Mask-Guided Region Attention Network for Person Re-Identification

Nanjing University of Posts and Telecommunications

Cong ZHOU (周聪) and Han YU* (余瀚)

Outline

- Motivation
- Methodology
- Numerical Tests & Analysis
- Conclusions
- References

Motivation

Person re-identification (ReID) is an important and practical task which identifies pedestrians across non-overlapping surveillance cameras based on their visual features. In general, ReID is an extremely challenging task due to complex background clutters, large pose variations and so on.

To improve ReID's performance, a robust and discriminative feature extraction methodology is particularly crucial.

(a) Background Clutters

(b) Pose Variations

Motivation

Recently, the feature alignment technique driven by human pose estimation (Kumar *et al.*, 2017), that is, matching two person images with their corresponding parts, increases the effectiveness of ReID to a certain extent.

Kumar, V., Namboodiri, A., Paluri, M., & Jawahar, C. V.: Pose-aware person recognition. In: CVPR, pp. 6223-6232. (2017).

Motivation

Problems:

1. Background Clutters
2. Extensive Noise From Adjacent Parts

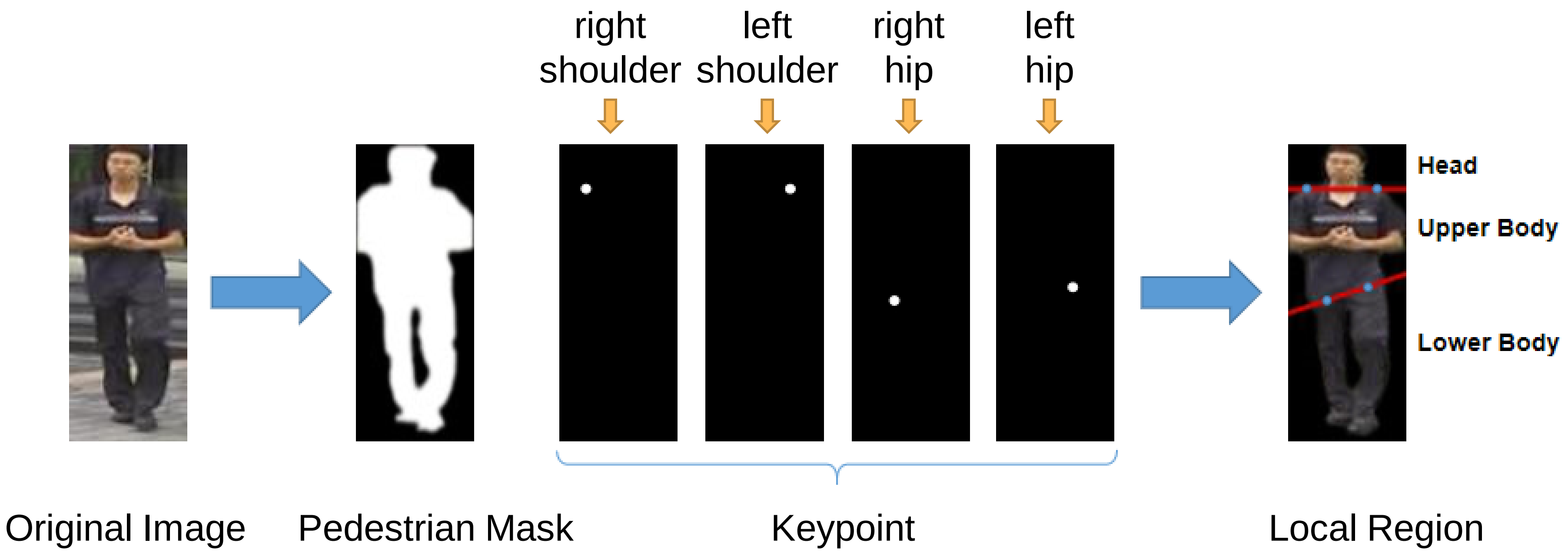
Patch (Zhao *et al.*, 2013)

Rectangular RoI (Xu *et al.*, 2018)

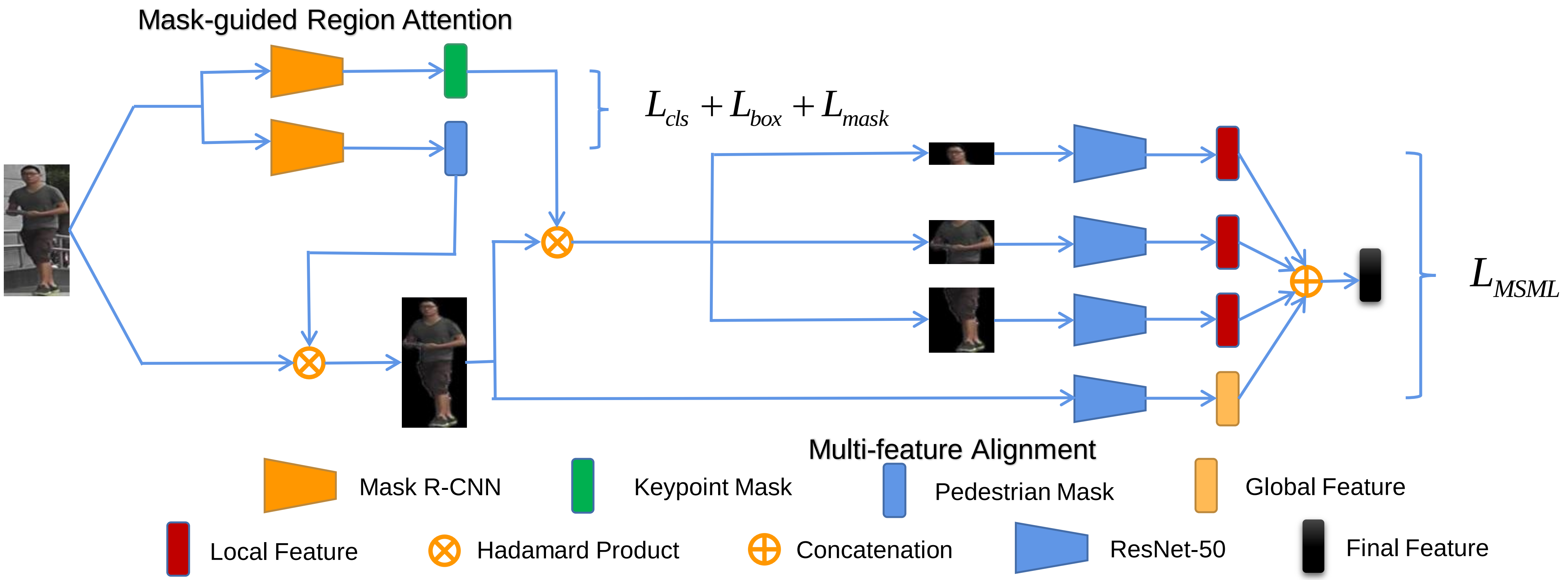
Zhao, R., Ouyang, W., & Wang, X.: Person re-identification by salience matching. In: ICCV, pp. 2528-2535. (2013).
Xu, J., Zhao, R., Zhu, F., Wang, H., & Ouyang, W.: Attention-aware compositional network for person re-identification. In: CVPR, pp. 2119-2128. (2018).

Methodology

Mask + Keypoint



Methodology



Mask-Guided Region Attention Network consists of two main components: Mask-guided Region Attention (**MRA**) and Multi-feature Alignment (**MA**).

MRA aims to generate two types of attention maps: pedestrian masks and human body keypoint masks.

MA aims to align the global feature and local features.

Methodology

1. The floating-number mask output is binarized at a threshold of 0.5.

2. In MA, the four ResNet-50 (He *et al.*, 2016) networks share the same parameters.

3. All the images input into Mask R-CNN (He *et al.*, 2017) and ResNet-50 are scaled with a factor of 1/256.

4. MRA and MA are trained independently.

He, K., Zhang, X., Ren, S., & Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770-778. (2016).
He, K., Gkioxari, G., Dollár, P., & Girshick, R.: Mask r-cnn. In: ICCV, pp. 2961-2969. (2017).

Numerical Tests & Analysis

Comparison results on Market-1501 dataset (Zheng *et al.*, 2015)

Market-1501	Single Query		Multiple Query	
	Rank-1	mAP	Rank-1	mAP
Spindle	76.90	-	-	-
DLPAR	81.00	63.40	-	-
MSCAN	80.31	57.53	-	-
SVDNet	82.30	62.10	-	-
PAN	82.81	63.35	88.18	71.72
Re-ranking	77.11	63.63	-	-
NFST	61.02	35.68	71.56	46.03
MGRAN (Ours)	88.91	78.03	90.07	81.30

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q.: Scalable person reidentification: A benchmark. In: ICCV, pp. 1116-1124. (2015).

Numerical Tests & Analysis

Comparison results on DukeMTMC-reID dataset (Zheng *et al.*, 2017)

DukeMTMC-reID	Rank-1	mAP
SVDNet	76.70	56.80
OIM	68.10	-
PAN	71.59	51.51
AACN	76.84	59.25
MGRAN (Ours)	78.12	63.57

Zheng, Z., Zheng, L., & Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: ICCV, pp. 3754-3762. (2017).

Numerical Tests & Analysis

Comparison results on CUHK03 dataset (Li *et al.*, 2014)

CUHK03	Labeled		Detected	
	Rank-1	Rank-5	Rank-1	Rank-5
PAR	85.40	97.60	81.60	97.30
NFST	62.55	90.05	54.70	84.75
SVDNet	81.80	-	-	-
DPFL	43.00	-	40.70	-
DaF	27.50	-	26.40	-
Transfer	85.40	-	84.10	-
MGRAN (Ours)	93.02	98.94	90.67	98.21

Li, W., Zhao, R., Xiao, T., & Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR, pp. 152-159. (2014).

Numerical Tests & Analysis

Features are aligned based on part-level and region-level respectively to verify the effectiveness of the proposed region-level feature alignment.

Ablation Analysis	CUHK03			
	Labeled		Detected	
	Rank-1	Rank-5	Rank-1	Rank-5
MGRAN-PL	91.83	97.41	89.35	96.75
MGRAN-RL	93.02	98.94	90.67	98.21

MGRAN-PL means aligning features based on part-level.

MGRAN-RL means aligning features based on region-level.

Conclusions

Advantage:
1. Novel Ideas (Mask and Region-Level Feature Alignment)
2. Better Accuracy

Disadvantage:
1. Low Detection Speed
2. Not Good Enough (Accuracy) : Imperfect Mask

References

[1] Kumar, V., Namboodiri, A., Paluri, M., & Jawahar, C. V.: Pose-aware person recognition. In: CVPR, pp. 6223-6232. (2017).

[2] Zhao, R., Ouyang, W., & Wang, X.: Person re-identification by salience matching. In: ICCV, pp. 2528-2535. (2013).

[3] Xu, J., Zhao, R., Zhu, F., Wang, H., & Ouyang, W.: Attention-aware compositional network for person re-identification. In: CVPR, pp. 2119-2128. (2018).

[4] He, K., Zhang, X., Ren, S., & Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770-778. (2016).

[5] He, K., Gkioxari, G., Dollár, P., & Girshick, R.: Mask r-cnn. In: ICCV, pp. 2961-2969. (2017).

[6] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q.: Scalable person reidentification: A benchmark. In: ICCV, pp. 1116-1124. (2015).

[7] Zheng, Z., Zheng, L., & Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: ICCV, pp. 3754-3762. (2017).

[8] Li, W., Zhao, R., Xiao, T., & Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR, pp. 152-159. (2014).

THANK YOU