Feature Selection Matters for Anchor-Free Object Detection

Chenchen Zhu

Carnegie Mellon University

04/29/2020

Overview

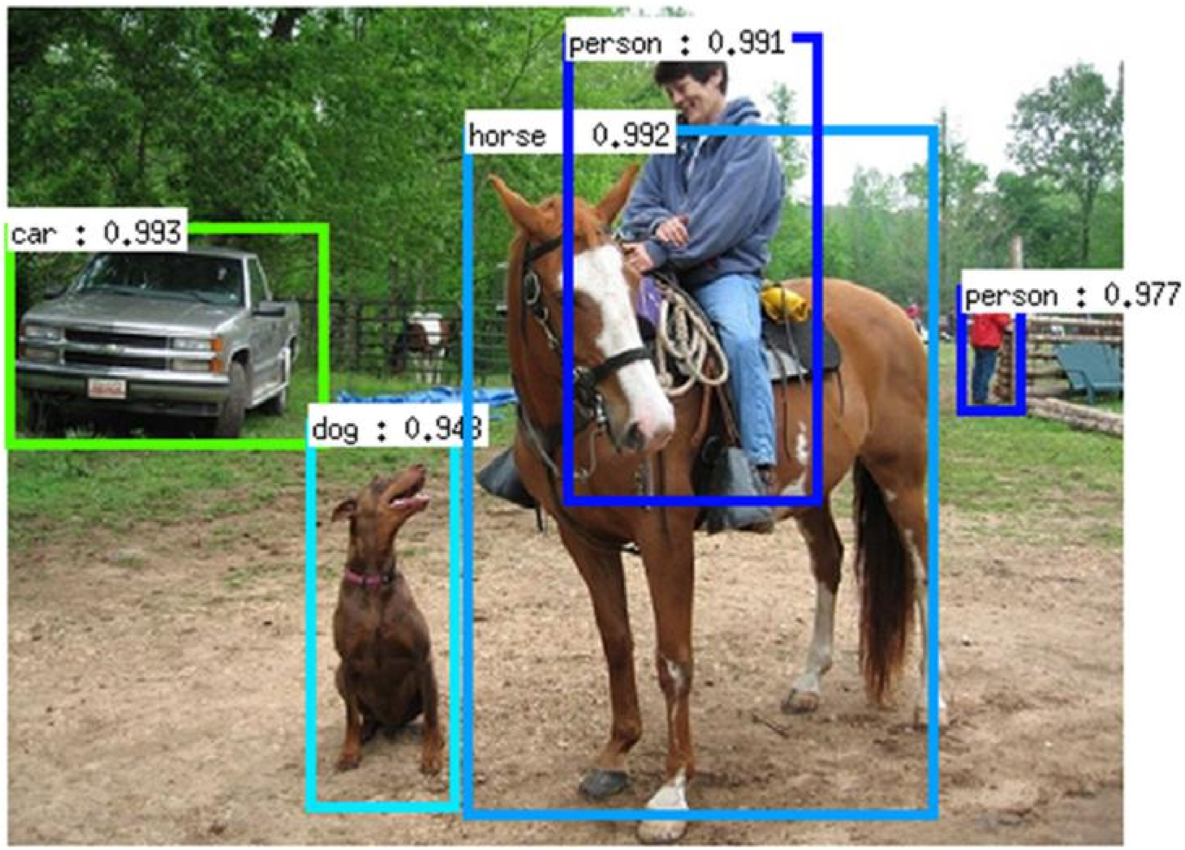
- Background
- Motivation
- Feature Selection in Anchor-Free Detection
  - General concept
  - Network architecture
  - Ground-truth and loss
  - Feature selection
- Experiments

Overview

- **Background**
- Motivation
- Feature Selection in Anchor-Free Detection
    - General concept
    - Network architecture
    - Ground-truth and loss
    - Feature selection
- Experiments

Background

A long-lasting challenge: scale variation



Background

Prior methods addressing scale variation

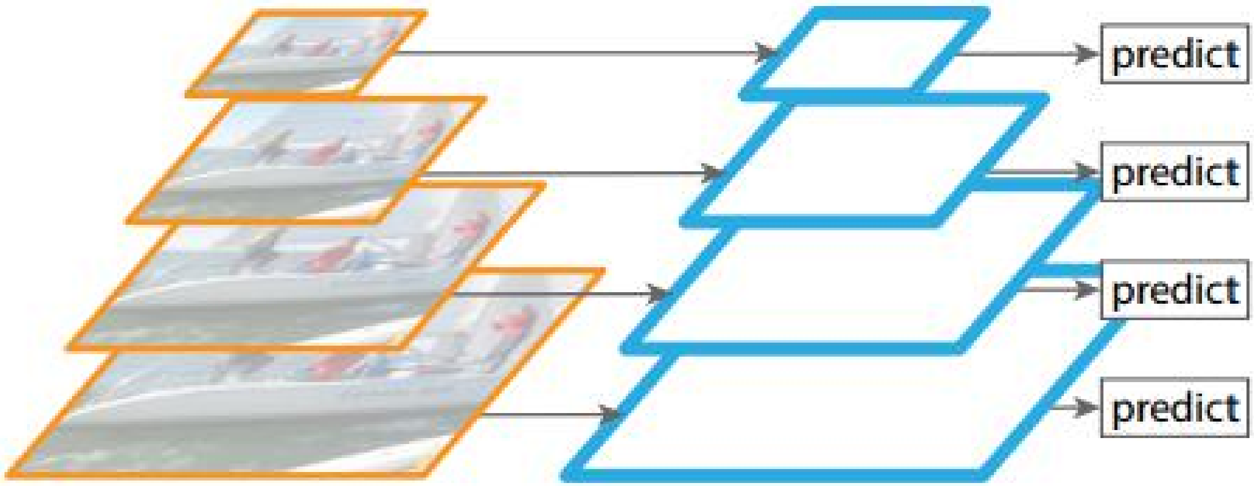
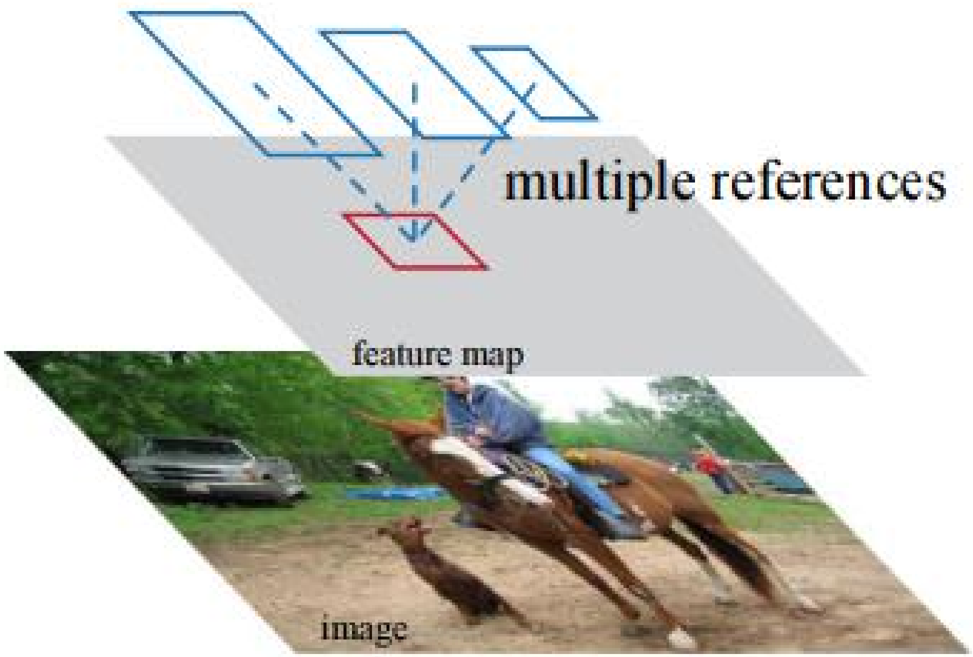


Image pyramid

Background

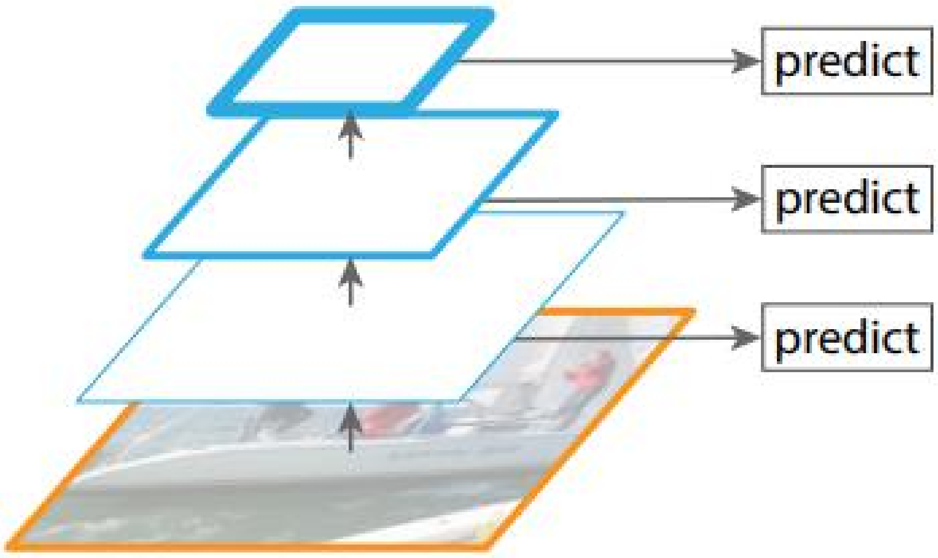
Prior methods addressing scale variation



Anchor boxes [Ren et al, Faster R-CNN]

Background

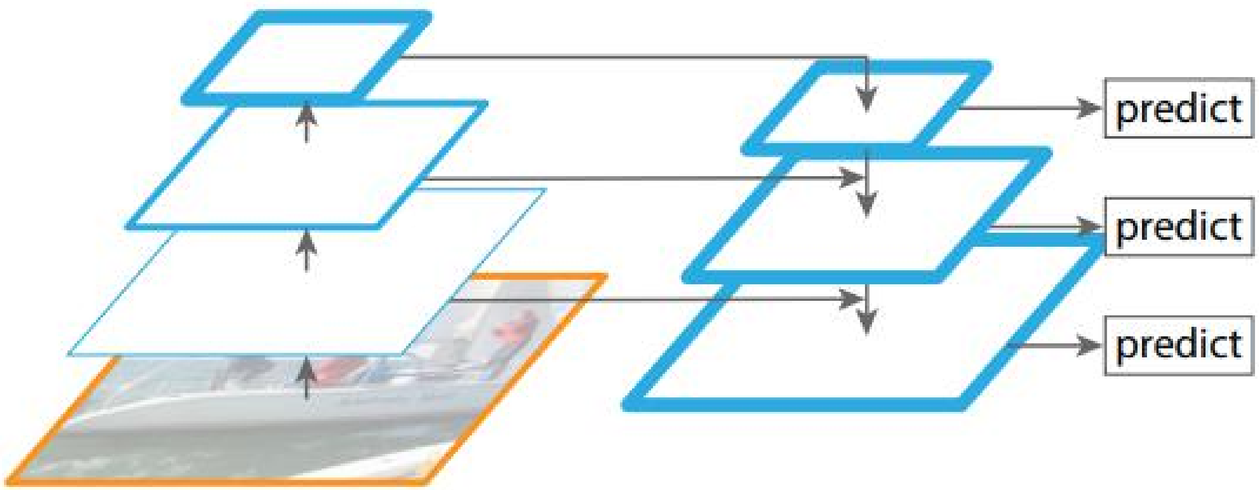
Prior methods addressing scale variation



Pyramidal feature hierarchy, e.g. [Liu et al, SSD]

Background

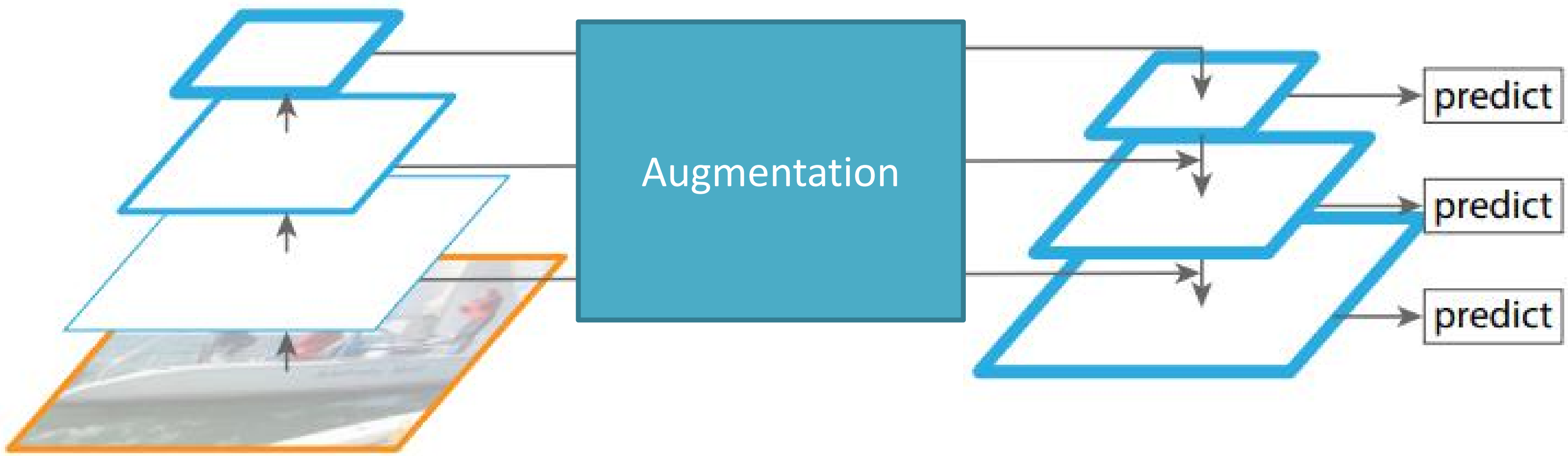
Prior methods addressing scale variation



Feature pyramid network [Lin et al, FPN, RetinaNet]

Background

Prior methods addressing scale variation

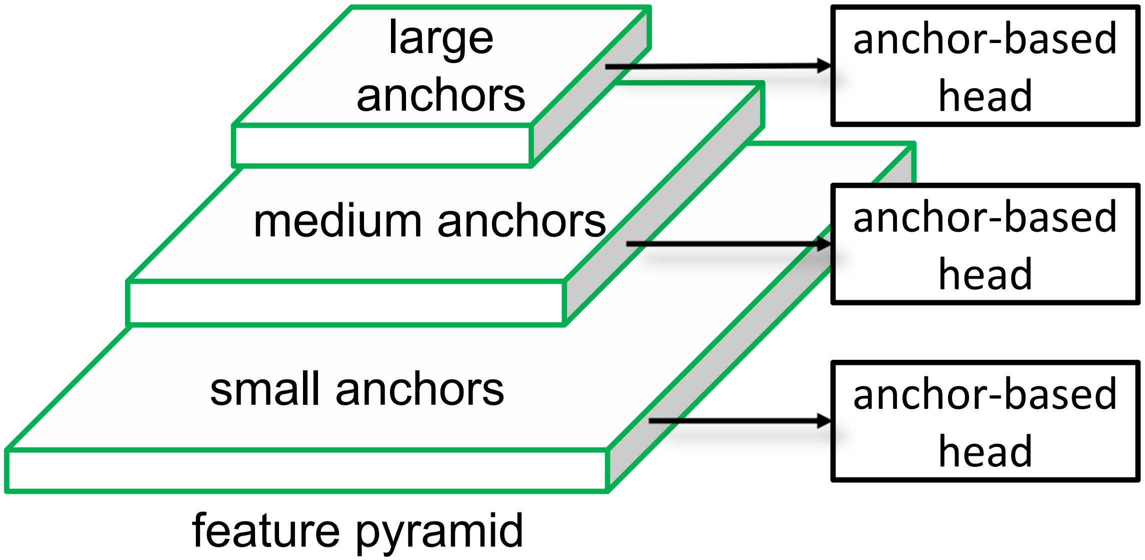


Balanced FPN [Pang et al, Libra R-CNN]
HRNet [Wang et al]
NAS-FPN [Ghiasi et al]
EfficentDet [Tan et al]

Background

Combining feature pyramid with anchor boxes

- **Smaller anchor associated with lower pyramid levels (local fine-grained information)**
- **Larger anchor associated with higher pyramid levels (global semantic information)**



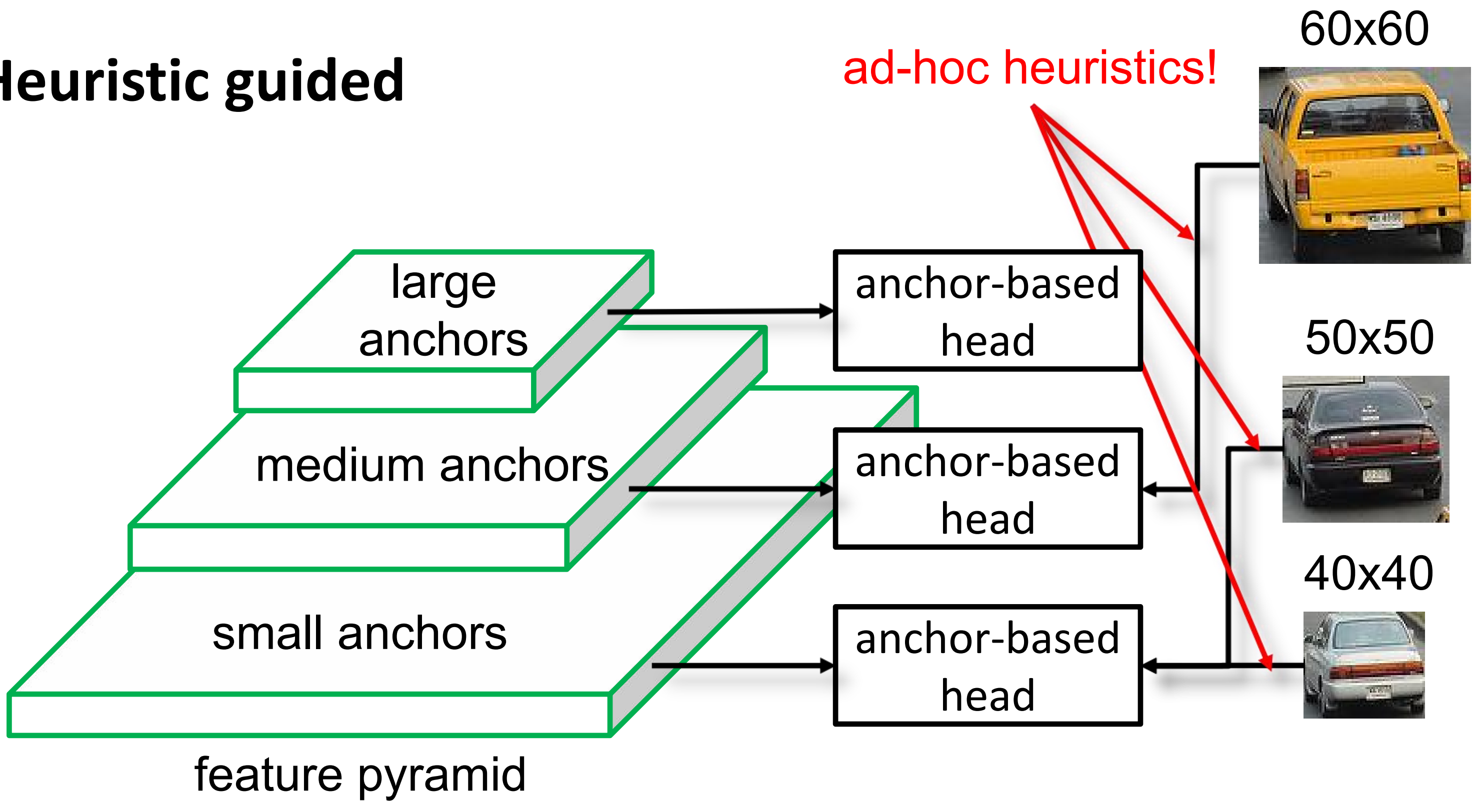
Overview

- **Background**
- **Motivation**
- Feature Selection in Anchor-Free Detection
  - General concept
  - Network architecture
  - Ground-truth and loss
  - Feature selection
- Experiments

Motivation

Implicit feature selection by anchor boxes

- IoU-based
- Heuristic guided



Motivation

Problem: feature selection by heuristics may not be optimal.

Question: how can we select feature level based on semantic information rather than just box size?

Answer: allowing arbitrary feature assignment by removing the anchor matching mechanism (using anchor-free methods), selecting the most suitable feature level/levels.

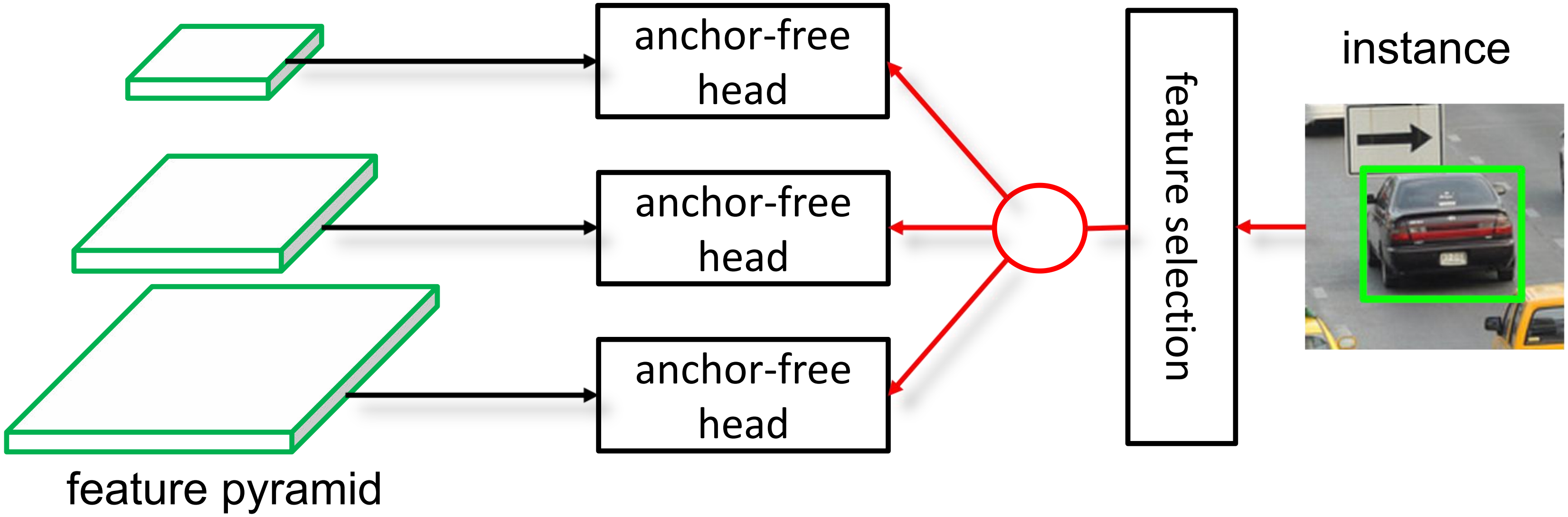
Overview

- Background
- Motivation
- Feature Selection in Anchor-Free Detection
  - General concept
  - Network architecture
  - Ground-truth and loss
  - Feature selection
- Experiments

Feature Selection in Anchor-Free Detection

The general concept

- Each instance can be *arbitrarily* assigned to a single or multiple feature levels.



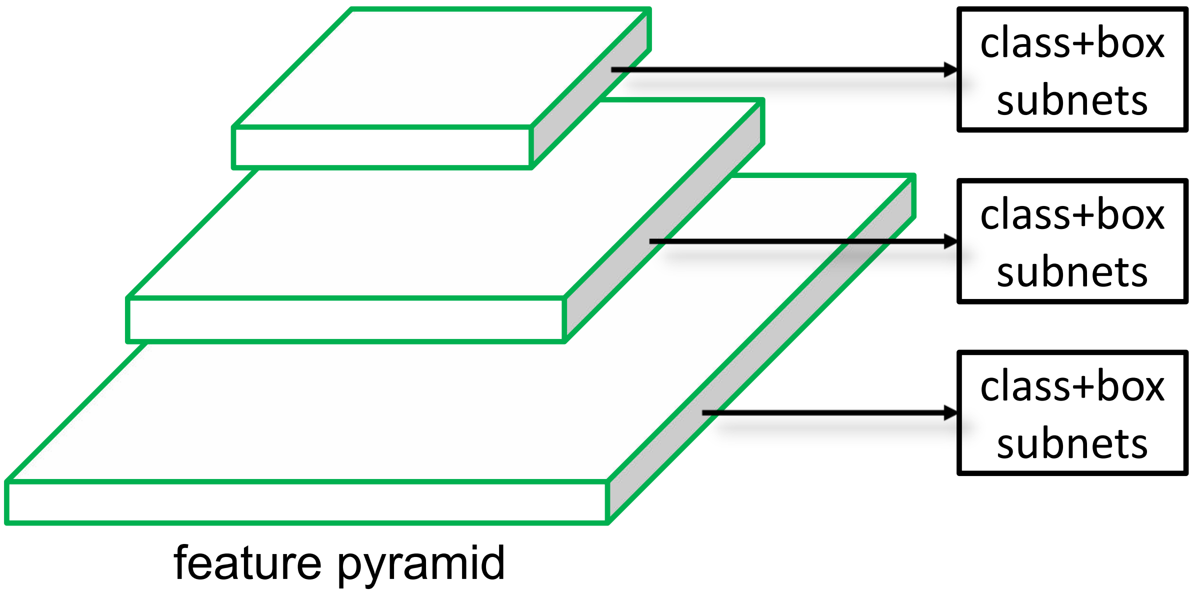
Feature Selection in Anchor-Free Detection

Instantiation

- Network architecture
- Ground-truth and loss
- Feature selection: heuristic guided vs. semantic guided

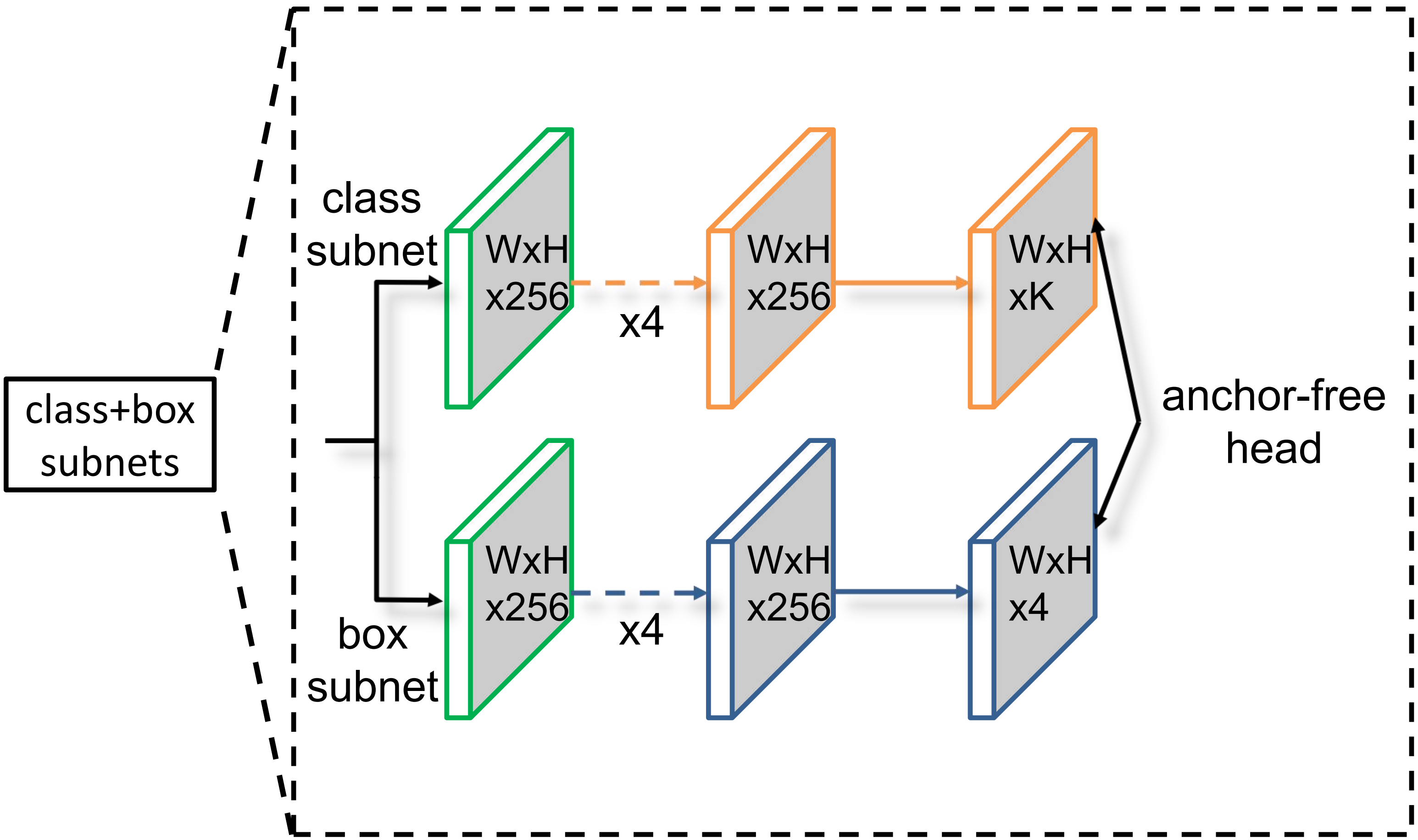
Feature Selection in Anchor-Free Detection

Network architecture (on RetinaNet)



Feature Selection in Anchor-Free Detection

Network architecture (on RetinaNet)



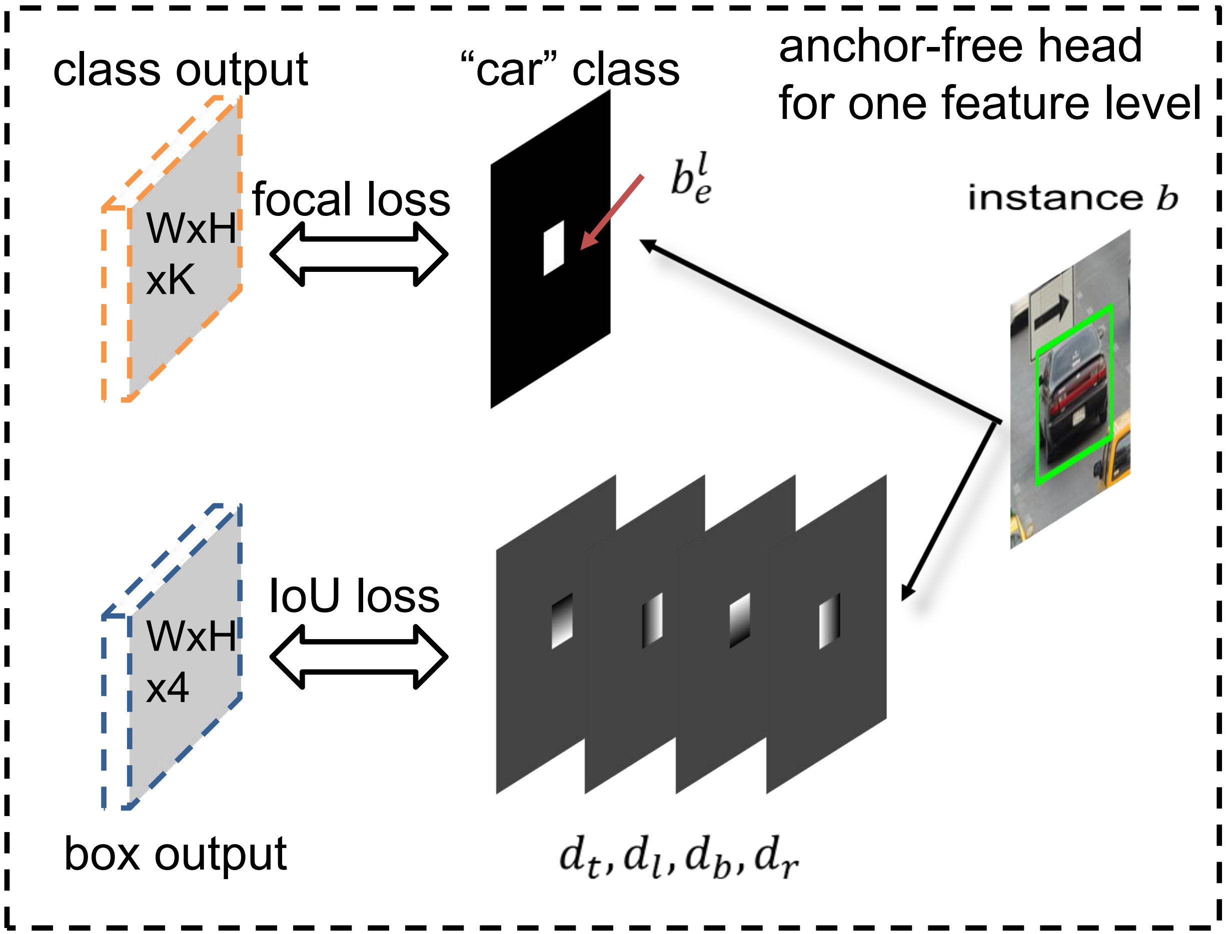
Feature Selection in Anchor-Free Detection

Ground-truth and loss

- **Definitions**
  - Instance box: $b = [x, y, w, h]$
  - Projected box on P_l: $b_p^l = \left[x_p^l, y_p^l, w_p^l, h_p^l\right] = b/2^l$
  - Effective box on P_l: $b_e^l = \left[x_p^l, y_p^l, \epsilon_e w_p^l, \epsilon_e h_p^l\right]$
  - For pixel (i, j) in b_e^l, $[d_{t_{i,j}}^l, d_{l_{i,j}}^l, d_{b_{i,j}}^l, d_{r_{i,j}}^l]$ are distances of (i, j) to the top, left, bottom, right boundaries of b_p^l, respectively.

Feature Selection in Anchor-Free Detection

Ground-truth and loss (similar to DenseBox [Huang et al])



Feature Selection in Anchor-Free Detection

Heuristic guided feature selection

$$l' = \lfloor l_0 + \log_2(\sqrt{wh}/224) \rfloor$$

where l_0 is the target level to which an instance with $w \times h = 224^2$ is mapped [Lin et al, FPN].

Feature Selection in Anchor-Free Detection

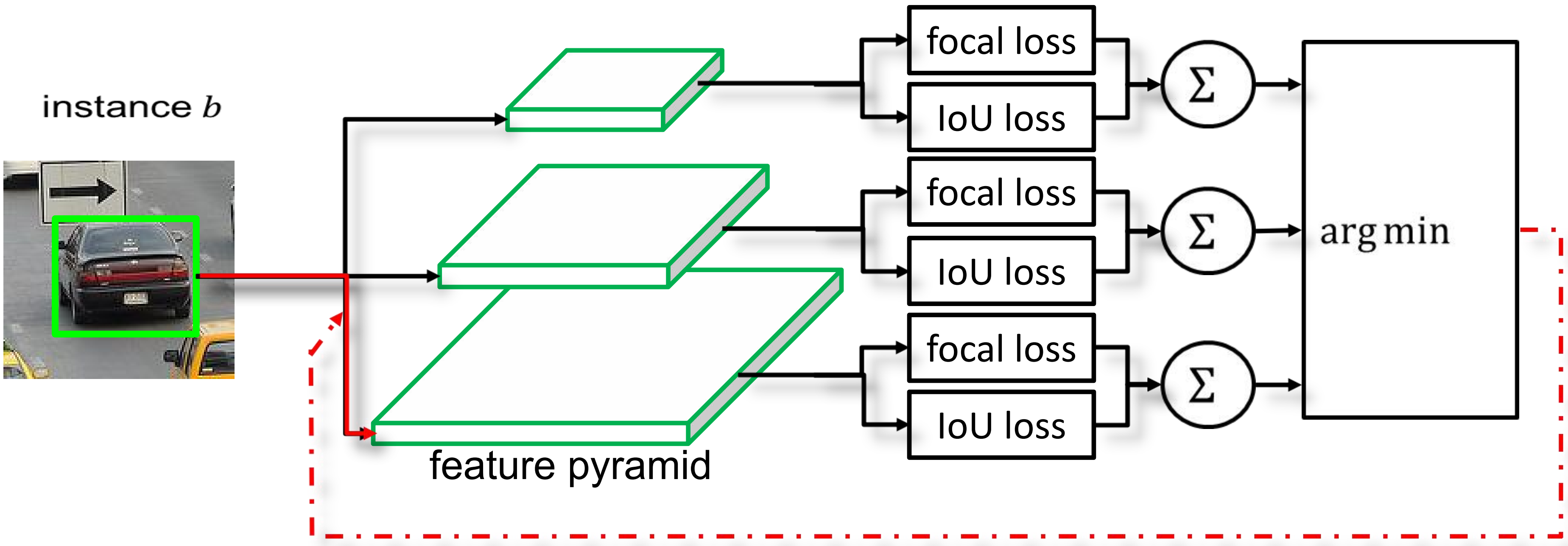
Question: what is a good representation of semantic information to guide feature selection?

Our assumption: semantic information is encoded in the network *loss*.

Feature Selection in Anchor-Free Detection

Semantic guided feature selection: hard version

$$l^* = \arg\min_l L_{FL}^b(l) + L_{IoU}^b(l)$$



Feature Selection in Anchor-Free Detection

Question: is it enough to select just one feature level for each instance?

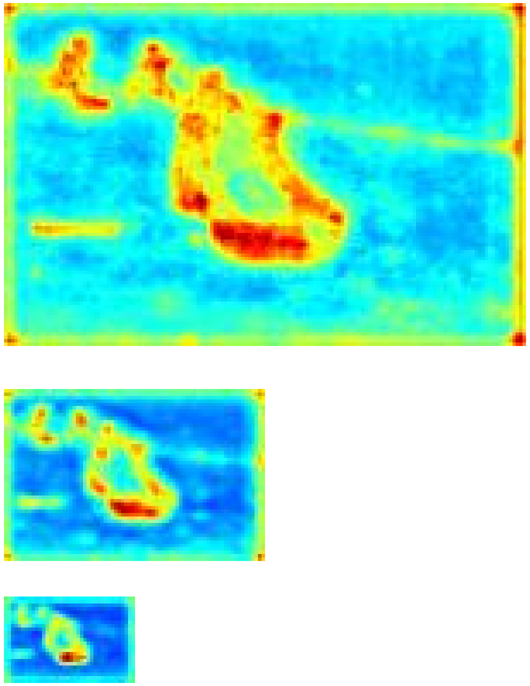
Feature Selection in Anchor-Free Detection

Can we use similar features from multiple levels to further improve the performance?

P_3

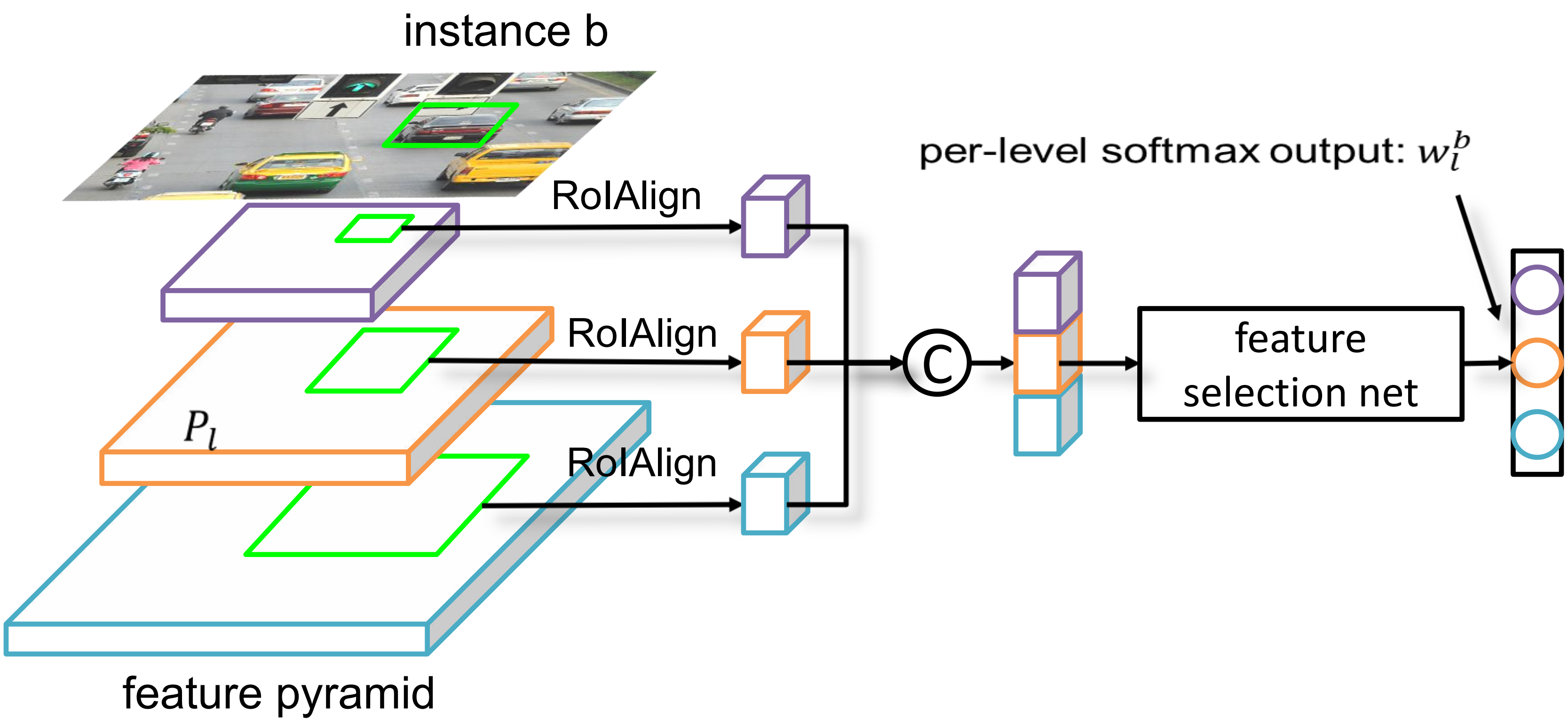
P_4

P_5



Feature Selection in Anchor-Free Detection

Semantic guided feature selection: soft version

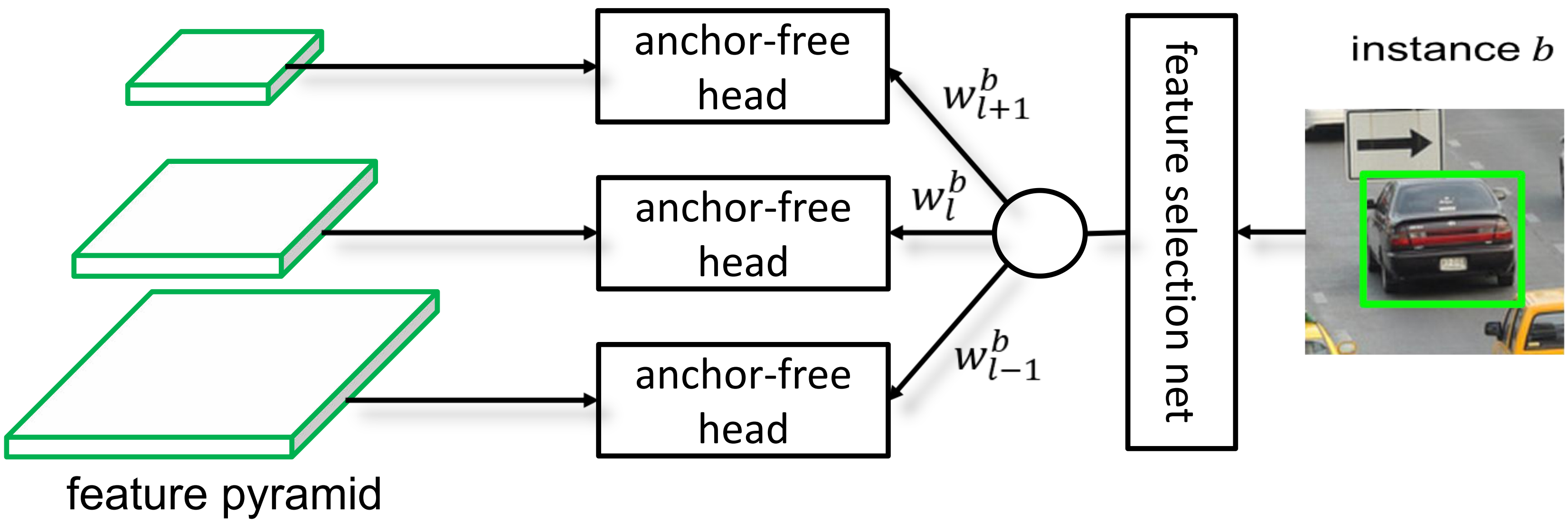


Feature Selection in Anchor-Free Detection

Semantic guided feature selection: soft version

$$Loss^b = \sum_l w_l^b \left[L_{FL}^b(l) + L_{IoU}^b(l)\right]$$

This can also be viewed as dividing an instance into several proportions and assigning each proportion to a level.



Overview

- Background
- Motivation
- Feature Selection in Anchor-Free Detection
  - General concept
  - Network architecture
  - Ground-truth and loss
  - Feature selection
- Experiments

Experiments

- **Data**
  - **COCO Dataset, train set:** train2017**, validation set:** val2017**, test set:** test-dev
- **Ablation study**
  - **Train on** train2017**, evaluate on** val2017
  - **ResNet-50 as backbone network**
- **Runtime analysis**
  - **Train on** train2017**, evaluate on** val2017
  - **Run on a single 1080Ti with CUDA 10 and CUDNN 7**
- **Compare with state of the arts**
  - **Train on** train2017 **with 2x iterations, evaluate on** test-dev

Experiments

Ablation study: the effect of feature selection

	Heuristic guided	Semantic guided		AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
		Hard selection	Soft selection						
RetinaNet (anchor-based)	✓			35.7	54.7	38.5	19.5	39.9	47.5
Ours (anchor-free)	✓			35.9	54.8	38.1	20.2	39.7	46.5
		✓		37.0	55.8	39.5	20.5	40.1	48.5
			✓	38.0	56.9	40.5	21.0	41.1	50.2

Experiments

Ablation study: the effect of feature selection

	Heuristic guided	Semantic guided		AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
		Hard selection	Soft selection						
RetinaNet (anchor-based)	✓			35.7	54.7	38.5	19.5	39.9	47.5
Ours (anchor-free)	✓			35.9	54.8	38.1	20.2	39.7	46.5
		✓		37.0	55.8	39.5	20.5	40.1	48.5
			✓	38.0	56.9	40.5	21.0	41.1	50.2

Anchor-free branches with heuristic feature selection can achieve comparable performance with anchor-based counterparts.

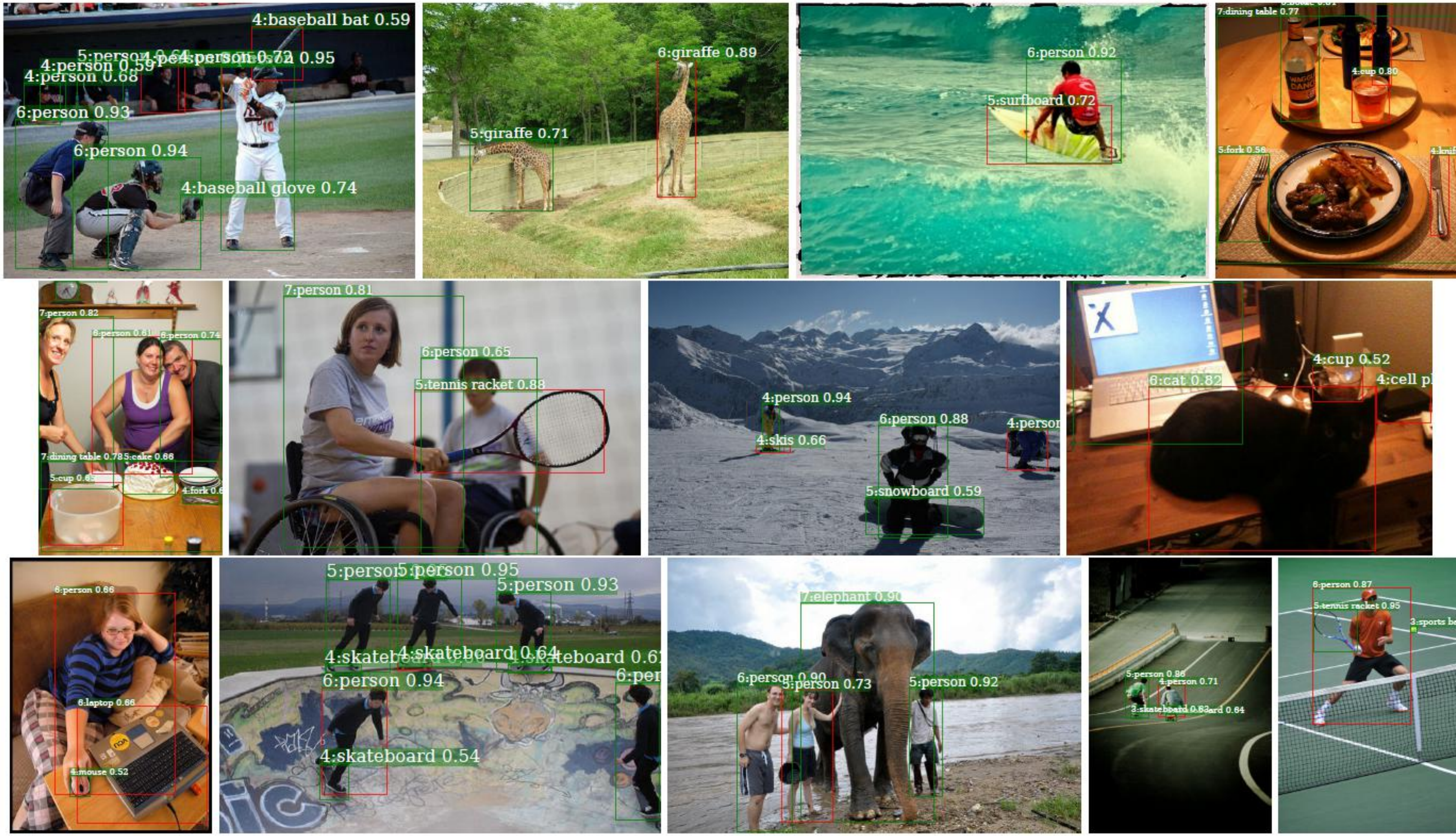
Experiments

Ablation study: the effect of feature selection

	Heuristic guided	Semantic guided		AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
		Hard selection	Soft selection						
RetinaNet (anchor-based)	✓			35.7	54.7	38.5	19.5	39.9	47.5
Ours (anchor-free)	✓			35.9	54.8	38.1	20.2	39.7	46.5
		✓		37.0	55.8	39.5	20.5	40.1	48.5
			✓	38.0	56.9	40.5	21.0	41.1	50.2

Hard version of semantic guided feature selection
chooses more suitable feature levels than heuristic guided selection.

Visualization of hard feature selection



Experiments

Ablation study: the effect of feature selection

	Heuristic guided	Semantic guided		AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
		Hard selection	Soft selection						
RetinaNet (anchor-based)	✓			35.7	54.7	38.5	19.5	39.9	47.5
Ours (anchor-free)	✓			35.9	54.8	38.1	20.2	39.7	46.5
		✓		37.0	55.8	39.5	20.5	40.1	48.5
			✓	**38.0**	**56.9**	**40.5**	**21.0**	**41.1**	**50.2**

Hard selection doesn't fully explore the network potential.
Using similarity from multiple features is helpful.

Visualization of soft feature selection

Visualization of soft feature selection

Experiments

Ablation study: the effect on different feature pyramids

Feature pyramid	Heuristic guided selection	Semantic guided selection	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FPN	✓		35.9	54.8	38.1	20.2	39.7	46.5
		✓	38.0	56.9	40.5	21.0	41.1	50.2
BFP	✓		36.8	57.2	39.0	22.0	41.0	45.9
		✓	**38.8**	**58.7**	**41.3**	**22.5**	**42.6**	**50.8**

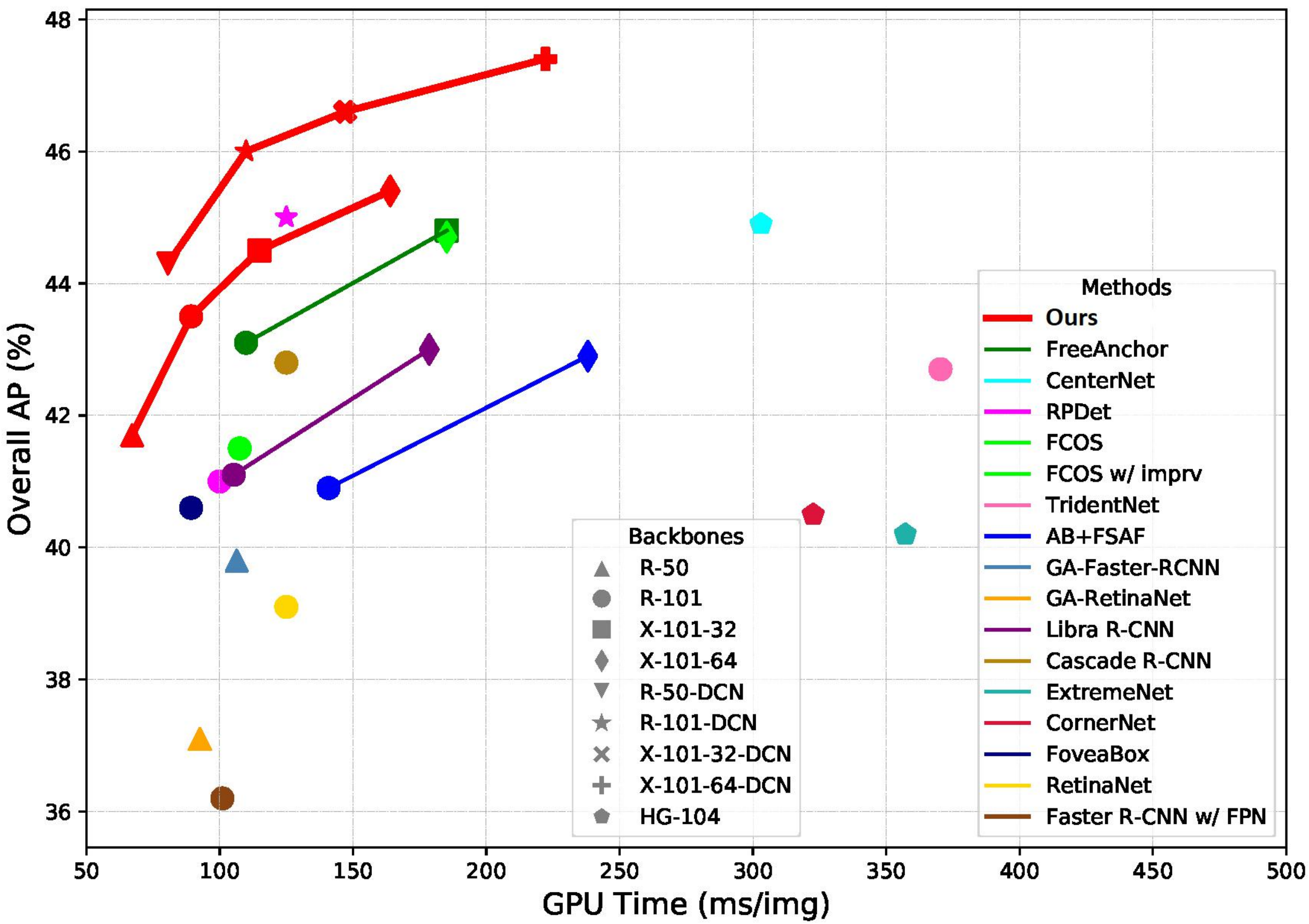
Experiments

Runtime analysis

Backbone	Method	AP	AP_{50}	Runtime (FPS)
ResNet-50	RetinaNet (anchor-based)	35.7	54.7	11.6
	Ours (anchor-free)	38.8	58.7	14.9
ResNet-101	RetinaNet (anchor-based)	37.7	57.2	8.0
	Ours (anchor-free)	41.0	60.7	11.2
ResNeXt-101	RetinaNet (anchor-based)	39.8	59.5	4.5
	Ours (anchor-free)	43.1	63.7	6.1

Experiments

Comparison with state of the arts



References

Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*. 2015.

Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.

Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.

Pang, Jiangmiao, et al. "Libra r-cnn: Towards balanced learning for object detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Wang, Jingdong, et al. "Deep high-resolution representation learning for visual recognition." arXiv preprint arXiv:1908.07919 (2019).

Ghiasi, Golnaz, et al. "Nas-fpn: Learning scalable feature pyramid architecture for object detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Tan, Mingxing, et al. "Efficientdet: Scalable and efficient object detection." arXiv preprint arXiv:1911.09070 (2019).

Huang, Lichao, et al. "Densebox: Unifying landmark localization with end to end object detection." *arXiv preprint arXiv:1509.04874* (2015).

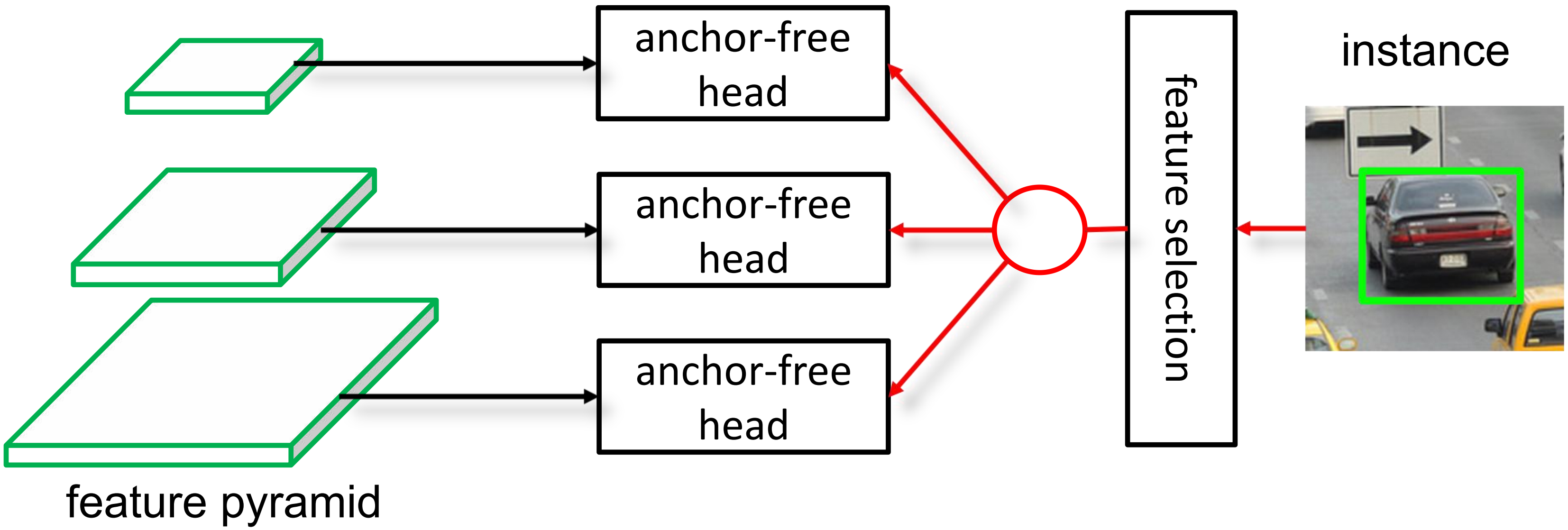
Zhu, Chenchen, Yihui He, and Marios Savvides. "Feature selective anchor-free module for single-shot object detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Zhu, Chenchen, et al. "Soft Anchor-Point Object Detection." *arXiv preprint arXiv:1911.12448* (2019).

Conclusion

Free feature selection is one of major differences between anchor-free and anchor-based methods.

Semantic guided feature selection is the key!



THANKS!