Pedestrian Alignment Network for Large-scale Person Re-Identification

Zhedong Zheng[1], Liang Zheng[2] and Yi Yang[1]

[1] University of Technology Sydney, Australia

[2] Australian National University, Australia

T-CSVT vol. 29(10), pp. 3037 - 3045. 2019

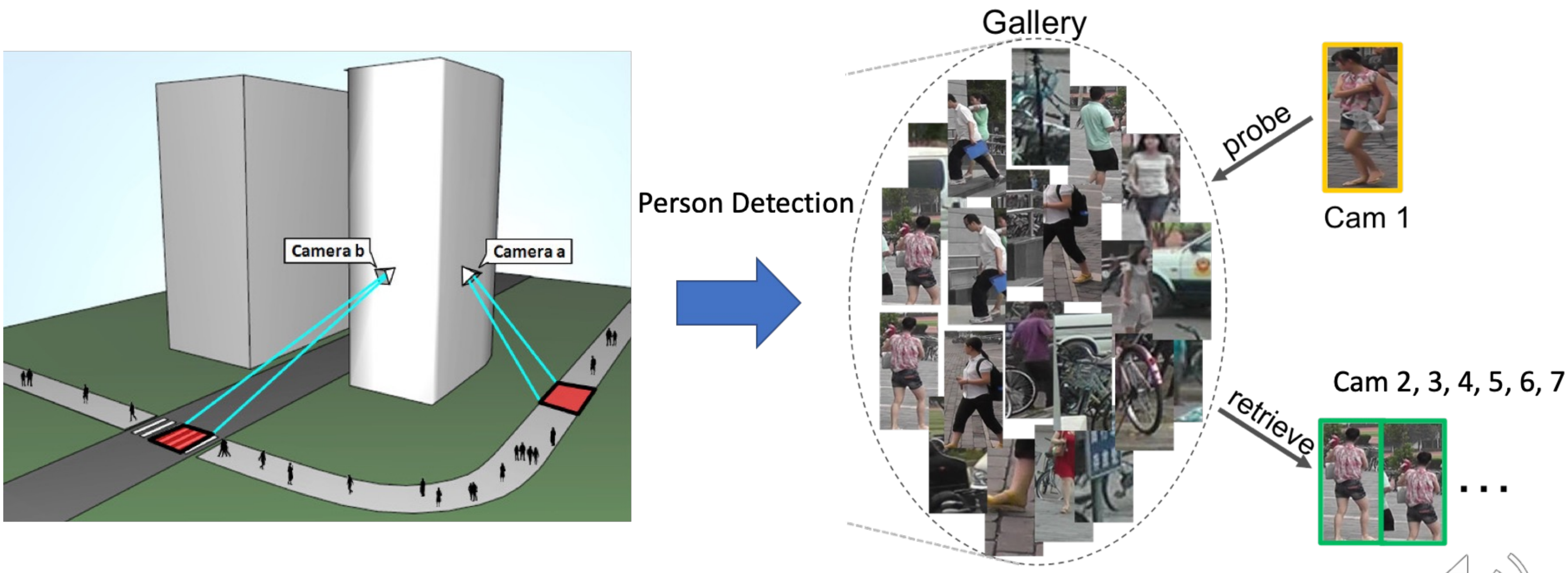
T-CSVT Youtube Channel: https://www.youtube.com/channel/UC46cVfjkp7XLdDXsgGuG_Lg

Outlines

- ❑ Problem Background
- ❑ Existing Methods
- ❑ Challenges
- ❑ Our Method & Contributions
- ❑ Experiments
- ❑ Conclusion and Academic Impacts

Problem Background

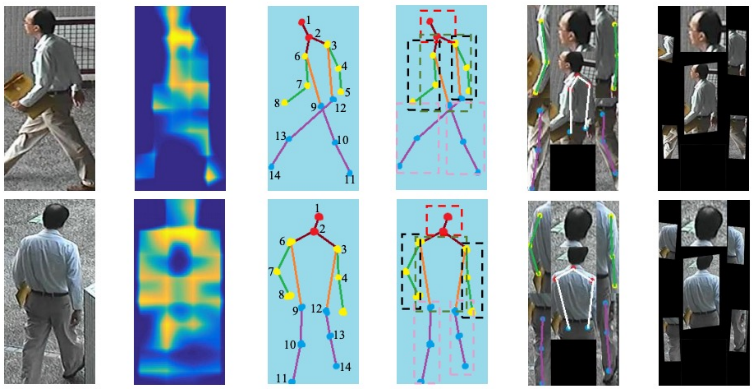
- ❑ Person re-identification (re-id) is to find the person of interest from different camera views.
- ❑ The emergence of this task can be attributed to 1) the increasing demand of public safety and 2) the widespread large camera networks in public space.



Existing Methods

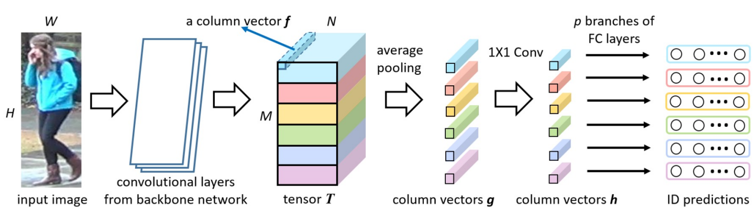
❑ Structure Information

- Part Alignment in the Pixel Level



Pose-driven CNN(Su et al., CVPR2017)

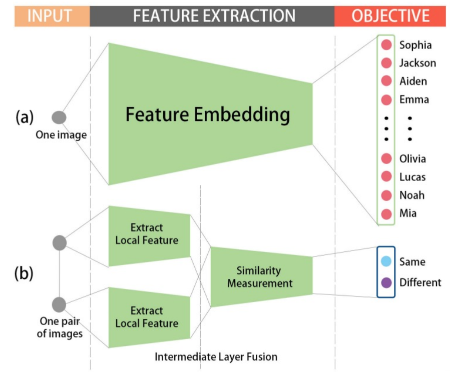
- Part Alignment in the Feature Level



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

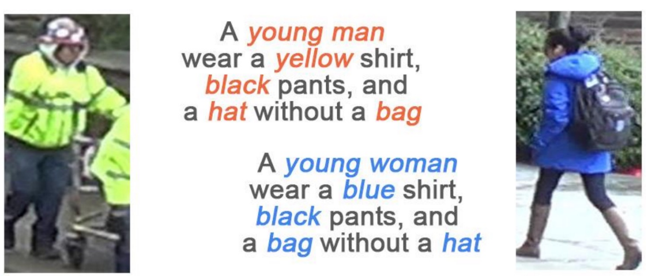
❑ Semantic Information

- Discriminative Losses



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

- Middle-level Features



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

Challenges

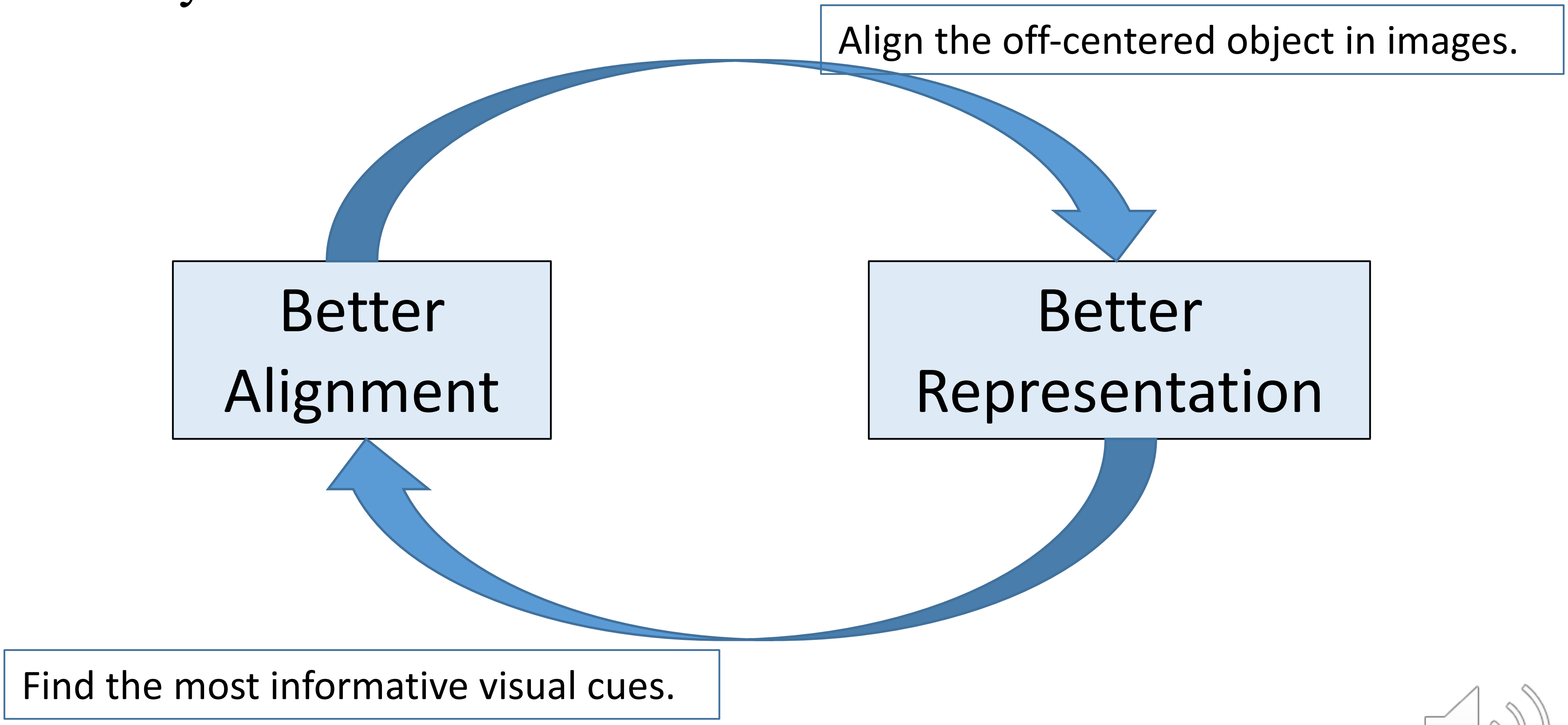
- However, one problem remains. In practice, person re-ID usually adopts automatic detectors to obtain cropped pedestrian images.

Is Input Pedestrian Image Good for Training?

Excessive Background **Part Missing**

Contribution

- We propose the pedestrian alignment network (PAN), which simultaneously aligns pedestrians within images and learns pedestrian descriptors. It addresses the misdetection problem and representation learning together, and improves the person re-ID accuracy without extra annotations.



Our Methods

❑ Architecture of the pedestrian alignment network (PAN)

➢ **Base Branch**

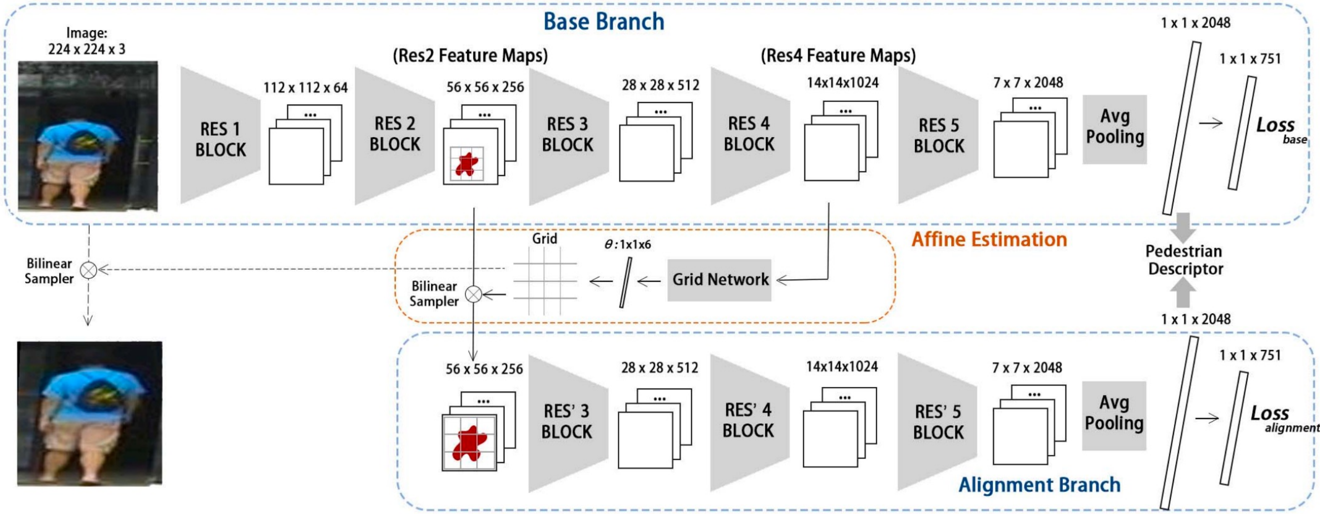
- Original Image → Identity Label

➢ **Affine Estimation**

- Feature Maps → Affine Grid

➢ **Alignment Branch**

- Aligned Features → Identity Label



Our Methods

❑ Architecture of the pedestrian alignment network (PAN)

➢ **Base Branch**

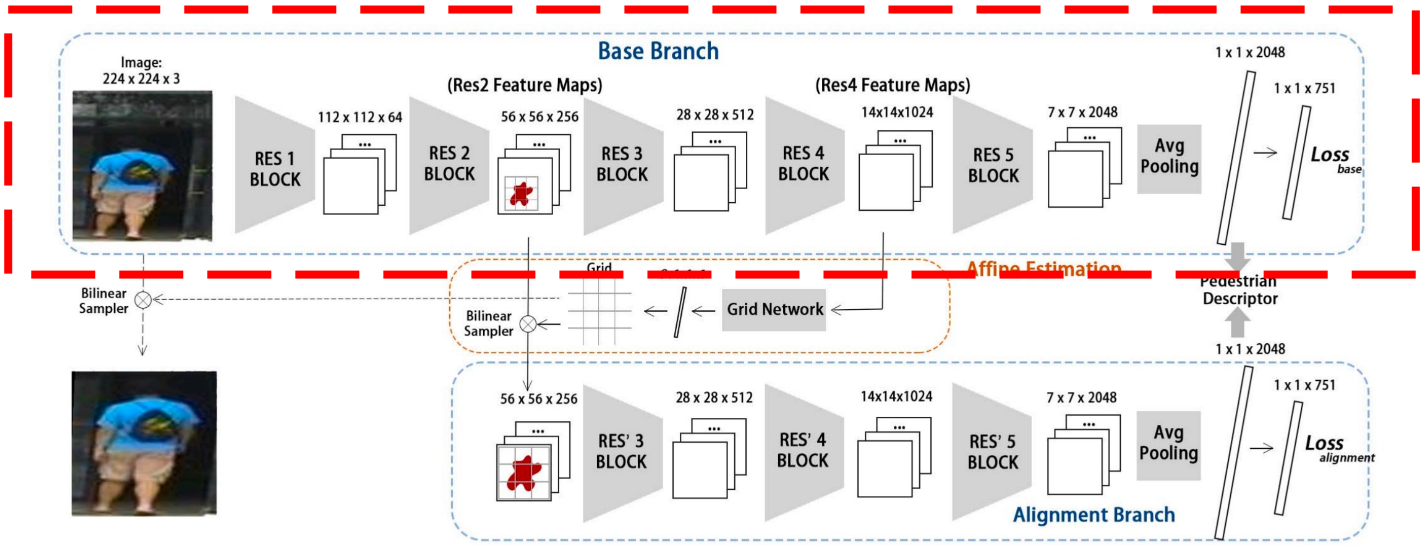
- Original Image → Identity Label

➢ **Affine Estimation**

- Feature Maps → Affine Grid

➢ **Alignment Branch**

- Aligned Features → Identity Label



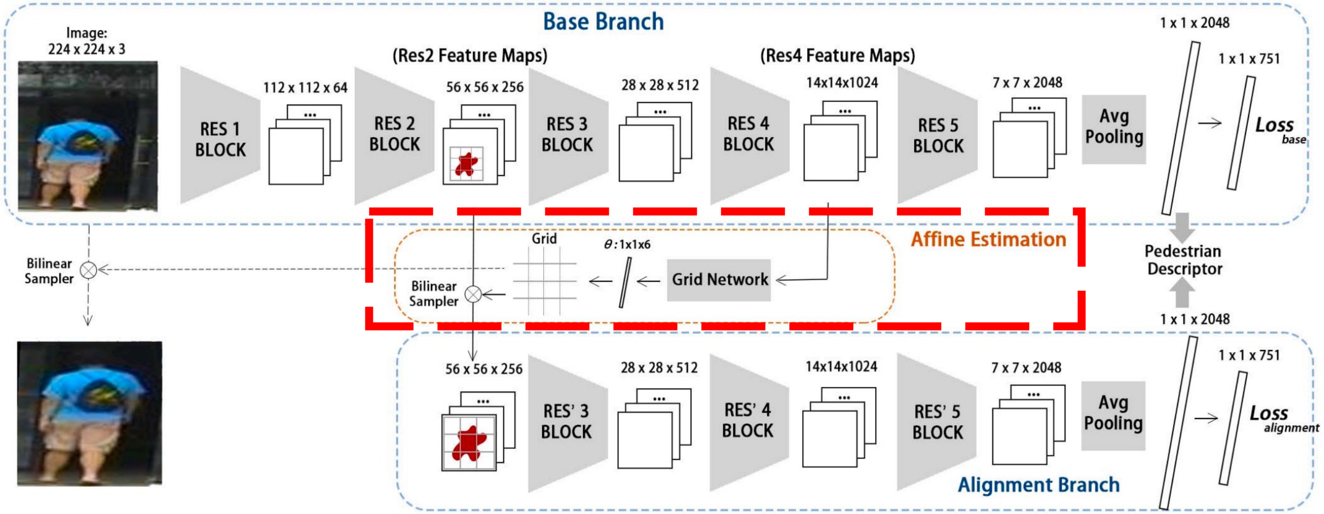
Our Methods

❑ Architecture of the pedestrian alignment network (PAN)

➢ **Base Branch**
- Original Image → Identity Label

➢ **Affine Estimation**
- Feature Maps → Affine Grid

➢ **Alignment Branch**
- Aligned Features → Identity Label



Our Methods

❑ Architecture of the pedestrian alignment network (PAN)

➢ **Base Branch**

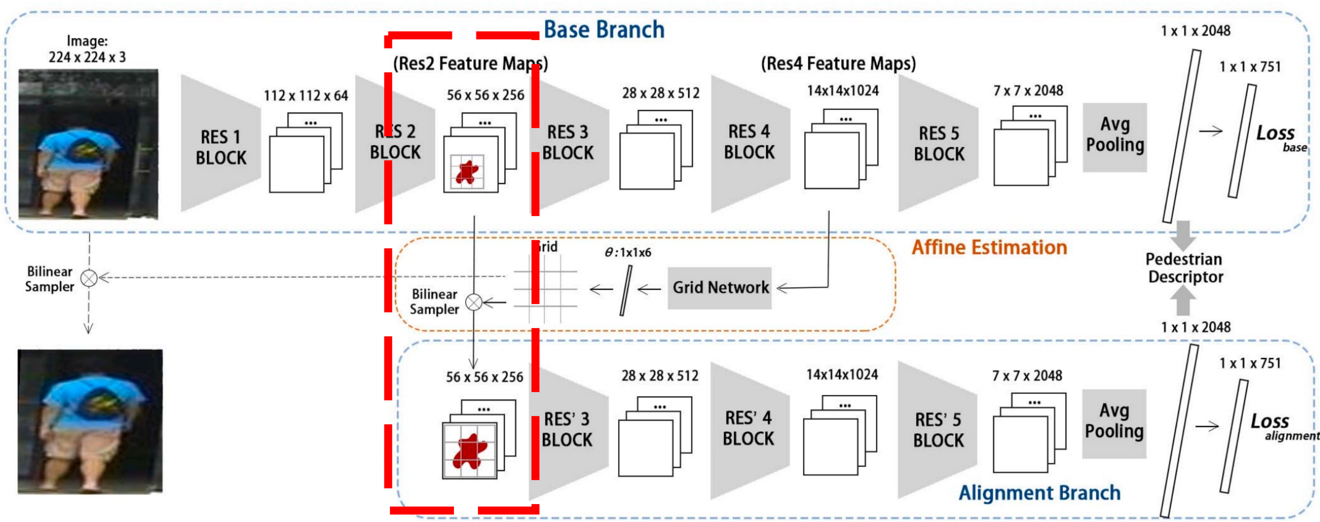
- Original Image → Identity Label

➢ **Affine Estimation**

- Feature Maps → Affine Grid

➢ **Alignment Branch**

- Aligned Features → Identity Label



Our Methods

❑ Architecture of the pedestrian alignment network (PAN)

➢ **Base Branch**

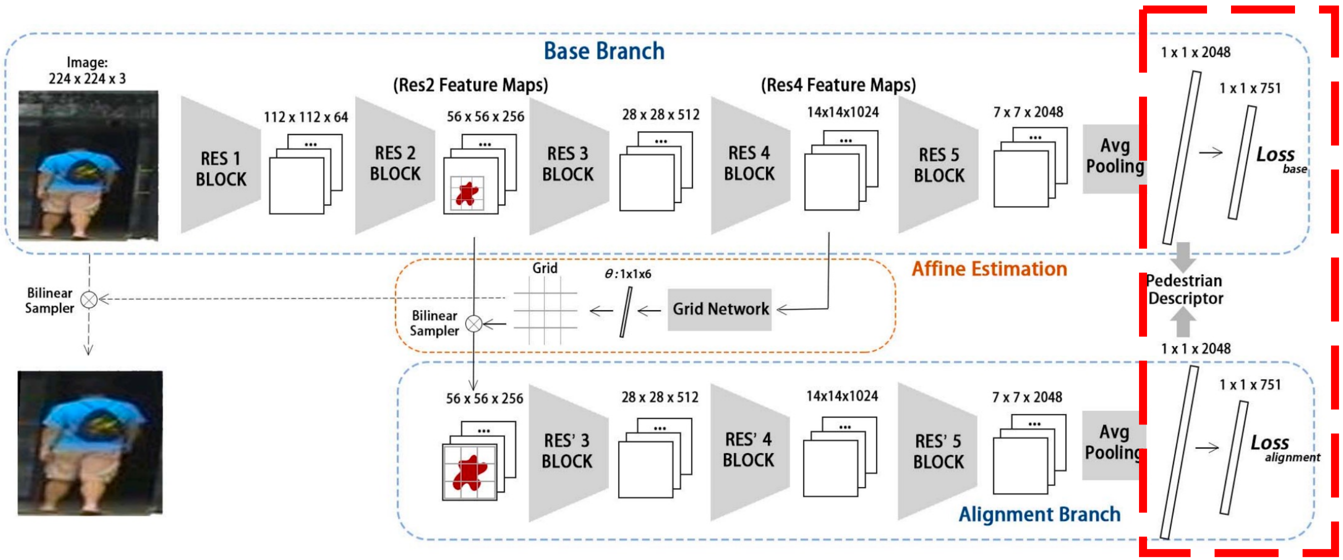
- Original Image → Identity Label

➢ **Affine Estimation**

- Feature Maps → Affine Grid

➢ **Alignment Branch**

- Aligned Features → Identity Label



Quantitative Experiments

- Datasets:
  - Market-1501
  - CUHK03-NP
  - DukeMTMC-reID
- Comparisons

TABLE III

RANK@1 ACCURACY (%) AND mAP (%) ON MARKET-1501. - THE RESPECTIVE PAPERS USE HAND-CRAFTED FEATURE, * THE RESPECTIVE PAPERS USE THEIR OWN SPECIFIC NETWORK

Methods	Backbone	Single Query Rank@1	Single Query mAP	Multi. Query Rank@1	Multi. Query mAP
BoW+kissme [9]	-	44.42	20.76	-	-
MST CNN [63]	*	45.1	-	55.4	-
FisherNet [64]	*	48.15	29.94	-	-
SL [65]	-	51.90	26.35	-	-
S-LSTM [33]	*	-	-	61.6	35.3
DNS [66]	-	55.43	29.87	71.56	46.03
Gated Reid [34]	*	65.88	39.55	76.04	48.45
CADL [67]	*	73.84	47.11	80.85	55.58
SOMAnet [68]	*	73.87	47.89	81.29	56.98
PIE [53]	Res50	78.65	53.87	-	-
Verif.-Identif. [44]	Res50	79.51	59.87	85.84	70.33
DCF [69]	*	80.31	57.53	86.79	66.70
DPR [70]	GoogLe	81.0	63.4	-	-
SSM [71]	Res50	82.21	68.80	88.18	76.18
SVDNet [72]	Res50	82.3	62.1	-	-
GAN+re-rank [47]	Res50	83.97	66.07	88.42	76.10
Basel.	Res50	80.17	59.14	87.41	72.05
Ours	Res50	82.81	63.35	88.18	71.72
Ours+re-rank	Res50	85.78	76.56	89.79	83.79
Ours (GAN)	Res50	86.67	69.33	90.88	76.32
Ours (GAN)+re-rank	Res50	**88.57**	**81.53**	**91.45**	**87.44**

TABLE IV

RANK@1 ACCURACY (%) AND mAP (%) ON CUHK03-NP. WE EVALUATE THE PROPOSED METHOD ON THE "DETECTED" AND "LABELED" SUBSETS ACCORDING TO THE NEW PROTOCOL IN [19]. - THE RESPECTIVE PAPERS USE HAND-CRAFTED FEATURE

Methods	Back bone	Detected Rank@1	Detected mAP	Labeled Rank@1	Labeled mAP
BOW+XQDA [9]	-	6.36	6.39	7.93	7.29
IDE [1]	Res50	21.3	19.7	22.2	21.0
IDE +DaF [74]	Res50	26.4	30.0	27.5	31.5
IDE+XQDA [19]	Res50	31.1	28.2	32.0	29.6
IDE+XQDA+re-rank [19]	Res50	34.7	37.4	38.1	40.3
Basel.	Res50	30.5	29.0	31.1	29.8
Ours	Res50	36.3	34.0	36.9	35.0
Ours+re-rank	Res50	**41.9**	**43.8**	**43.9**	**45.8**

TABLE V

RANK@1 ACCURACY (%) AND mAP (%) ON DukeMTMC-reID. WE FOLLOW THE EVALUATION PROTOCOL IN [47]. - THE RESPECTIVE PAPERS USE HAND-CRAFTED FEATURE

Methods	Backbone	Rank@1	mAP
BoW+kissme [9]	-	25.13	12.17
LOMO+XQDA [16]	-	30.75	17.04
Gan [47]	Res50	67.68	47.13
Verif.-Identif. [44]	Res50	68.9	49.3
APR [43]	Res50	70.69	51.88
Basel. [47]	Res50	65.22	44.99
Ours	Res50	71.59	51.51
Ours+re-rank	Res50	**75.94**	**66.74**

Ablation Studies

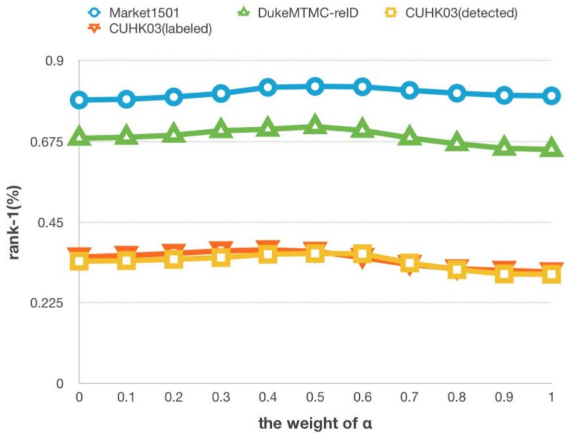
❑ Base Branch vs. Alignment Branch

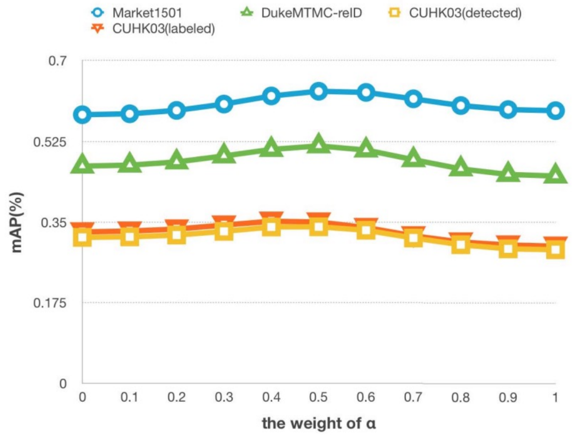
- Alignment Branch works well on the dataset with much detection noise.
- Base Branch and Alignment Branch are complementary.

Methods	dim	*Market-1501*				*CUHK03-NP (detected)*				*CUHK03-NP (labeled)*				*DukeMTMC-reID*			
		1	5	20	mAP	1	5	20	mAP	1	5	20	mAP	1	5	20	mAP
Base	2,048	80.17	91.69	96.59	59.14	30.50	51.07	71.64	29.04	31.14	52.00	74.21	29.80	65.22	79.13	87.75	44.99
Alignment	2,048	79.01	90.86	96.14	58.27	34.14	54.50	72.71	31.71	35.29	53.64	72.43	32.90	68.36	81.37	88.64	47.14
PAN	4,096	82.81	93.53	97.06	63.35	36.29	55.50	75.07	34.00	36.86	56.86	75.14	35.03	71.59	83.89	90.62	51.51

❑ Hyper-parameter

- alpha selection

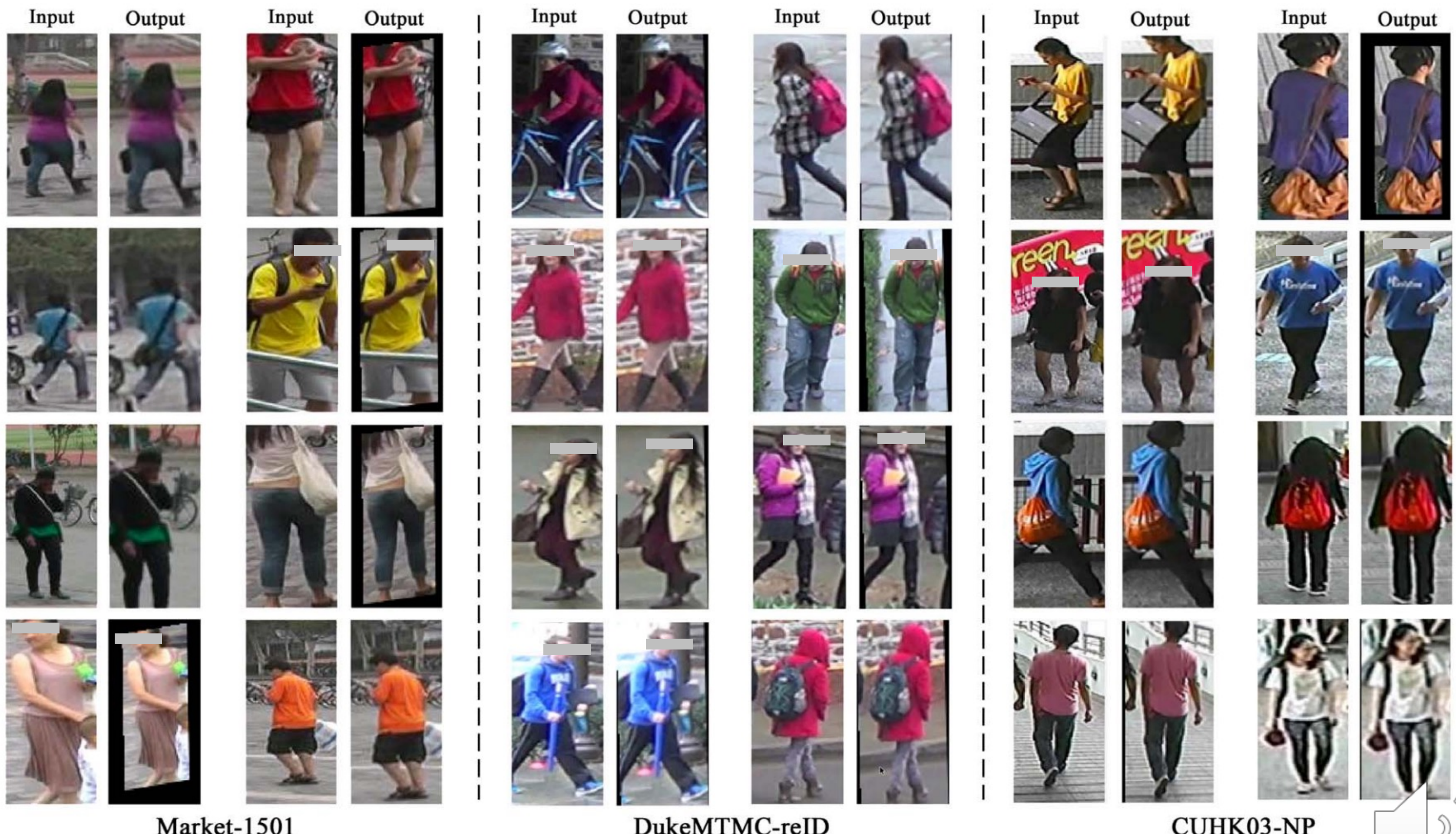




Qualitative Experiments

❑ Alignment Results:

- No extra location annotations are used in our work.



Conclusion and Academic Impacts

- Upon publishing (2018, early access), the paper reported ~5% improvement in rank-1 accuracy over the state-of-the-art methods on three benchmarks. After that, the proposed system has become a must-be-compared one in the community, and the idea of end-to-end alignment has been adopted in the highest-performing object re-identification systems. This paper has been cited for ~290 times according to Google Scholar.
- Besides, the winning solution of the AICity Challenge (in conjunction with CVPR 2020) used a mostly identical module as this paper, which outperformed the second-best team by a large margin (6% in mean average precision).

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	0.6899
5	30	BestImage	0.6684
6	44	BeBetter	0.6683
7	72	UMD_RC	0.6668
8	7	Ainnovation	0.6561
9	46	NMB	[illegible]
10	81	Shahe	0.6191

Zheng Z, Jiang M, Wang Z, et al. Going beyond real data: A robust visual representation for vehicle re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 598-599.

Thanks a lot for your attention.

Our code is available at

https://github.com/layumi/Pedestrian_Alignment