TARGET DRIVEN ADAPTIVE LOSS FOR INFRARED SMALL TARGET DETECTION

Supplementary Material

Table 3: We compared our proposed loss and existing losses with the metrics of IoU (%), $P_d(\%)$, and $F_a(10^{-6})$. The best and second values are highlighted in bold and underline, respectively.

Method	U-Net [24]			UIUNet [11]			SCTransNet [13]		
	IoU↑	P_d↑	F_a↓	IoU↑	P_d↑	F_a↓	IoU↑	P_d↑	F_a↓
BCE loss [2]	64.76	88.65	**10.32**	64.56	88.77	10.38	67.03	89.00	**6.98**
IoU loss [18]	63.14	91.40	<u>19.58</u>	<u>63.70</u>	88.43	22.88	67.38	88.60	18.97
Dice loss [19]	62.19	91.75	12.37	62.27	90.81	22.64	67.72	<u>89.69</u>	16.64
SLS loss [9]	<u>66.24</u>	92.43	38.10	<u>65.20</u>	<u>90.81</u>	25.02	<u>68.22</u>	88.65	21.18
SLS loss + TDA loss (Ours)	**69.48**	**94.84**	14.04	**66.96**	**92.17**	**9.56**	**68.59**	**91.75**	<u>9.41</u>

The proposed TDA loss can be applied to prior segmentation models developed for IRSTD and enhance their detection performance. To validate this claim, we evaluated TDA loss on three other architectures: a simple U-Net, and two state-of-the-art models, UIU-Net and SCTransNet [11, 13, 24]. U-Net has a simple encoder and decoder architecture, and it is able to capture both global context and fine-grained features. UIUNet consists of a U-Net embedded with smaller U-Nets. UIUNet effectively extracts multi-scale information using a nested architecture. SCTransNet introduces spatial-channel cross transformer blocks. SCTransNet encodes global context information and improves detection performance for targets with high similarity to the background.

We used Adam as the optimizer for all models. For U-Net, we adopted the settings from the open-source implementation[2]: learning rate 0.0005, batch size 16, 400 epochs, with the learning rate halved at epochs 200 and 300. For UIUNet and SCTransNet, we followed their original papers, using a learning rate of 0.001. UIUNet used batch size 3 and 500 epochs; SCTransNet used batch size 16 and 1000 epochs, with cosine annealing to 0.00001.

Table 3 shows the detection performance of our TDA loss and existing losses with three different architectures: UNet, UIUNet, and SCTransNet. We used IRSTD-1k dataset for training and evaluation. The results show that our TDA loss enhances the detection performance for three different networks. These results indicate that the proposed method enables the model to extract target shapes more accurately and improves detection performance for more challenging targets.

To verify the applicability of our TDA loss to existing loss functions, we evaluated its performance in combination with IoU loss, Dice loss, and BCE loss. We used MSHNet as a backbone model. Table 4 compares the detection performance with and without TDA loss on the IRSTD-1k dataset. The result demonstrates that our TDA loss comprehensively improves performance and can be effectively integrated with various existing losses for IRSTD.

Table 4: Comparison of detection performance on IRSTD-1k with and without the proposed TDA loss. The proposed TDA loss consistently improves performance when combined with BCE, IoU, and Dice losses. The better values for each metric are highlighted in bold.

Method	IoU↑	P_d↑	F_a↓
BCE loss	63.32	90.72	**6.37**
BCE loss + TDA loss	**65.13** (+1.81)	**96.21** (+5.49)	18.52 (+12.15)
IoU loss	65.57	87.62	**6.83**
IoU loss + TDA loss	**67.31** (+1.74)	**91.06** (+3.44)	9.71 (+2.88)
Dice loss	64.60	89.00	8.50
DIce loss + TDA loss	**67.34** (+2.74)	**90.72** (+1.72)	**7.43** (-1.07)

[2] https://github.com/XinyiYing/BasicIRSTD