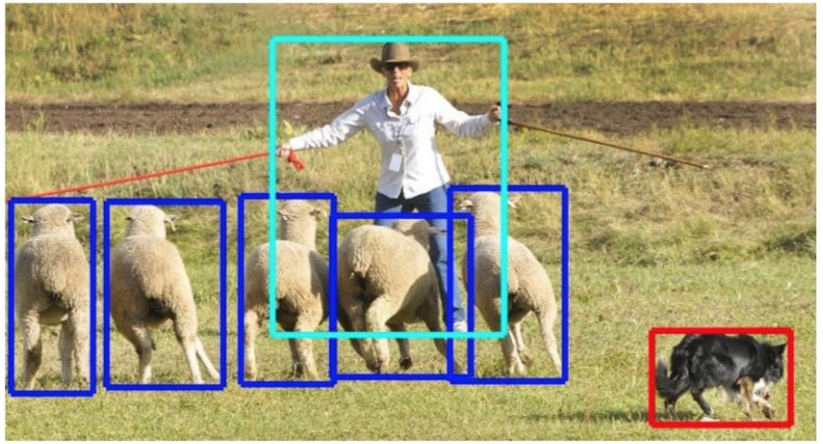
Recent Progress in Object Detection

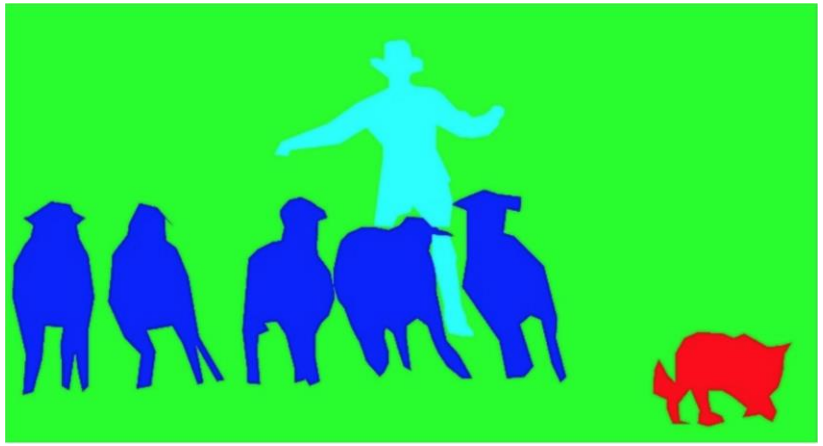
Jiaqi Wang
Multimedia Laboratory
The Chinese University of Hong Kong

Task definition

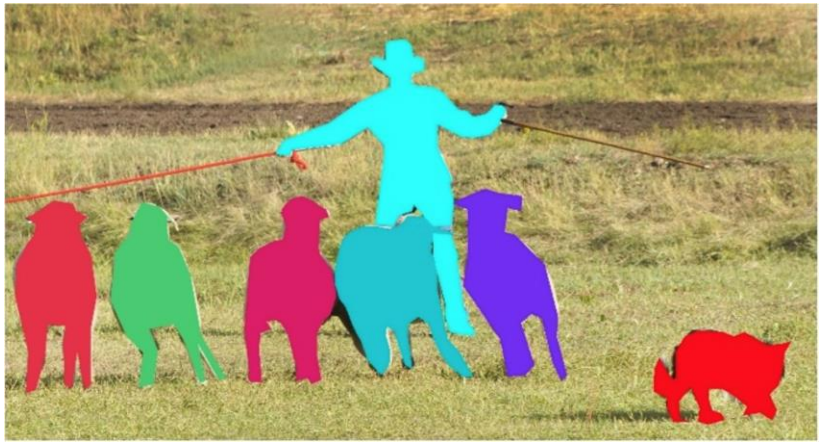
Image classification



Object detection

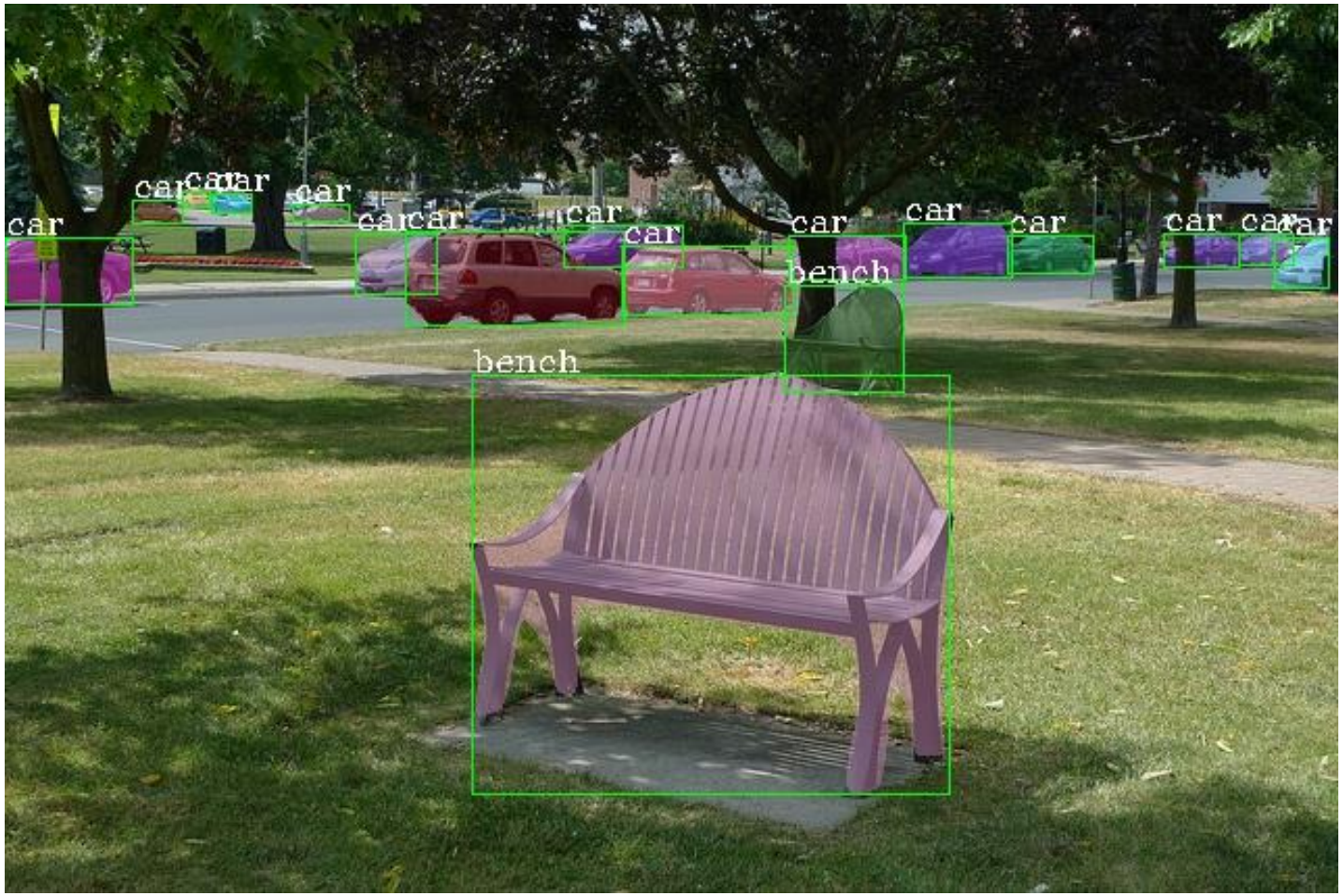


Semantic segmentation

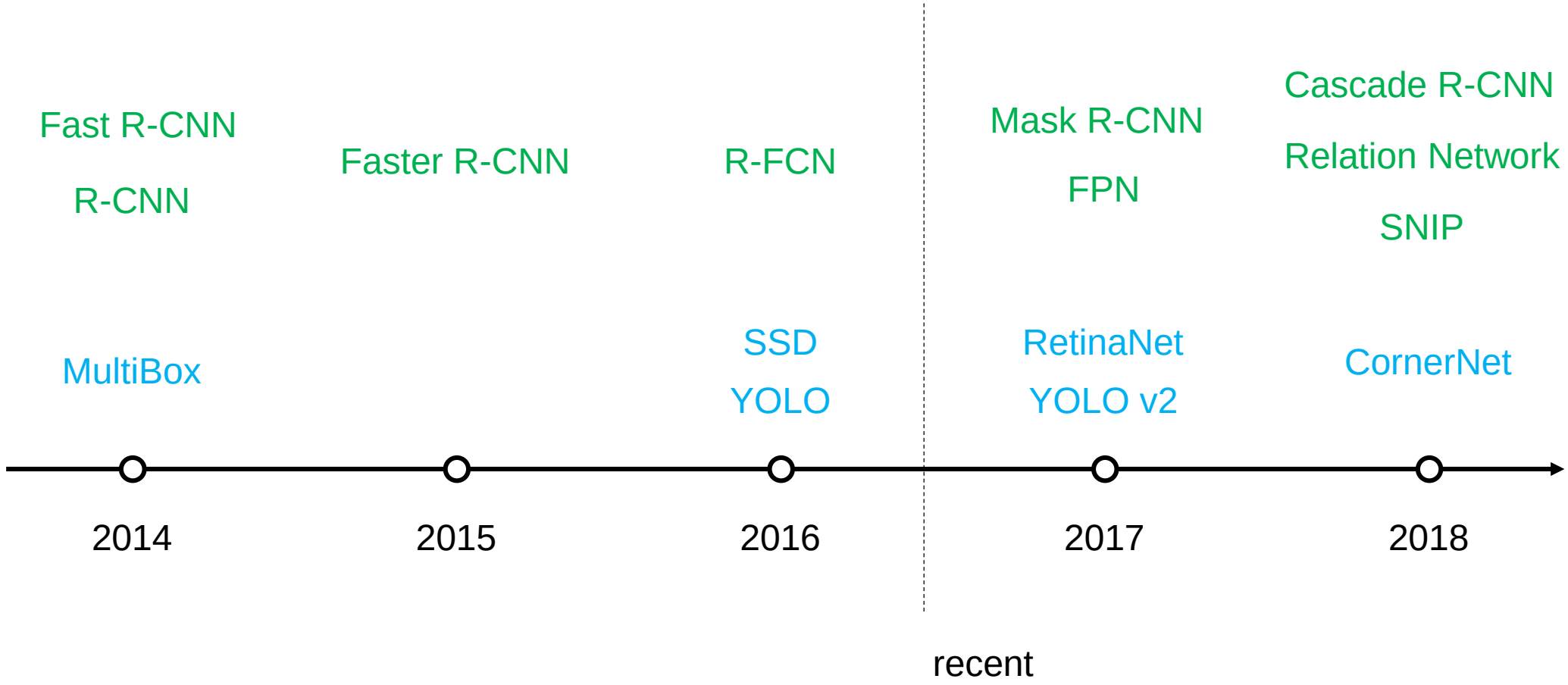


Instance segmentation

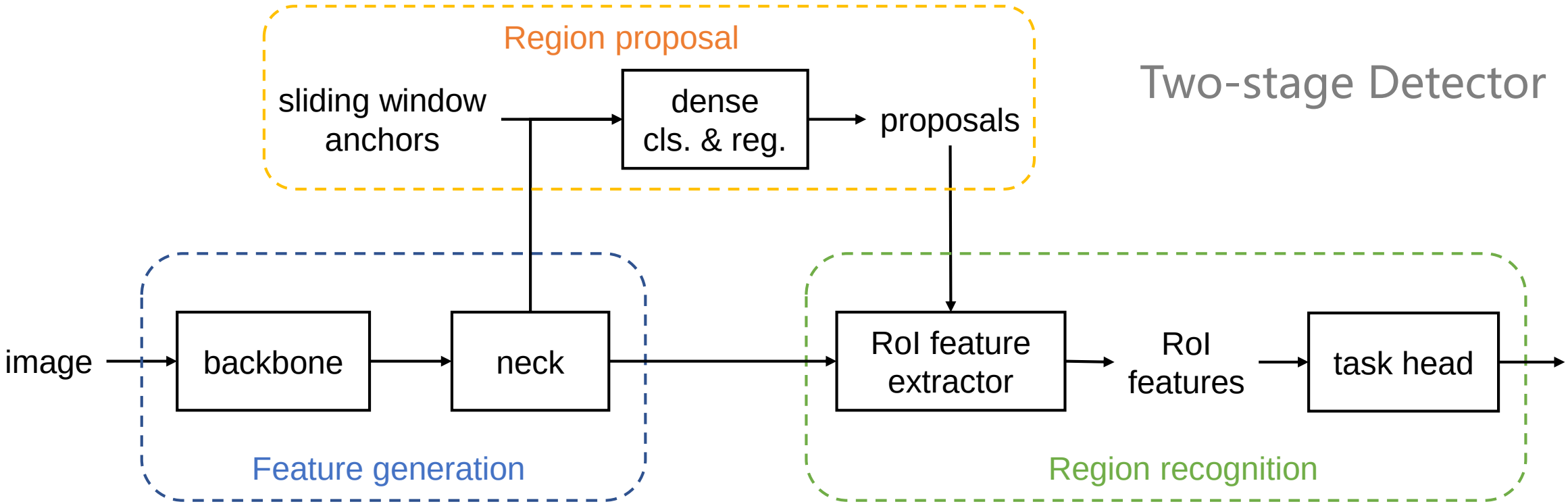
Task definition



Progress

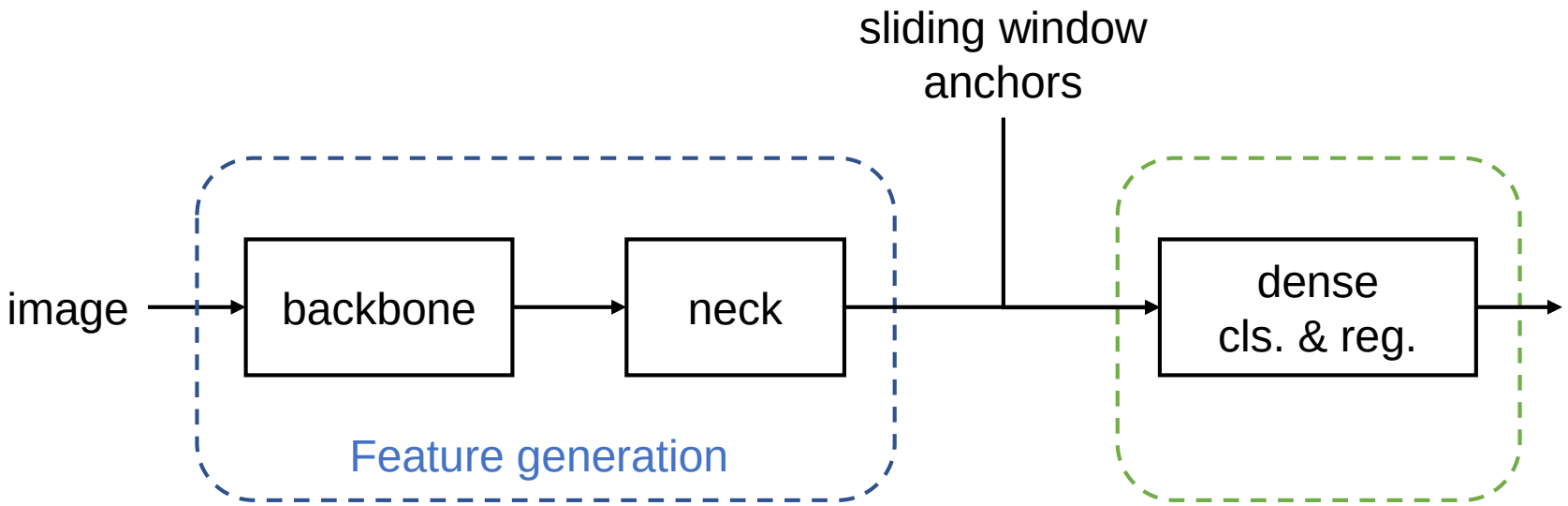

Fast R-CNN
R-CNN
Faster R-CNN
R-FCN
Mask R-CNN
FPN
Cascade R-CNN
Relation Network
SNIP
MultiBox
SSD
YOLO
RetinaNet
YOLO v2
CornerNet
2014
2015
2016
2017
2018
recent

General pipeline



General pipeline

Single-stage Detector

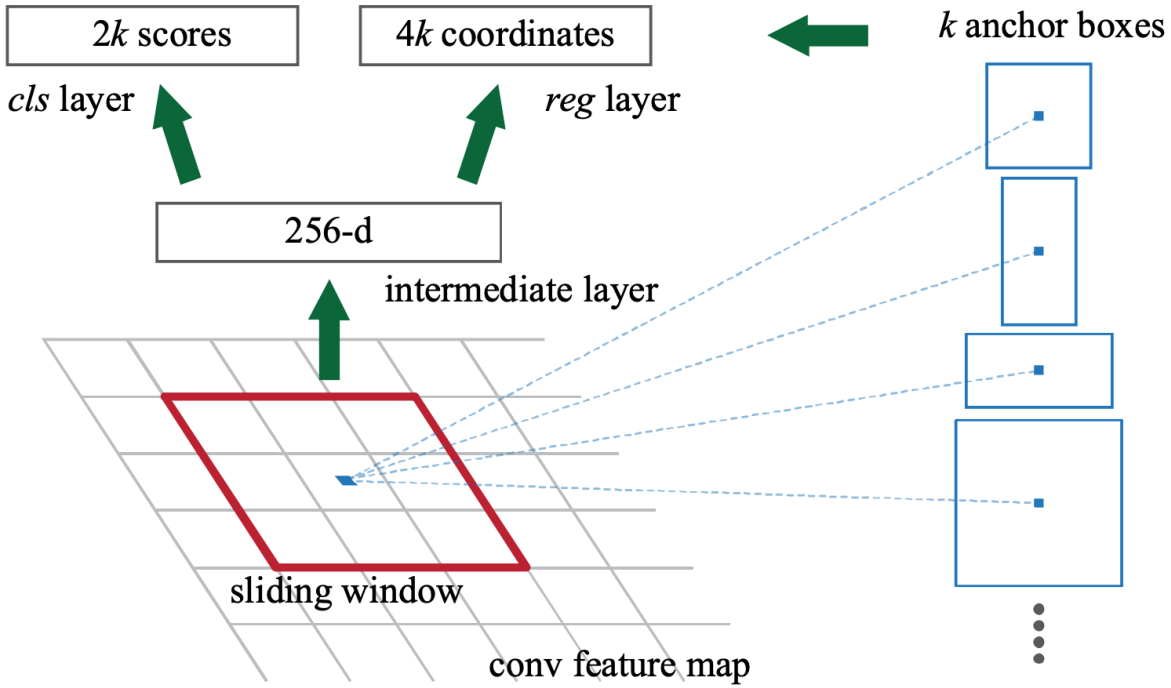


Faster R-CNN

- Region Proposal Network (RPN)
- Training pipeline

Faster R-CNN

- RPN

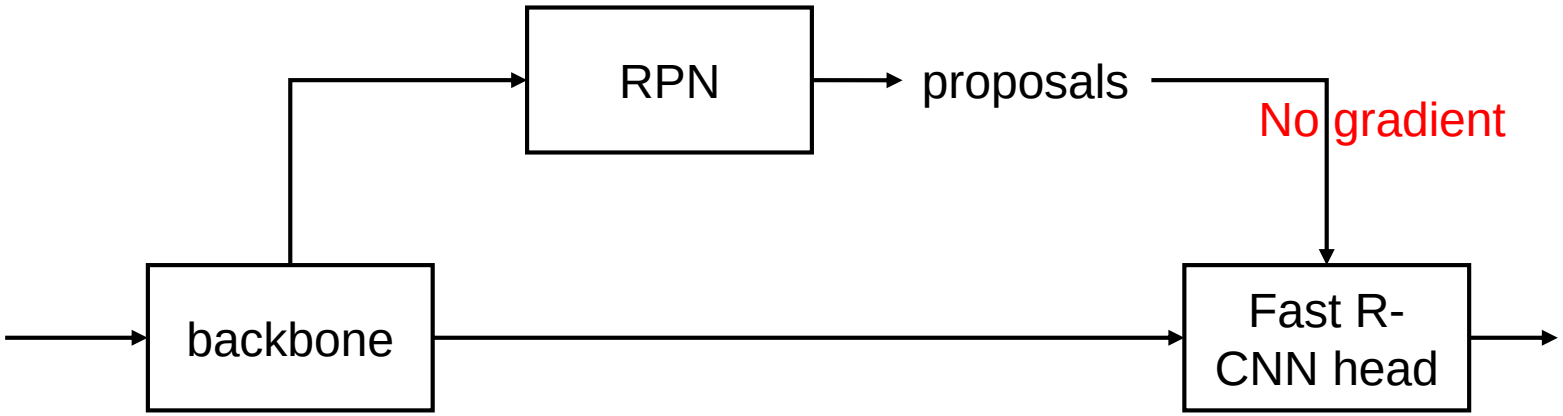


Faster R-CNN

Training pipeline

- Joint training: multi-task

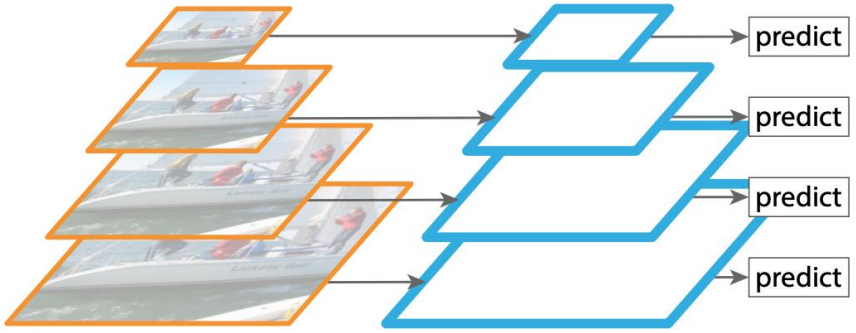
  preferred



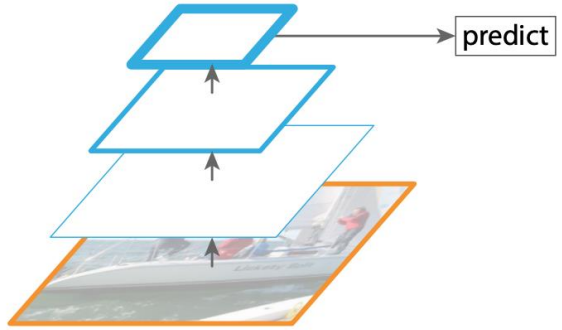
Feature Pyramid Network (FPN)

- Top-down pathway
- Multi-level prediction

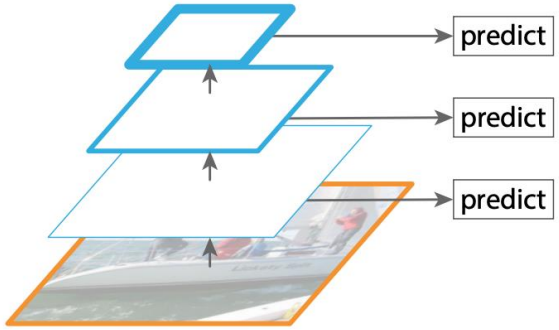
Feature Pyramid Network (FPN)



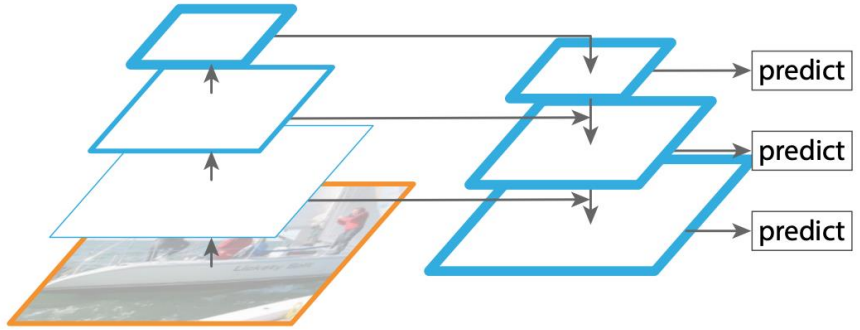
(a) Featurized image pyramid



(b) Single feature map

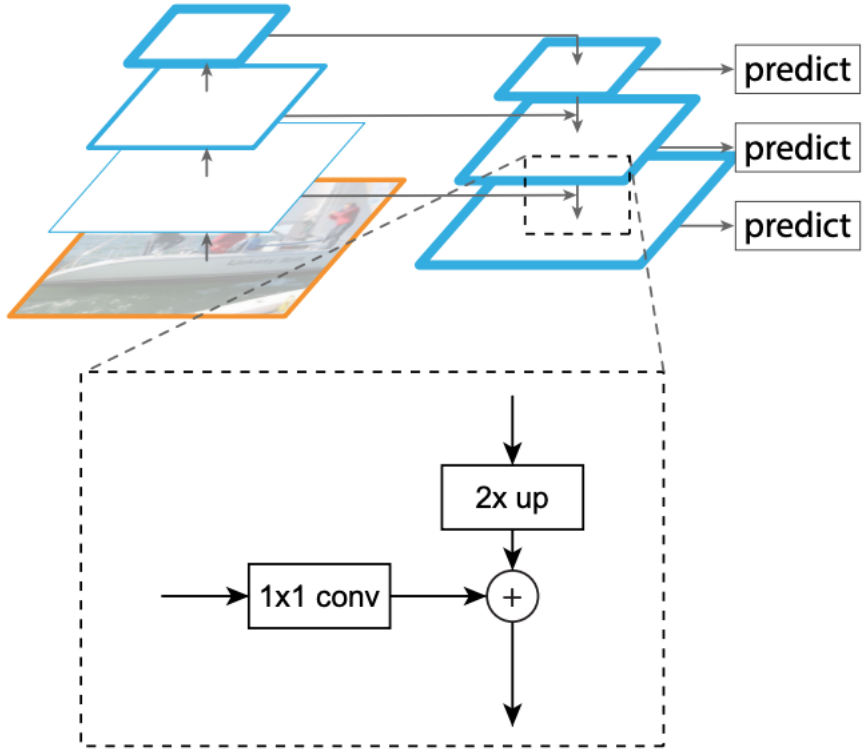


(c) Pyramidal feature hierarchy



(d) Feature Pyramid Network

Feature Pyramid Network (FPN)

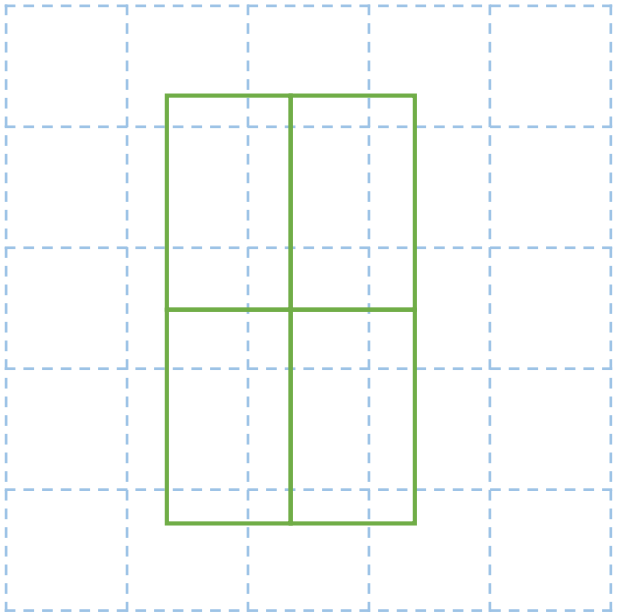


Mask R-CNN

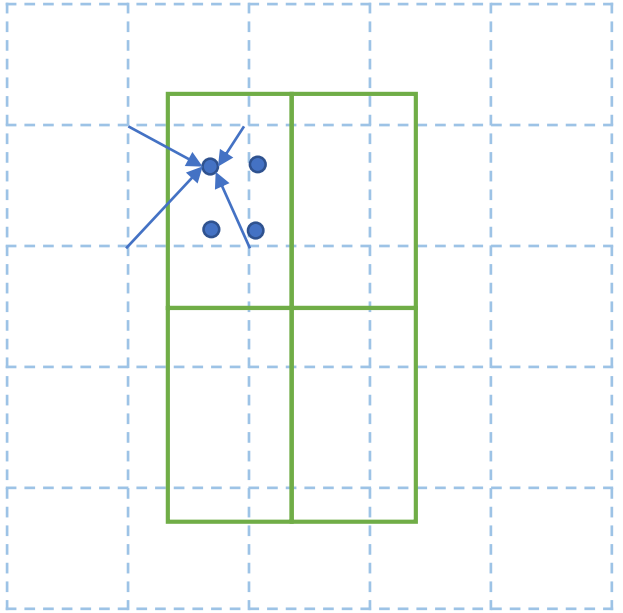
- RoIAlign
- Mask branch

Mask R-CNN

RoI Pooling

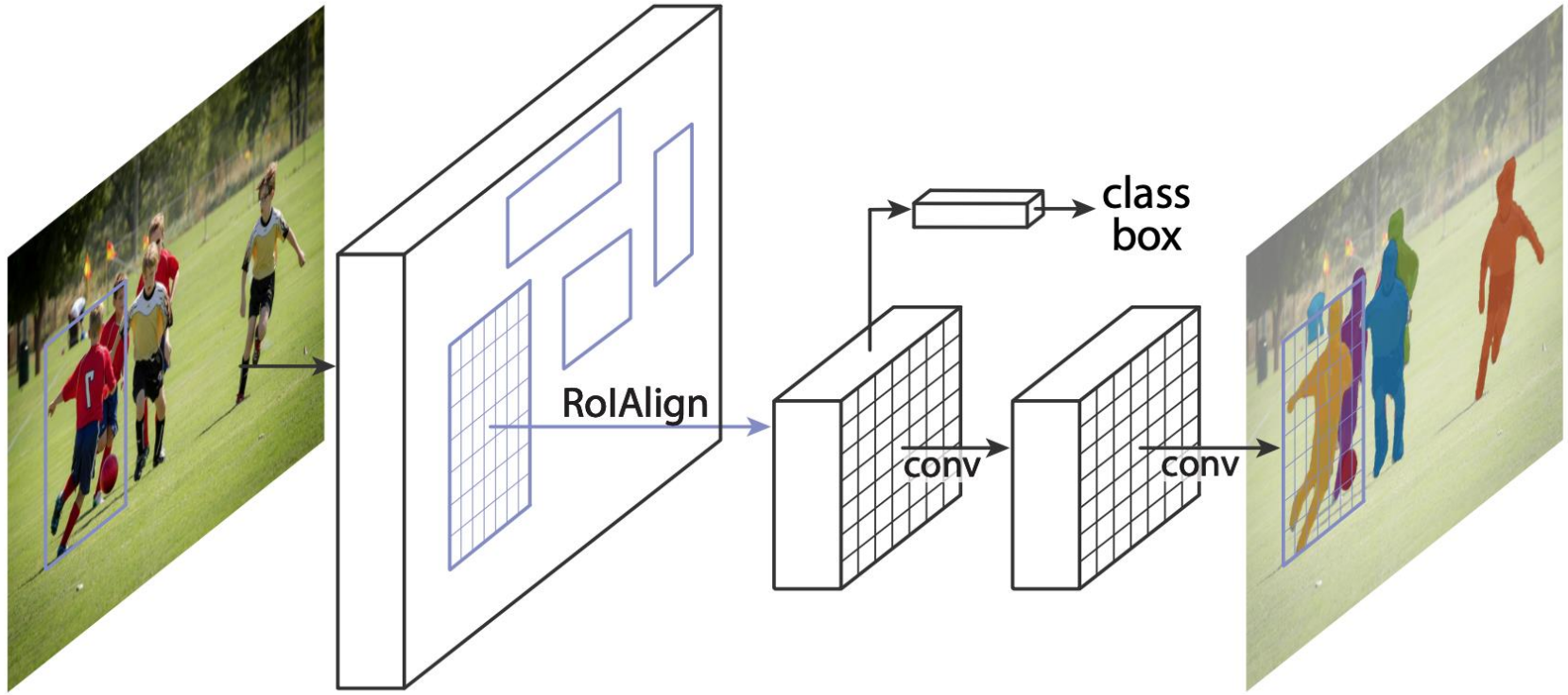


RoI Align



Mask R-CNN

Mask branch

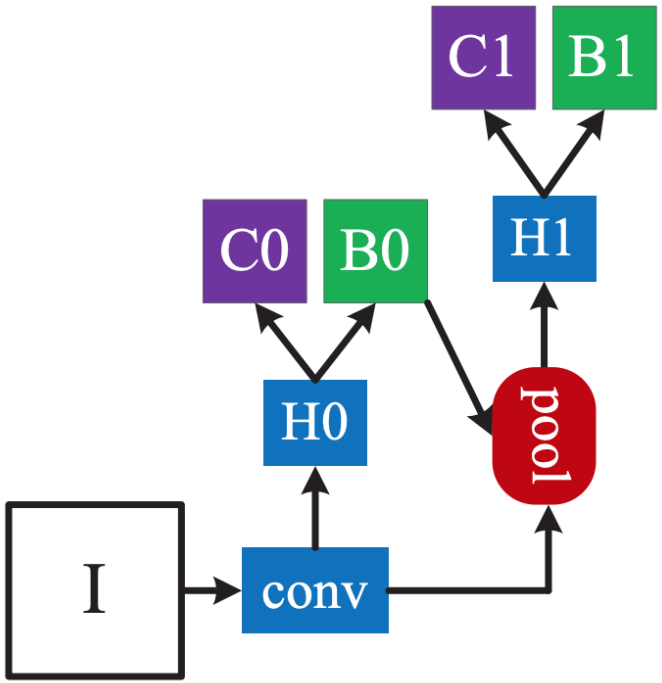


Cascade R-CNN

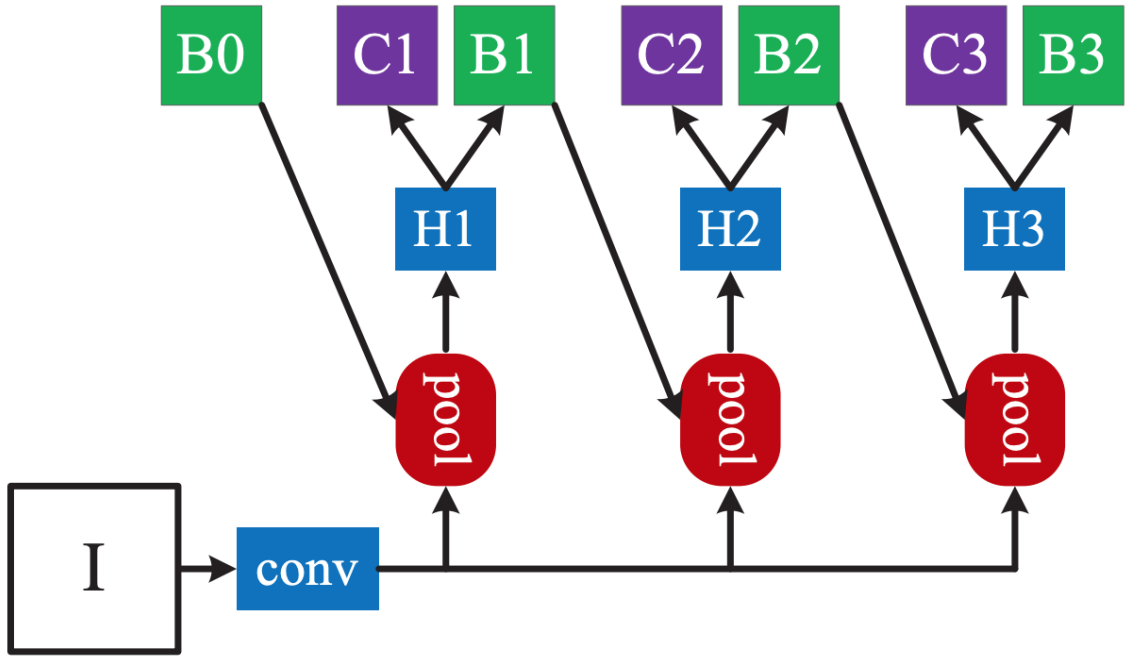
- Cascade architecture
- Training distribution

Cascade R-CNN

Cascade architecture



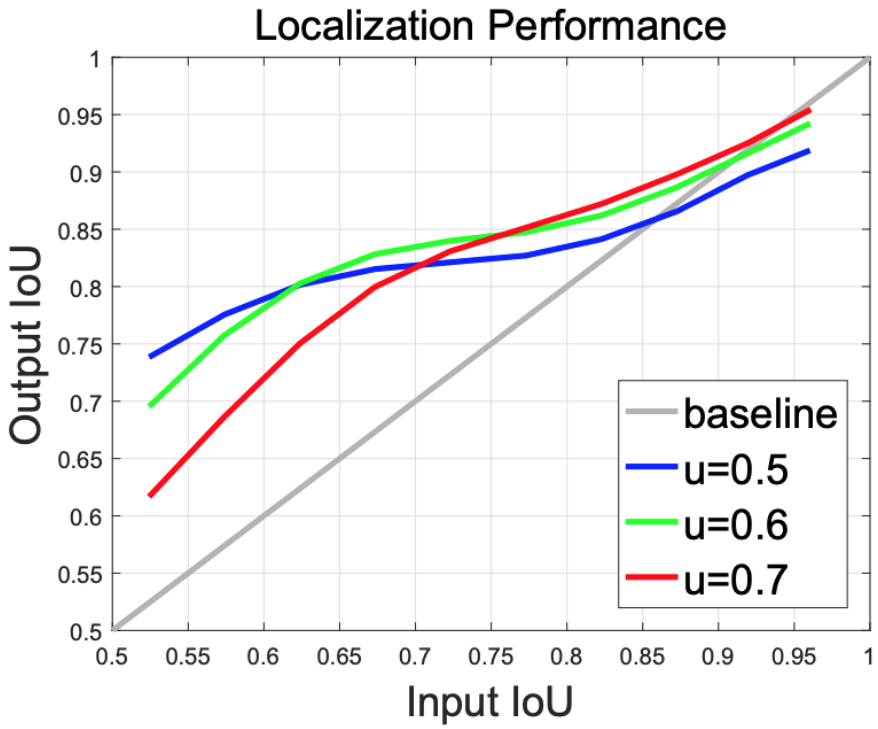
Faster R-CNN



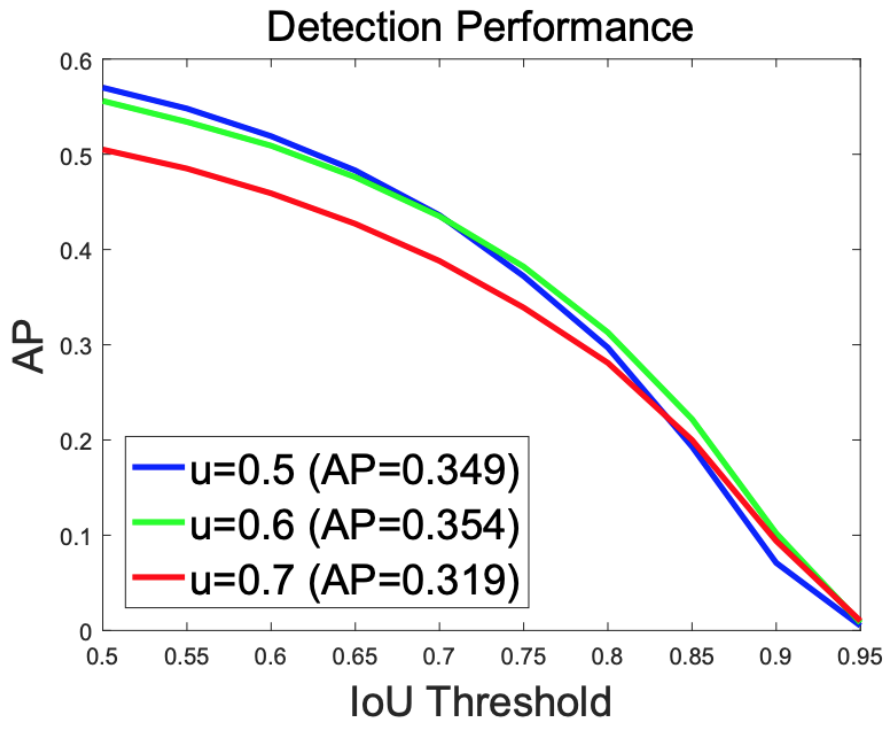
Cascade R-CNN

Cascade R-CNN

Training distribution



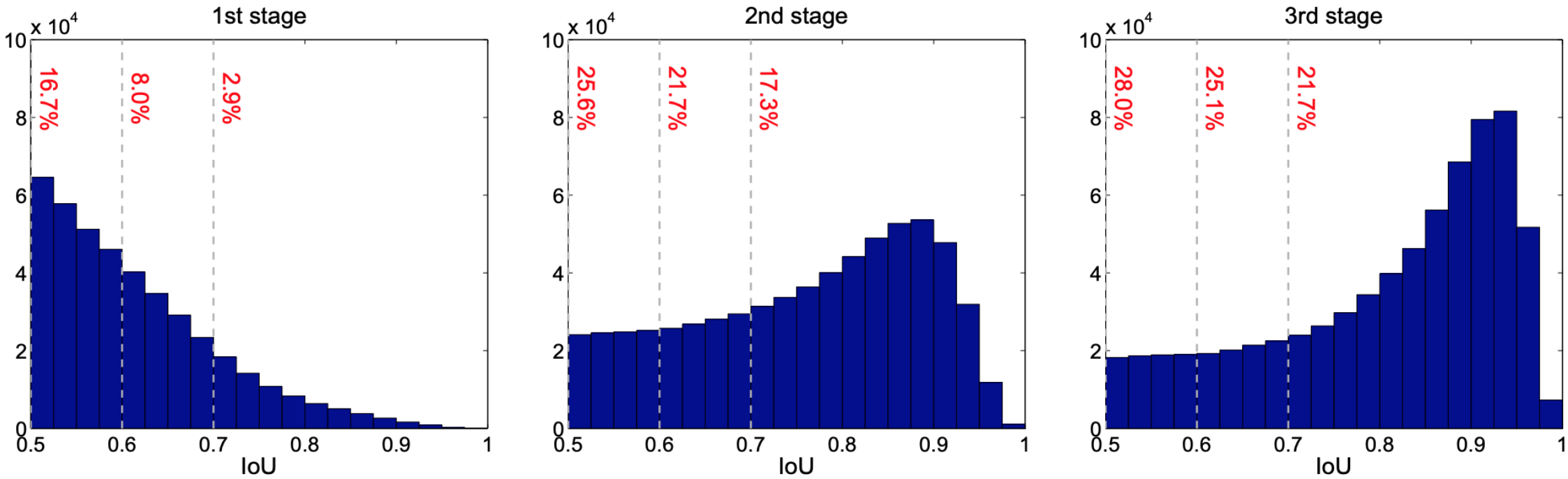
Regressor



Detector

Cascade R-CNN

Training distribution



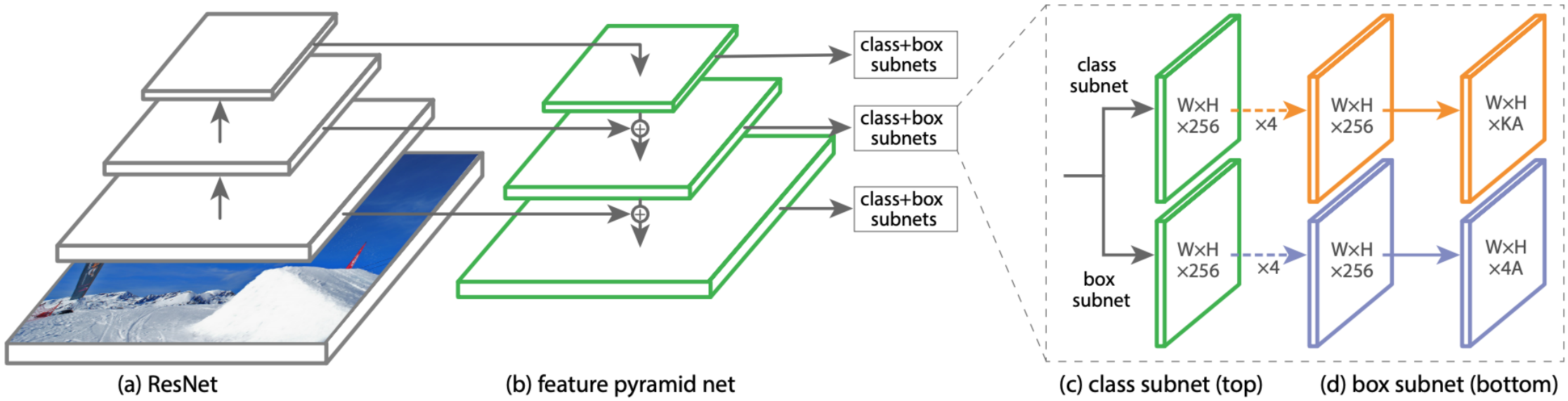
RetinaNet

- FPN
- Focal Loss

RetinaNet

FPN

heavier head than
SSD / Faster R-CNN



RetinaNet

Focal Loss

- Problem: class imbalance
  - inefficient training
  - loss is overwhelmed by negative samples

Model	Solution
Two-stage detectors	1) proposal 2) mini-batch sampling
SSD	Hard negative mining
RetinaNet	Focal loss

RetinaNet

Focal Loss

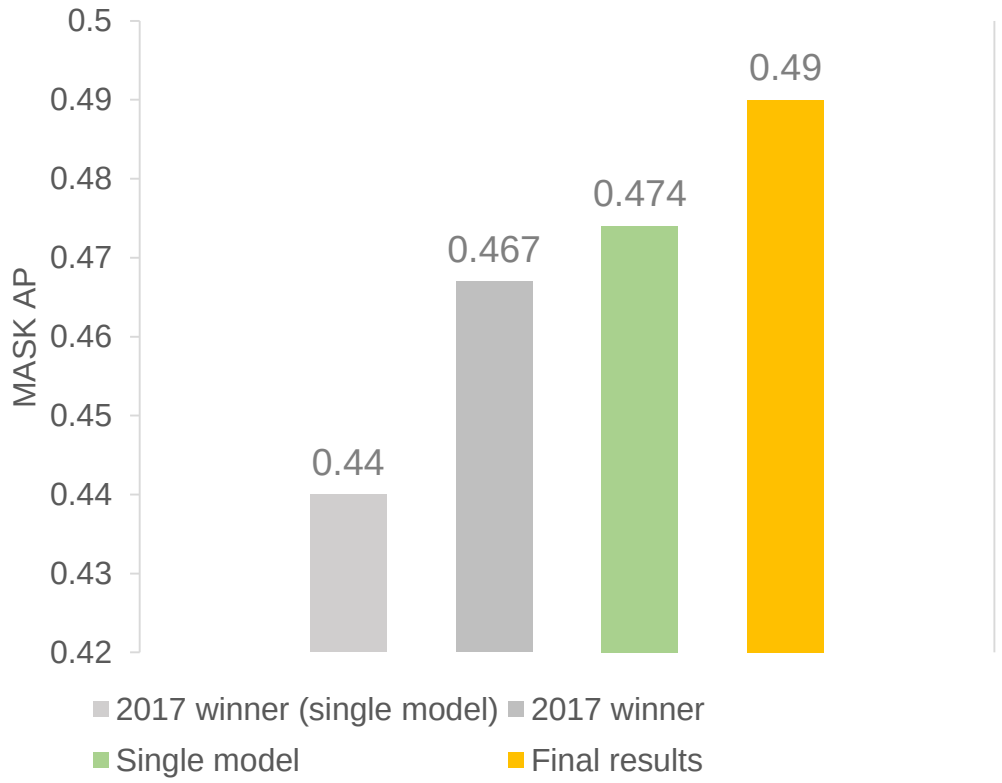
- Solution: high confidence -> small loss

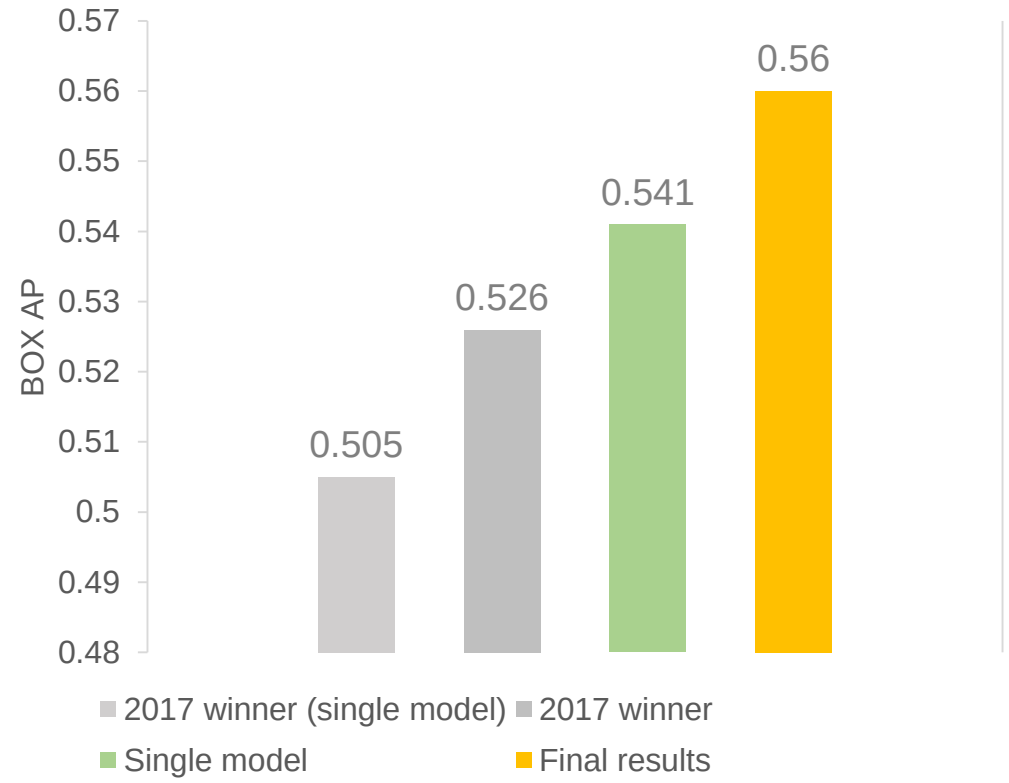
$$p_{\mathrm{t}} = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}$$

$$\mathrm{FL}(p_{\mathrm{t}}) = -\alpha_{\mathrm{t}}(1 - p_{\mathrm{t}})^{\gamma} \log(p_{\mathrm{t}})$$

COCO Challenge 2018

Comparison of our approach with 2017 winning entries on COCO test-dev.





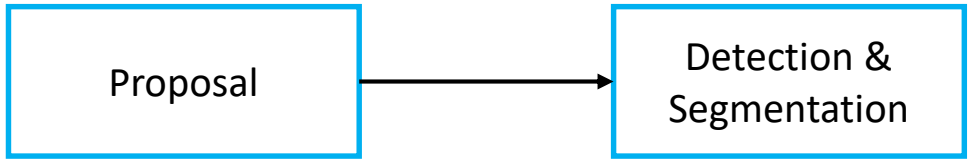
COCO Challenge 2018

1. We developed a hybrid task cascade framework for detection and segmentation.

Detection &
Segmentation

COCO Challenge 2018

1. We developed a hybrid task cascade framework for detection and segmentation.
2. We proposed a feature guided anchoring scheme to improve the average recall (AR) of RPN by 10 points.



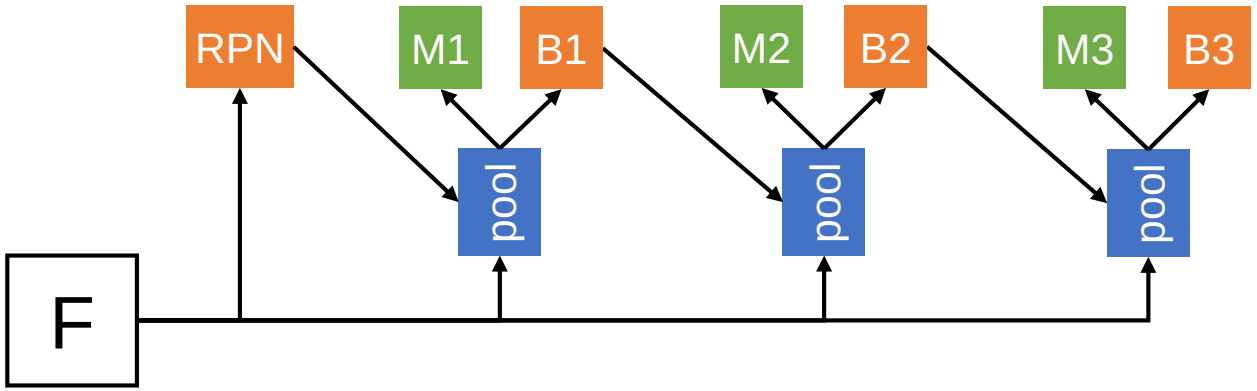
COCO Challenge 2018

1. We developed a hybrid task cascade framework for detection and segmentation.
2. We proposed a feature guided anchoring scheme to improve the average recall (AR) of RPN by 10 points.
3. We designed a new backbone FishNet.

COCO Challenge 2018

Hybrid Task Cascade (HTC)

- Cascade Mask R-CNN (Cascade R-CNN + Mask R-CNN)

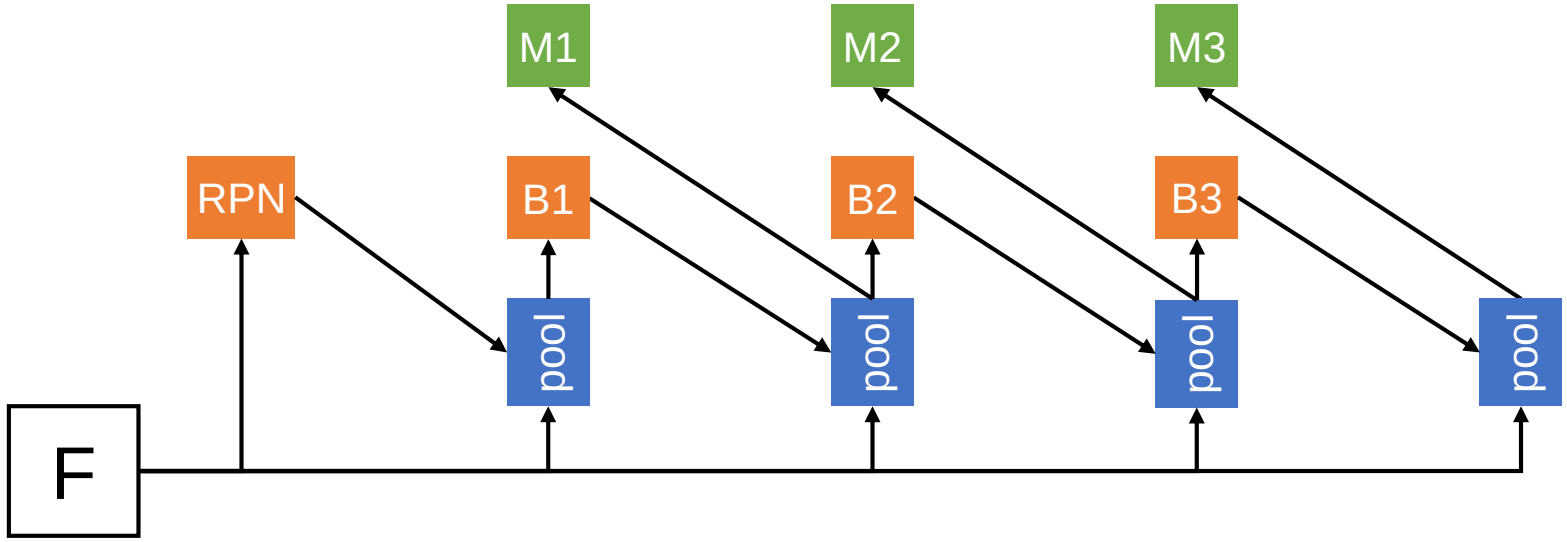


Problem: Two branches at each stage are executed in parallel, without interaction.

COCO Challenge 2018

Hybrid Task Cascade (HTC)

- Interleaved execution

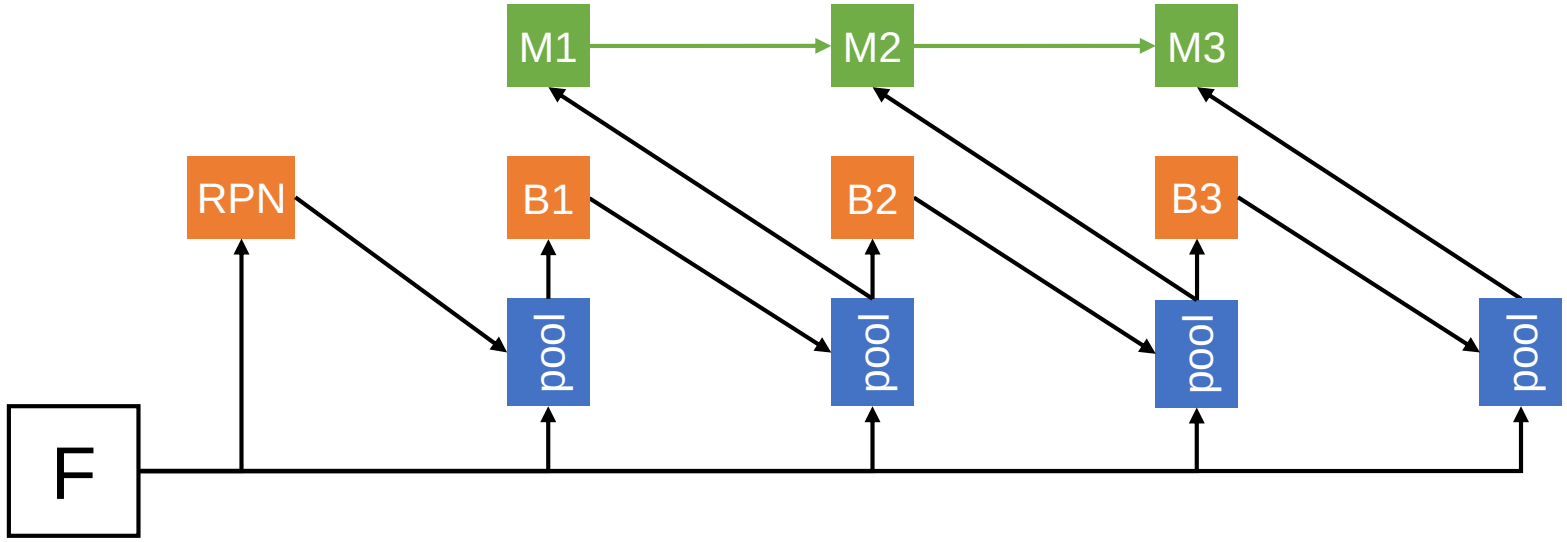


Problem: No direct information flow between mask branches at different stages.

COCO Challenge 2018

Hybrid Task Cascade (HTC)

- Mask Information Flow

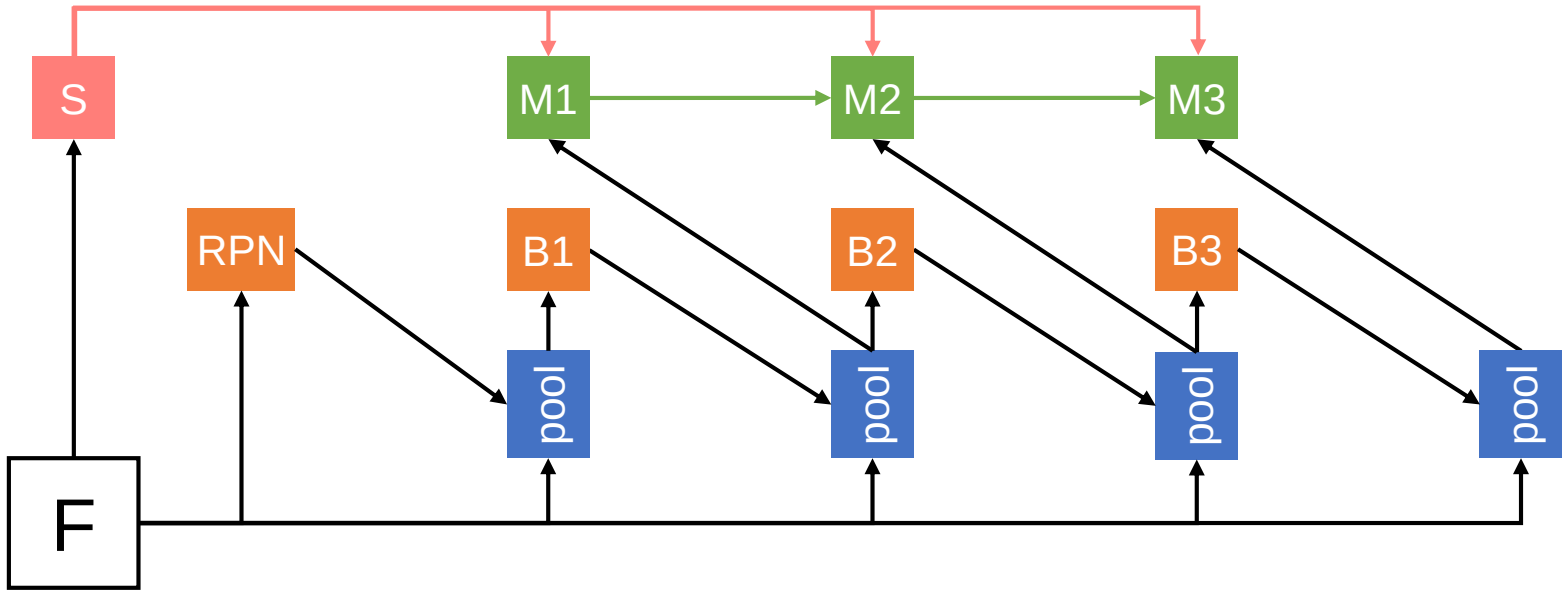


Problem: Spatial context is not much explored.

COCO Challenge 2018

Hybrid Task Cascade (HTC)

- Spatial context



COCO Challenge 2018

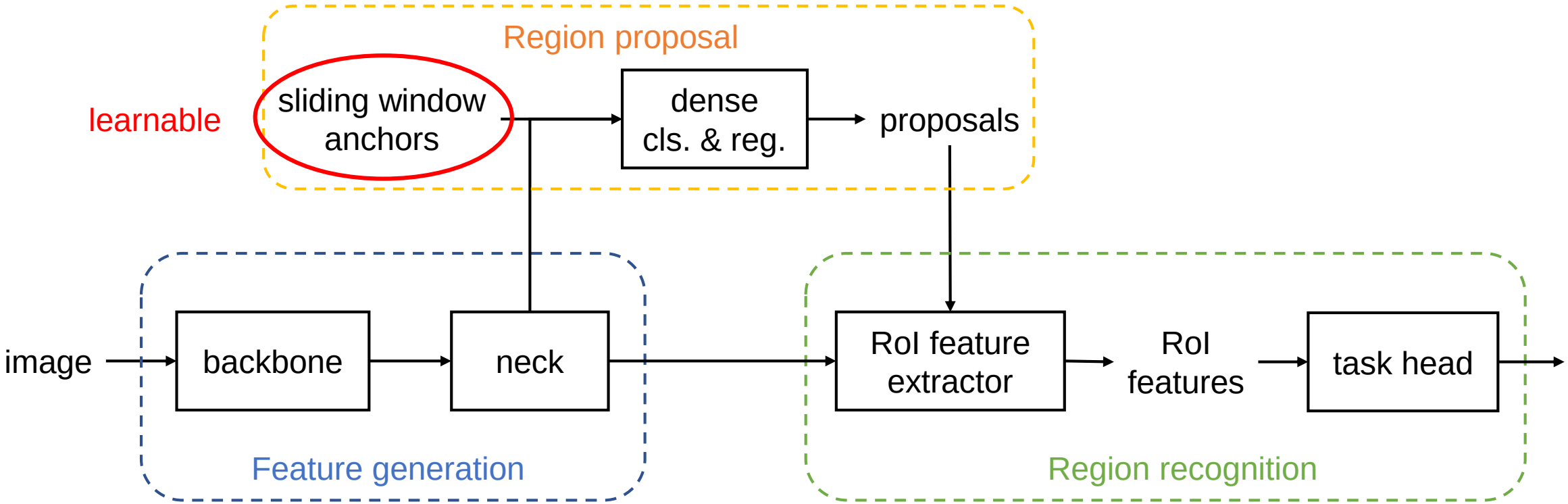
Hybrid Task Cascade (HTC)

Table 1: Comparison with state-of-the-art methods on COCO test-dev dataset.

Method	Backbone	box AP	mask AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	runtime (fps)
Mask R-CNN [18]	ResNet-50-FPN	39.1	35.6	57.6	38.1	18.7	38.3	46.6	5.3
PANet[27]	ResNet-50-FPN	41.2	36.6	58.0	39.3	16.3	38.1	52.4	-
Cascade Mask R-CNN	ResNet-50-FPN	42.7	36.9	58.6	39.7	19.6	39.3	48.8	3.0
Cascade Mask R-CNN	ResNet-101-FPN	44.4	38.4	60.2	41.4	20.2	41.0	50.6	2.9
Cascade Mask R-CNN	ResNeXt-101-FPN	46.6	40.1	62.7	43.4	22.0	42.8	52.9	2.5
HTC (ours)	ResNet-50-FPN	43.6	**38.4**	60.0	41.5	20.4	40.7	51.2	2.5
HTC (ours)	ResNet-101-FPN	45.3	**39.7**	61.8	43.1	21.0	42.2	53.5	2.4
HTC (ours)	ResNeXt-101-FPN	47.1	**41.2**	63.9	44.7	22.8	43.9	54.6	2.1

COCO Challenge 2018

Guided anchoring



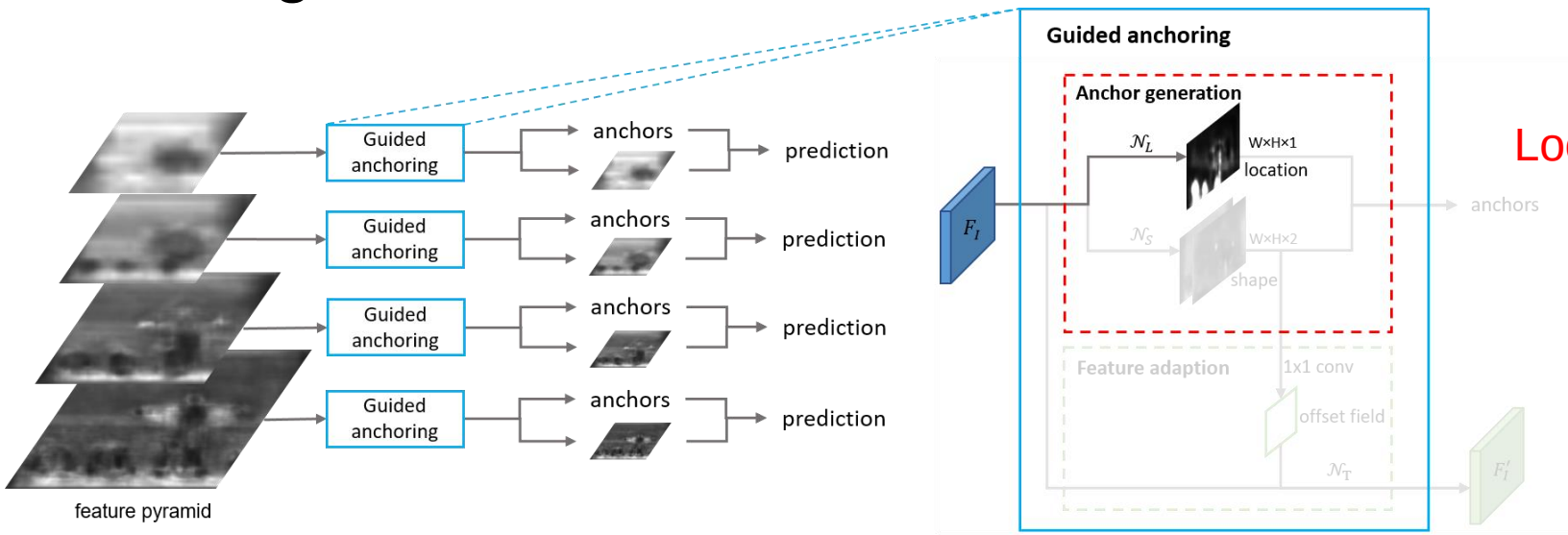
COCO Challenge 2018

Guided anchoring

- Our goal
  - Sparse
  - Arbitrary shape
- General rules for anchor design
  - Alignment
  - Consistency

COCO Challenge 2018

Guided anchoring



Location prediction

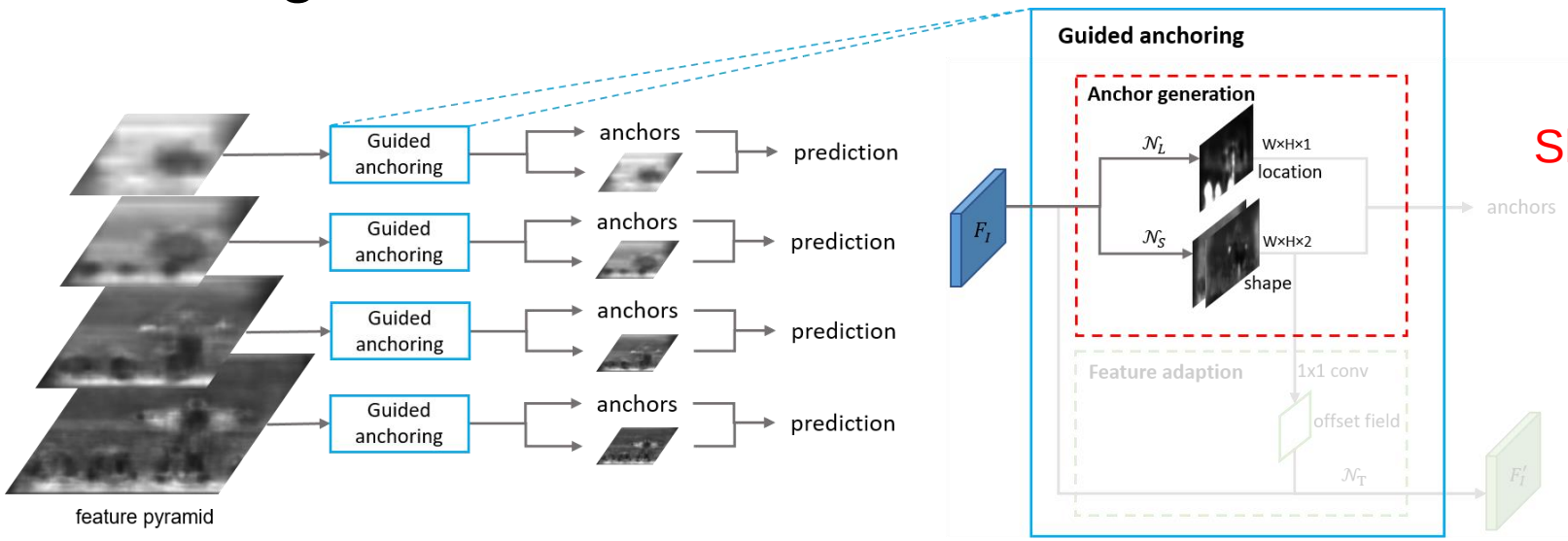
predicted anchor probabilities

predicted anchor aspect ratios

predicted anchors

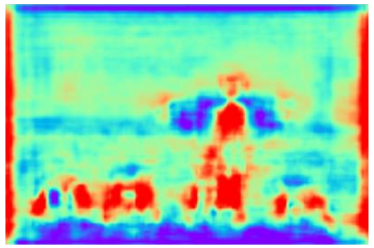
COCO Challenge 2018

Guided anchoring



Shape prediction

predicted anchor probabilities

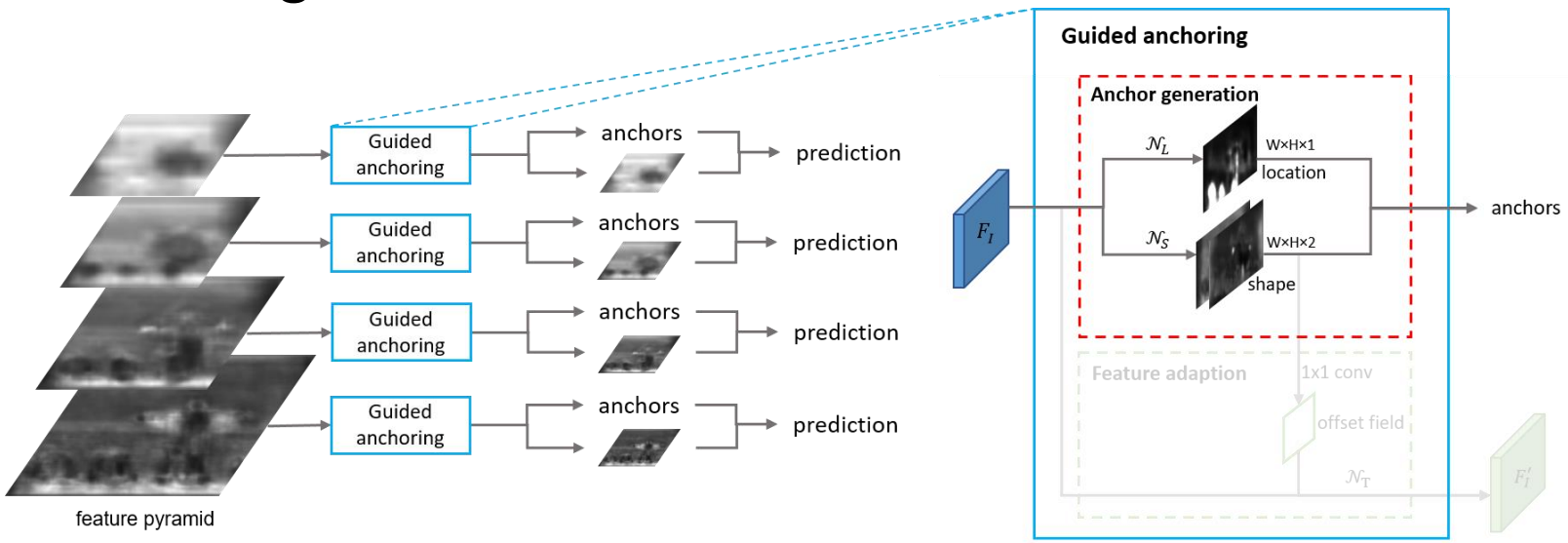


predicted anchor aspect ratios

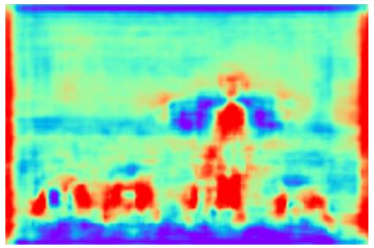
predicted anchors

COCO Challenge 2018

Guided anchoring



predicted anchor probabilities

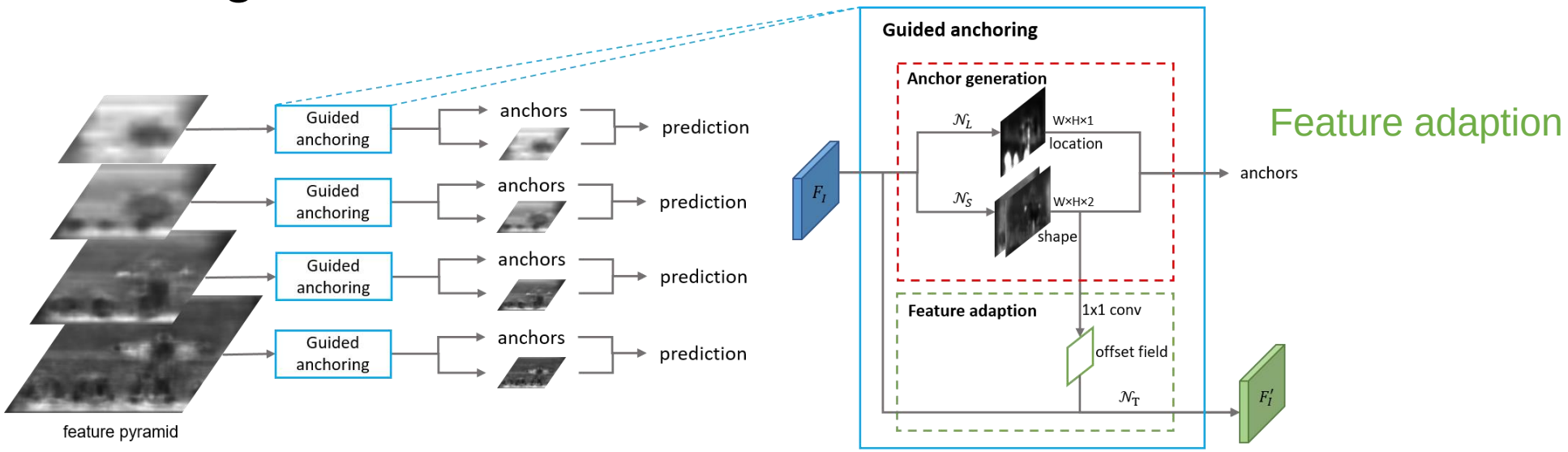


predicted anchor aspect ratios

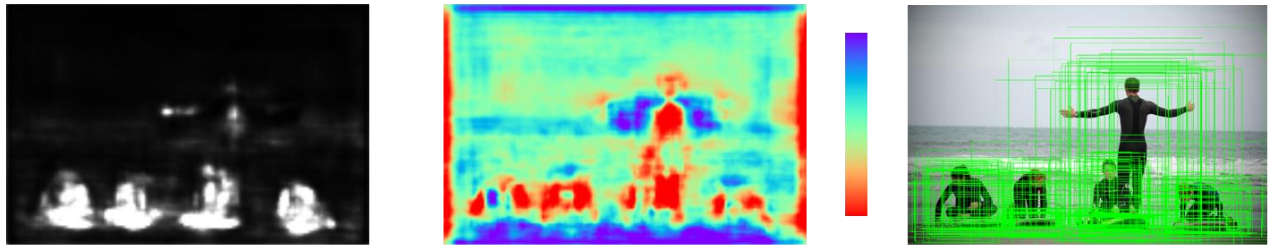
predicted anchors

COCO Challenge 2018

Guided anchoring



Feature adaption



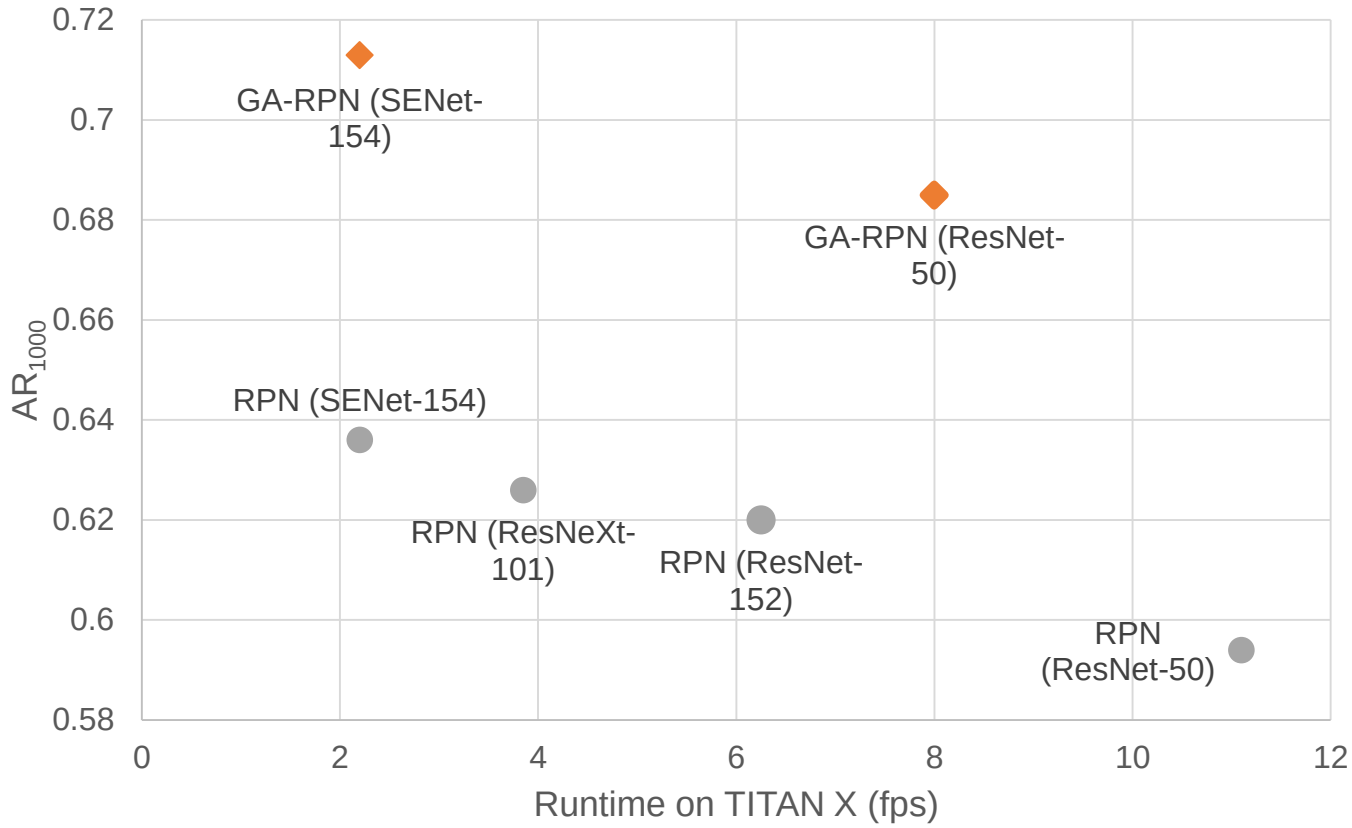
predicted anchor probabilities

predicted anchor aspect ratios

predicted anchors

COCO Challenge 2018

Guided anchoring



COCO Challenge 2018

Guided anchoring

Table 2: Detection results on MS COCO 2017 *test-dev*.

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Fast R-CNN	37.1	59.6	39.7	20.7	39.5	47.1
GA-Fast-RCNN	**39.4**	59.4	42.8	21.6	41.9	50.4
Faster R-CNN	37.1	59.1	40.1	21.3	39.8	46.5
GA-Faster-RCNN	**39.8**	59.2	43.5	21.8	42.6	50.7
RetinaNet	35.9	55.4	38.8	19.4	38.9	46.5
GA-RetinaNet	**37.1**	56.9	40.0	20.1	40.1	48.0

COCO Challenge 2018

Implementation details

1. Training scales
   - short edge: random sampled from 400 ~ 1400
   - long edge: 1600
2. Test scales
   - (600, 900), (800, 1200), (1000, 1500), (1200, 1800), (1400, 2100)
3. Pipeline
   - Joint training
   - Finetune with GA-RPN proposals
   - Test with GA-RPN proposals
4. Resources
   - 32 Tesla V100 GPUs (16GB) for 3 days

COCO Challenge 2018

Implementation details

Backbones

- SENet-154
- ResNeXt101 (64*4d)
- ResNeXt101 (32*8d)
- DPN-107
- FishNet

~0.8 points higher

comparable

COCO Challenge 2018

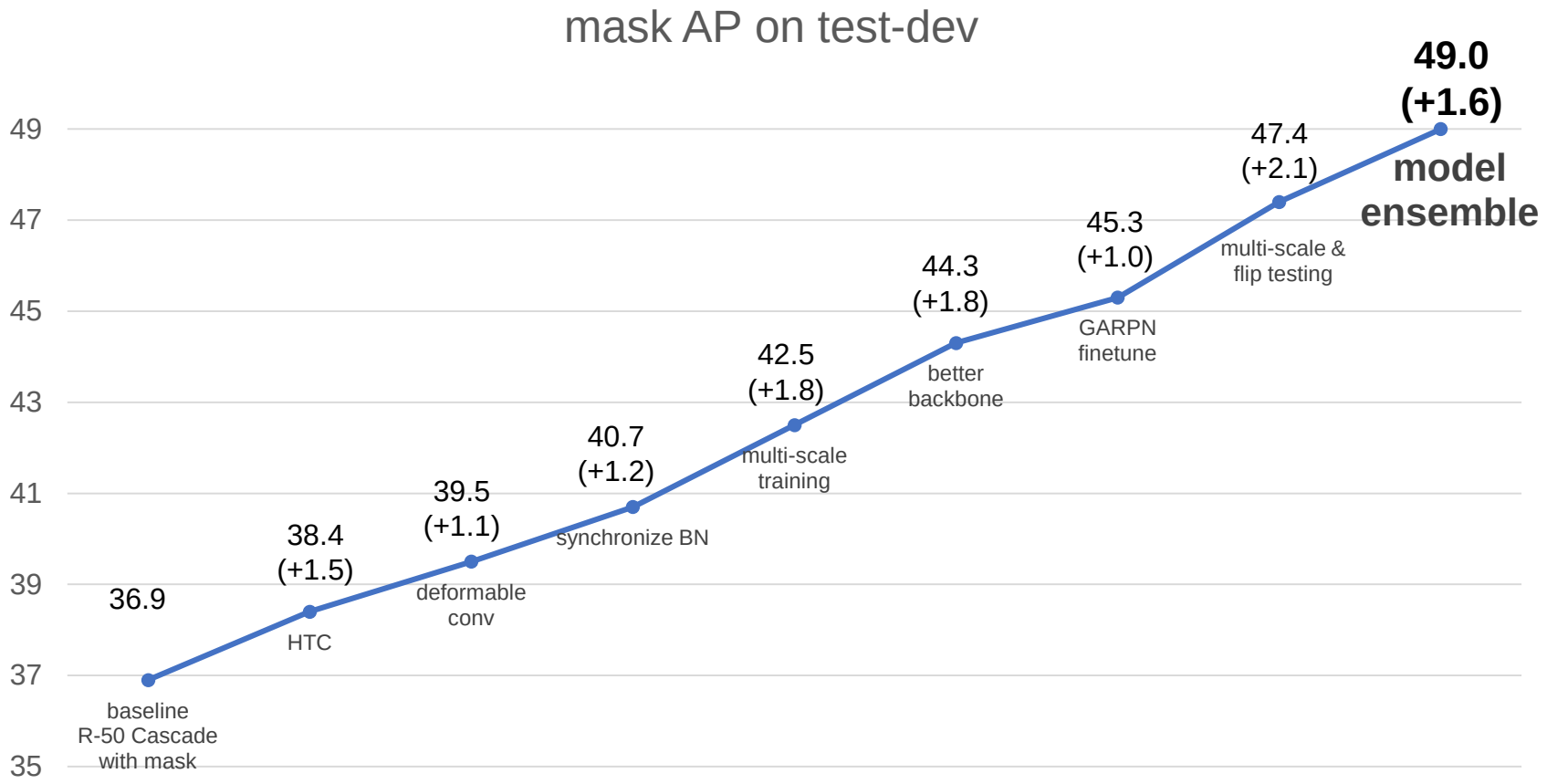
Implementation details

Other tricks

- w/ SoftNMS
- w/o OHEM
- w/o classwise balance sampling
- w/o voting for bbox or mask

COCO Challenge 2018

- With bells and whistles



mmdetection

- **Comprehensive**
  - ☑ RPN
  - ☑ Mask R-CNN
  - ☑ Cascade R-CNN
  - ☐ More ··· ···
  - ☑ Fast/Faster R-CNN
  - ☑ FPN
  - ☑ RetinaNet

- **High performance**
  - ☑ Better performance
  - ☑ Optimized memory consumption
  - ☑ Faster speed

- **Handy to develop**
  - ☑ Written with PyTorch
  - ☑ Modular design

GitHub: mmdetection

Hybrid Task Cascade for Instance Segmentation (Accepted to CVPR 2019)
https://arxiv.org/abs/1901.07518

Region Proposal by Guided Anchoring (Accepted to CVPR 2019)
https://arxiv.org/abs/1901.03278

FishNet: A Versatile Backbone for Image, Region, and Pixel Level Prediction (Accepted to NIPS 2018)
https://arxiv.org/abs/1901.03495