Bidirectional Matrix Feature Pyramid Network for Object Detection

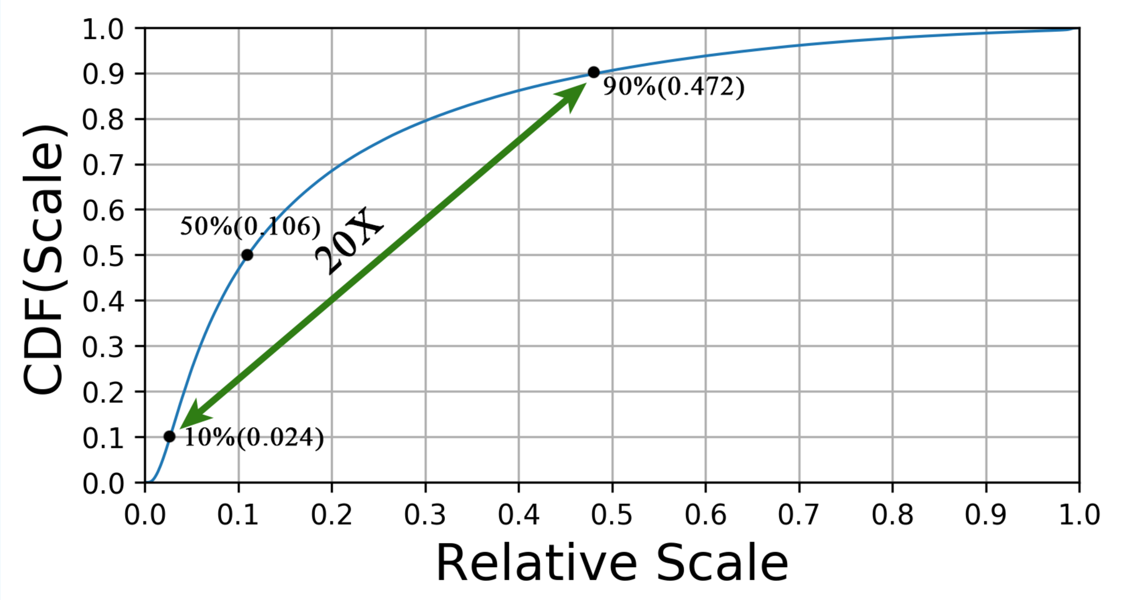
Wei Xu, Yi Gan, Jianbo Su
Shanghai Jiao Tong University
{xuweiVG, 627780002ganyi, jbsu}@sjtu.edu.cn

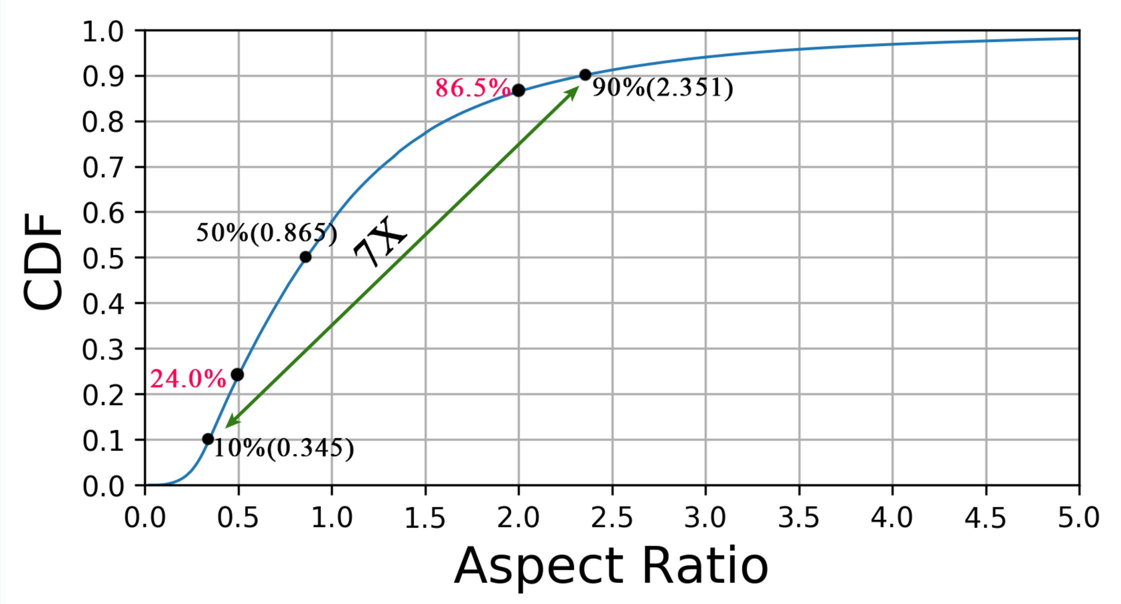
Introduction

- Most feature pyramids just map the objects to feature maps with relevant square receptive fields, but rarely pay attention to the aspect ratio variation. It will lead to a poor match between rectangular objects and assigned features with square receptive fields, thus preventing from accurate recognition and location.
- The information propagation among feature layers is sparse, namely, each feature in the pyramid may mainly or only contain single-level information, which is not representative enough for classification and localization sub-tasks.
- Bidirectional Matrix Feature Pyramid Network (BMFPN) is proposed to address these issues in this paper. The receptive fields of the designed feature layers have various scales and aspect ratios. Objects can be assigned to appropriate and representative feature maps with relevant receptive fields. Moreover, it forms bidirectional and reticular information flow, which effectively fuses multi-level information.

Problems

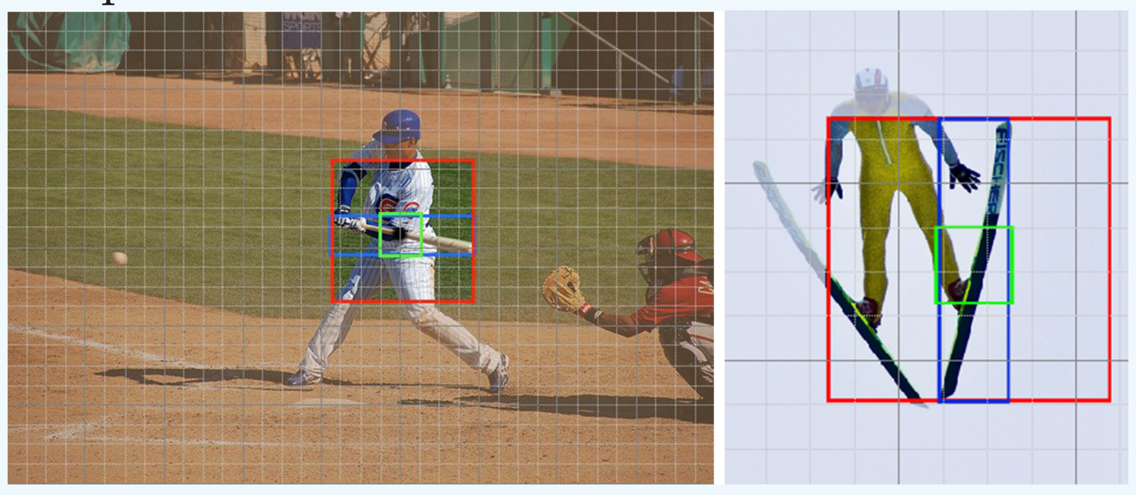
1. Statistical analysis of annotated bounding boxes (bboxes) in MS COCO.





There are quite a few slender and stubby objects in MS COCO.Not only the scale variation but also the aspect ratio variation should also be taken into account.

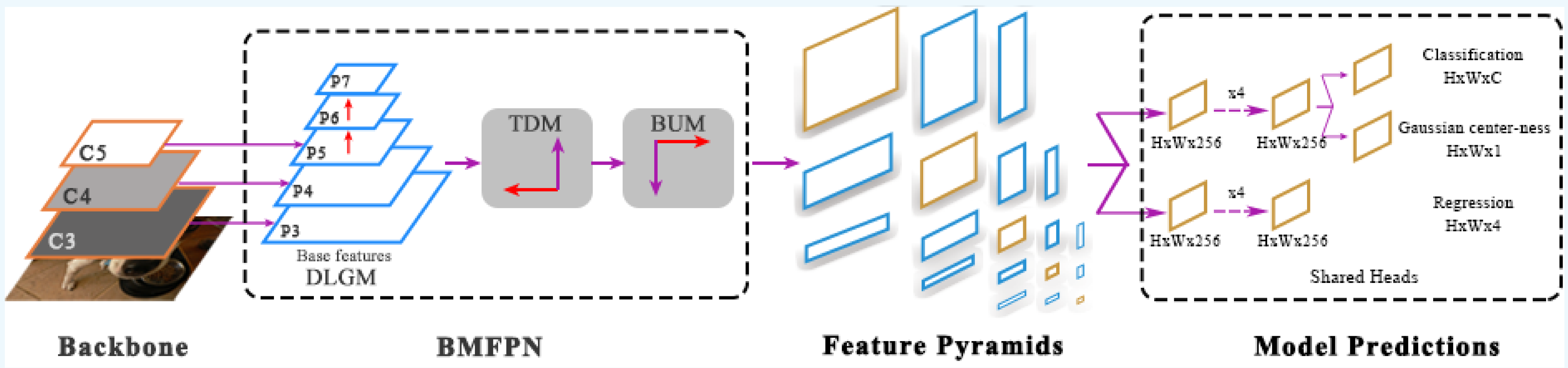
2. Examples of objects assignment when a slender object is mapped to a layer with rectangular receptive filed.



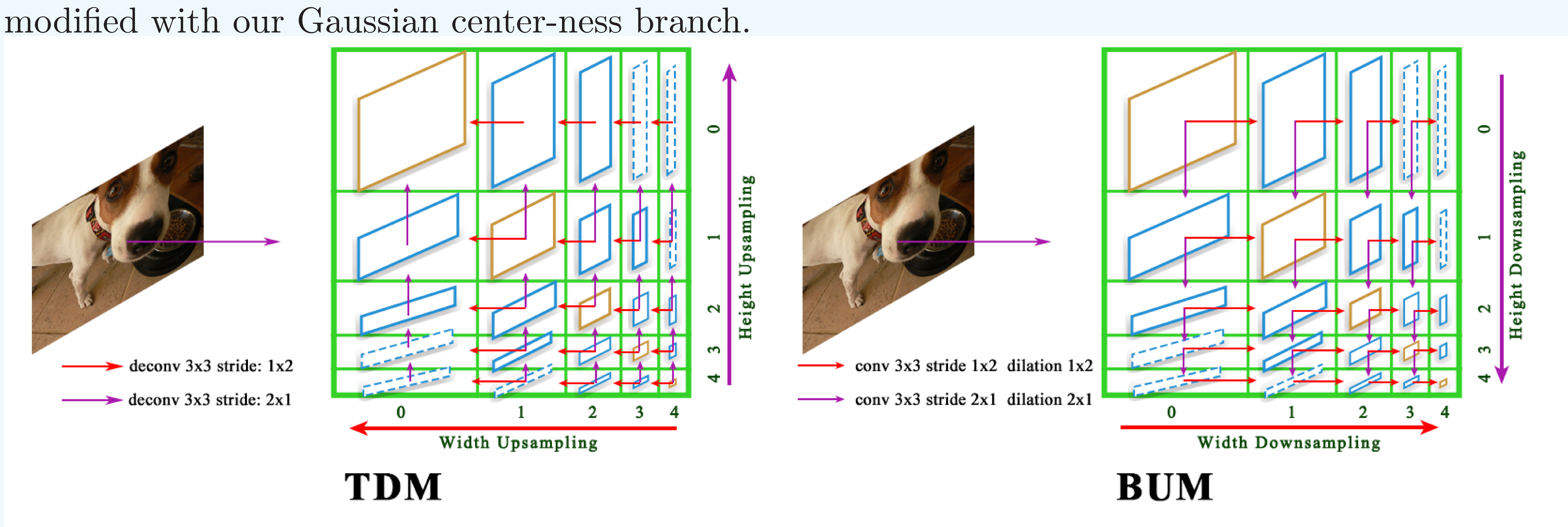
The poor match between objects and assigned features is bound to occur among a rectangular and a square receptive filed.

3. The information propagation among feature layers is sparse in most typical feature pyramids. For example, FPN [1] only forms a top-down pathway to propagate high-level information. Each feature in the pyramid may mainly or only contain single-level information, thus limiting the detection performance.

Methods



The features C3-C5 extracted from backbone are fed into our BMFPN. Then the feature pyramids are used for further predictions. Our methods is based on FCOS [2] and its center-ness branch is modified with our Gaussian center-ness branch.



TDM builds multiple top-down pathways to propagate high-level information from deep layers to shallow layers with de-convolution. Then only five diagonal layers in TDM are fed into BUM for computation reduction. BUM aims to construct a feature pyramid whose receptive field has different scales and aspect ratios. Owing to the asymmetric strides in dilated convolutions, the receptive fields of non-diagonal feature layers change to a rectangular region.

Results

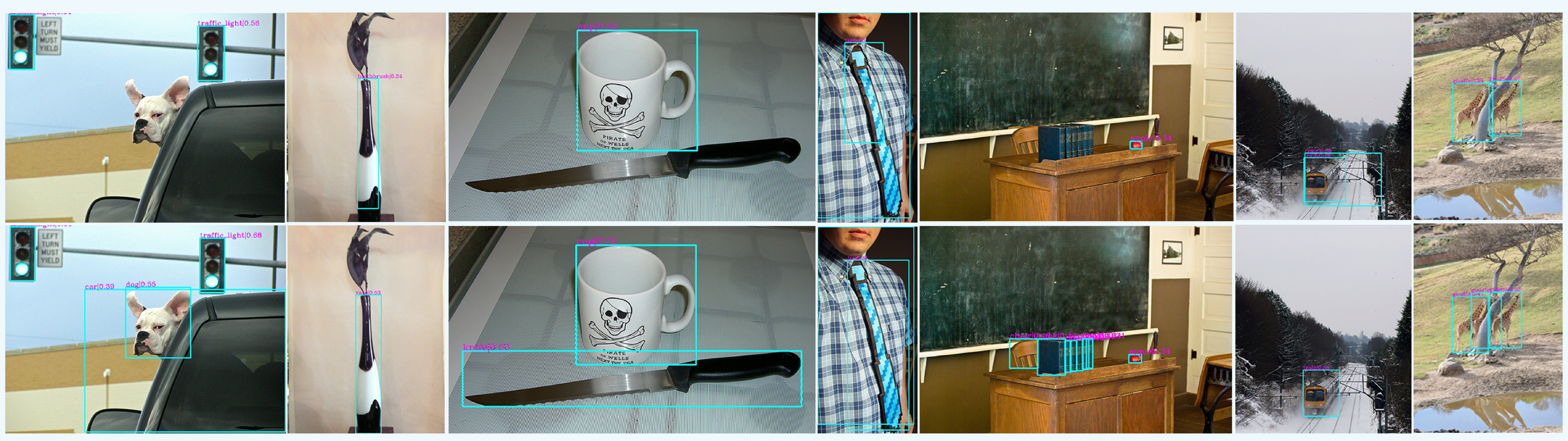
GCB	DLGM	TDM	BUM	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
				36.7	55.8	39.2	21.0	40.7	48.4
✓				37.0	55.9	39.5	21.3	40.9	48.6
	✓			33.2	49.9	35.4	14.5	36.1	47.8
	✓	✓		37.6	56.7	40.3	21.0	41.5	49.6
	✓		✓	36.5	53.0	39.3	16.9	40.6	52.2
	✓	✓	✓	39.7	57.7	42.8	22.4	44.4	53.0
✓	✓	✓	✓	**40.0**	**57.7**	**43.1**	**22.8**	**44.5**	**53.4**

With all these components added to FCOS [2], improvement on AP is 3.3% over baselines. The results shows that large size instances contribute most(+5.0%). And it makes more accurate detection with 3.9% improvement on AP75. Besides, as is shown in the visualization results, our model is better at finding challenging instances, such as very thin and obscured objects.

Conclusion

In this paper, the popular feature pyramids used in advanced object detectors is reinvestigated. It is found that a poor match between objects and feature maps exists in these approaches, thus preventing from accurate recognition and location. Besides, we reveal that the sparse information flow in the pyramid can not provide representative enough information for model prediction. Bidirectional Matrix Feature Pyramid Network is proposed to address these issues. It constructs a feature pyramid whose receptive fields have various scales and aspect ratios. And it forms bidirectional and reticular information flow, which effectively fuses multi-level information. To evaluate our proposed architecture, an anchor-free detector is designed and trained by integrating BMFPN into FCOS [2]. Extensive experiments demonstrate the effectiveness of the proposed architecture and novel modules.

Visualization results



References

[1] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[2] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 9627–9636, 2019.