Pooling Body Parts on Feature Maps for Misalignment Robust Person Re-Identification

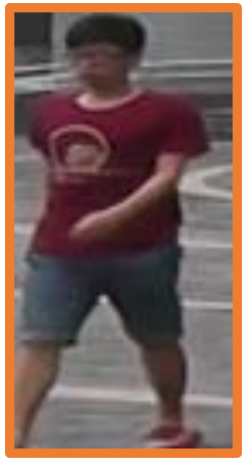
Yuting Liu, Qijun Zhao, Zhihong Wu

College of Computer Science, Sichuan University, Chengdu, China

ISBA 2018 January 12, 2018

Person Re-identification

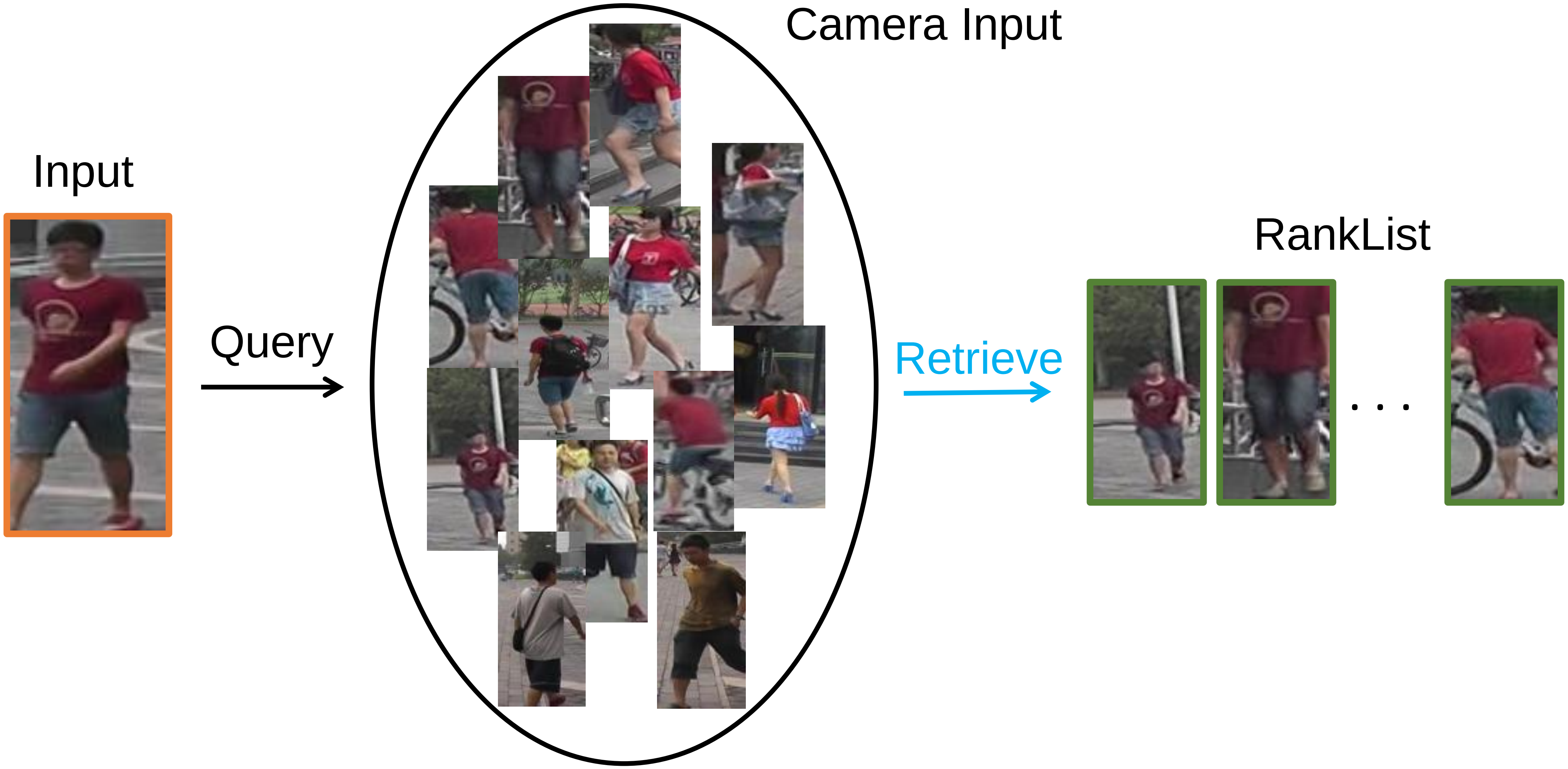
Input



Query

Camera Input

Person Re-identification



Classical Method

Query

Feature Extraction

(HSV, SILTP, …
Deep Features)

Feature Matching

(Euclidean,
KISSME,
XQDA,…)

Feature Extraction

(HSV, SILTP,…
Deep Features)

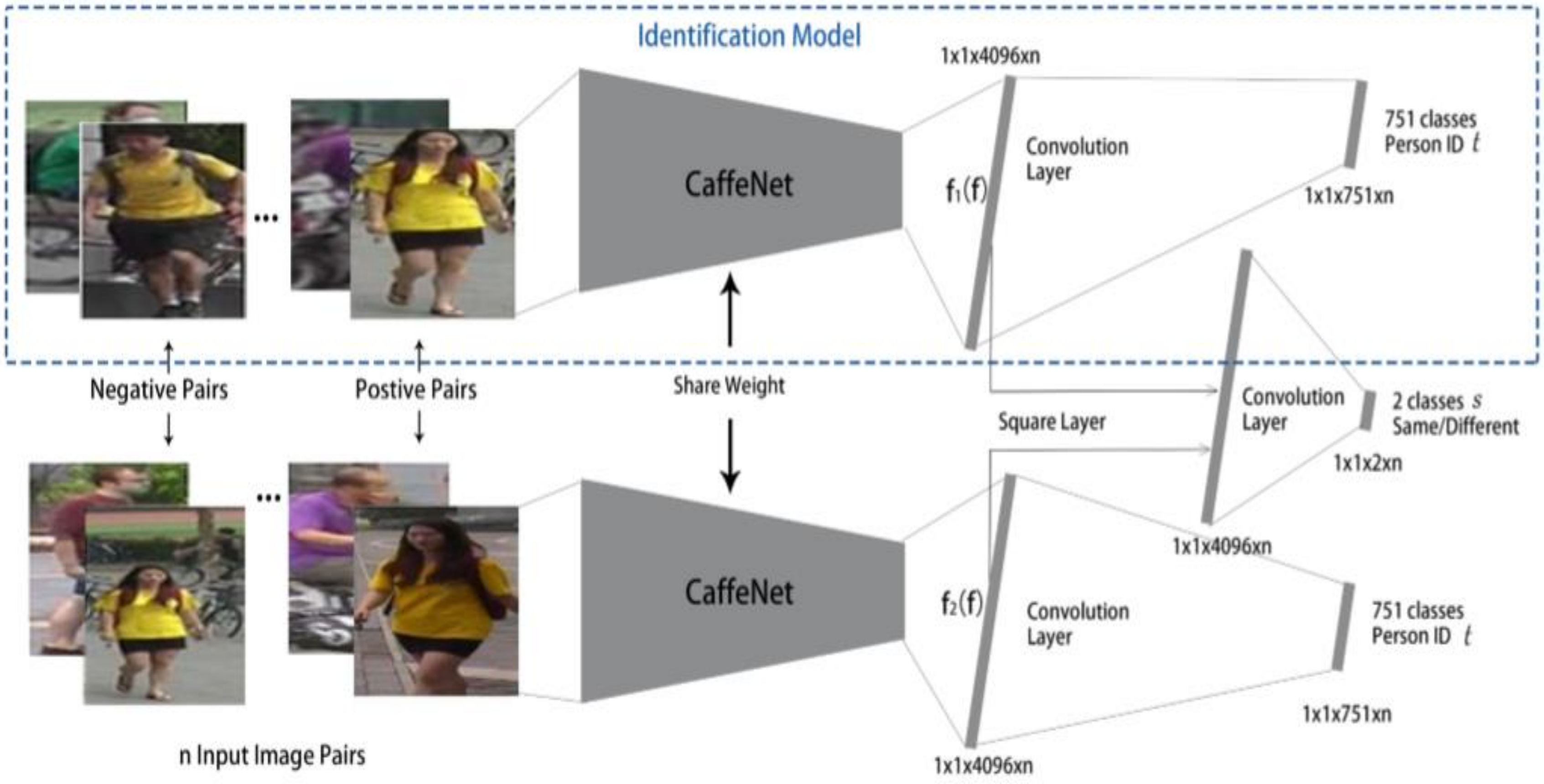
Gallery

Misalignment

Feature Learning

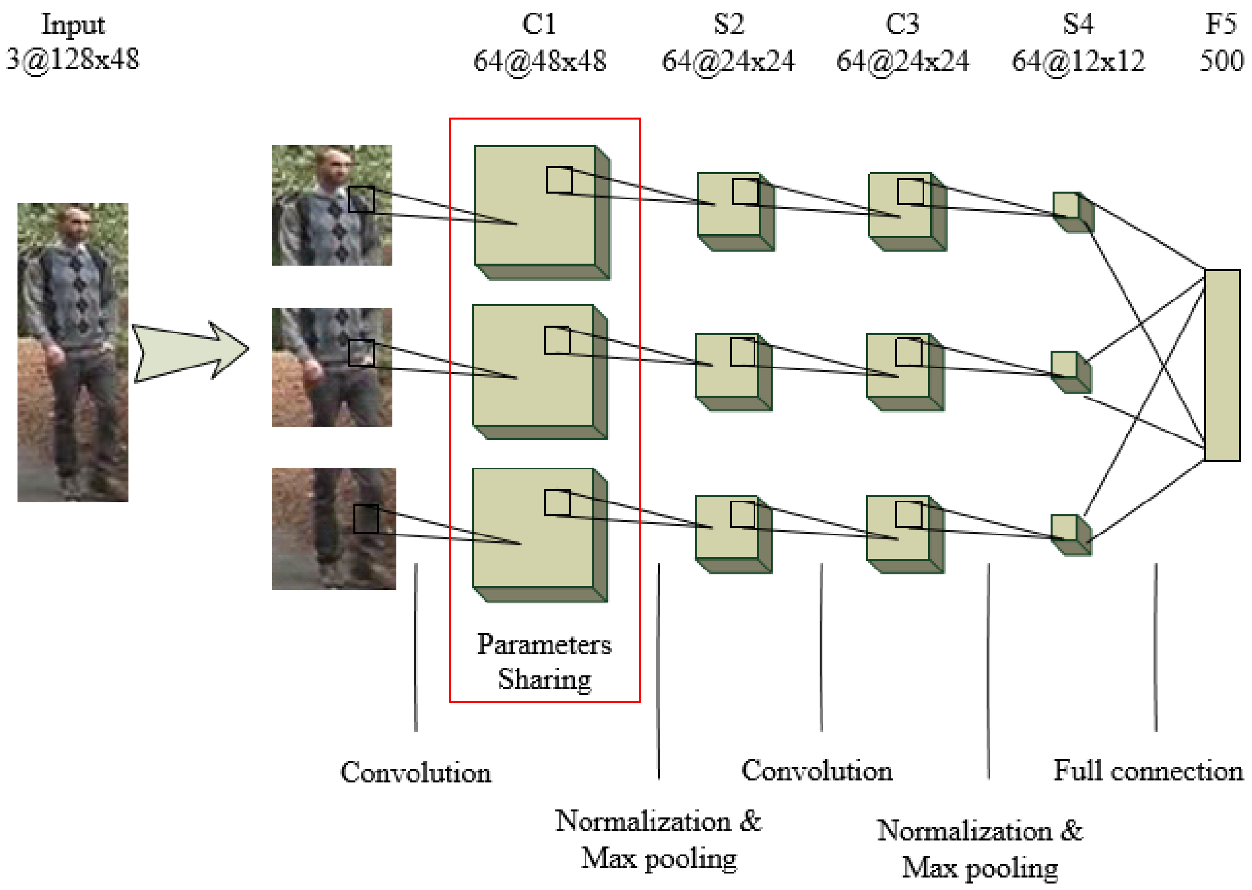
Feature Matching

Existing DL Works



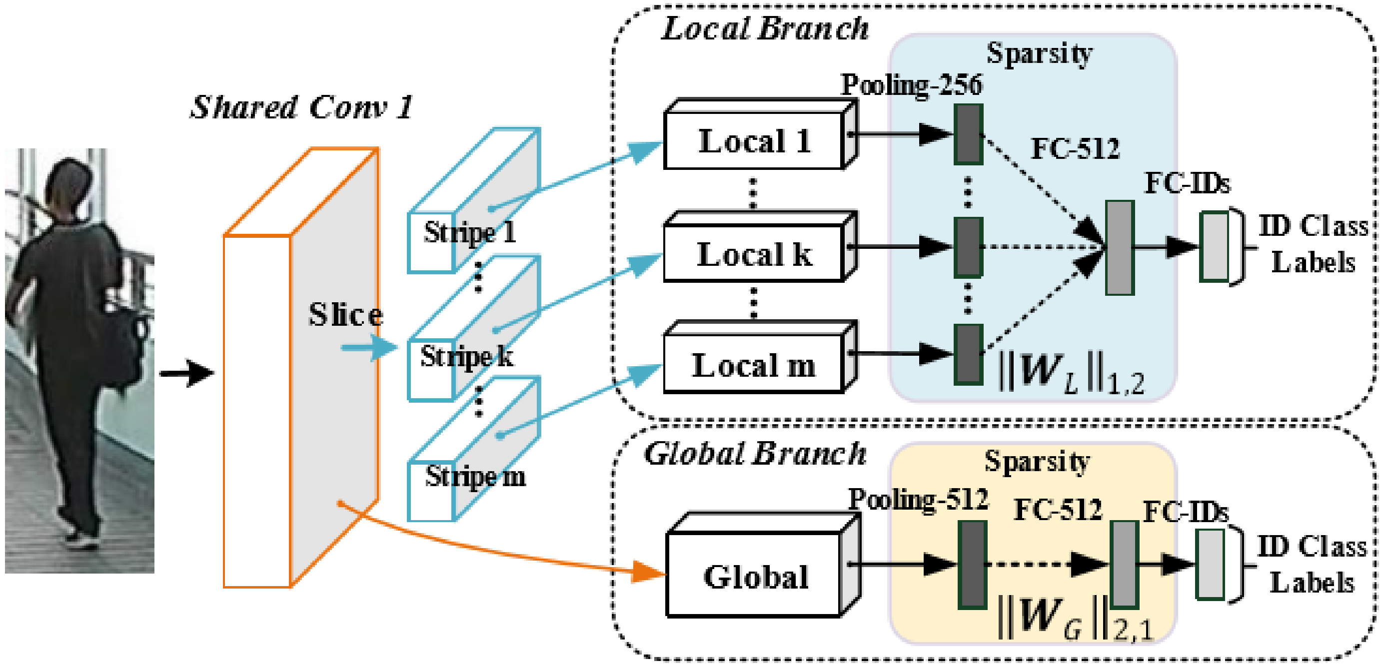
Z. Zheng, L. Zheng, Y. Yang. A Discriminatively Learned CNN Embedding for Person Reidentification. ACM Transactions on Multimedia Computing, Communications, and Applications, 2016.

Existing DL Works



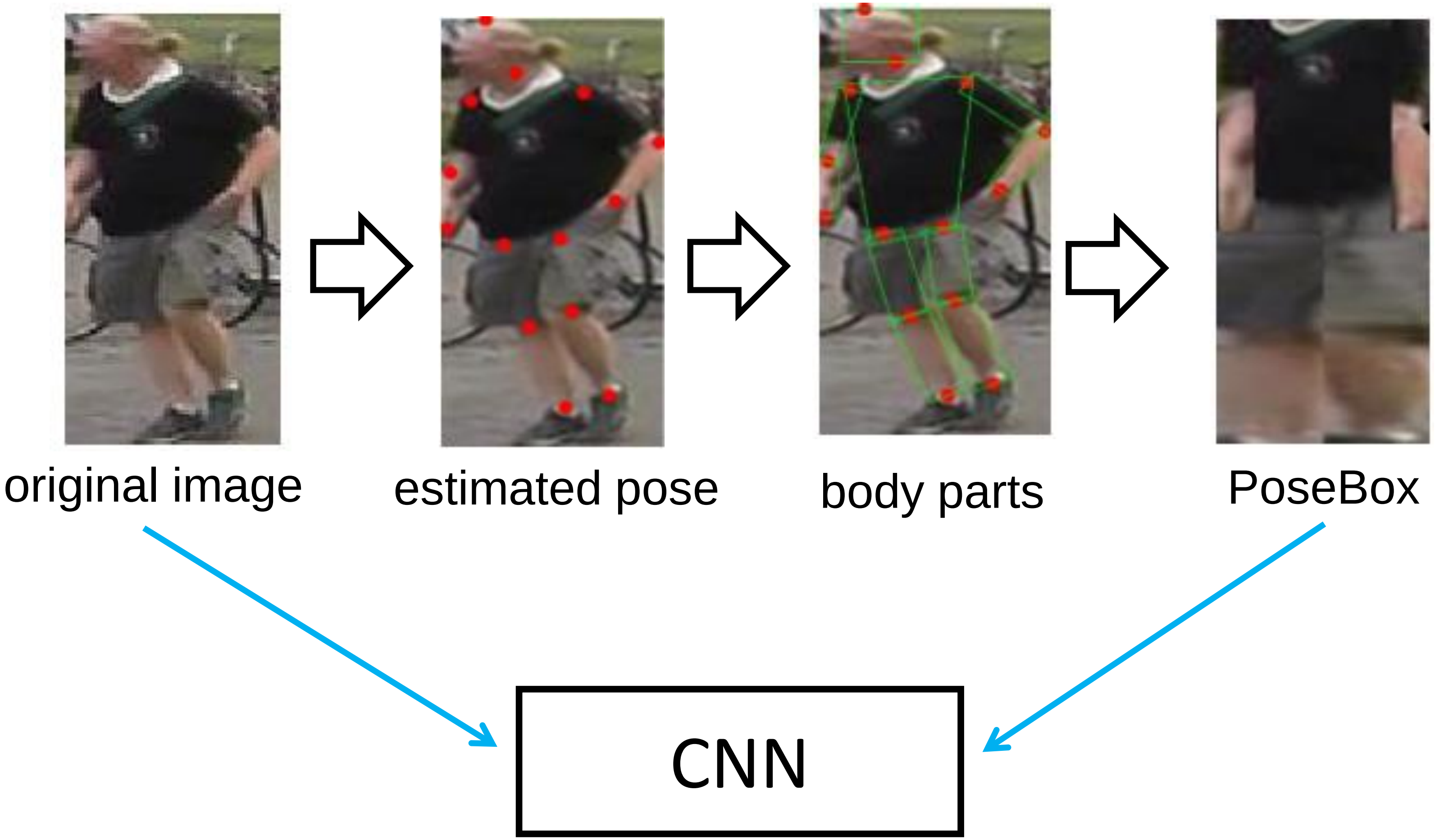
D. Yi, Z. Lei, S. Liao, S. Z. Li. Deep Metric Learning for Person Re-identification. In *ICPR, 2014.*

Existing DL Works



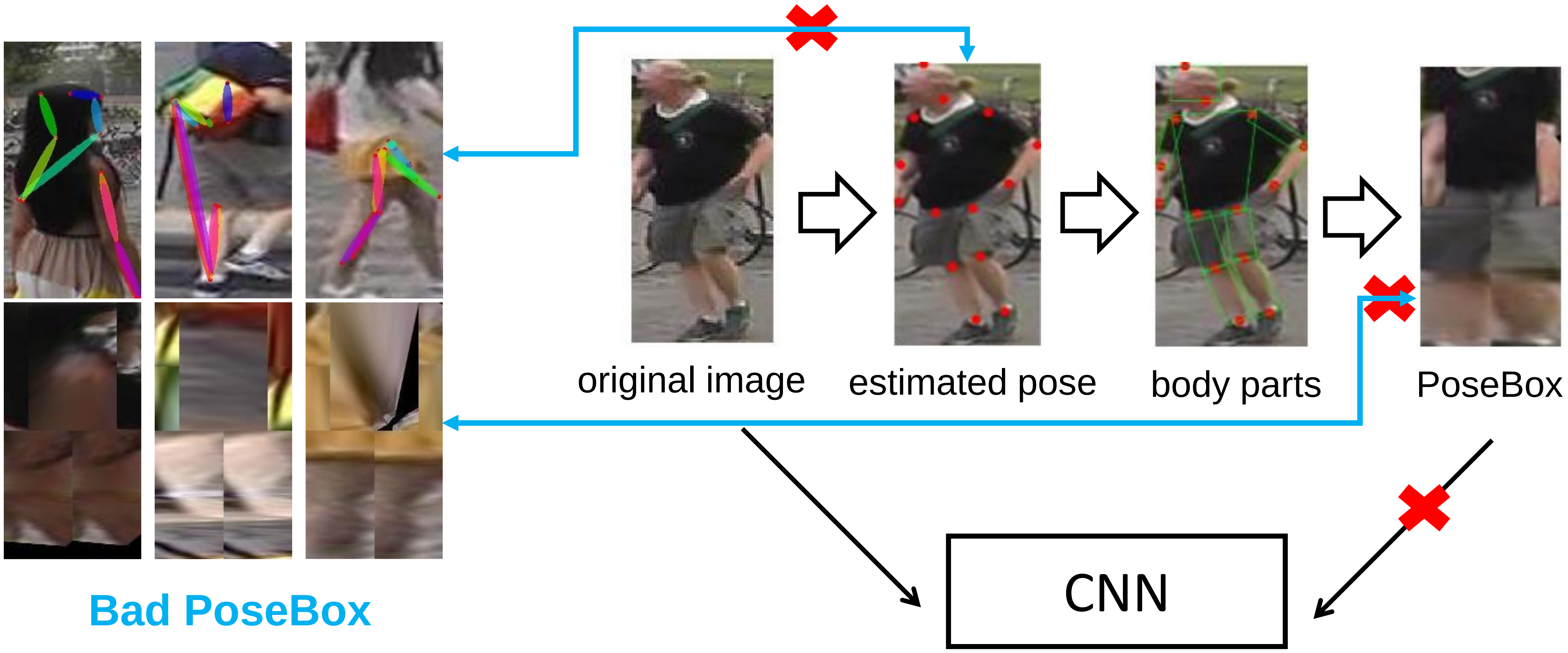
W. Li,X. Zhu,S. Gong. Person Re-Identification by Deep Joint Learning of Multi-Loss Classification. In IJCAI, 2017

Pose Invariant Embedding(PIE)



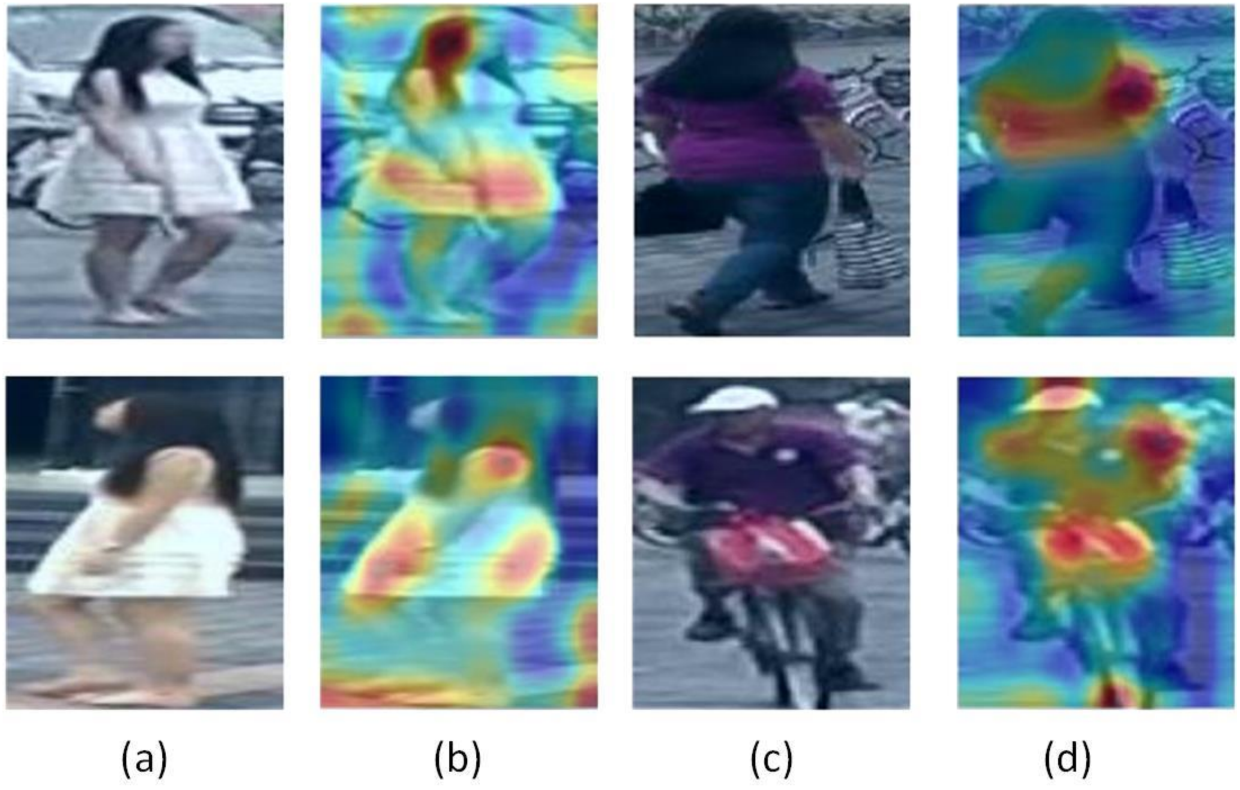
L. Zheng, Y Huang, H. Lu, and Y. Yang. Pose Invariant Embedding for Deep Person Re-identification. ArXiv preprint arXiv:1701.07732, 2017.

Pose Invariant Embedding(PIE)



Main Idea

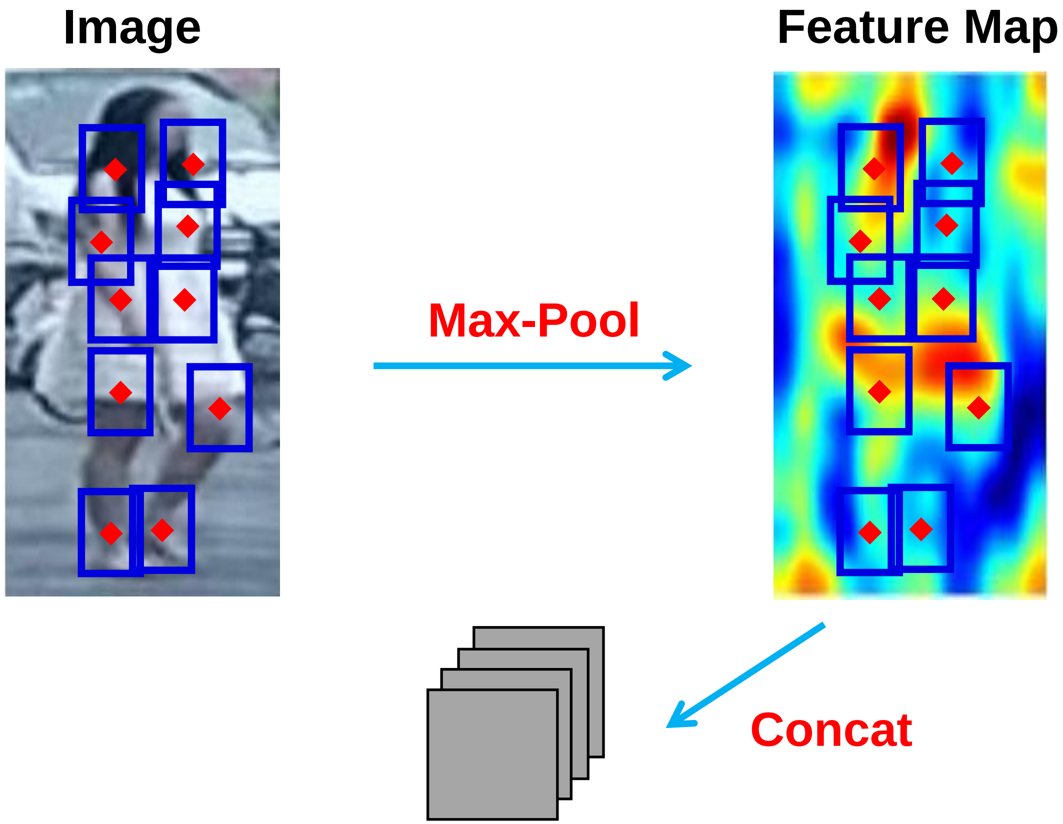
- Salient Local Features



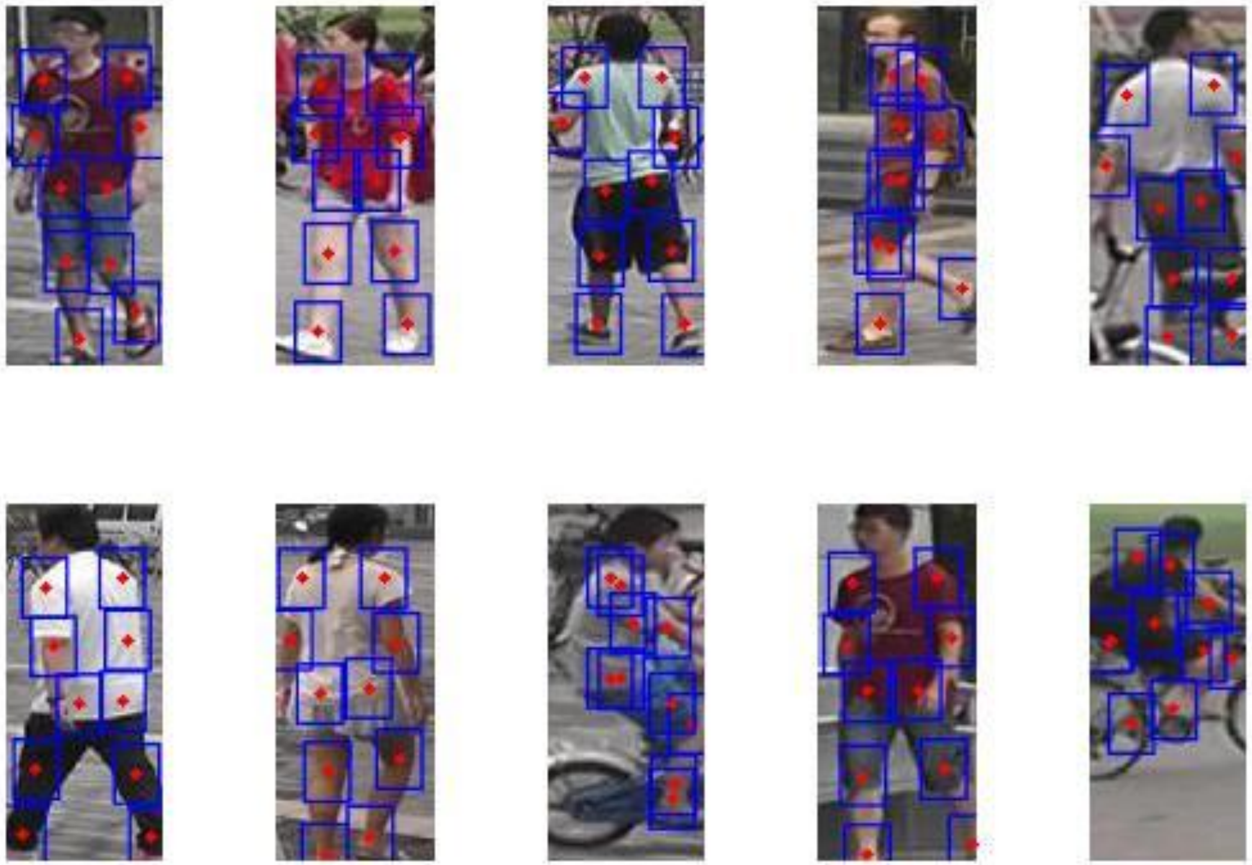
Main Idea

- Pooling Body Parts on Feature Maps

Image Level ~~Affine~~ Transformation



Body Part Segmentation

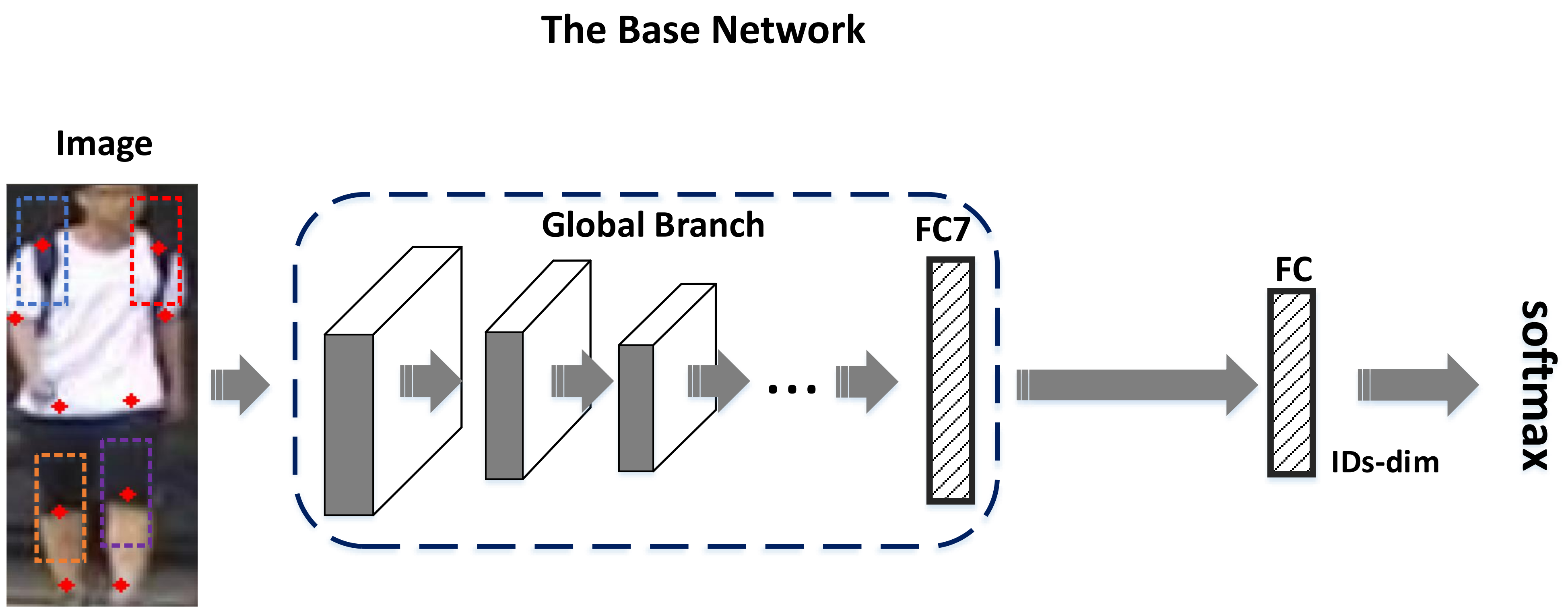


By Convolution Pose Mechine (CPM) Method

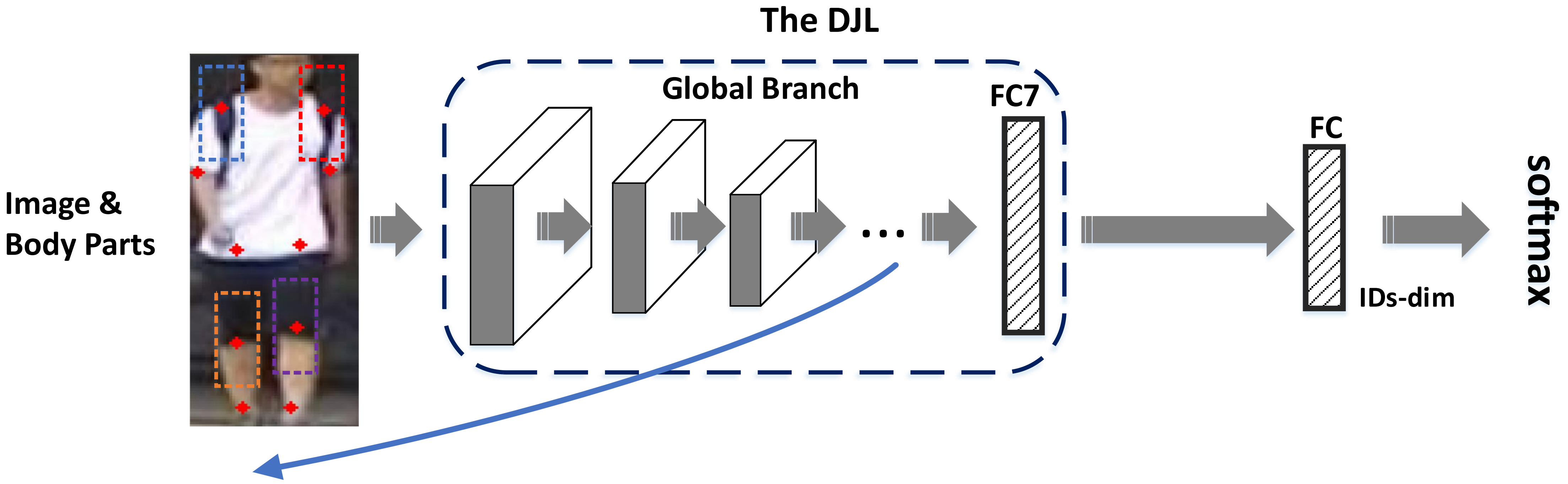
10 Body Parts

- Left Shoulders
- Right Shoulders
- Left Elbows
- Right Elbows
- Left Wrists
- Right Wrists
- Left Hips
- Right Hips
- Left Knees
- Right Knees
- Left Ankles
- Right Ankles

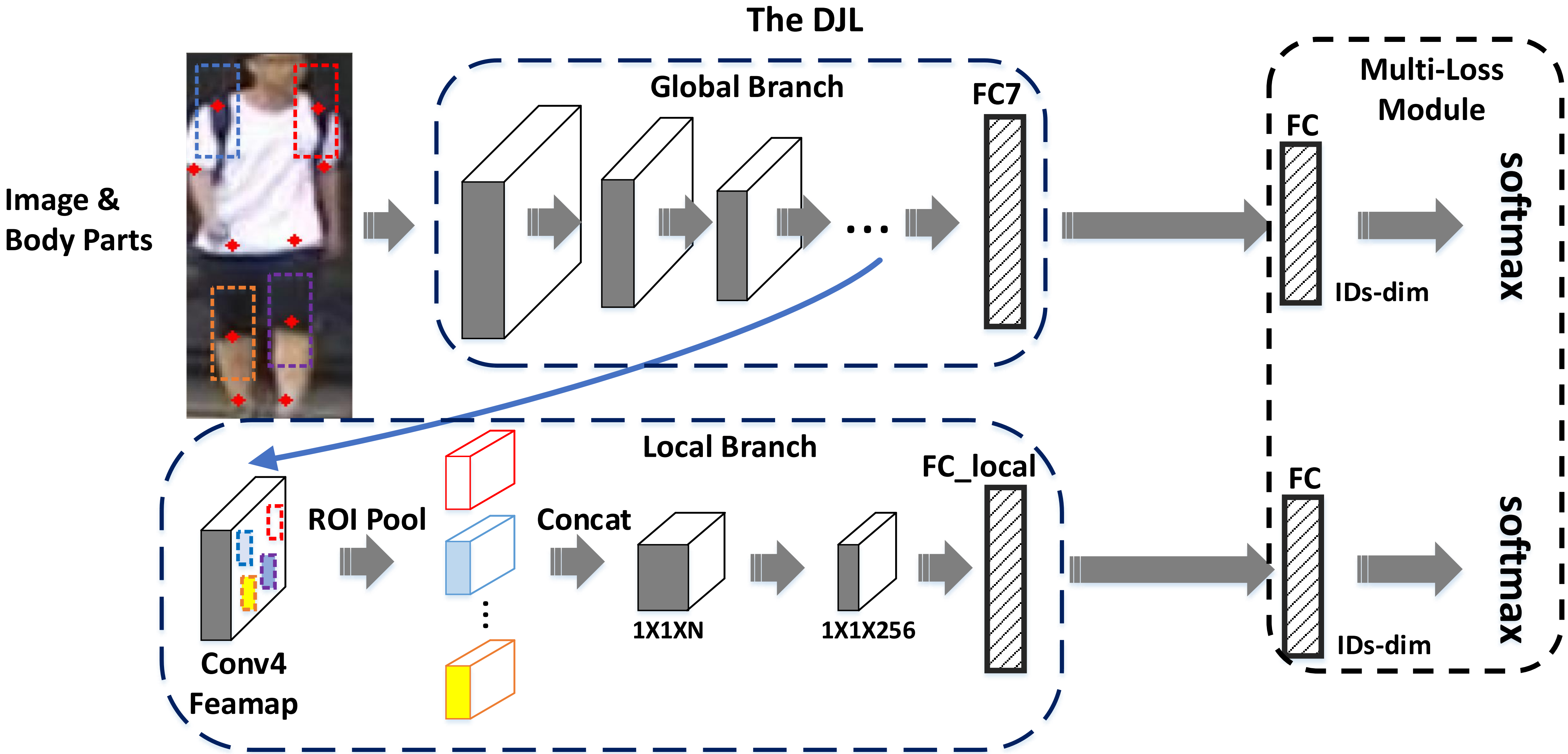
The Deep Joint Learning Network



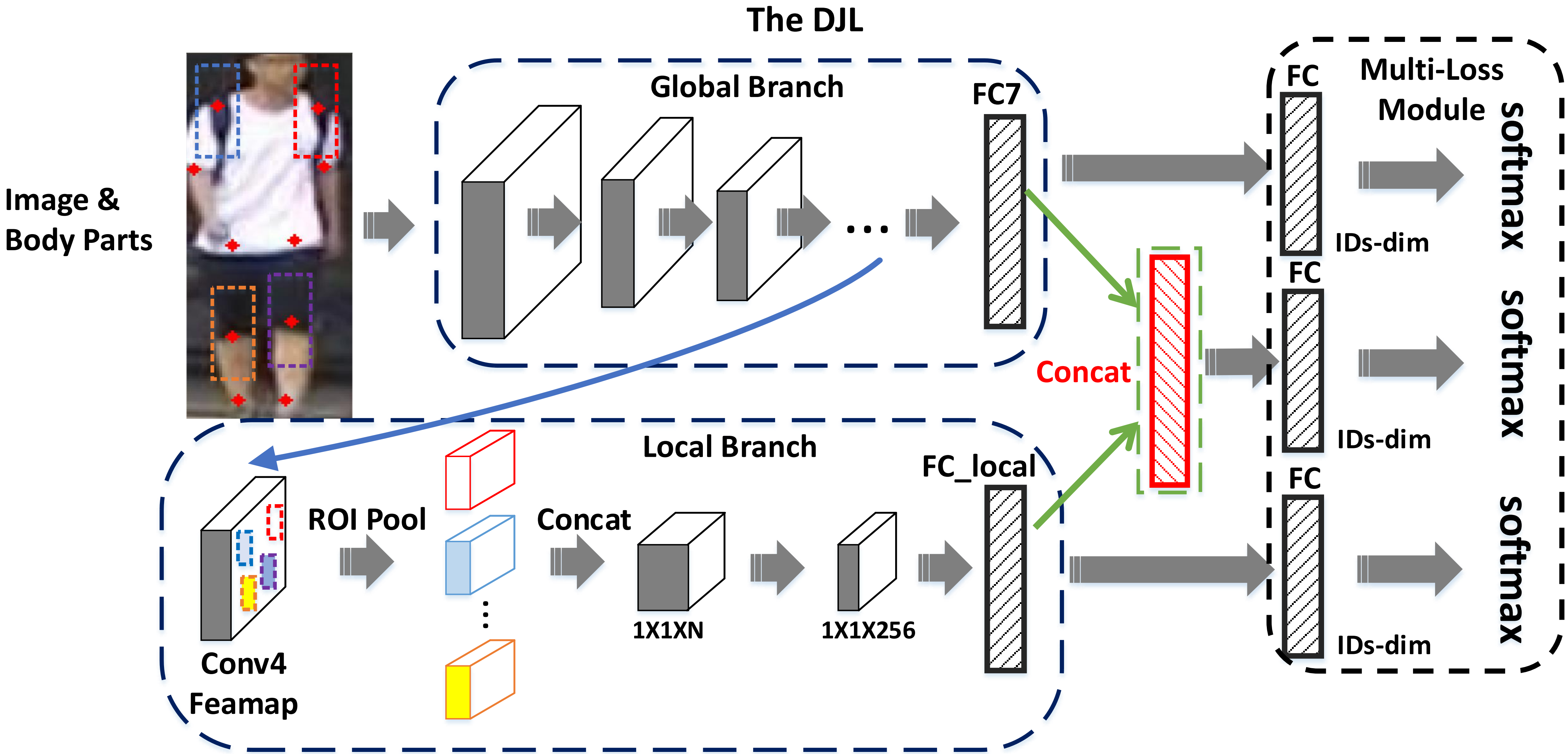
The Deep Joint Learning Network



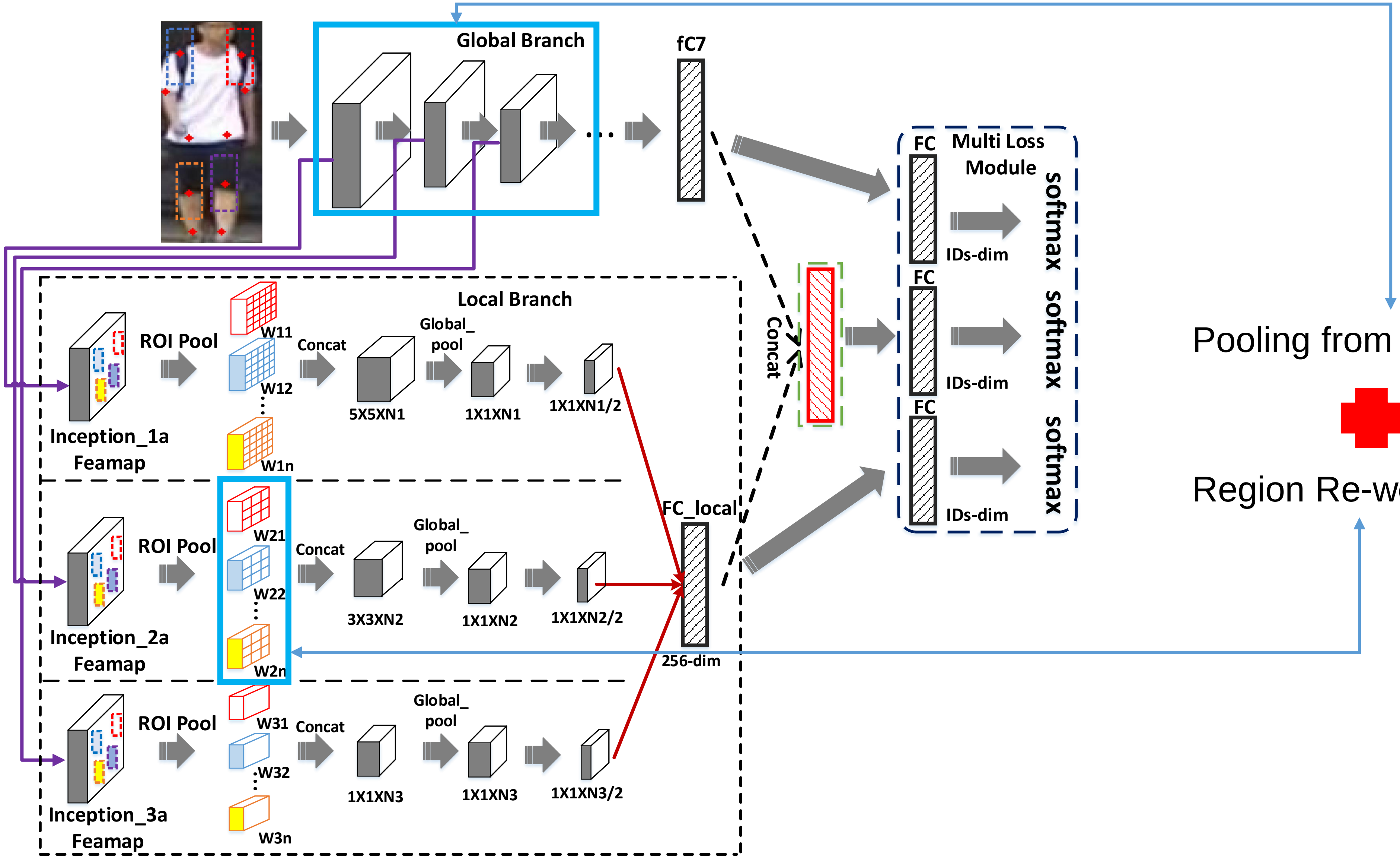
The Deep Joint Learning Network



The Deep Joint Learning Network



The Deep Joint Learning Network



Pooling from Multi-layers

Region Re-weighting

Experimental Results

- Comparison with the three base networks

Method	Market-1501					CUHK03			
	Rank-1	Rank-5	Rank-10	Rank-20	mAP	Rank-1	Rank-5	Rank-10	Rank-20
AlexNet	57.75	77.52	84.47	89.46	33.80	53.03	79.53	87.82	94.21
Residual-50	72.42	86.49	91.03	94.42	48.01	61.79	85.46	92.31	97.86
InceptionNet	79.66	91.51	94.54	96.50	56.59	80.85	95.90	98.17	99.48
Proposed (AlexNet)	67.64	84.80	89.88	93.53	43.60	71.95	90.30	94.91	98.16
Proposed (Residual-50)	78.86	90.38	93.91	96.35	55.49	80.83	95.92	98.66	99.54
Proposed (InceptionNet)	**85.12**	**93.91**	**95.69**	**97.51**	**64.82**	**84.25**	**97.40**	**98.86**	**99.67**

- Comparison with PIE

Base Network	Market-1501		CUHK03	
	DJL	PIE	DJL	PIE
AlexNet	+9.89	+9.12	+18.92	+2.65
Residual-50	+6.44	+5.66	+19.04	+5.50

Experimental Results

- Comparison with state-of-the-art methods on Market1501 and CUHK03

Methods	Rank-1	mAP
Gate-SCNN [29]	65.88	39.55
PIE [34]	78.65	53.87
DLCE [35]	79.51	59.87
MSCAN [19]	80.31	57.53
APR [21]	84.29	64.67
JLML [19]	85.10	65.50
DJL	**85.12**	64.82
DJL+RRW+Mul_s	**85.99**	65.65

(a) Comparison with state-of-the-art methods on Market1501. Rank-1 accuracy (%) and mAP (%) are shown.

Methods	Rank-1
PIE [34]	62.60
MSCAN [19]	67.99
Gate-SCNN [29]	68.10
X-Corr [27]	72.00
JLML [19]	80.60
DJL	**84.25**
DJL+RRW+Mul_s	**85.90**

(b) Comparison with state-of-the-art methods on CUHK03. Rank-1 accuracy (%) is shown.

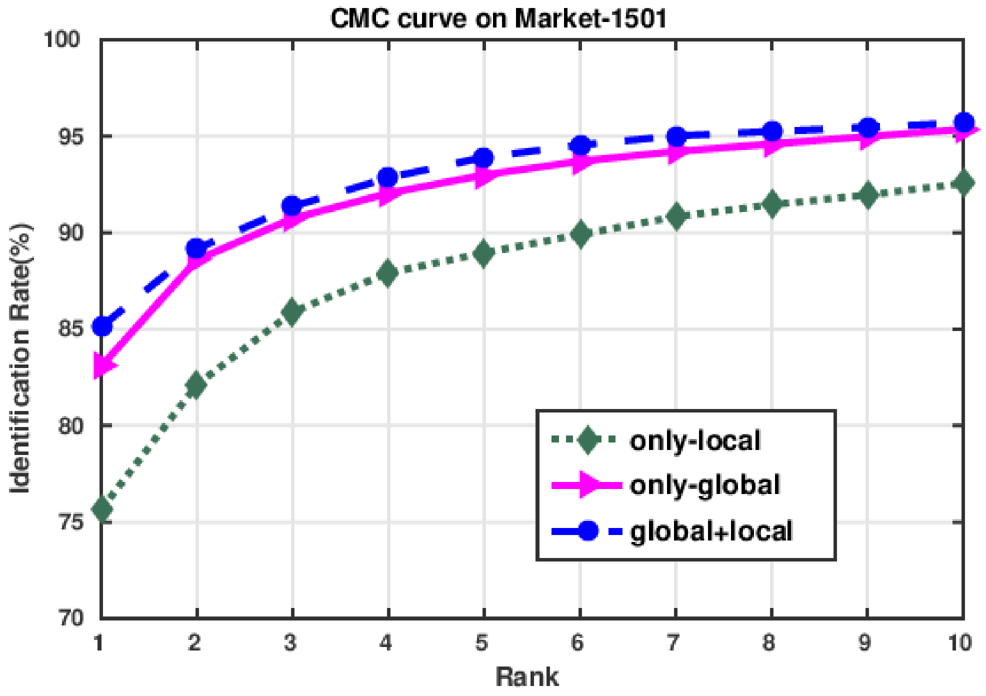
Experimental Results

- The impact of body part segmentation error

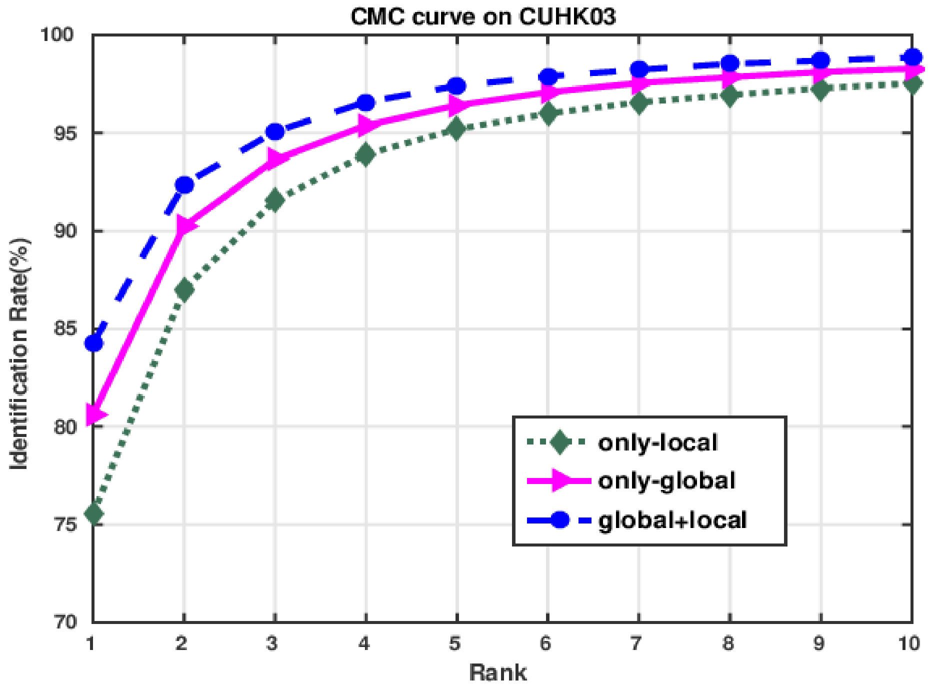
Base Network	Experiment setting	Market-1501				
		Rank-1	Rank-5	Rank-10	Rank-20	mAP
AlexNet	-	67.64	84.80	89.88	93.53	43.60
	Disturb-small	68.82	84.95	89.31	93.50	44.89
	Disturb-violent	64.79	82.21	88.15	92.22	40.84
Residual-50	-	78.86	90.38	93.91	96.35	55.49
	Disturb-small	77.76	89.88	92.96	96.02	54.62
	Disturb-violent	75.95	88.60	92.37	95.19	52.71
InceptionNet	-	85.12	93.91	95.69	97.51	64.82
	Disturb-small	84.53	93.79	95.93	97.54	64.89
	Disturb-violent	83.61	93.65	95.99	97.60	63.44

Experimental Results

- Complementary effects



(a)



(b)

Experimental Results

- Retrieval examples

Conclusion

- Propose a DJL network to learn better feature representation from both entire image and local body parts.
- Do feature level alignment by pooling the features around the body parts to learn misalignment robust local features.
- Jointly optimizes the global and aligned local features to further enhance the discriminative capability of learned feature representations.
- The simple and efficient DJL pipeline can be easily integrated with other state-of-the-art person re-ID networks

Thank you!

2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)