A Solution for Densely Annotated Large Scale Object Detection Task

Yuan Gao, Hui Shen, Donghong Zhong, Jian Wang,

Zeyu Liu, Ti Bai, Xiang Long and Shilei Wen

Baidu 百度

百度视觉技术部

COMPUTER VISION TECHNOLOGY

PaddlePaddle

Object 365 Dataset

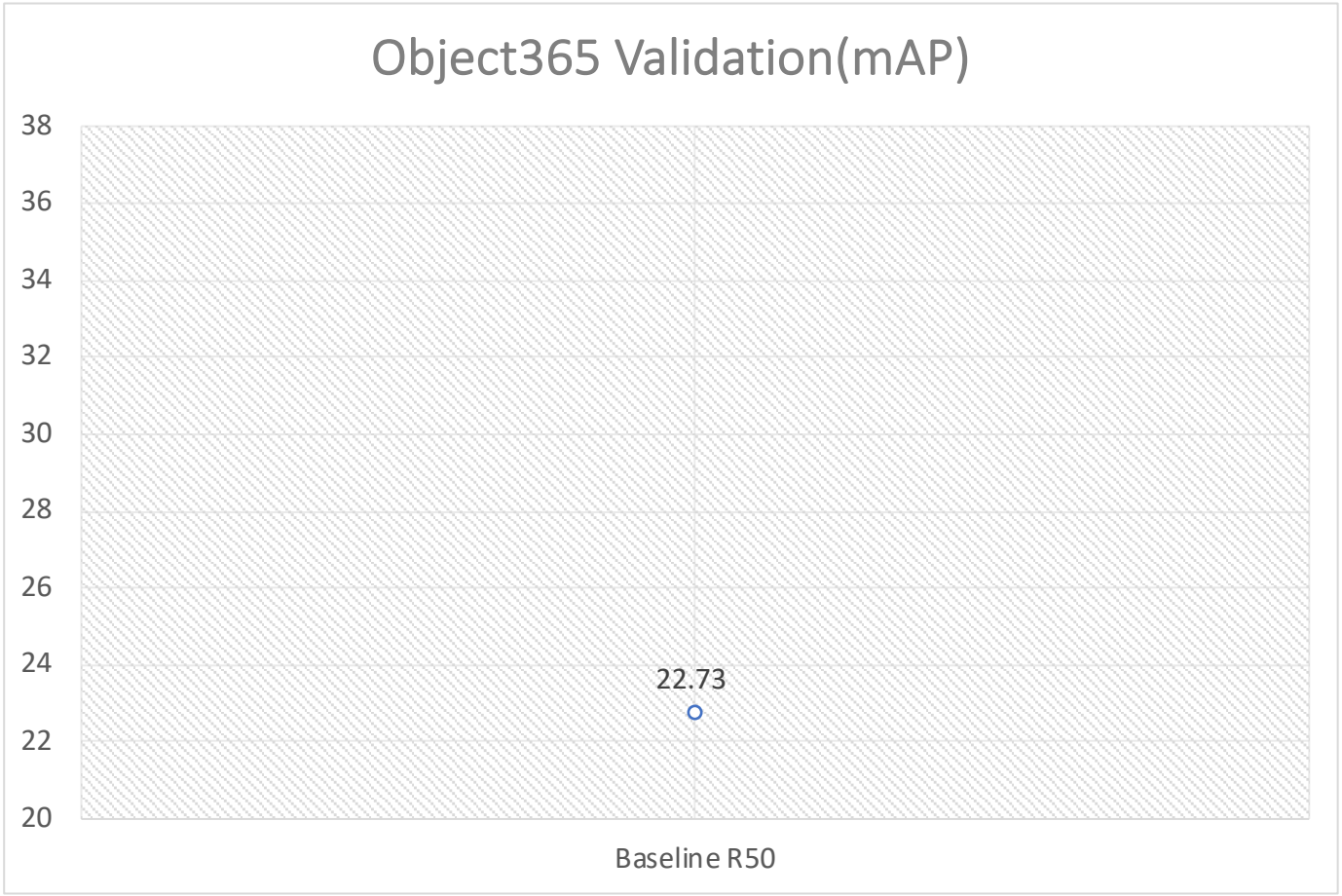
Pretrain	# Class	# Image	# Box in Total	Box Num Avg	Image Height Avg	Image Width Avg	Box Area Avg (Pixel)	Max # Box
COCO17 (Train)	80	118287	0.86M	7.27	484	577	12025	93
Object 365 (Train)	365	608606	9.62M	15.81	536	662	14074	835

Full Track

百度视觉技术部
COMPUTER VISION TECHNOLOGY

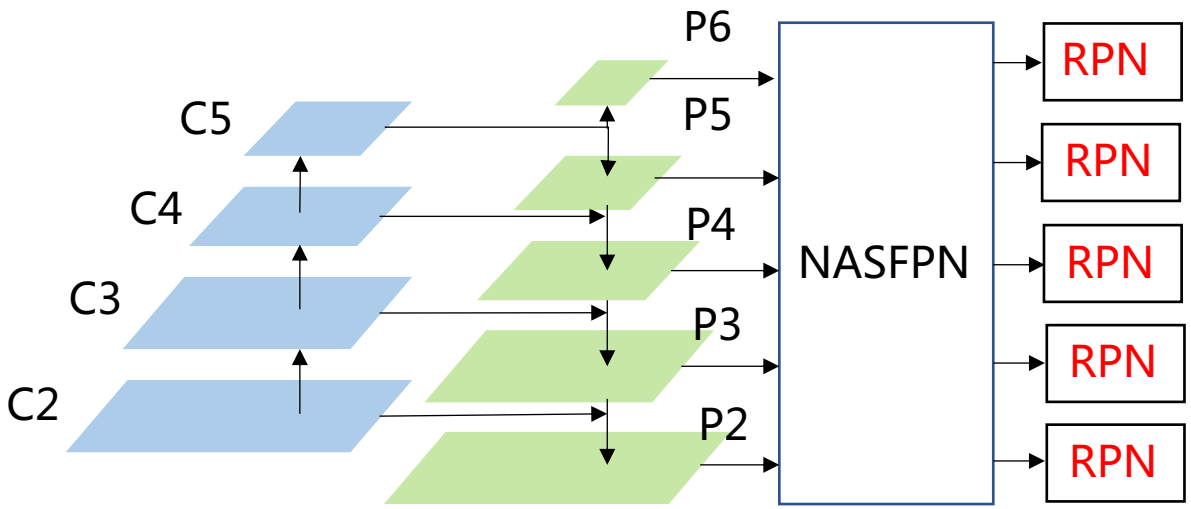
PaddlePaddle

R50 Cascade RCNN



Neural Architecture Search

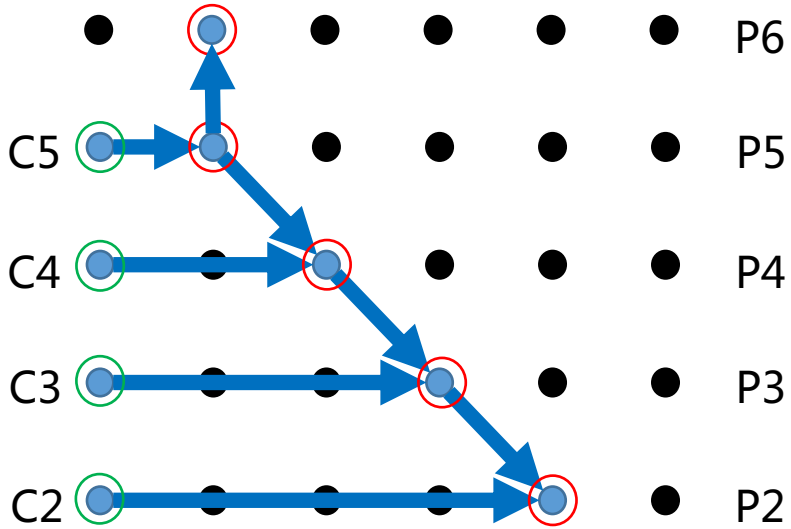
- An RL based Neural Architecture Search is adopted.
- The NAS-FPN module is directly cascaded behind the original FPN module.



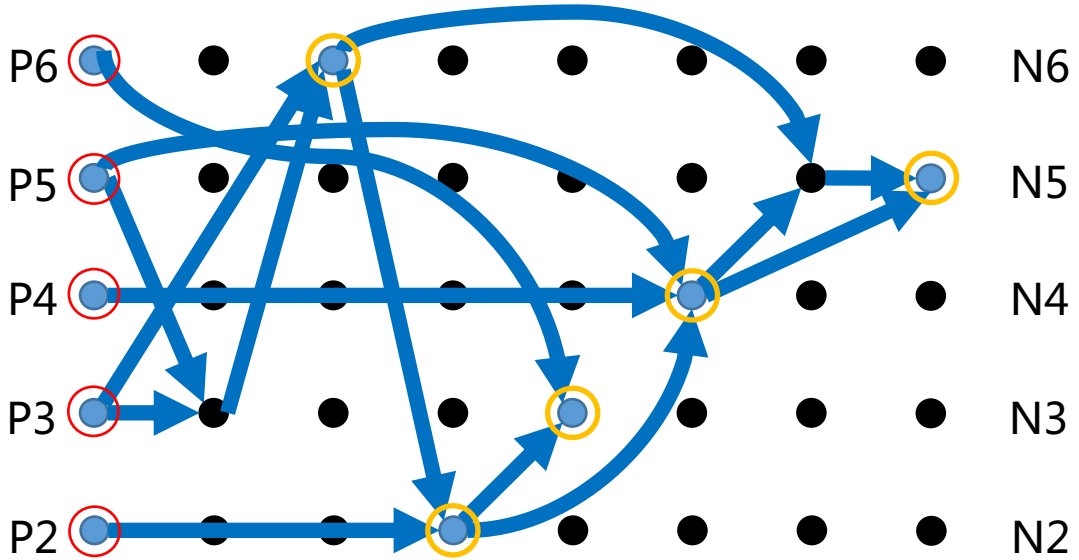
- A strong architecture found by prior knowledge[1] is used to initialized the NAS-FPN searching procedure.

[1] Ghiasi G, Lin T Y, Pang R, et al. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. arXiv:1904.07392, 2019.

Neural Architecture Search



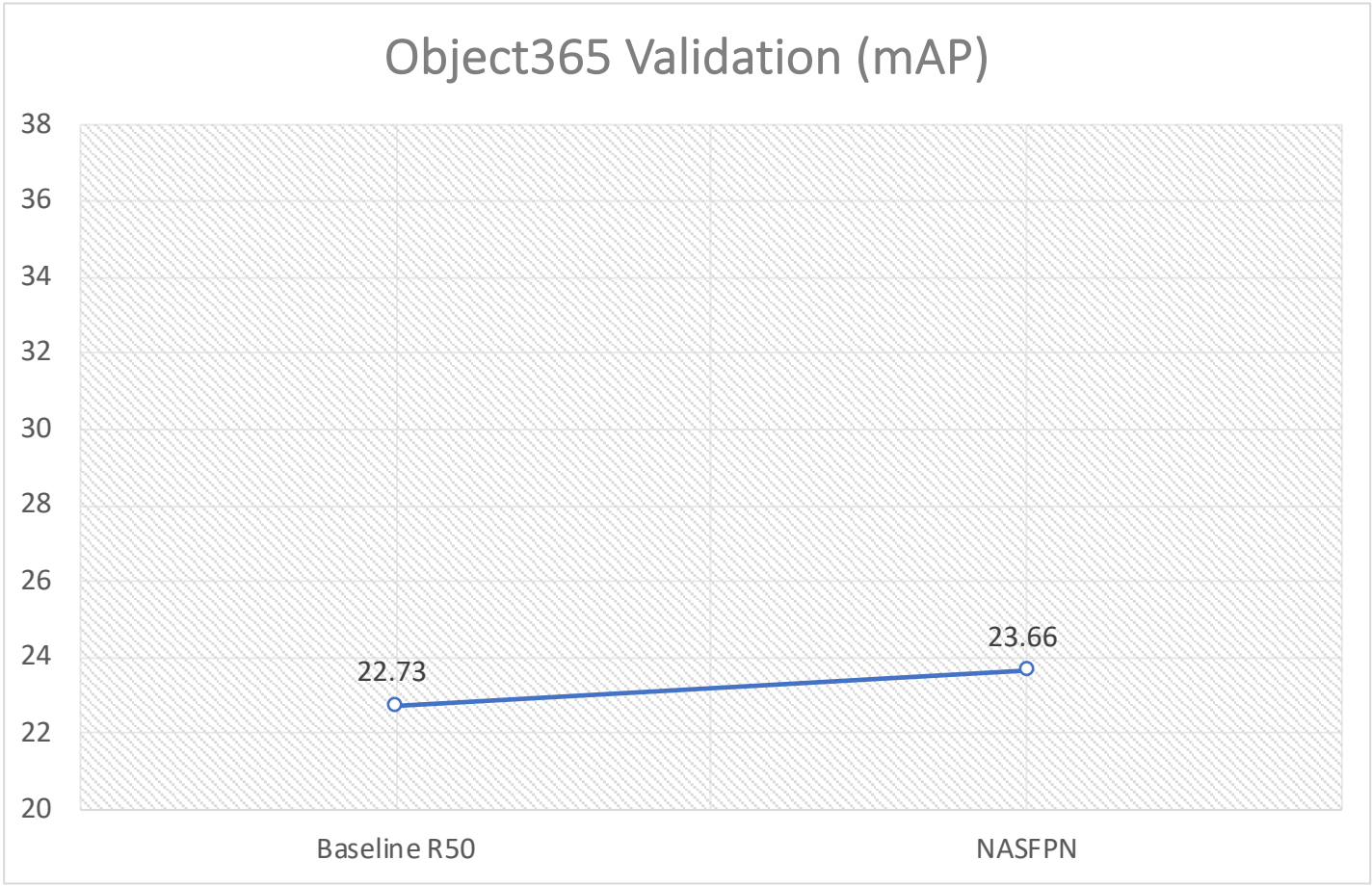
(a) FPN



(b) Our NAS-FPN

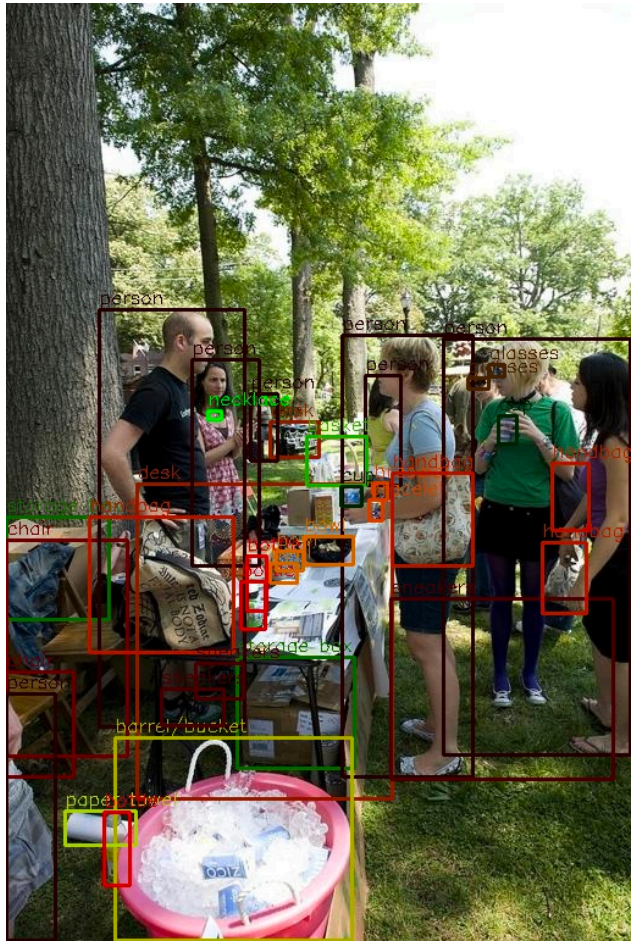
The architecture graph of original FPN and NAS-FPN after ~400 episodes

Neural Architecture Search



Class Diversity Sensitive Sampling

15 Classes

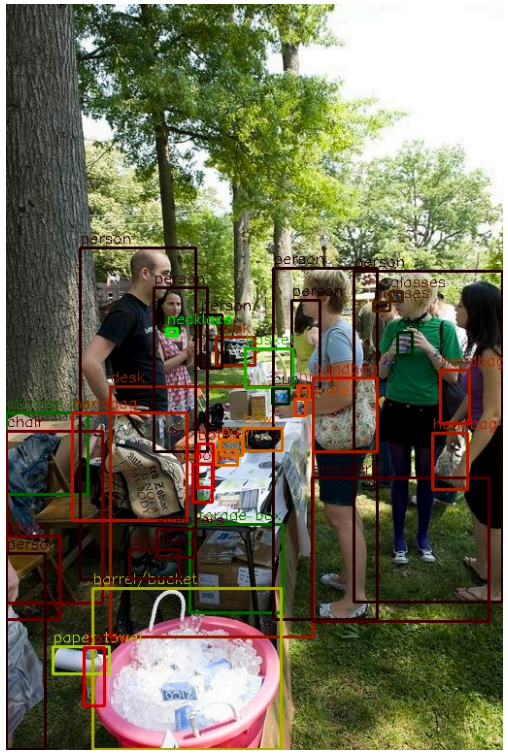


4 Classes

Sampling probability equally is not appropriate.

Class Diversity Sensitive Sampling

The i th image contains 15 Classes



$$W_i = \ln(C_i + \varepsilon) \sum_{c=1}^{N} P_c H_c$$

W_i = Sampling weight of the i th image.

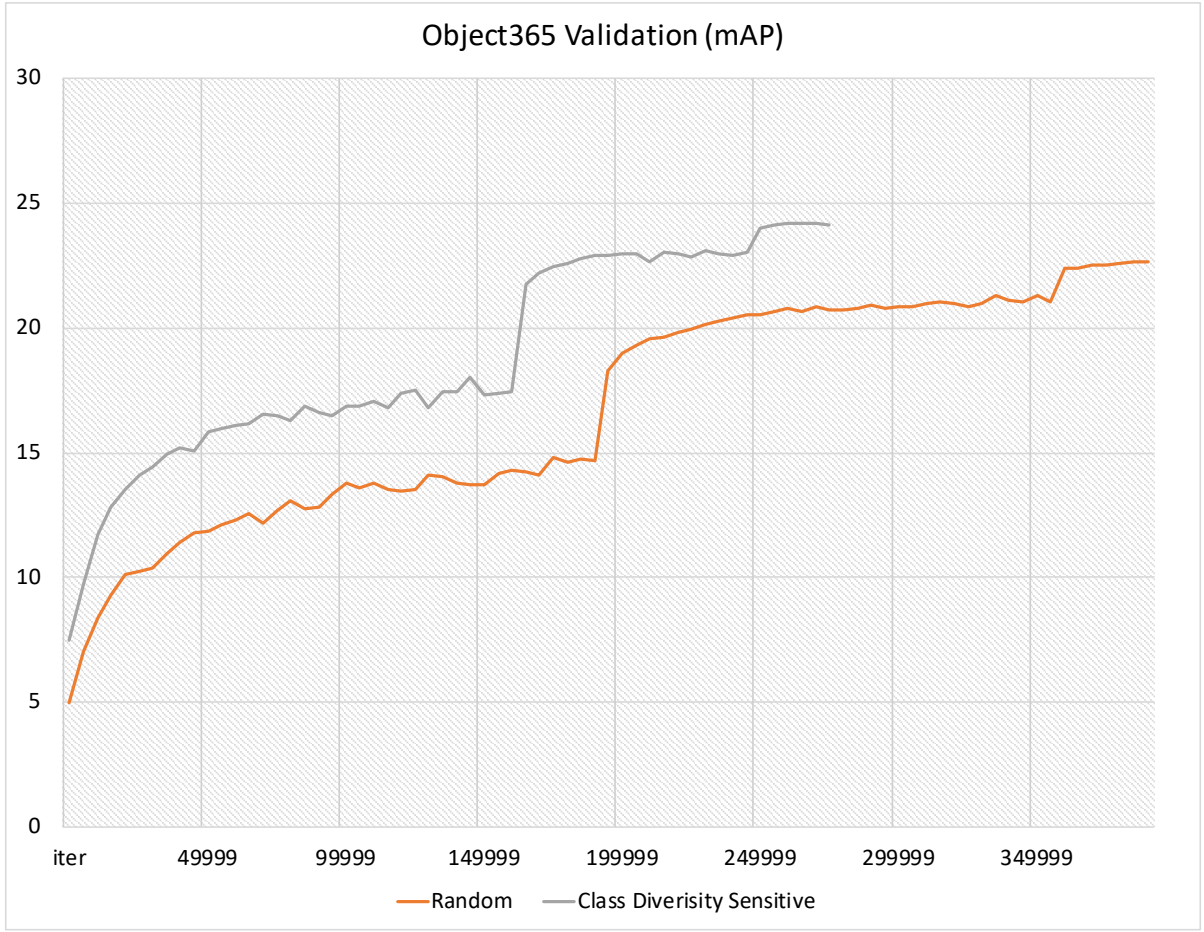
C_i = Total number of the classes of the i th image.

N = Total number of the classes of the dataset.

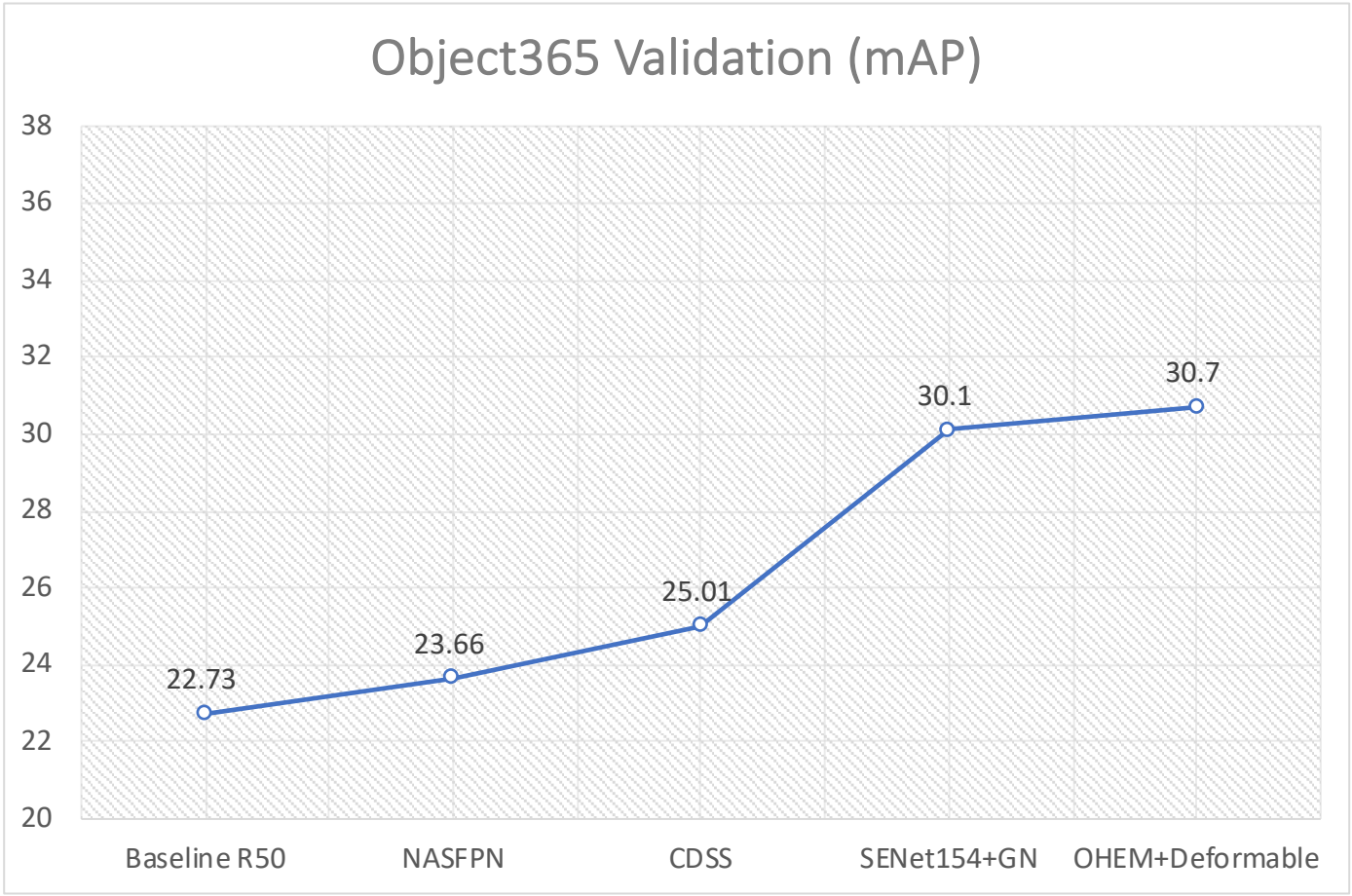
P_c = The c th class prior probability, according to the total box number of the dataset.

H_c = 1 if the images contains class c or 0.

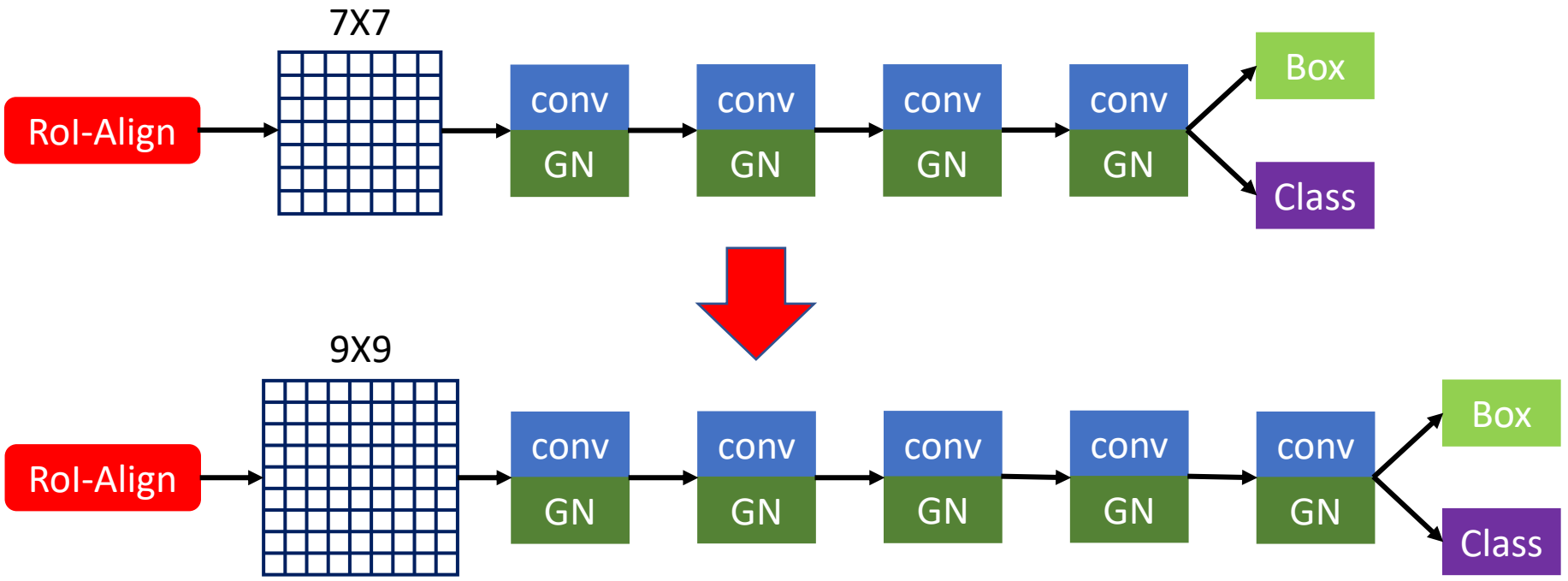
Class Diversity Sensitive Sampling



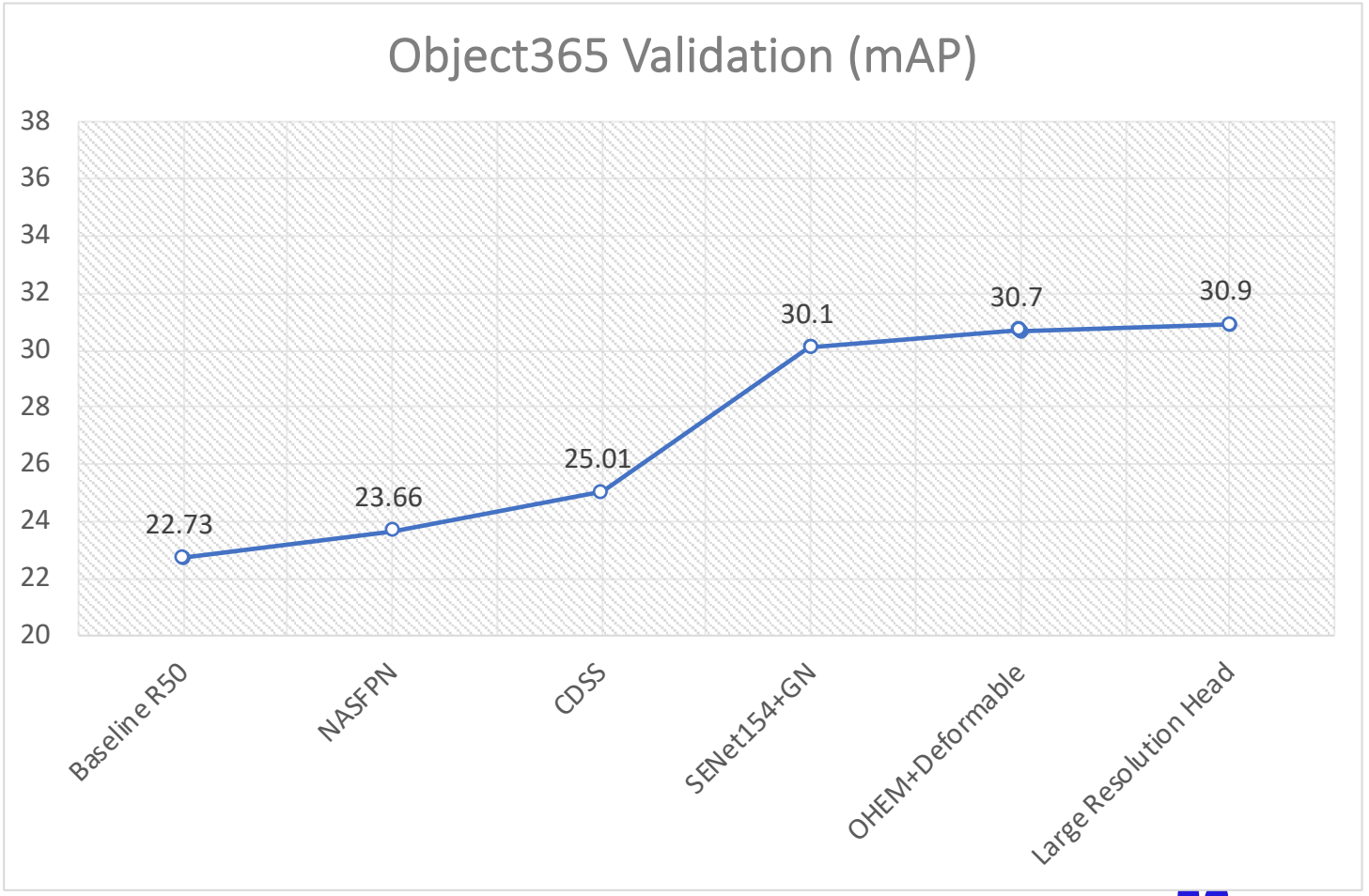
Class Diversity Sensitive Sampling



Large Resolution Box Head

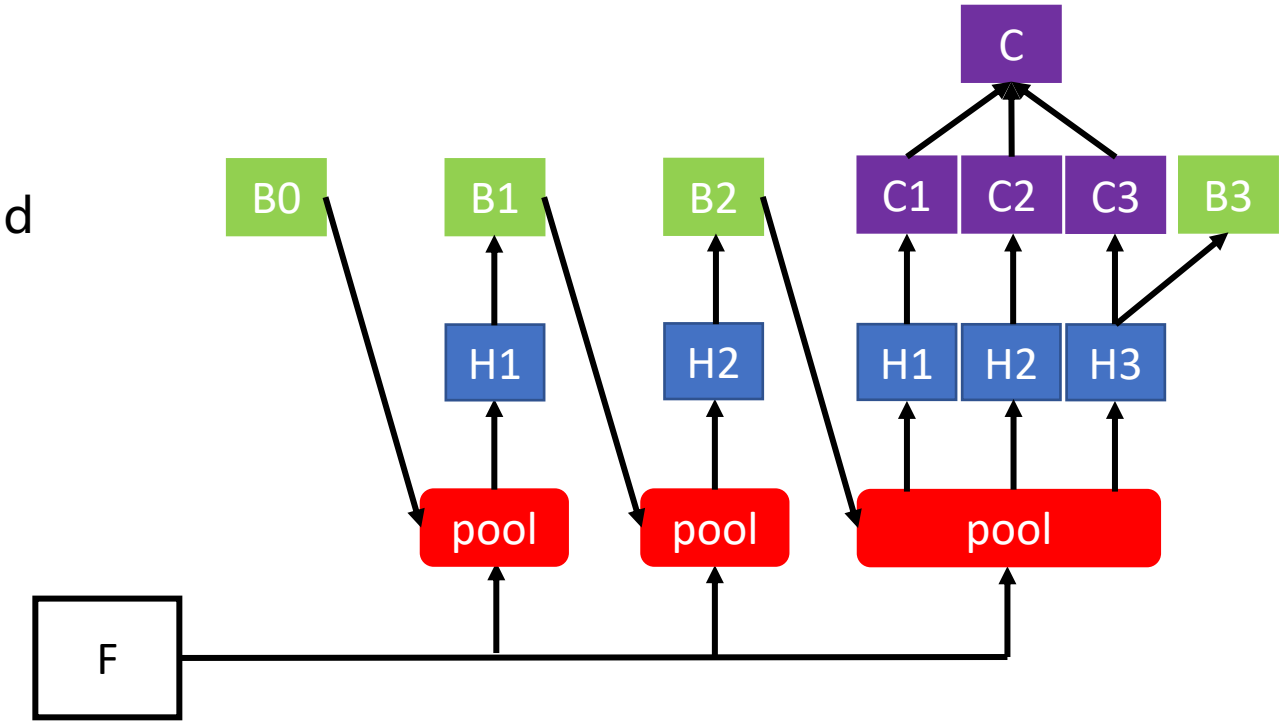


Large Resolution Box Head



Cascade RCNN Testing

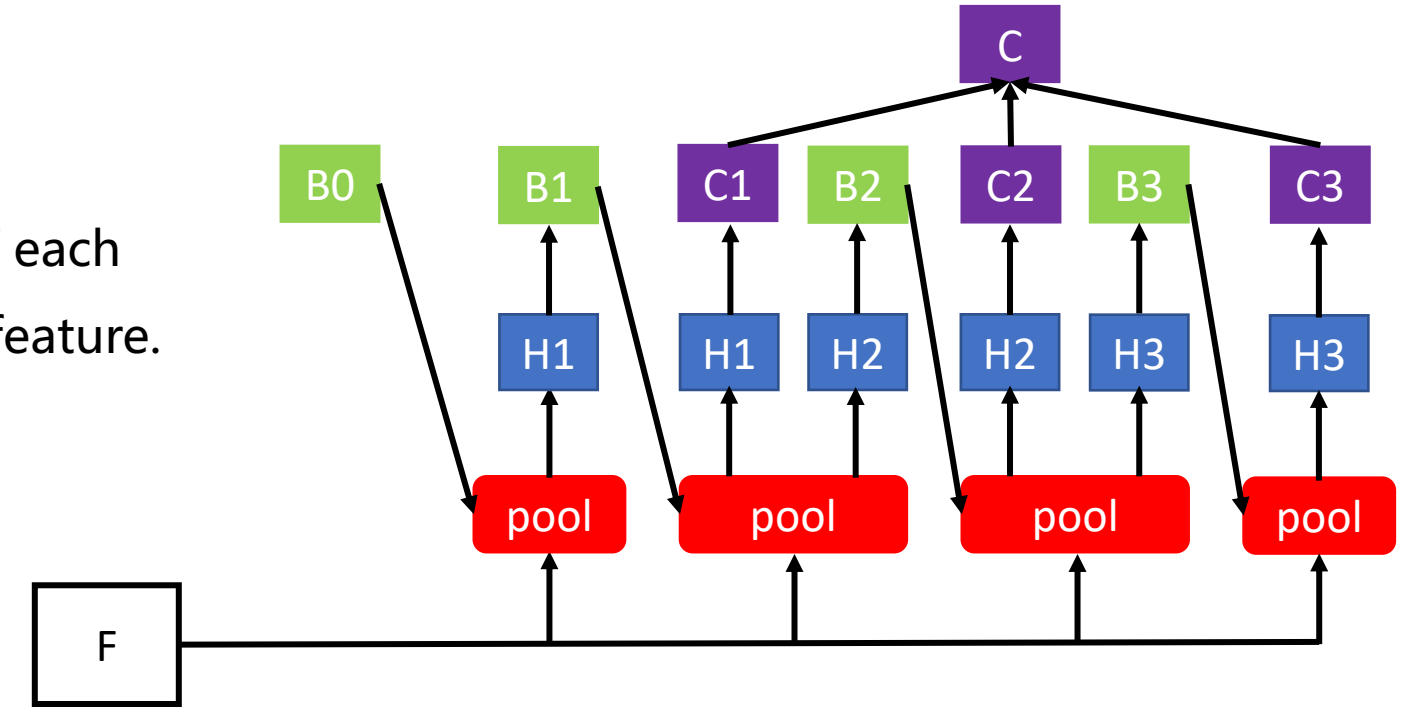
- Use the predicted bbox of the 2nd stage to extract the feature. (Standard Cascade RCNN)



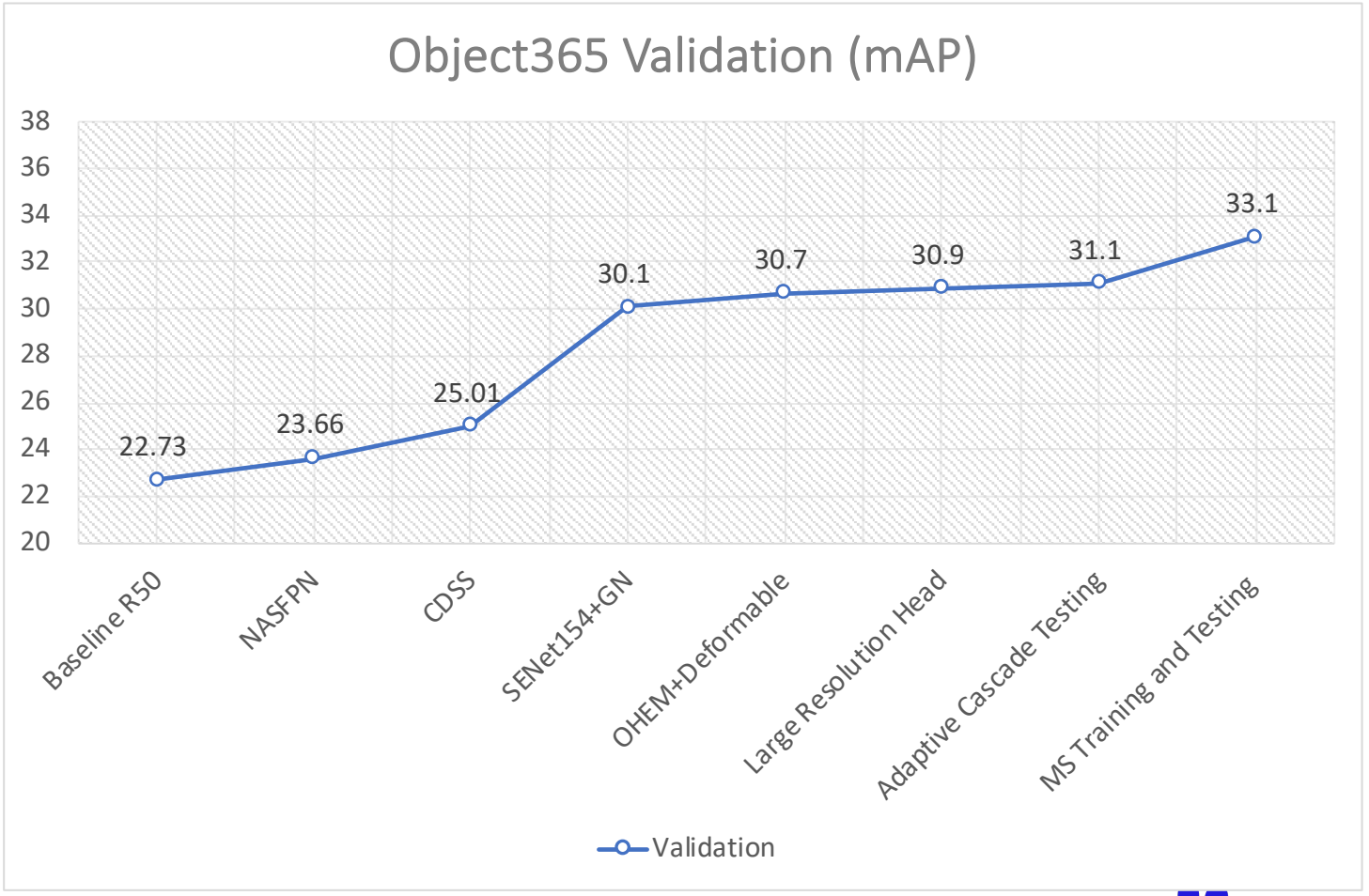
[1] Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. CVPR.2018

Cascade RCNN Adaptive Testing

- Use the predicted bbox of each stage itself to extract the feature.



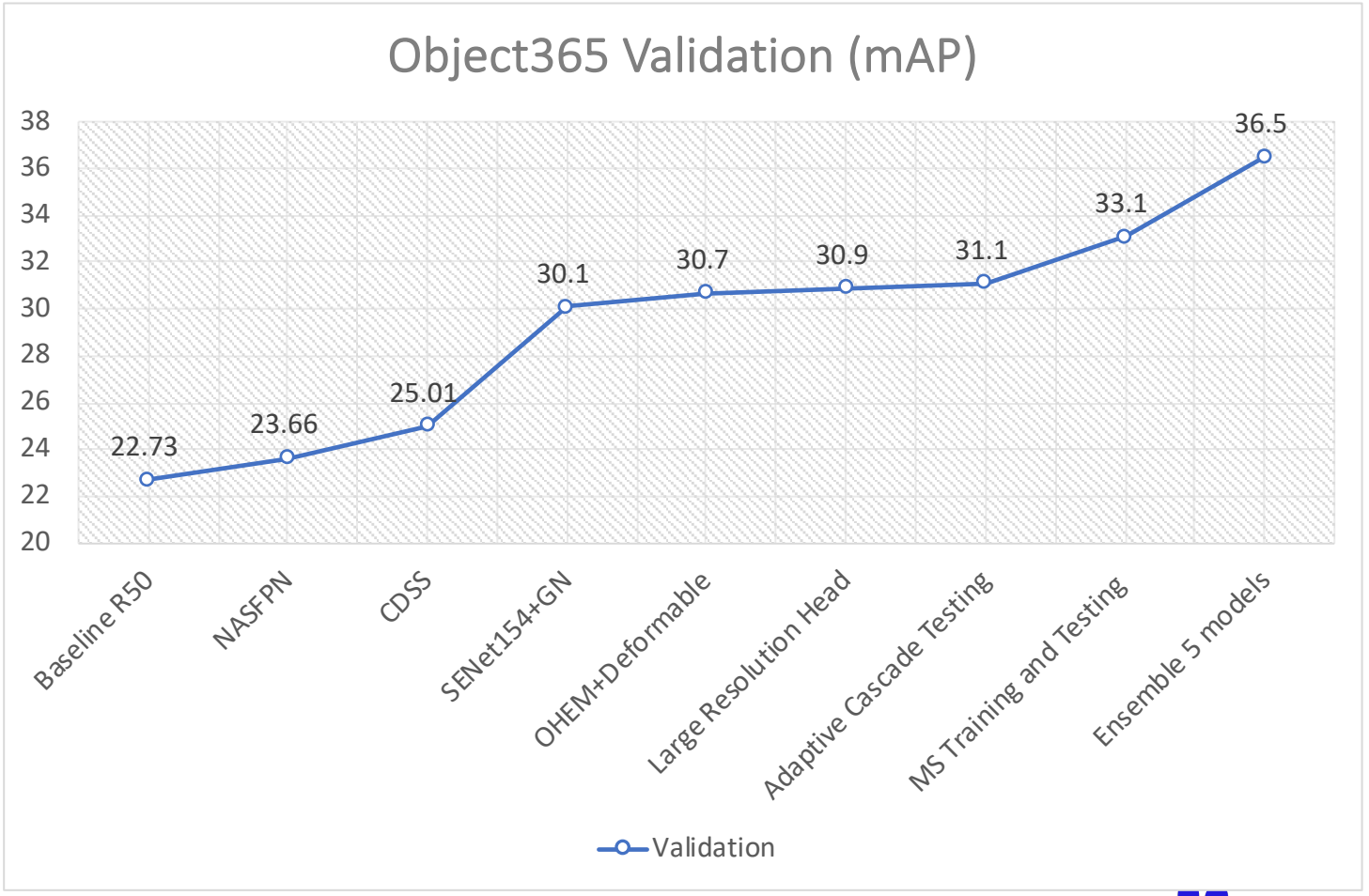
Cascade RCNN Adaptive Testing



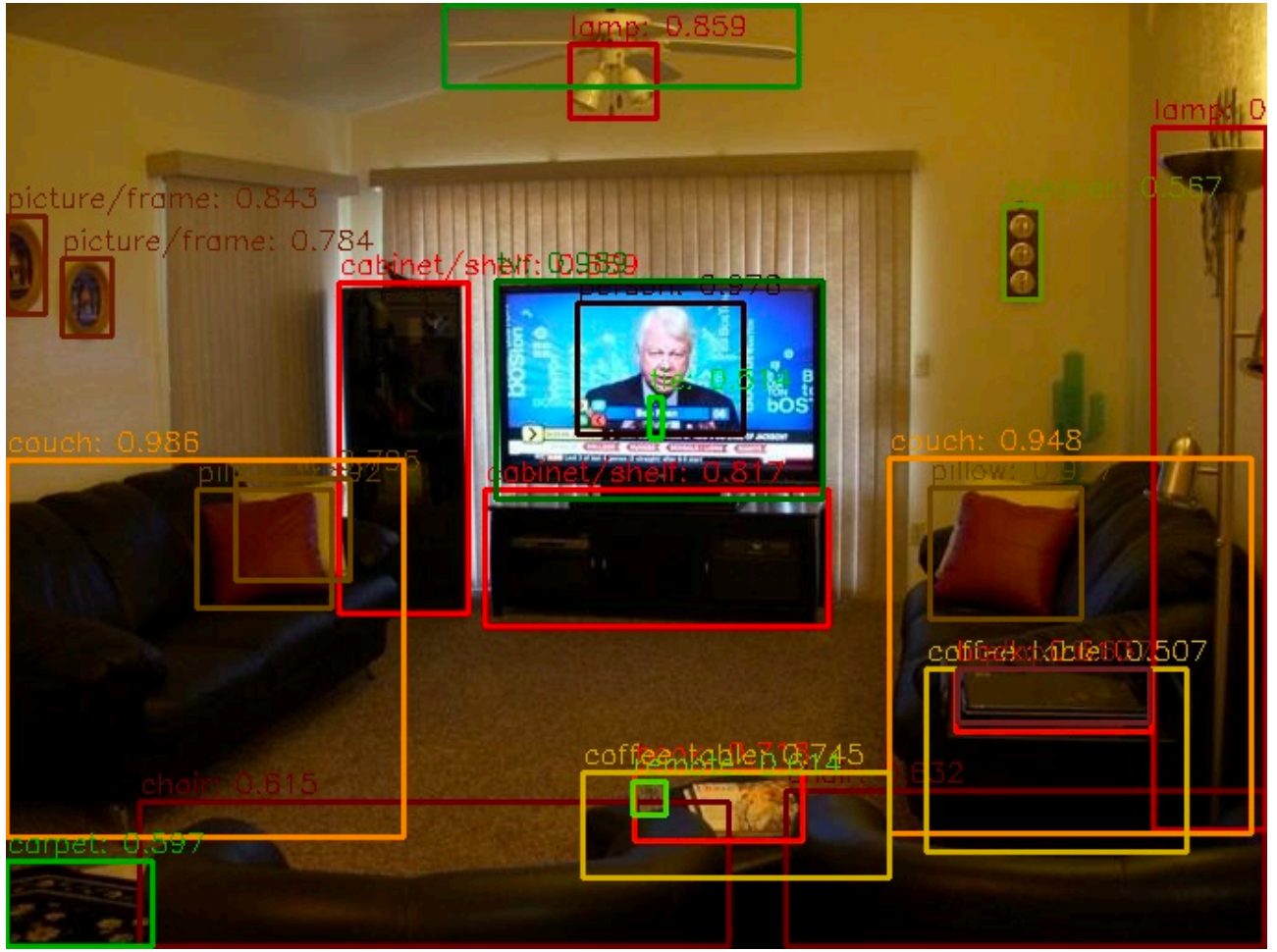
Implementation Details

- Use COCO Pretrained model, mAP 52.9 on COCO17 minival.
- Training multiscale size (400, 1400), max size 1600.
- Testing multiscale size (400, 1400), max size 2100.
- 8 V100(32GB) x 2 for 7 days.
- Weight Standardization brings model diversity.
- SoftNMS is adopted.

Implementation Details



Visualization



Visualization

Tiny Track

百度视觉技术部
COMPUTER VISION TECHNOLOGY

PaddlePaddle

Full Track Pretrain

Pretrain	Full Val mAP	Tiny Val mAP	Gain	Tiny Test mAP
COCO Pretrain	-	28.9	-	-
Obj365 Full Pretrain	30.7	33.0	+4.1	-
Obj365 Full Pretrain	32.9	34.8	+5.9	-
Ensemble 8 models	-	37.6	+8.7	29.0

- Multiscale input with flip in Training and Testing

Visualization

- Full Track Pretrained

- COCO17 Pretrained

Paddle Paddle Detection

- Fast/Faster R-CNN, FPN, Mask RCNN, Cascade R-CNN, Yolo v3, RetinaNet, SSD ……
- GN, SyncBN, Deformable Conv v1/v2 ……
- https://github.com/PaddlePaddle/models/tree/develop/PaddleCV/object_detection
- Training framework will be released soon.

PaddlePaddle

Thank you!

Please feel free to contact us, if you have any questions.

gaoyuan18@baidu.com

Baidu 百度

百度视觉技术部
COMPUTER VISION TECHNOLOGY

PaddlePaddle