UNCERTAINTY-GUIDED PERSON SEARCH MODEL WITH AUXILIARY SHALLOW FEATURE EXPLORATION

Zongyi Li[1], Zhongyang Li[1], Yuxuan Shi*, Hefei Ling[1], Jiazhong Chen[1], Runsheng Wang[1], Ping Li[1]
[1]Department of Computer Science and Technology, Huazhong University of Science and Technology

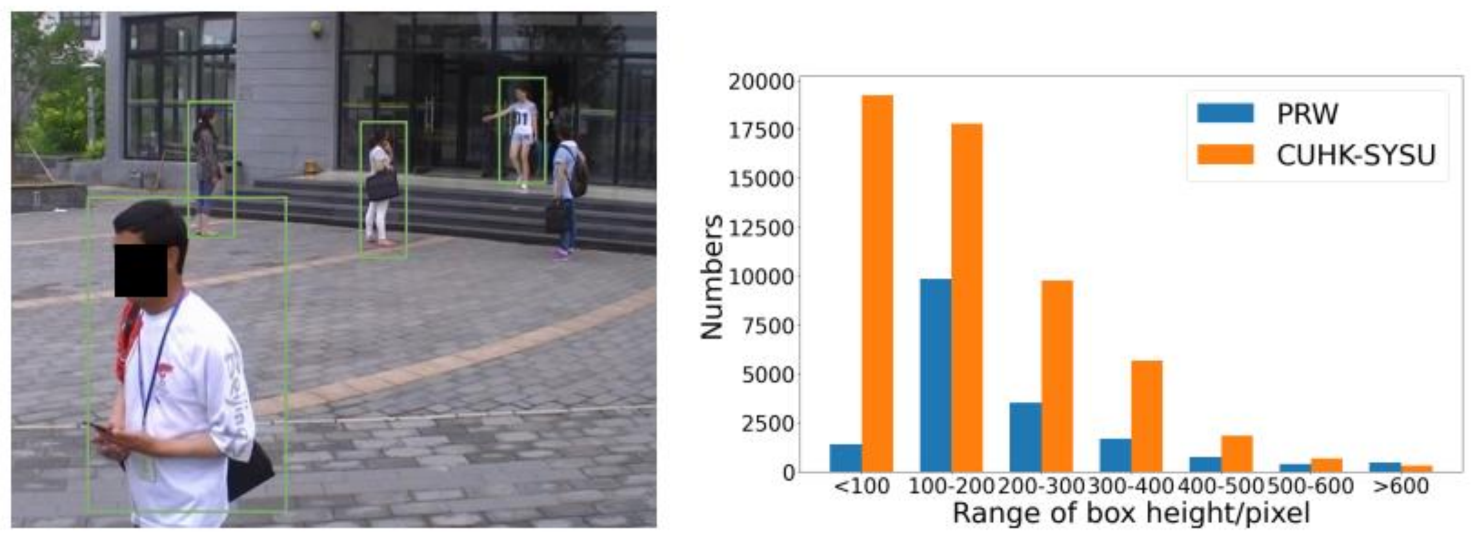
華中科技大學

One-step Person Search

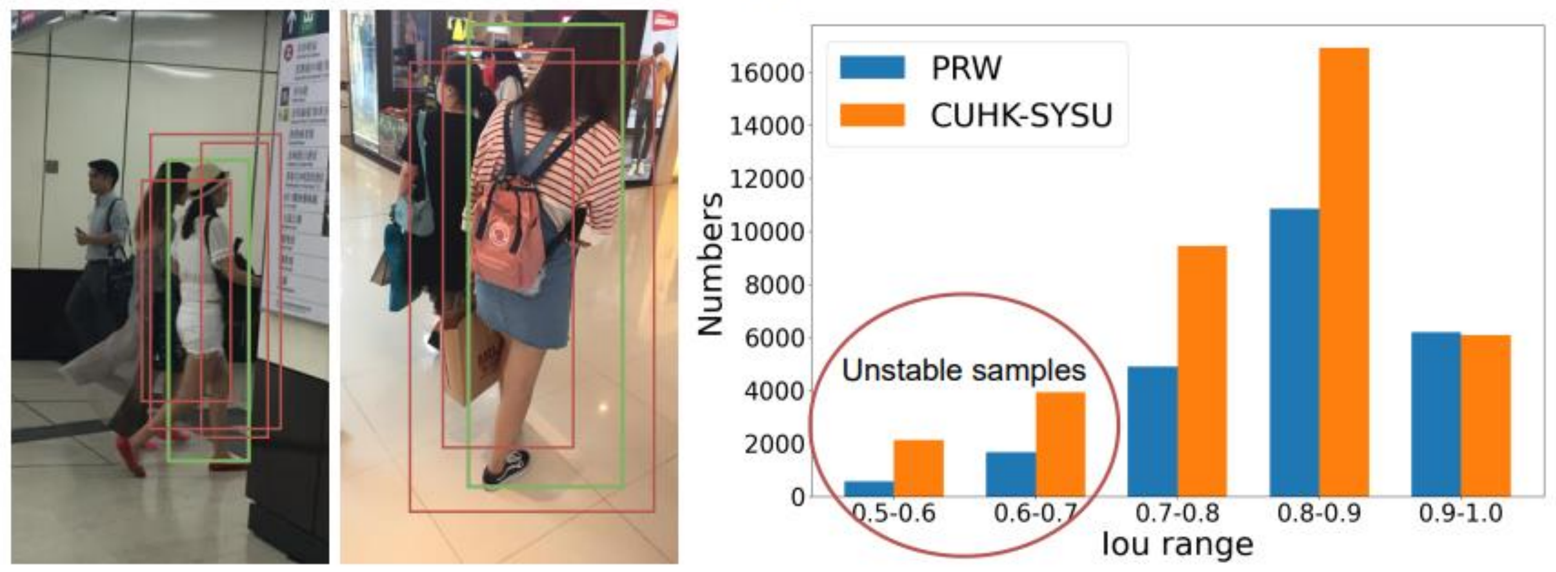
Person search aims to localize and identify a target person in the whole uncropped scene images. It can be widely applied in real life such as in video surveillance and lost people's retrieval. Challenges: Different scales, inaccurate detection boxs and occlusion.

Illustration of diverse scale problem

There are different scale regions in one scene image. (a).The distributions of the height of boxes in PRW and CUHK-SYSU datasets are significantly diverse. (b).The distribution of iou which is calculated between predicted boxes and ground-truths



(a)

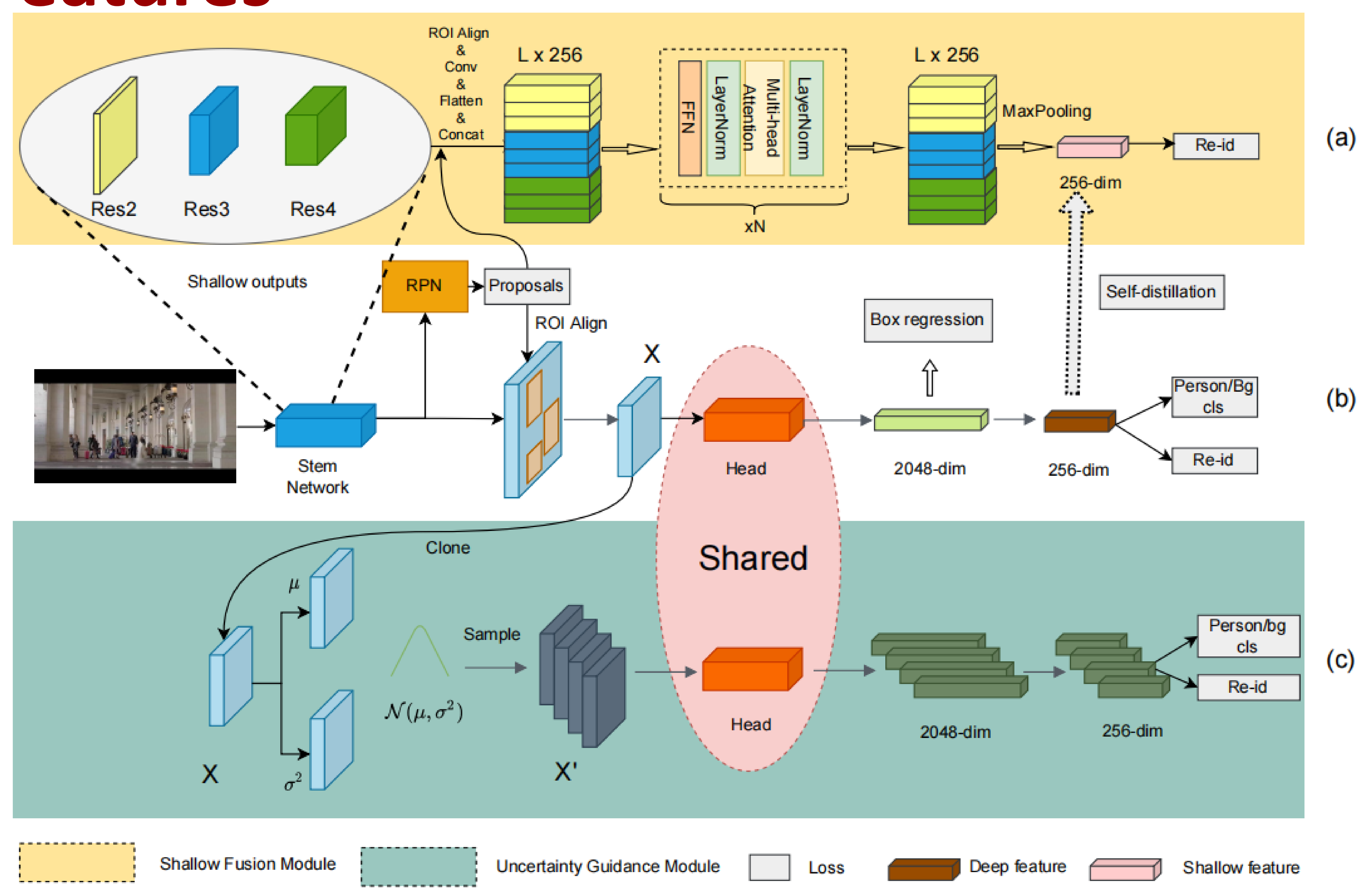


(b)

Contributions:

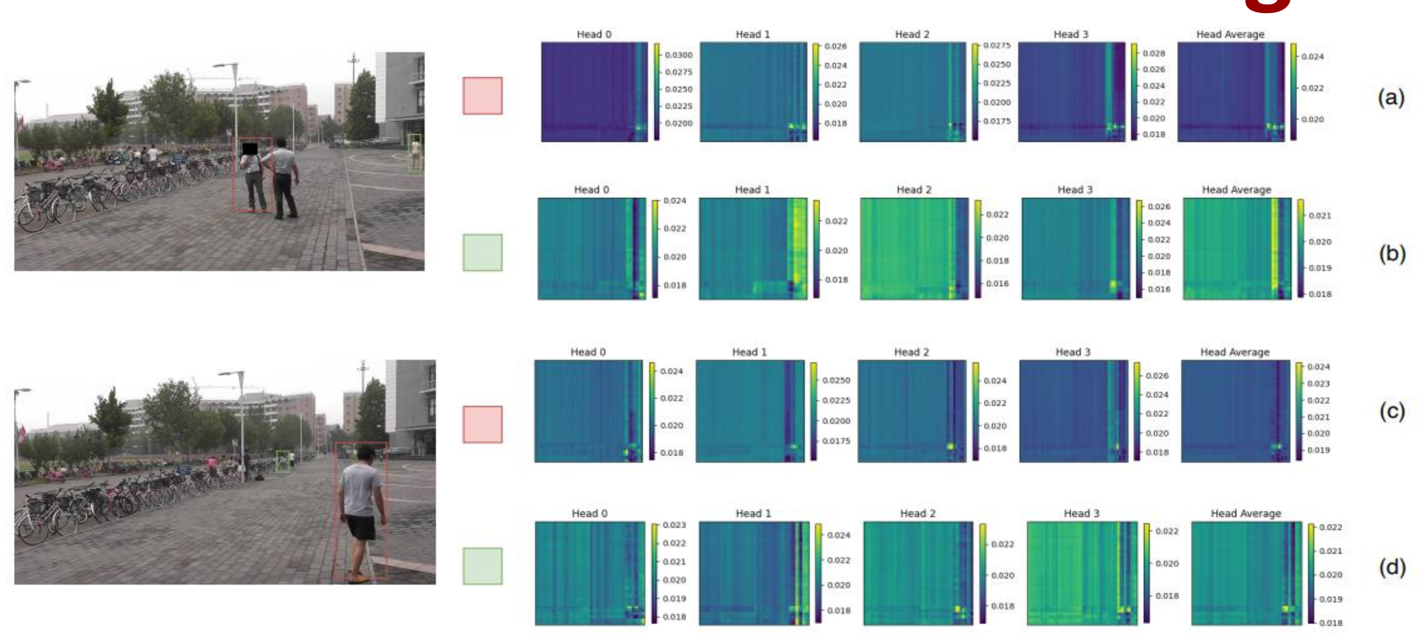
1. A simple but effective Shallow Fusion Module (SFM) is proposed to extract multi-scale fusion features to improve re-id performance.
2. An Uncertainty Guidance (UG) module is introduced to reduce the impact of coarse samples on the model without incurring inference cost.
3. Extensive experiments on PRW and CUHK-SYSU datasets show the efficiency of our proposed modules.

Uncertainty Guided with Auxiliary Shallow Features



- Shallow Feature Fusion Module is proposed to extract multi-scale features and fuse them into a concise feature vector, thereby enhancing the matching performance.
- A self-distillation loss is added between deep and shallow branches to compensate for the missing deep semantic information of shallow features
- An Uncertainty Guidance (UG) module is proposed to mitigate the coarse embeddings' influence on the model.
- The concept of uncertainty treats each sample embedding as a Gaussian distribution, and it can reduce the impact of coarse embeddings through the generated variance associated with each distribution.

The visualization of attention weights



Comparison with state-of-the-arts (CUHK-SYSU, PRW)

Method	Backbone	CUHK-SYSU mAP	CUHK-SYSU Top-1	PRW mAP	PRW Top-1
Two-step					
IGPN [17]	ResNet50	90.3	91.4	**47.2**	87.0
TCTS [18]	ResNet50	**93.9**	**95.1**	46.8	**87.5**
One-step with two-stage detector					
OIM [3]	ResNet50	75.5	78.7	21.3	49.4
BINet [19]	ResNet50	90.0	90.7	45.3	81.7
NAE [5]	ResNet50	91.5	92.4	43.3	80.9
SeqNet [8]	ResNet50	93.8	94.6	46.7	83.4
AlignPS [6]	ResNet50	93.1	93.4	45.9	81.9
OIMNet++ [20]	ResNet50	93.1	94.1	47.7	84.8
PSTR [9]	ResNet50	**93.5**	**95.0**	**49.5**	**87.8**
Ours					
NAE*	ResNet50	91.3	92.7	43.2	78.7
UGAS (NAE)	ResNet50	**92.4**	**93.5**	**48.1**	**84.1**
SeqNet*	ResNet50	93.6	94.1	46.5	84.2
UGAS (SeqNet)	ResNet50	**94.3**	**94.8**	**51.9**	**85.5**

Ablation study

Num	SFM	SD	UG	PRW mAP	PRW top-1
(a)	-	-	-	46.5	84.2
(b)	✓	-	-	48.4	83.3
(c)	✓	✓	-	50.1	84.0
(d)	-	-	✓	48.9	85.1
(e)	✓	✓	✓	**51.9**	**85.5**

Ablation study on the effectiveness of components in our method.

Model	PRW (L) mAP	PRW (L) Top-1	CUHK-SYSU (L) mAP	CUHK-SYSU (L) Top-1
SeqNet*	38.8	76.6	85.4	85.5
SeqNet+SFM	42.6	77.2	86.8	86.9

The performance on low-resolution datasets

Feature	with/o dis mAP	with/o dis Top-1	with dis mAP	with dis Top-1
Only deep	45.5	81.8	47.8	82.5
Only shallow	44.1	80.4	46.8	82.1
Deep & Shallow (mean)	45.9	81.8	47.2	81.9
Deep & Shallow (concat)	48.4	83.3	50.1	84.0

The efficiency of SeqNet with shallow feature fusion module