

# Wider Pedestrian Challenge 2018

VIPL Team

Institute of Computing Technology, Chinese Academy of Sciences

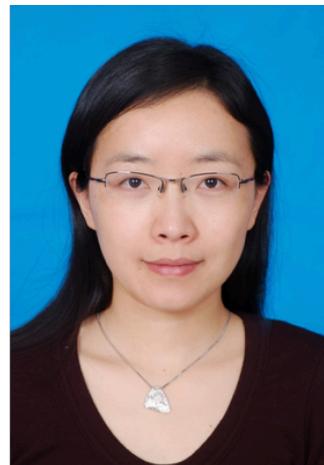
# Team members



Hongkai Zhang



Bingpeng Ma



Hong Chang



Shiguang Shan



Xilin Chen

Institute of Computing Technology, Chinese Academy of Sciences

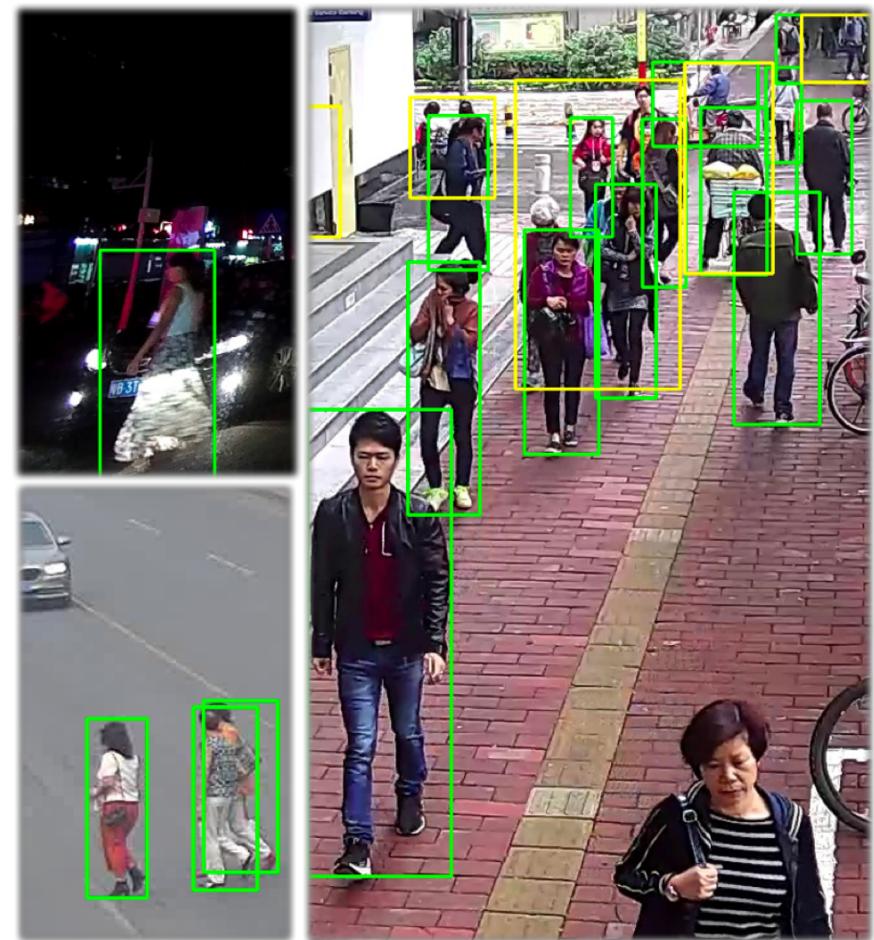


# Overview

- Data analysis
- Architecture
  - Base model
  - FPN with Cascade R-CNN
  - Useful modules
- Training
- Testing
- Results
- Conclusion

# Data analysis

- Different brightness and scenes
- Various size of persons
- Occlusion
- Unlabelled person



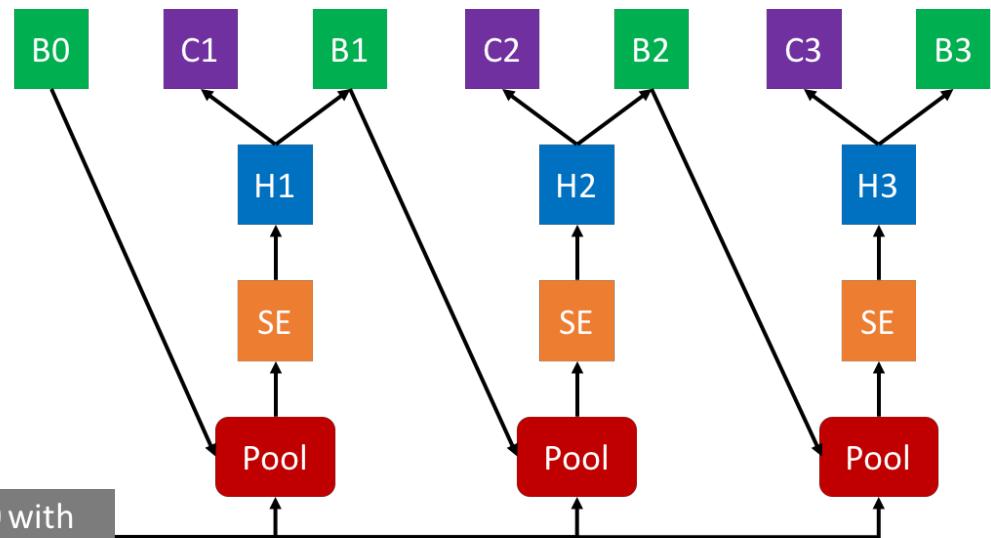
# Architecture



Random Crop

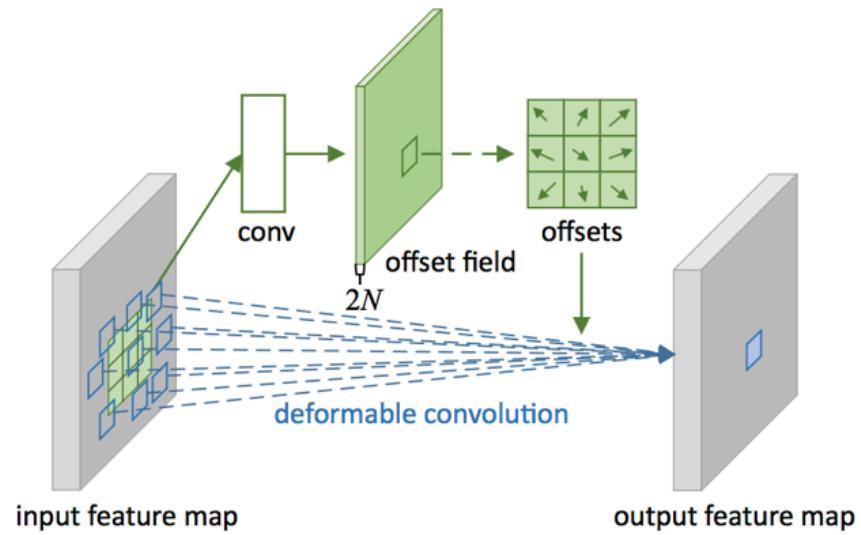


Resnet-50 with  
Deformable Conv



# Base model

- Resnet-50[K. He, CVPR'16]
  - A very powerful and popular base model
- Deformable convolution[J. Dai, ICCV'17]
  - Learning offsets to focus on the important positions for better feature extraction
  - Occlusion handling



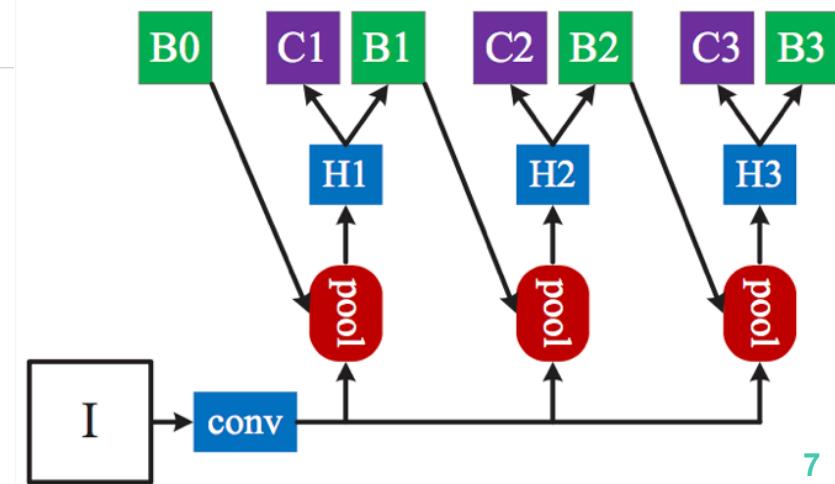
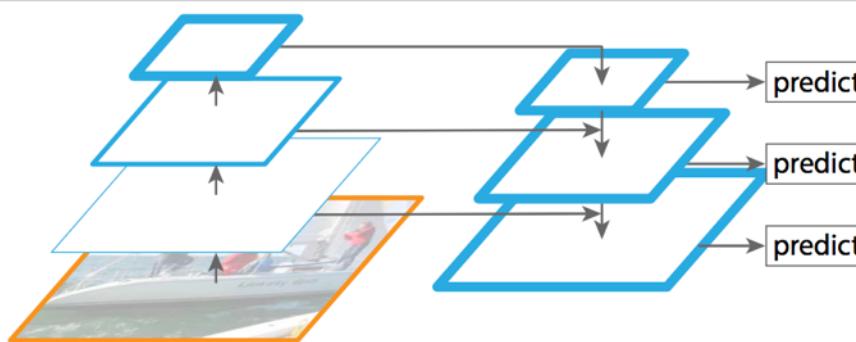
# FPN with Cascade R-CNN

## ■ FPN[T. Lin, CVPR'17]

- Form a feature pyramid to use different levels of features
- Handling different scales of person

## ■ Cascade R-CNN[Z. Cai, CVPR'18]

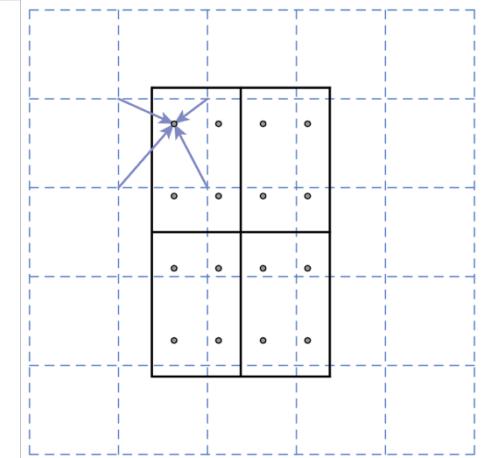
- Add extra R-CNNs with different IOU threshold
- More accurate localization performance



# Useful modules

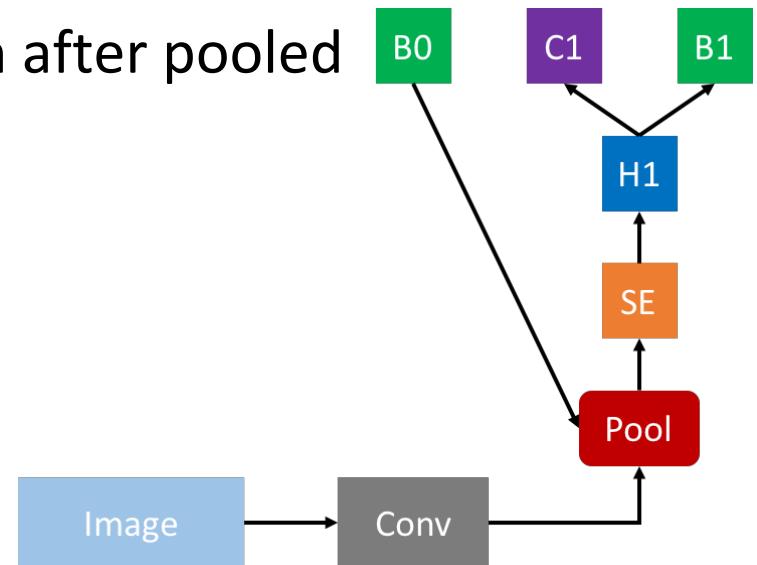
## ROI-Align[K. He, ICCV'17]

- Use bilinear interpolation instead of quantization when pooling features
- Better for feature extraction



## Re-weight Pool5/SE[S. Zhang, CVPR'18]

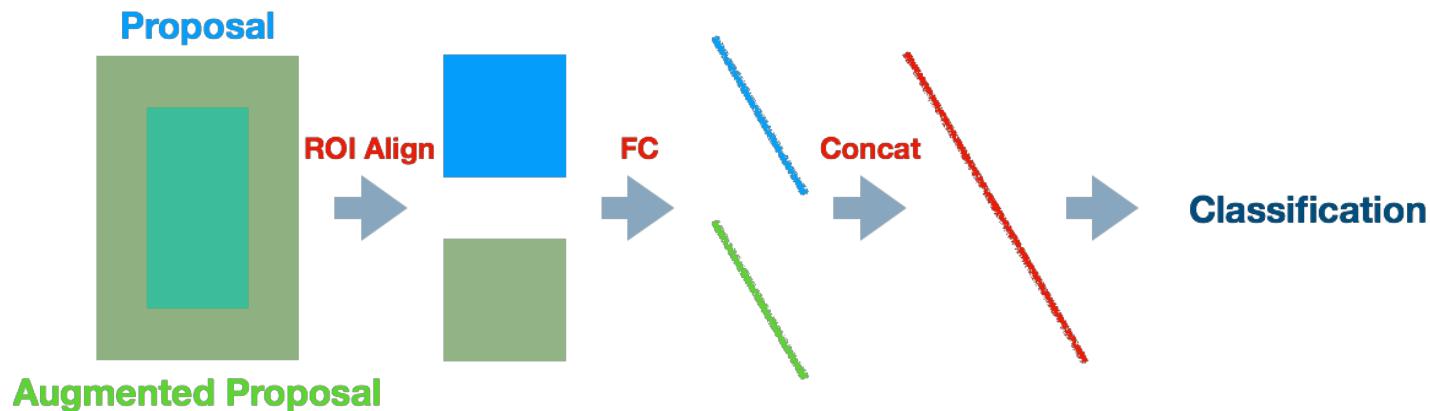
- Add channel-wise attention after pooled feature
- Occlusion handling



# Useful modules

## Context information

- Context is useful for classification
- Concatenate FCs
- Concatenate feature maps may lead to some misalignment problems



# Training

## ■ Data Augmentation

- Change gamma, saturation, gaussian blur/noise, etc.
- Random crop
  - Sparse and unlabelled pedestrians
  - Larger batch size can benefit BN[S. Ioffe, ICML'15] training

## ■ Multi-label

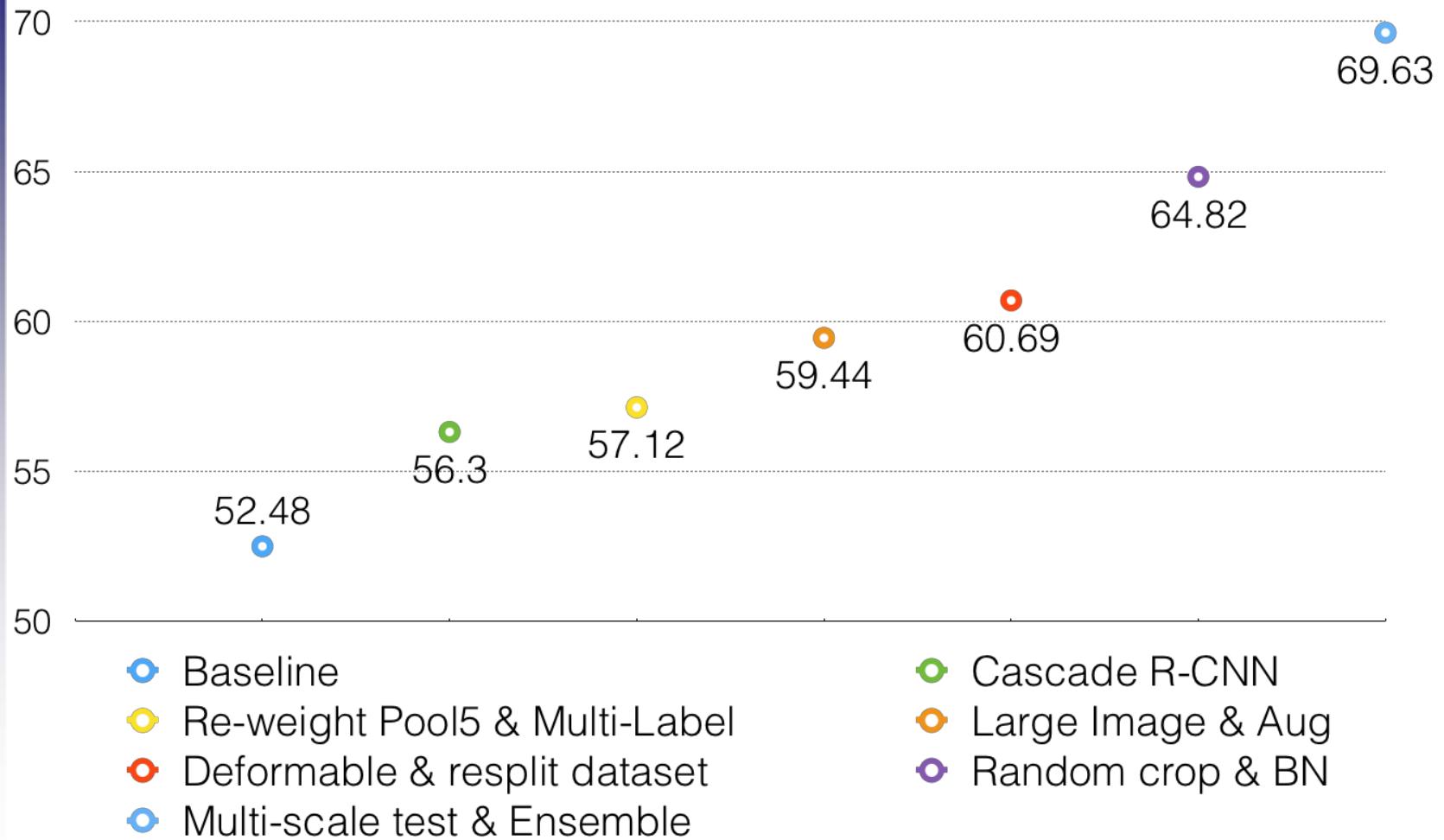
- Regarding pedestrian and cyclist as different labels when training
- Learning more discriminative features

# Testing

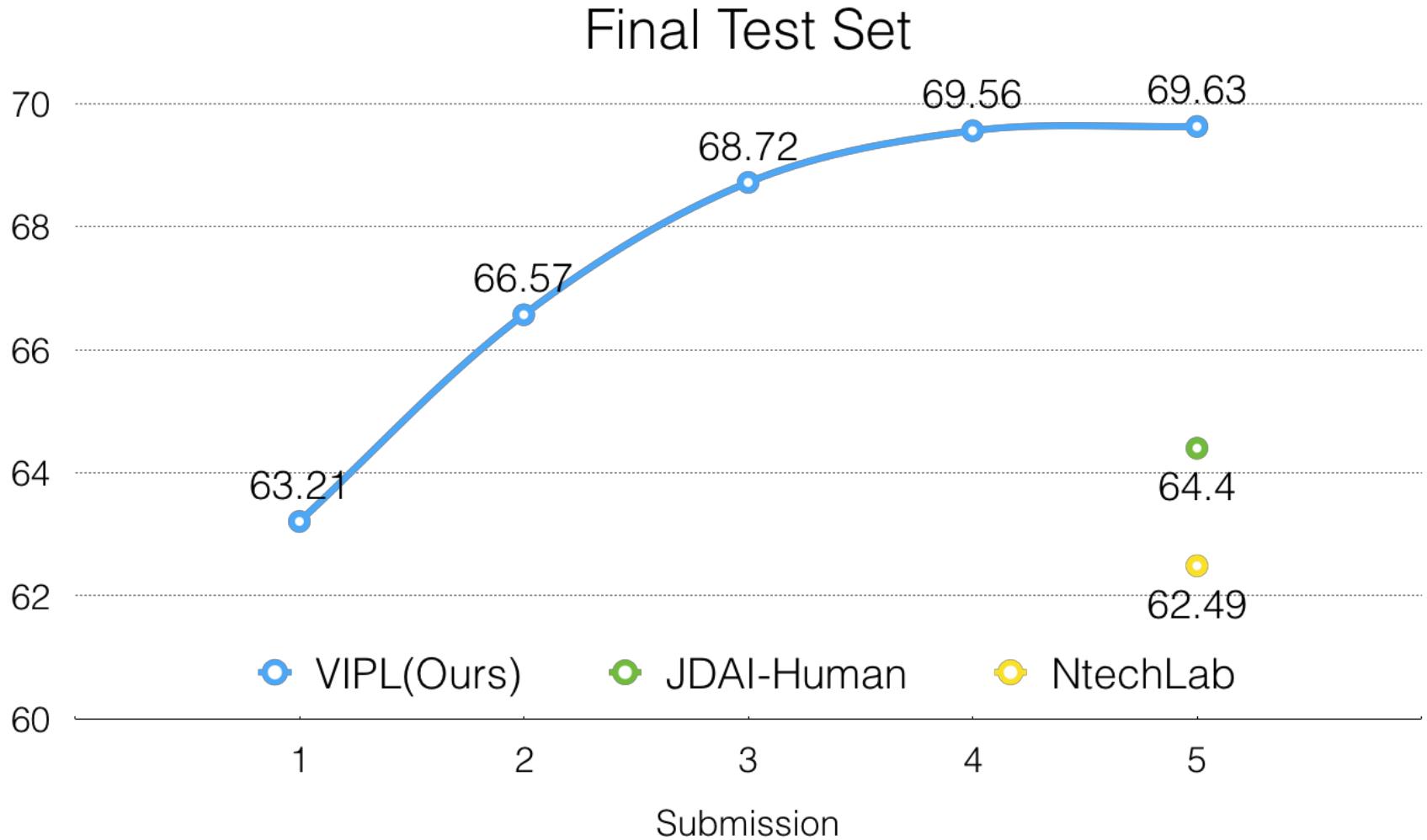
- Multi scale testing(4 scale + flip)
  - Merge results from different scales, then do soft-nms[N. Bodla, CVPR'18]
  - Box-voting: Averaging coordinates between result boxes and candidates
- Ensemble
  - Split the network into RPN-net and RCNN-net
  - Select proposals from all RPN-net, put them into RCNN-net and get results
  - Averaging score and coordinates

# Results

Validation Set



# Results





# Conclusion

- According to the AP metric, Cascade RCNN is adopted for better localization performance
- Random cropping is not only a method for data augmentation but also handles sparse/unlabelled persons and brings gains because of larger batch size
- Testing tools are very powerful, but you have to be careful since the devil is in the detail



Q & A

THANKS

# References

- [K. He, CVPR'16] Kaiming He, et al. Deep Residual Learning for Image Recognition. CVPR16.
- [J. Dai, ICCV'17] Jifeng Dai, et, al. Deformable Convolutional Networks. ICCV17.
- [T. Lin, CVPR'17] Tsung-Yi Lin, et, al. Feature Pyramid Networks for Object Detection. CVPR17.
- [Z. Cai, CVPR'18] Zhaowei Cai, et, al. Cascade R-CNN: Delving into High Quality Object Detection. CVPR18.
- [K. He, ICCV'17] Kaiming He, et, al. Mask R-CNN. ICCV17.
- [S. Zhang, CVPR'18] Shanshan Zhang, et, al. Occluded Pedestrian Detection Through Guided Attention in CNNs. CVPR18.
- [S. Ioffe, ICML'15] Sergey Ioffe, et al. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML15.
- [N. Bodla, CVPR'18] Navaneeth Bodla, et al. Soft-NMS -- Improving Object Detection With One Line of Code. CVPR18.