Recurrent Deep Attention Network for Person Re-Identification

Changhao Wang[1], Jun Zhou[2] , Xianfei Duan[2], Guanwen Zhang[1]*, Wei Zhou[1]

1 School of Electronics and Information, Northwestern Polytechnical University, Xi' an, China

2 CNPC logging Co.,Ltd, China

January 10，2021

CONTENTS

I. Introduction

II. The Proposed Method

III. Experiments

IV. Conclusions

INTRODUCTION

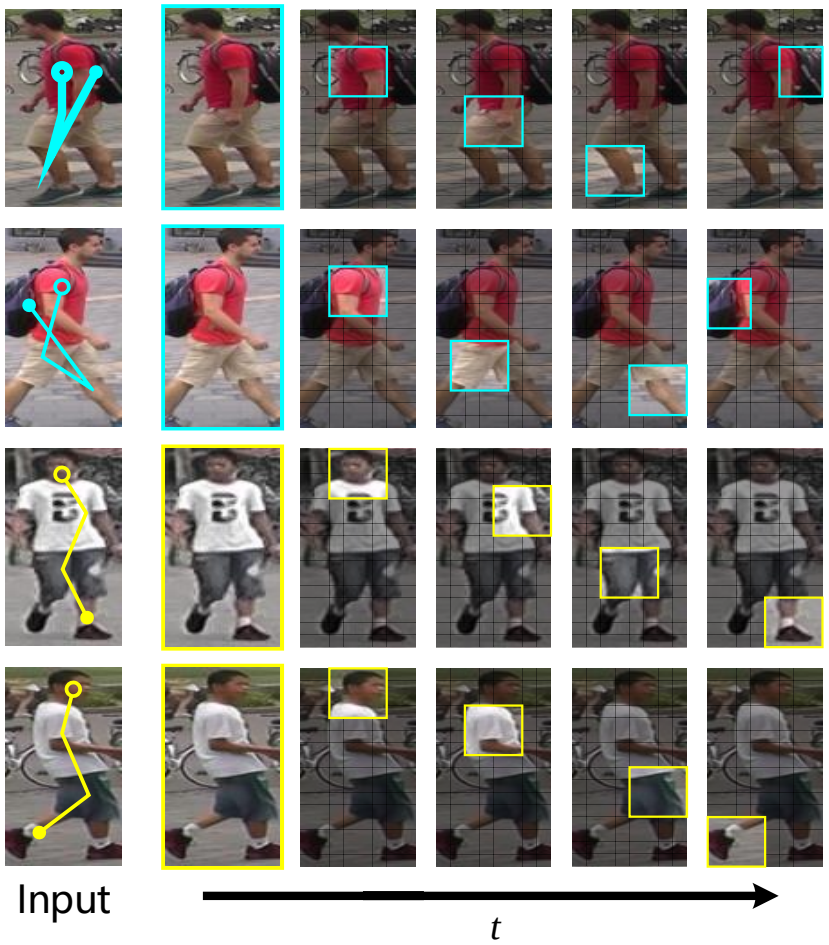
Challenges in Person Re-Identification

- Dramatic changes on individual appearance
- Occluded
- Complicated background clutters
- Inaccurate bounding box

INTRODUCTION

Our Solution

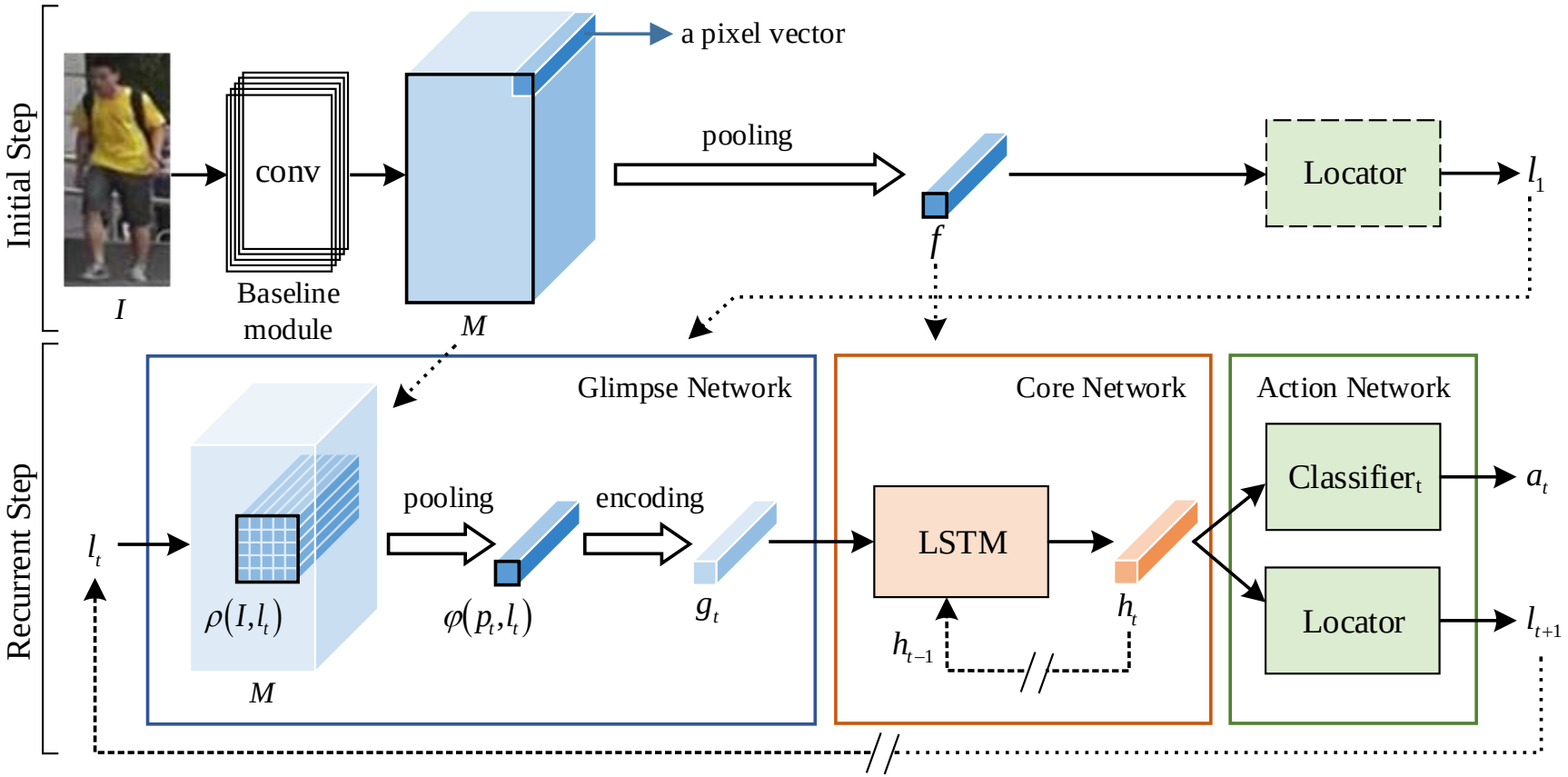
- We propose an attention selection mechanism based on reinforcement learning for person re-id task.
- The proposed model can focus on identity-sensitive regions to build up internal cognition of individuals progressively.



THE PROPOSED METHOD

An Overview of RDAN

- A Baseline Module
- A Glimpse Module
- A Core Network
- An Action Network



THE PROPOSED METHOD

The Markov Decision Process

- **State** The hidden state $\boldsymbol{h}_t$ of the core network
- **Action** The location l_t produced by the locator
- **Reward** r_t^{id}, r_t^{rc}, r_t^{gc}

- Identity Reward

$$r_t^{id} = \begin{cases} 1, a_t = y \\ 0, a_t \neq y \end{cases}$$

- Relative Comparison Reward

$$r_t^{rc} = d_t^{a,n} - d_t^{a,p}$$

- Gradual Comparison Reward

$$r_t^{gc} = (d_t^{a,n} - d_t^{a,p}) - (d_{t-1}^{a,n} - d_{t-1}^{a,p})$$

where, $d_t^{a,p} = \|\boldsymbol{h}_t^a - \boldsymbol{h}_t^n\|_2$, $d_t^{a,n} = \|\boldsymbol{h}_t^a - \boldsymbol{h}_t^n\|_2$

THE PROPOSED METHOD

Optimization

- Loss function for supervised learning

$$L_{id} = -\sum_{t=1}^{T}\sum_{i=1}^{N} 1\{i = y\}\log P_t^i$$

$$L_{tri} = \sum_{t=1}^{T} \max\left(\left(d_t^{a,p} - d_t^{a,n} + \alpha\right), 0\right)$$

- Object function for reinforcement learning

$$J(\theta) = \mathbb{E}_\pi[R] = \mathbb{E}_\pi[\sum_{t=1}^{T} r_t]$$

EXPRIMENTS

Experiment Results

Method	Market-1501		DukeMTMC-reID	
	Rank-1(%)	mAP(%)	Rank-1(%)	mAP(%)
PCB[1]	92.3	77.4	81.9	65.3
PCB + RPP [1]	93.8	81.6	83.3	69.2
VPM [2]	93.0	80.8	83.6	72.6
HA-CNN [3]	91.2	75.7	80.5	63.8
Mancs [4]	93.1	82.3	84.9	71.8
CASN(IDE) [5]	92.0	78.0	84.5	67.0
CASN(PCB) [5]	94.4	82.8	87.7	73.7
RDAN	**94.6**	**85.4**	**88.0**	**75.2**

EXPRIMENTS

Experiment Results

Method	CUHK03-NP(Detected)		CUHK03-NP(Labeled)	
	Rank-1(%)	mAP(%)	Rank-1(%)	mAP(%)
PCB [1]	61.3	54.2	-	-
PCB + RPP [1]	63.7	57.5	-	-
HA-CNN [3]	41.7	39.6	44.4	41.0
Mancs [4]	65.5	60.5	69.0	63.9
CASN(IDE) [5]	57.4	50.7	58.9	52.2
CASN(PCB) [5]	**71.5**	64.4	73.7	68.0
RDAN	69.5	**64.5**	**74.2**	**69.4**

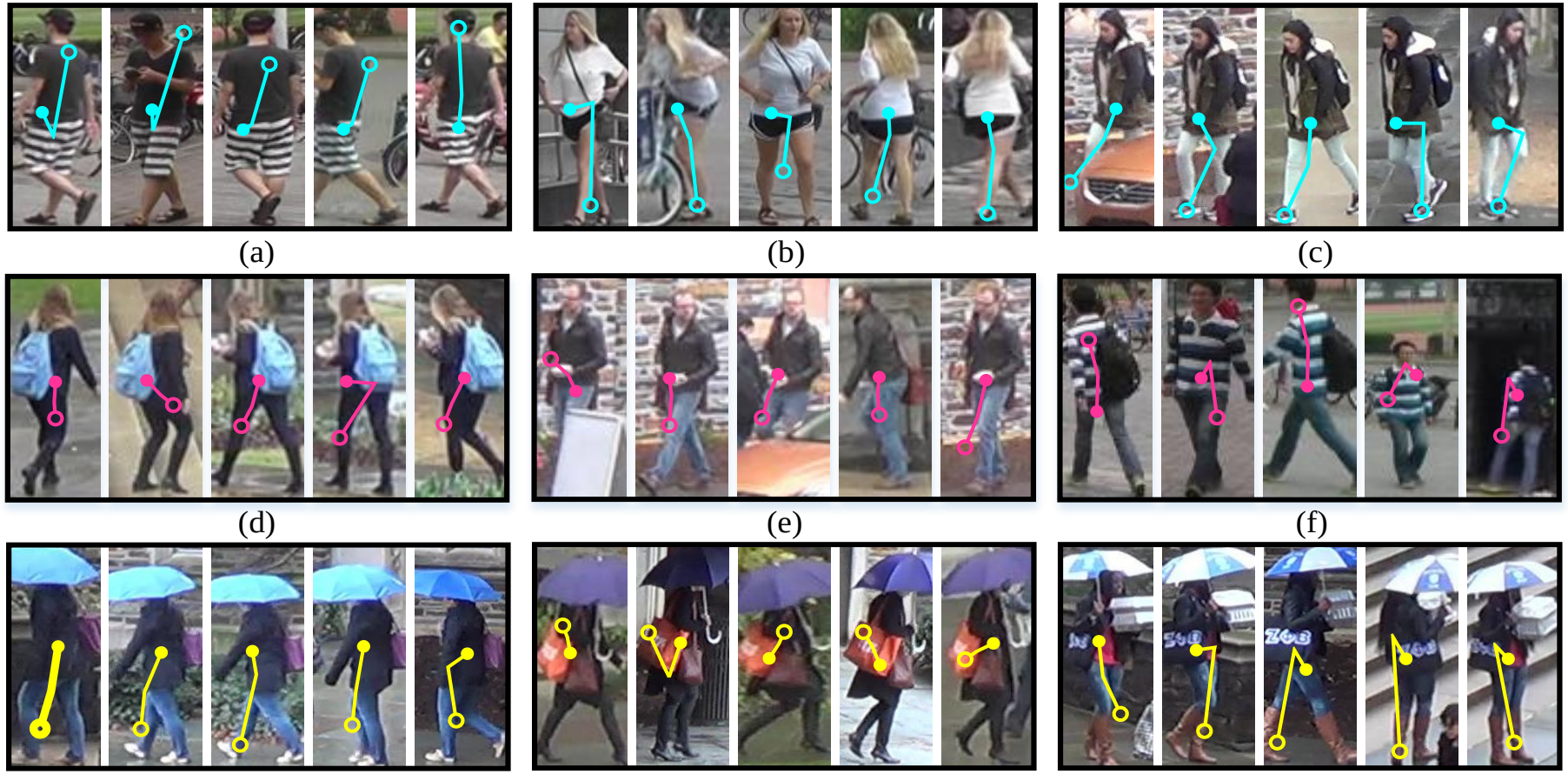
EXPRIMENTS

Ablation Study

	Market-1501		DukeMTMC-reID	
	Rank-1(%)	mAP(%)	Rank-1(%)	mAP(%)
IDE Baseline	90.0	77.5	83.0	68.4
L_{id} +Random Policy	92.5	82.0	86.1	71.8
$L_{id} + r_t^{id}$	92.9	82.3	86.3	72.6
$L_{id} + r_t^{rc}$	93.9	84.4	87.4	74.7
$L_{id} + r_t^{gc}$	93.6	84.4	87.8	74.5
$L_{id} + L_{tri} + r_t^{rc}$	94.0	**85.7**	87.7	**75.2**
$L_{id} + L_{tri} + r_t^{gc}$	**94.6**	85.4	**88.0**	**75.2**

EXPRIMENTS

Visualization



CONCLUSION

- we propose a Recurrent Deep Attention Network (RDAN) that **embeds convolutional architecture** in **recurrent attention model** and propose an attention selection mechanism based on **reinforcement learning** for person re-id.

- The proposed RDAN selects attention on the convolutional feature maps, and combines **global** and **local features** together as the internal representation of inputs.

REFERENCE

[1] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline)," in ECCV, 2018, pp. 501–518.

[2] Y. Sun, Q. Xu, Y. Li, C. Zhang, Y. Li, S. Wang, and J. Sun, "Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 393–402.

[3] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 2285–2294.

[4] C. Wang, Q. Zhang, C. Huang, W. Liu, and X. Wang, "Mancs: A multi-task attentional network with curriculum sampling for person reidentification," in ECCV, 2018, pp. 384–400.

[5] M. Zheng, S. Karanam, Z. Wu, and R. J. Radke, "Re-identification with consistent attentive siamese networks," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 5735–5744.

THANK YOU FOR WATCHING