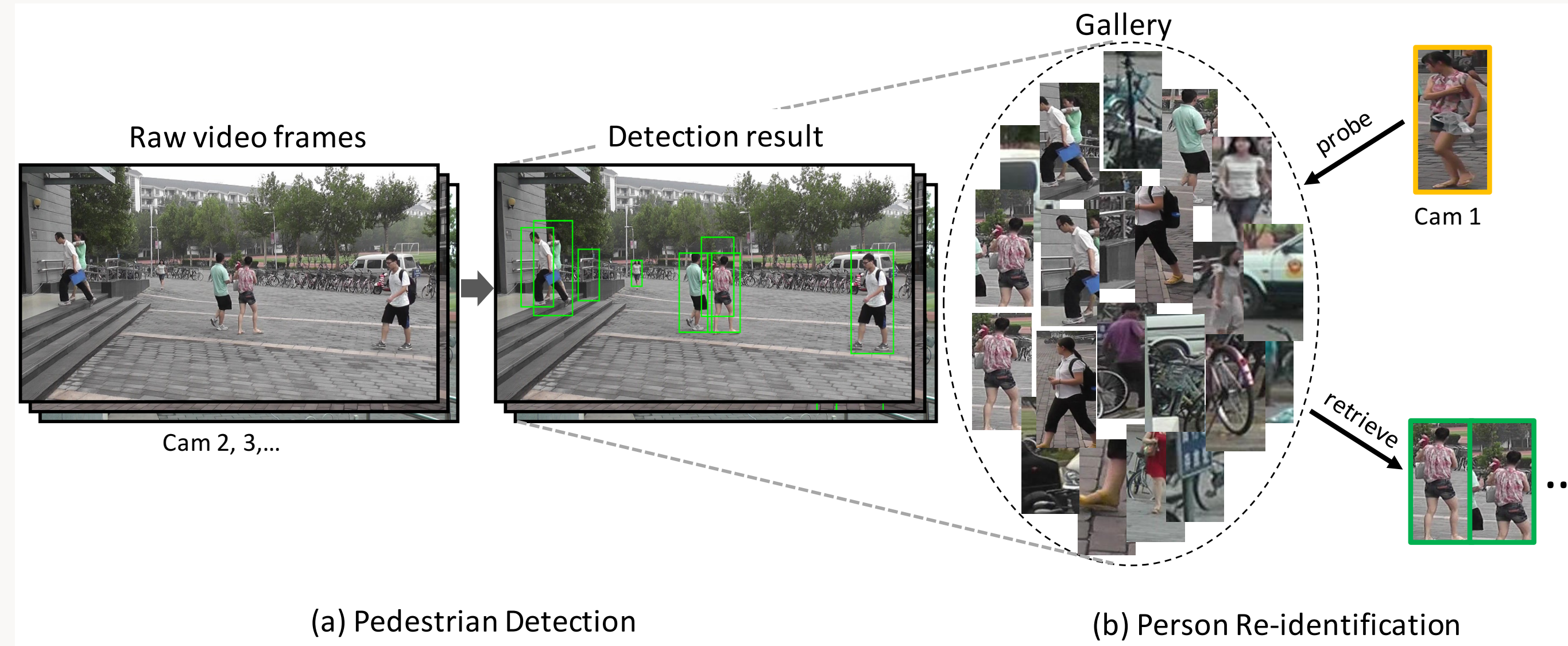


Person Search via A Mask-Guided Two-Stream CNN Model

Di Chen¹ Shanshan Zhang^{1*} Wanli Ouyang² Jian Yang^{1*} Ying Tai³

¹Nanjing University of Science and Technology, ²The University of Sydney, ³YouTu Lab, Tencent

Task



Person Search = Pedestrian Detection + Person Re-ID [1]

Motivation

- It is not appropriate to share representations between the detection and re-ID tasks, as their goals contradict with each other.
- It is more suitable to consider a compromised strategy of *paying extra attention on the foreground person* while also using the background as a complementary cue.

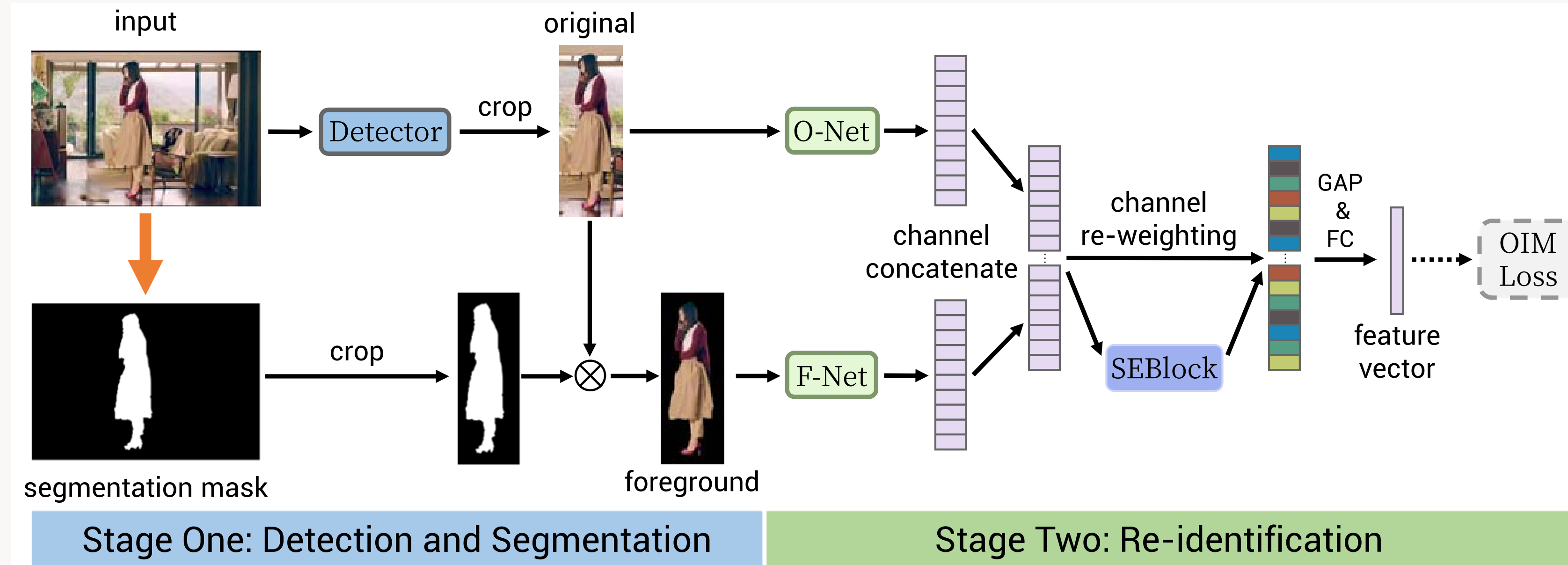
Datasets

- CUHK-SYSU [2]
 - 11,206 images, 5,532 identities, 55,272 bboxes for training.
 - 2,900 probes, 6,978 gallery images.
 - Several gallery subsets with different sizes.
- PRW [1]
 - 5,134 images, 482 identities, 18,301 bboxes for training.
 - 2,057 probes, 6,112 gallery images.

References

- [1] Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., Tian, Q.: Person re-identification in the wild. In: CVPR. (2017)
- [2] Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. In: CVPR. (2017)

Mask-Guided Two-Stream CNN Model



- Two stages are trained SEPARATELY.
- RoI expansion by a ratio of γ is conducted while cropping proposals.
- Detector:** Faster R-CNN based on VGG16.
- Segmentation Mask:** FCIS pre-trained on COCO.
- O-Net and F-Net:** ResNet50.

Ablation Study

- Separation > Integration

Method	Joint	AP	Recall	Method	Joint	mAP	top-1
OIM-ours	✓	69.5	75.6	GT + OIM [2]	✓	77.9	80.5
CNN [2]	✗	78.0	75.7	GT + IDNetOIM	✗	78.5	81.7

- Visual Component Study

O F B E	mAP	top-1
✓	78.5	81.7
✓	75.3	78.7
✓	34.2	35.9
✓	77.7	81.1
✓	38.7	40.0
✓	89.1	90.0

- O:** Original image
- F:** Foreground person only
- B:** Background only
- E:** Expand RoI by a ratio of γ
 - Hard discarding BG hurts.
 - Hard expansion on BG hurts.
 - Two-stream modeling boosts a lot.

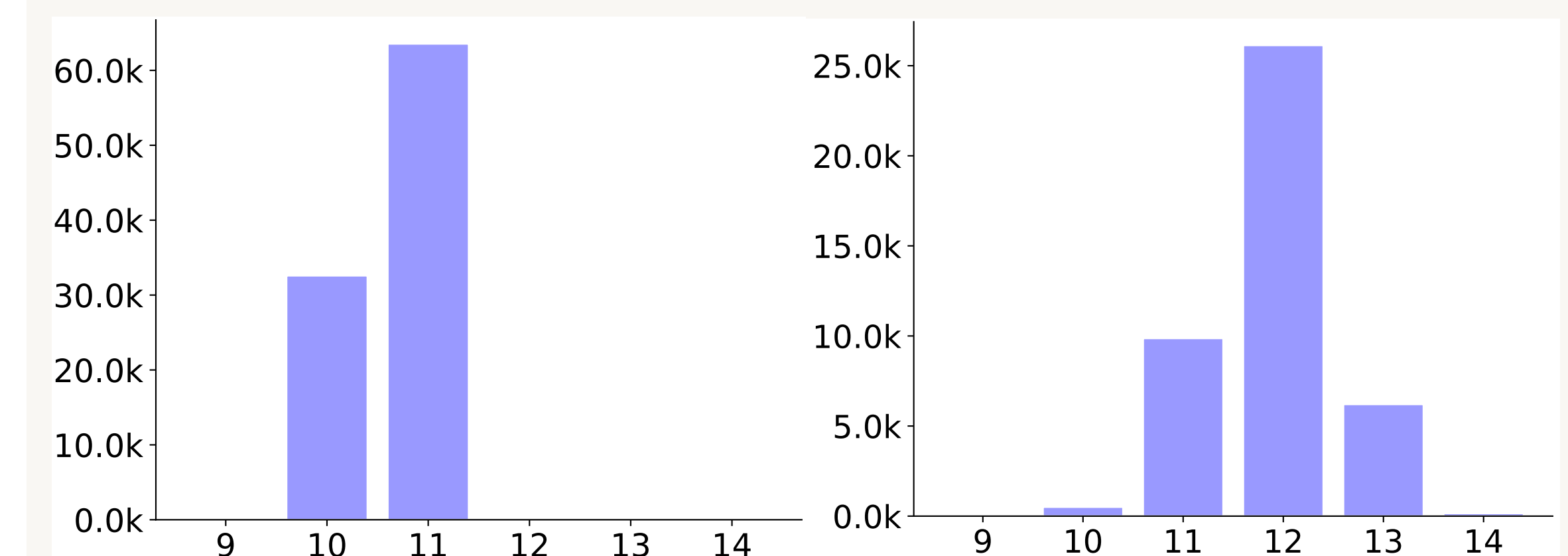
Comparison with State-of-the-Art Methods

Quantitative Results

Method	CUHK-SYSU		PRW	
	mAP	top-1	mAP	top1
OIM	75.5	78.7	21.3	49.9
IAN	76.3	80.1	23.0	61.9
NPSM	77.9	81.2	24.2	53.1
Ours(CNN _v + IDNetOIM)	75.8	79.5	28.2	66.7
Ours(CNN_v + MGTS)	83.0	83.7	32.6	72.1

SEBlock Weights Inspection

- Average weights for sample i : $Avg_i(F) > Avg_i(O)$
- Number of F stream weights among the top 20: $N^{20}(F)$
 - Most informative cues are from the foreground patch.
 - Context information contained in the original image patch is helpful.



Re-ID with Different γ

Value of γ	1.0	1.1	1.2	1.3	1.4	1.5
mAP	85.6	85.4	88.9	89.1	87.8	87.1
top-1	86.6	86.2	89.8	90.0	88.2	87.8

- A proper amount of context is better than no context.
- Too much background could be harmful.



THE UNIVERSITY OF SYDNEY

