Feature Information Driven Position Gaussian Distribution Estimation for Tiny Object Detection (supplementary material)

Jinghao Bian[1] Mingtao Feng[1,3*] Weisheng Dong[1] Fangfang Wu[1] Jianqiao Luo[2]
Yaonan Wang[2] Guangming Shi[1]
[1]Xidian University, [2]Hunan University
[3]Jiangxi Communication Terminal Industrial Technology Research Institute

1. Network Architecture Details

We list the detailed components of Position Gaussian Distribution Prediction module in Tab. 1, which are mainly composed of convolution and transposed convolution layers [2].

Layer name	Components	
Up&Conv1	4×4 TransConv	[3×3 Conv, BN, Relu]×2
Up&Conv2	4×4 TransConv	[3×3 Conv, BN, Relu, 3×3 Conv]×1
Up&Conv3	4×4 TransConv	[3×3 Conv, BN, Relu, 3×3 Conv]×1
Conv4	[3×3 Conv, BN, Relu]×2	3×3 Conv

Table 1. Detailed components of Position Gaussian Distribution Prediction module.

2. Additional Ablation Study

Impact on the performance of different $\mathcal{L}_{IE}$ hyper-parameter λ_1. We investigate the performance impact of varying the $\mathcal{L}_{IE}$ hyper-parameter λ_1, while keeping λ_2 fixed at 1.0. As shown in Tab. 2, among the empirically tested hyper-parameter values, $\lambda_1 = 0.01$ yields the best performance. From the perspective of minimizing the Information Entropy loss $\mathcal{L}_{IE}$, the objective is to reduce the mean amount of feature information to achieve the lowest overall encoding cost. However, assigning excessive optimization weight to $\mathcal{L}_{IE}$ causes the network to overlook abundant object details to minimize the Information Entropy loss, which is particularly detrimental for tiny objects with extremely limited pixel counts. Conversely, assigning too little optimization weight results in insufficient exploration of the image's spatial structure, leading to a suboptimal information map for feature enhancement.

λ_1	AP	$AP_{0.5}$	AP_{vt}	AP_t	AP_s
0.001	**28.3**	48.4	**3.5**	12.1	26.0
0.01 (Ours)	**28.3**	**48.5**	**3.5**	**12.6**	**26.1**
0.1	27.6	47.5	3.2	12.1	25.5
1.0	27.6	47.4	3.0	12.0	25.5

Table 2. Performance of different λ_1, where λ_2 is fixed to 1.0.

*Corresponding author

Impact on the performance of different $\mathcal{L}_{pred}$ hyper-parameter λ_2. We evaluate the performance with varying $\mathcal{L}_{pred}$ hyper-parameter λ_2, keeping λ_1 fixed at 0.01. Empirically, different values of λ_2 are selected for training. As shown in Tab. 3, smaller values of λ_2 impair the network's ability to predict the distribution map, while larger values intensify competition among different loss terms, negatively affecting the final classification and regression of tiny objects. Based on these results, we select $\lambda_2 = 1.0$ as the optimal value.

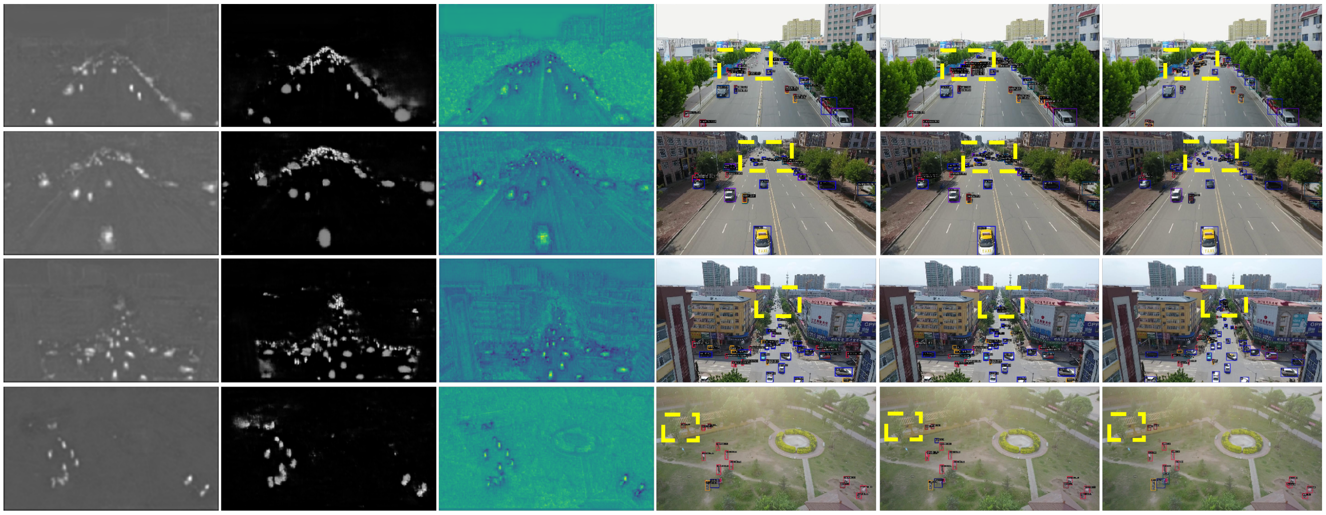
λ_2	AP	$AP_{0.5}$	AP_{vt}	AP_t	AP_s
0.01	27.9	47.7	2.7	12.4	26.0
0.1	**28.3**	48.3	3.4	**12.9**	25.8
1.0 (Ours)	**28.3**	**48.5**	**3.5**	12.6	26.1
10.0	**28.3**	**48.5**	3.2	12.2	**26.5**

Table 3. Performance of different λ_2, where λ_1 is fixed to 0.01.

3. More Visualization Results

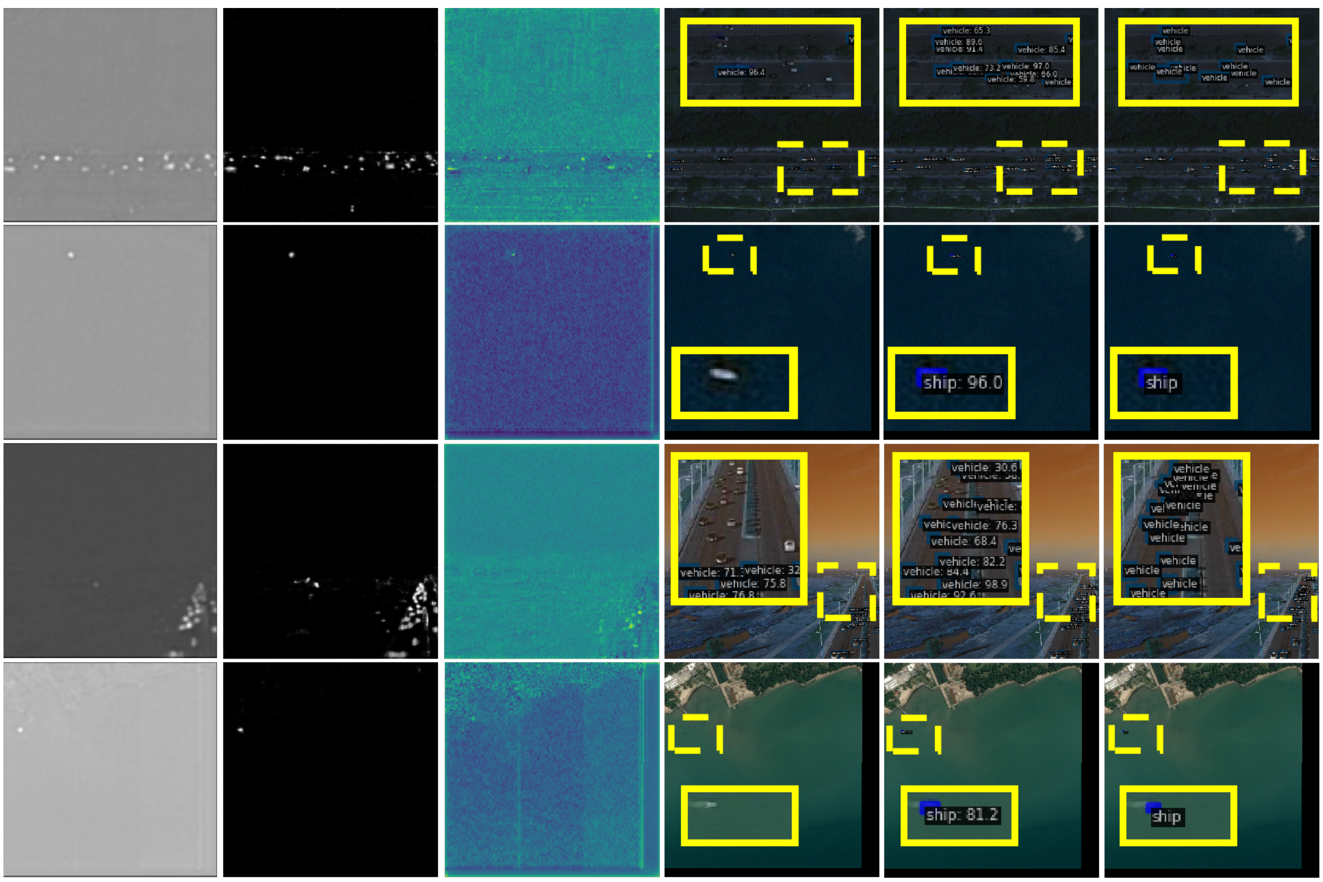
More Visualization Results on VisDrone2019. We provide additional visualization results on the VisDrone2019 dataset in Fig. 1. These qualitative results demonstrate that the information map and distribution map derived from our proposed method effectively highlight the salient regions in the image. The information map extensively exploits the spatial structure of the targets, while the distribution map enables tiny objects to stand out prominently from general objects and the background. Guided by these two maps, the enhanced feature $P_2^{'}$ allows tiny objects to be clearly distinguishable from the background, resulting in improved classification and regression performance. Consequently, our method successfully detects tiny objects at significant distances from the drone's perspective, whereas the baseline method fails to capture these tiny objects due to weak representations. These detection results further validate the effectiveness of our approach in enhancing tiny object representations.

Visualization Results on AI-TOD and AI-TODv2. We present visualizations of qualitative results on AI-TOD and



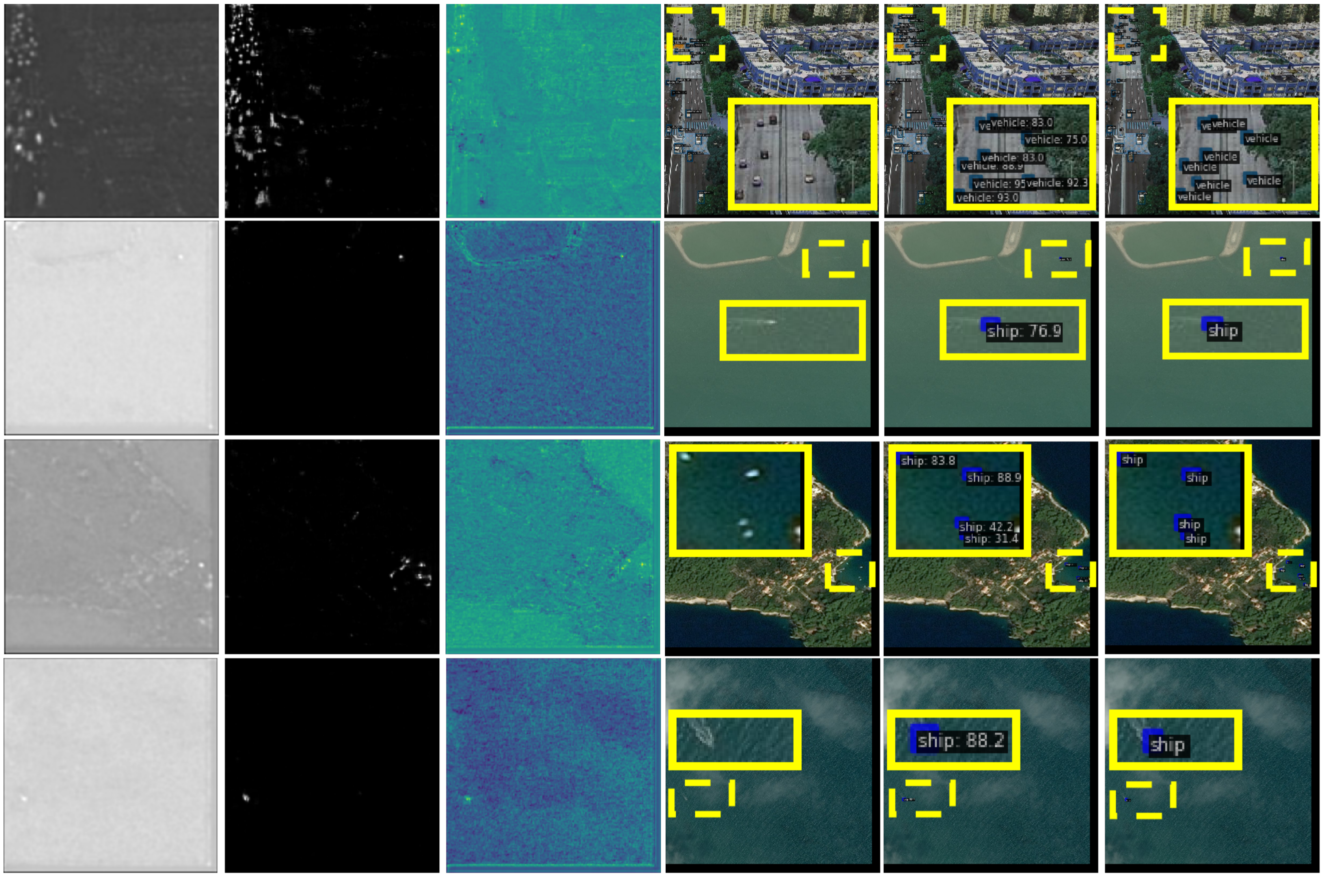
(a) Information map (b) Distribution map (c) Enhanced feature (d) DetectoRS (e) DetectoRS w/ ours (f) GT

Figure 1. More visualization of qualitative results on VisDrone2019. (a)-(c) are estimated information map, predicted distribution map and final enhanced feature, (d)-(f) are the detection results of DetectoRS [1], our method and GT, dotted boxes are drawn for better comparison. Best viewed on screen with zoom.



(a) Information map (b) Distribution map (c) Enhanced feature (d) DetectoRS (e) DetectoRS w/ ours (f) GT

Figure 2. More visualization of qualitative results on AI-TOD. Images are padded if necessary. Dotted boxes are drawn for better comparison, and dotted boxes are amplified for clearer observation due to targets' extremely limited pixel occupancy. Best viewed on screen with zoom.



(a) Information map **(b) Distribution map** **(c) Enhanced feature** **(d) DetectoRS** **(e) DetectoRS w/ ours** **(f) GT**

Figure 3. More visualization of qualitative results on AI-TODv2. Images are padded if necessary. Dotted boxes are drawn for better comparison, and dotted boxes are amplified for clearer observation due to targets' extremely limited pixel occupancy. Best viewed on screen with zoom.

AI-TODv2, as shown in Fig. 2 and Fig. 3, respectively. The annotated targets in these datasets are extremely tiny, posing significant challenges to generic detectors. As seen in Fig. 2(d) and Fig. 3(d), the baseline generic detector fails to detect most tiny objects, primarily due to the difficulty in extracting features from objects with extremely limited pixel information. In contrast, the information map and distribution map generated by our method effectively highlight tiny objects, identifying regions for subsequent feature enhancement. Notably, as illustrated in Fig. 2(c) and Fig. 3(c), the enhancement process restores the visibility of previously overlooked tiny representations, contributing to the improved detection results within the amplified boxes. These qualitative results further demonstrate the generalization capability of our proposed method for tiny object detection.

References

[1] Siyuan Qiao, Liang-Chieh Chen, and Alan Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10213–10224, 2021. 2

[2] Matthew D Zeiler, Graham W Taylor, and Rob Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *2011 international conference on computer vision*, pages 2018–2025. IEEE, 2011. 1