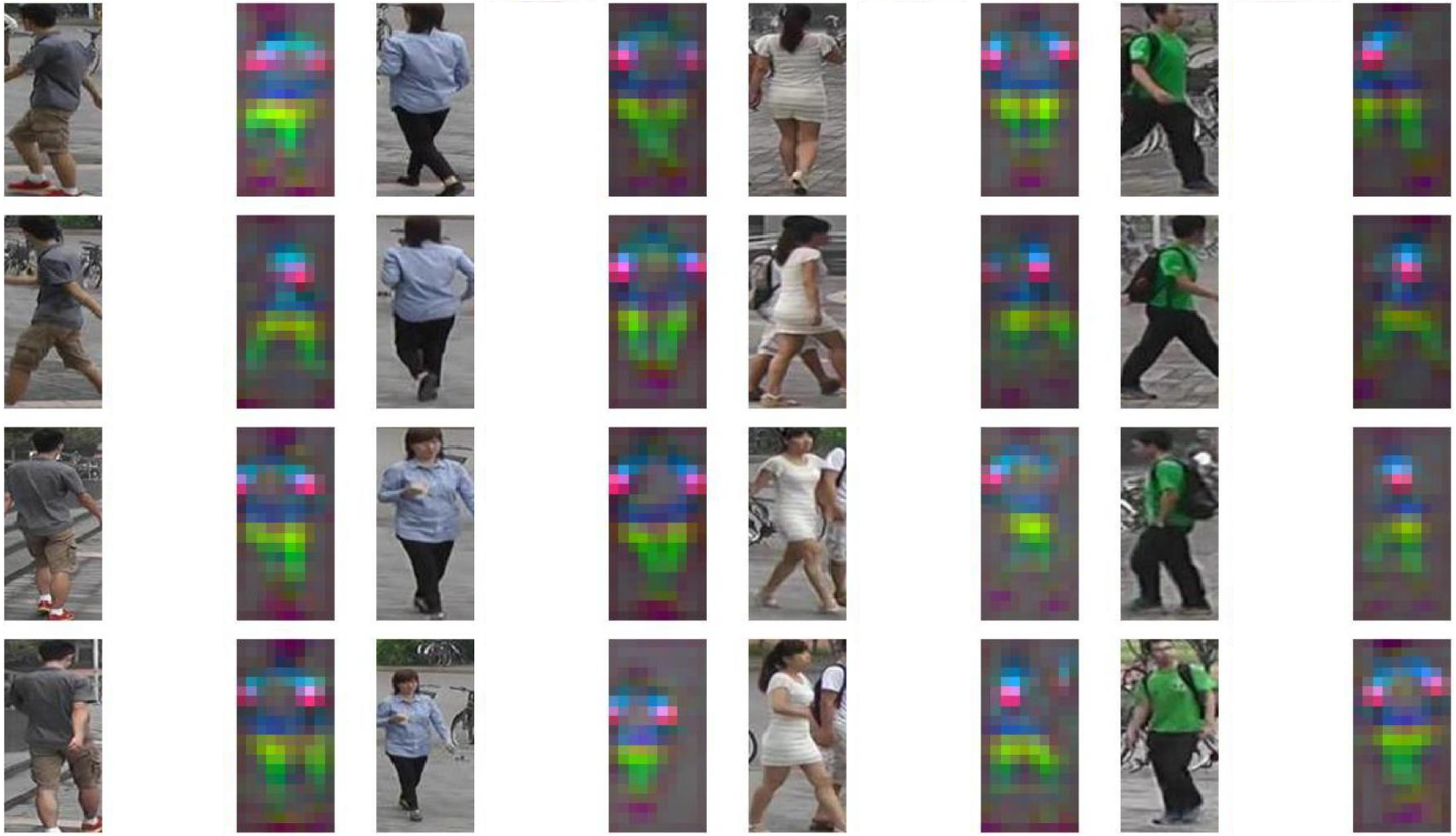
Part Descriptors Clustered by Body Parts



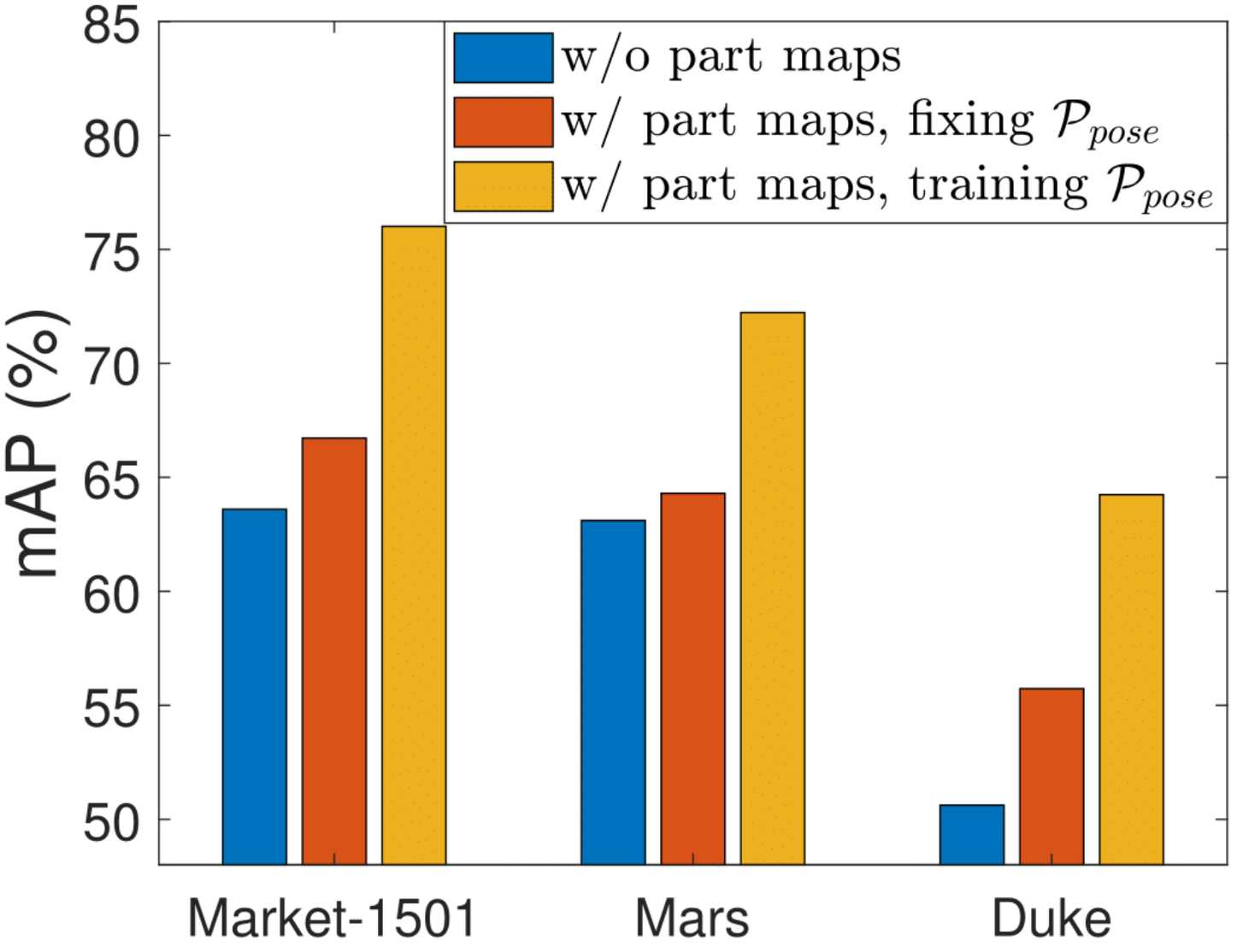
Appearance & Part Descriptors are Informative



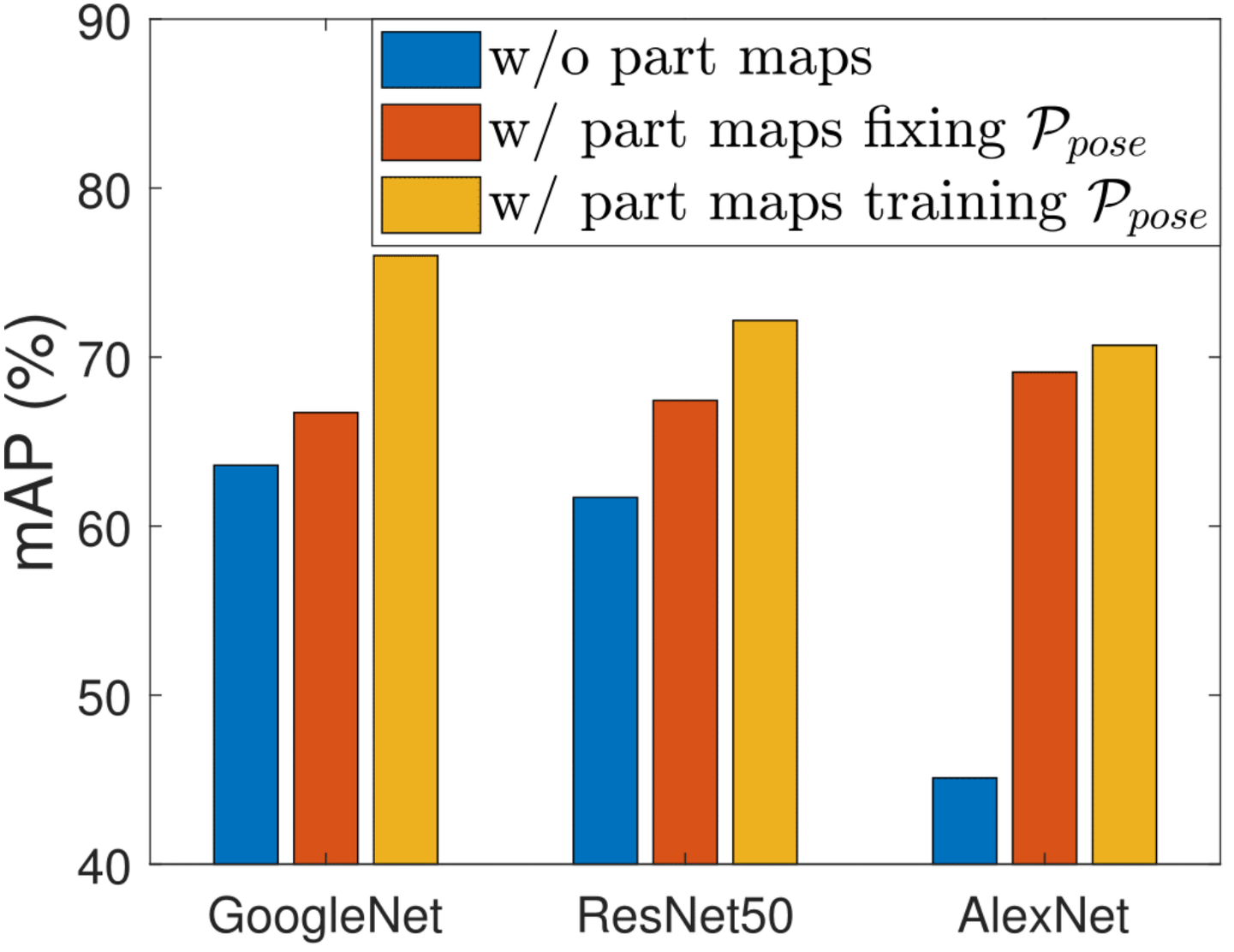
Appearance & Part Descriptors are Informative



Effectiveness of Part Descriptors on Various datasets



Effectiveness of Part Descriptors on Various Networks



Market-1501

	Sinlge Query					Multi Query				
Rank	1	5	10	20	mAP	1	5	10	20	mAP
Varior et al. 2016 [52]	61.6	-	-	-	35.3	-	-	-	-	-
Zhong et al. 2017 [71]	77.1	-	-	-	63.6	-	-	-	-	-
Zhao et al. 2017 [18]	80.9	91.7	94.7	96.6	63.4	-	-	-	-	-
Sun et al. 2017 [73]	82.3	92.3	95.2	-	62.1	-	-	-	-	-
Geng et al. 2016 [74]	83.7	-	-	-	65.5	89.6	-	-	-	73.8
Lin et al. 2017 [70]	84.3	93.2	95.2	97.0	64.7	-	-	-	-	-
Bai et al. 2017 [75]	82.2	-	-	-	68.8	88.2	-	-	-	76.2
Chen et al. 2017 [76]	72.3	88.2	91.9	95.0	-	-	-	-	-	
Hermans et al. 2017 [72]	84.9	94.2	-	-	69.1	90.5	96.3	-	-	76.4
+ re-ranking	86.7	93.4	-	-	81.1	91.8	95.8	-	-	87.2
Zhang et al. 2017 [77]	87.7	-	-	-	68.8	91.7	-	-	-	77.1
Zhong et al. 2017 [78]	87.1	-	-	-	71.3	-	-	-	-	-
+ re-ranking	89.1	-	-	-	83.9	-	-	-	-	-
Chen et al. 2017 [79] (MobileNet)	90.0	-	-	-	70.6	-	-	-	-	-
Chen et al. 2017 [79] (Inception-V3)	88.6	-	-	-	72.6	-	-	-	-	-
Ustinova et al. 2017 [56] (Bilinear)	66.4	85.0	90.2	-	41.2	-	-	-	-	-
Wei et al. 2017 [80] (Pose)	89.9	-	-	-	73.9	-	-	-	-	-
Zheng et al. 2017 [15] (Pose)	79.3	90.8	94.4	96.5	56.0	-	-	-	-	-
Zhao et al. 2017 [16] (Pose)	76.9	91.5	94.6	96.7	-	-	-	-	-	-
Su et al. 2017 [14] (Pose)	84.1	92.7	94.9	96.8	65.4	-	-	-	-	-
Proposed (Inception-V1, R-CPM)	88.8	95.6	97.3	98.6	74.5	92.9	97.3	98.4	99.1	81.7
Proposed (Inception-V1, OpenPose)	**90.2**	**96.1**	**97.4**	**98.4**	**76.0**	**93.2**	**97.5**	**98.4**	**99.1**	**82.7**
+ dilation	**91.7**	**96.9**	**98.1**	**98.9**	**79.6**	**94.0**	**98.0**	**98.8**	**99.3**	**85.2**
+ re-ranking	**93.4**	**96.4**	**97.4**	**98.2**	**89.9**	**95.4**	**97.5**	**98.2**	**98.9**	**93.1**

Market-1501 + 500K

	Gallery size				
	metric	19732	119732	219732	519732
Zheng et al. 2017 [81]	rank-1	79.5	73.8	71.5	68.3
	mAP	59.9	52.3	49.1	45.2
Linet al. 2017 [70]	rank-1	84.0	79.9	78.2	75.4
	mAP	62.8	56.5	53.6	49.8
Hermans et al. 2017 [72]	rank-1	84.9	79.7	77.9	74.7
	mAP	69.1	61.9	58.7	53.6
Proposed (Inception V1, OpenPose)	rank-1	**91.7**	**88.3**	**86.6**	**84.1**
	mAP	**79.6**	**74.2**	**71.5**	**67.2**

CUHK

	CUHK03								CUHK01							
	Detected				Manual				100 test IDs				486 test IDs			
Rank	1	5	10	20	1	5	10	20	1	5	10	20	1	5	10	20
Shi et al. [11]	52.1	84.0	92.0	96.8	61.3	88.5	96.0	99.0	69.4	90.8	96.0	-	-	-	-	-
SIR-CIR [51]	52.2	-	-	-	-	-	-	-	71.8	91.6	96.0	98.0	-	-	-	-
Varior et al. [52]	68.1	88.1	94.6	98.8	-	-	-	-	-	-	-	-	-	-	-	-
Bai et al. [75]	72.7	92.4	96.1	-	76.6	94.6	98.0	-	-	-	-	-	-	-	-	-
Zhang et al. [10]	-	-	-	-	80.2	97.7	99.2	99.8	89.6	97.8	98.9	99.7	76.5	94.2	**97.5**	-
Sun et al. [73]	81.8	95.2	97.2	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhao et al. [18]	81.6	97.3	98.4	**99.5**	85.4	97.6	99.4	**99.9**	88.5	**98.4**	**99.6**	**99.9**	74.7	92.6	96.2	98.4
Geng et al. [74]	84.1	-	-	-	85.4	-	-	-	**93.2**	-	-	-	77.0	-	-	-
Chen et al. [76]	87.5	97.4	98.7	**99.5**	-	-	-	-	-	-	-	-	74.5	91.2	94.8	97.1
Ustinova et al. [56] (Bilinear)	63.7	89.2	94.7	97.5	69.7	93.4	98.9	99.4	-	-	-	-	52.9	78.1	86.3	92.6
Wei et al. 2017 [80] (Pose)	82.2	95.8	97.6	98.7	85.0	97.9	99.1	99.6	-	-	-	-	-	-	-	-
Zheng et al. [15] (Pose)	67.1	92.2	96.6	98.1	-	-	-	-	-	-	-	-	-	-	-	-
Zhao et al. [16] (Pose)	-	-	-	-	88.5	97.8	98.6	99.2	-	-	-	-	79.9	**94.4**	97.1	98.6
Su et al. [14] (Pose)	78.3	94.8	97.2	98.4	88.7	98.6	99.2	99.7	-	-	-	-	-	-	-	-
Proposed	**88.0**	**97.6**	**98.6**	**99.0**	**91.5**	**99.0**	**99.5**	**99.9**	90.4	97.1	98.1	98.9	**80.7**	**94.4**	**97.3**	**98.6**

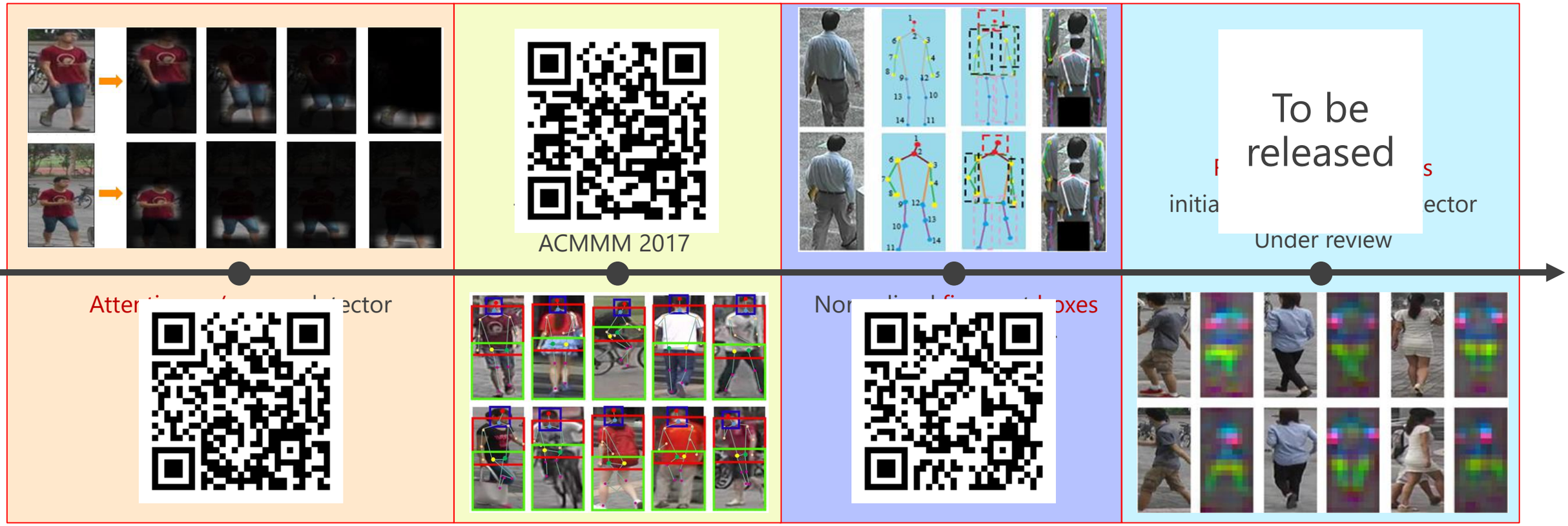
DukeMTMC

Rank	1	5	10	20	mAP
Zheng et al. [82]	67.7	-	-	-	47.1
Tong et al. [83]	68.1	-	-	-	-
Lin et al. [70]	70.7	-	-	-	51.9
Schumann et al. [84]	72.6	-	-	-	52.0
Sun et al. [73]	76.7	86.4	89.9	-	56.8
Chen et al. [79] (MobileNet)	77.6	-	-	-	58.6
Chen et al. [79] (Inception-V3)	79.2	-	-	-	60.6
Zhun et al. [78]	79.3	-	-	-	62.4
+ re-ranking	84.0	-	-	-	78.3
Proposed (Inception V1, OpenPose)	82.1	90.2	92.7	95.0	64.2
+ dilation	**84.4**	**92.2**	**93.8**	**95.7**	**69.3**
+ re-ranking	**88.3**	**93.1**	**95.0**	**96.1**	**83.9**

Video: Mars

Rank	1	5	10	20	mAP
Xu et al. [85] (Video)	44	70	74	81	-
McLaughlin et al. [86] (Video)	45	65	71	78	27.9
Zheng et al. [4] (Video)	68.3	82.6	-	89.4	49.3
Liu et al. [87] (Video)	68.3	81.4	-	90.6	52.9
Zhou et al. [88]	70.6	90.0	-	97.6	50.7
Li et al. [17]	71.8	86.6	-	93.1	56.1
+ re-ranking	83.0	93.7	-	97.6	66.4
Liu et al. [89]	73.7	84.9	-	91.6	51.7
Hermans et al. [72]	79.8	91.4	-	-	67.7
+ re-ranking	81.2	90.8	-	-	77.4
Proposed (Inception V1, OpenPose)	83.0	92.8	95	96.8	72.2
+ dilation	**84.7**	**94.4**	**96.3**	**97.5**	**75.9**
+ re-ranking	**85.1**	**94.2**	**96.1**	**97.4**	**83.9**

Part-Aligned Representation Learning



Thanks!
QA