MFA: Multi-layer Feature-aware Attack for Object Detection (Supplementary Material)

Wen Chen[1] **Yushan Zhang**[3] **Zhiheng Li**[1] **Yuehuan Wang**[1,2*]

[1]School of Artificial Intelligence and Automation , Huazhong University of Science and Technology, Wuhan, China
[2]National Key Lab of Science and Technology on Multi-spectral Information Processing , Wuhan, China
[3]Shanghai Institute of Satellite Engineering , Shanghai, China

A COMPUTATIONAL COST ANALYSIS

The time cost of generating adversative camouflage is mainly concentrated in the training phase. We compared MFA with the FCA[Wang et al., 2022b] on an NVIDIA RTX 1080Ti 12GB GPU. Table 1 lists the results of computational cost statistics. our method is nearly equivalent to FCA in terms of speed. This is because FCA modifies the three loss functions of yolov3[Redmon and Farhadi, 2018] to serve as adversarial losses. Its training process is consistent with that of yolov3 and does not involve post-processing operations. In contrast, our method employs post-processing to filter out target objects, avoiding redundancy of targets. Our method includes some additional operations, such as feature attribution, but the reduced number of objects being operated on results in no significant additional time

Table 1: The results of computational cost statistics.

Method	Epoch1	Epoch2	Epoch3	Epoch4	Epoch5	**Total**
FCA	3:06:50	3:06:11	3:06:18	3:06:56	3:06:51	15:33:06
MFA	3:05:22	3:03:03	3:02:50	3:03:04	3:03:15	15:17:34

B SOURCE MODEL EXPERIMENT

We experimented with yolov5 as another source model to strengthen the effectiveness of the MFA. We select nine models as target models to evaluate black box attacks. Specifically, SSD[Liu et al., 2016], Faster RCNN[Ren et al., 2015], Mask RCNN[He et al., 2017], Cornernet[Law and Deng, 2018], FCOS[Tian et al., 2019], Swin Transformer[Liu et al., 2021], TOOD[Feng et al., 2021], VFNet[Zhang et al., 2021] and yolov7[Wang et al., 2022a]. The comparison result of different source models in the digital space are shown in table 2.

Table 2: The comparison result of different source models in the digital space.

Source model	ASR(%)									
	SSD	Faster	Mask	Corner	FCOS	Swin	TOOD	VFNet	yolov7	**mASR**
yolov3	96.39	92.69	92.98	85.62	98.86	87.81	94.01	82.64	89.96	91.22
yolov5	92.19	95.26	92.81	92.16	88.06	97.01	98.32	89.68	94.84	93.37

*Corresponding Author

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

References

Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3490–3499. IEEE Computer Society, 2021.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In *Proceedings of the European conference on computer vision (ECCV)*, pages 734–750, 2018.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

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

Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022a.

Donghua Wang, Tingsong Jiang, Jialiang Sun, Weien Zhou, Zhiqiang Gong, Xiaoya Zhang, Wen Yao, and Xiaoqian Chen. Fca: Learning a 3d full-coverage vehicle camouflage for multi-view physical adversarial attack. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2414–2422, 2022b.

Haoyang Zhang, Ying Wang, Feras Dayoub, and Niko Sunderhauf. Varifocalnet: An iou-aware dense object detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8514–8523, 2021.