Supplementary Information

MSAL-YOLO: a YOLOv8-based detector for small and densely distributed object detection in UAV aerial imagery

Fei Ding, Xiufu Du, Haining Zhang, Haibin Liu, Liguo Han

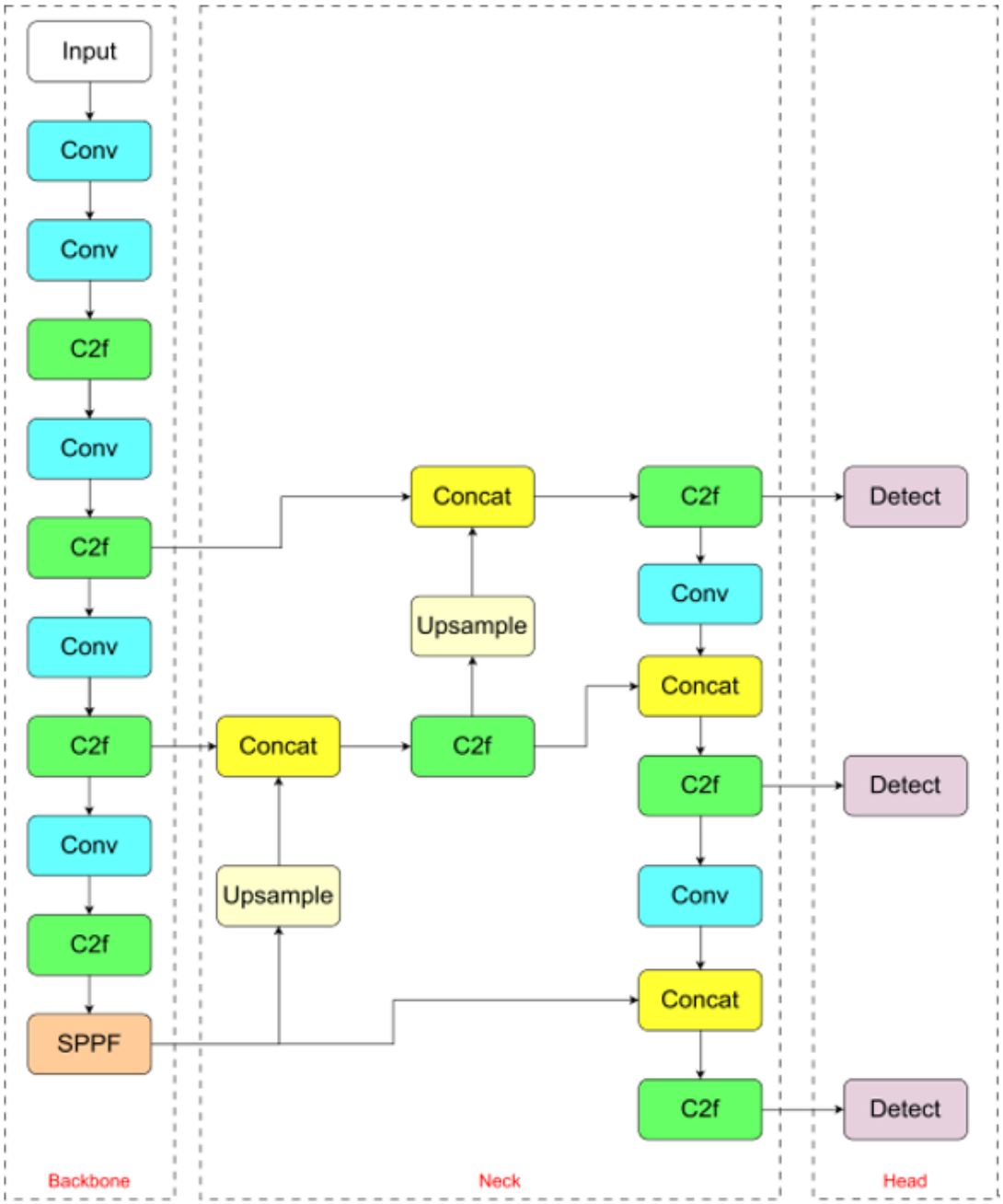
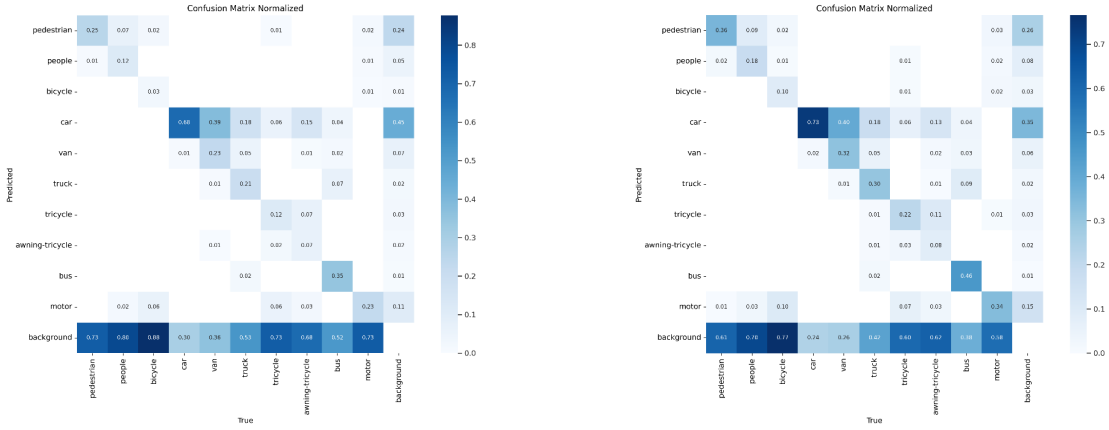


Figure S1: **Supplementary Fig. S1.** Overall architecture of the baseline YOLOv8n network used in this study. The model consists of a backbone, a neck, and a decoupled detection head.

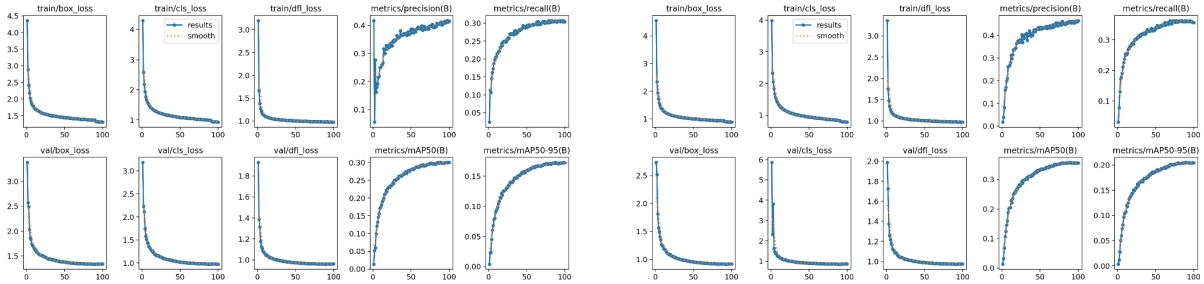
Table S1: Class-wise AP_{50} comparison on the VisDrone2019 validation set.

Category	YOLOv8n (%)	MSAL-YOLO (%)	Improvement (%)
Pedestrian	31.2	38.0	+6.8
People	25.3	28.9	+3.6
Bicycle	6.2	9.3	+3.1
Car	73.9	77.4	+3.5
Van	35.3	42.9	+7.6
Truck	25.7	32.5	+6.8
Tricycle	18.6	23.9	+5.3
Awning-tricycle	10.5	12.7	+2.2
Bus	39.8	52.5	+12.7
Motor	33.3	38.9	+5.6



(a) YOLOv8n (b) MSAL-YOLO

Figure S2: **Supplementary Fig. S2.** Comparison chart of confusion matrices.



(a) YOLOv8n (b) MSAL-YOLO

Figure S3: **Supplementary Fig. S3.** Comparison of training process curves for YOLOv8n and MSAL-YOLO on the VisDrone2019 dataset.

Figure S4: **Supplementary Fig. S4.** Failure cases under extreme conditions. Representative examples where performance degradation occurs due to severe object occlusion, extreme target density, background similarity, and extremely small object scales. Red arrows indicate typical missed detections or inaccurate localizations.