R4Det: 4D Radar-Camera Fusion for High-Performance 3D Object Detection

Supplementary Material

A. Additional ablation of the PDF module

Table 8 further presents the experimental process of adjusting the loss weight of the PDF module. Sensitivity analysis (C1–C4) shows that slight variations of λ_1, λ_3, λ_{dense}, or λ_{abs} lead to minor performance drops, confirming that the final setting (C5) achieves a balanced trade-off between stability and accuracy.

Table 8. **Ablation of the Panoramic Depth Fusion module.** Each setting corresponds to different combinations of loss components.

Setting	λ_1	λ_{abs}	λ_{dense}	λ_3	BEV mAP↑
A	0.1	0.01	–	–	45.15
B (+λ_{dense})	0.1	0.01	0.03	–	46.08
C1 (↓λ_1)	0.05	0.01	0.03	0.05	46.64
C2 (↓λ_3)	0.1	0.01	0.03	0.02	46.30
C3 (↑λ_{dense})	0.1	0.01	0.05	0.05	46.45
C4 (↑λ_{abs})	0.1	0.03	0.03	0.05	46.12
C5 (ours)	0.1	0.01	0.03	0.05	**46.86**

B. Additional ablation of the DGTF module

We also examine temporal depth sensitivity (Table 9): using the immediate predecessor $(t-1)$ achieves the best accuracy and stability, whereas fusing more distant frames ($t-2$, $t-3$) leads to noise accumulation.

Table 9. **Sensitivity to historical frame gap.** Evaluation of different historical fusion depths. Single-step $(t-1)$ achieves the best accuracy and stability.

Fusion Depth	BEV mAP	3D mAP
t & $t-3$	49.20	42.74
t & $t-2$	49.87	43.12
$\boldsymbol{t}$ & $\boldsymbol{t-1}$	**50.41**	**44.86**

C. Qualitative Analyses of the DGTF module

R4Det is the first to validate the feasibility of temporal modeling on these two datasets. As shown in the Figure 8, the learned offsets Δp (red arrows) can align with object motion, reconstructing relative motion flow, while the mask m precisely suppresses the background. This physically proves DGTF performs explicit, physics-based motion alignment.

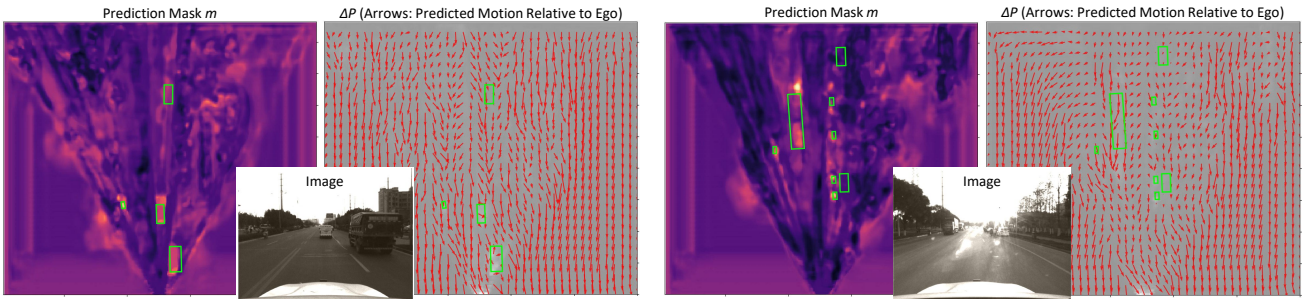

Figure 8. **DGTF Visualization.** Learned offsets (Δp, red arrows) align strictly with vehicle motion, while the mask (m, heatmap) suppresses background, proving explicit temporal alignment.