

# Supplementary Material for Meta-ZSDETR: Zero-shot DETR with Meta-learning

Lu Zhang<sup>1</sup> Chenbo Zhang<sup>1</sup> Jiajia Zhao<sup>2</sup> Jihong Guan<sup>3</sup> Shuigeng Zhou<sup>1\*</sup>

<sup>1</sup>Shanghai Key Lab of Intelligent Information Processing, and School of Computer Science, Fudan University, China

<sup>2</sup>Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, China

<sup>3</sup>Department of Computer Science & Technology, Tongji University, China

{l\_zhang19, sgzhou}@fudan.edu.cn, cbzhang21@m.fudan.edu.cn

zhaojiajia1982@gmail.com, jhguan@tongji.edu.cn

## 1. Additional experimental results

Here, we report the class-wise AP in 65/15 split of MS COCO. The results are shown in Tab. 1. As we can see, Meta-ZSDETR achieves the best results in most categories, and boosts the mAP to above 20 for the first time, proving the effectiveness of our method.

65/15	airp	trai	metr	cat	bear	scse	frbe	snrd	fork	swic	hdog	tlet	mose	tstr	hier	mAP
PL [3]	20	48.2	0.6	28.3	13.8	12.4	<b>21.8</b>	15.1	8.9	8.5	0.9	5.7	0.0	1.7	0.0	12.4
SU [1]	10.1	48.7	1.2	64.0	64.1	12.2	0.7	28	16.4	19.4	0.1	18.7	1.2	0.5	0.2	19.0
Robust-Syn [2]	20.8	53.0	1.3	<b>64.3</b>	55.5	11.6	0.4	<b>31.3</b>	18.0	20.3	0.1	15.2	<b>4.2</b>	0.5	0.6	19.8
Meta-ZSDETR	<b>22.1</b>	<b>58.1</b>	<b>5.9</b>	60.6	<b>69.2</b>	<b>15.4</b>	3.2	28.0	<b>23.4</b>	<b>23.2</b>	<b>2.3</b>	<b>20.7</b>	1.5	<b>2.9</b>	<b>1.0</b>	<b>22.5</b>

Table 1. Class-wise AP and mAP on unseen classes of MS COCO dataset in ZSD setting with IoU=0.5.

## References

- [1] Nasir Hayat, Munawar Hayat, Shafin Rahman, Salman Khan, Syed Waqas Zamir, and Fahad Shahbaz Khan. Synthesizing the unseen for zero-shot object detection. In *ACCV*, 2020. 1
- [2] Peiliang Huang, Junwei Han, De Cheng, and Dingwen Zhang. Robust region feature synthesizer for zero-shot object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7622–7631, 2022. 1
- [3] Shafin Rahman, Salman Khan, and Nick Barnes. Improved visual-semantic alignment for zero-shot object detection. In *AAAI*, volume 34, pages 11932–11939, 2020. 1
- [4] Caixia Yan, Xiaojun Chang, Minnan Luo, Huan Liu, Xiaoqin Zhang, and Qinghua Zheng. Semantics-guided contrastive network for zero-shot object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [5] Ye Zheng, Ruoran Huang, Chuanqi Han, Xi Huang, and Li Cui. Background learnable cascade for zero-shot object detection. In *ACCV*, 2020.

---

\*correspondence author