*2025 International Conference on Intelligent Computing*
*July 26-29, Ningbo, China*
https://www.ic-icc.cn/2025/index.php

# Light but Mighty: When Lightweight Meets Stable Detection in Infrared Small Target Detection

Pengyuan Zhang[1], Cheng Zhang[1], Xubing Yang[1], Yan Zhang[1] and Li Zhang[1]

[1] College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China

{Arrosw, zcheng, xbyang, zhangyan, lizhang}@njfu.edu.cn

**Abstract.** With the development of deep learning, infrared small target detection methods have yielded promising results benefiting from the powerful feature extraction capability of deep neural networks. However, these methods with large numbers of parameters are often impractical due to hardware limitations, while existing lightweight models tend to be unstable as they struggle to effectively capture small targets. To cope with these challenges, we propose a novel light but mighty network (Limi-Net), which maintains lightweight while having a stable ability to capture small targets. First, since the targets are rare in infrared images, we propose an Infrared Target Simulator that generates pseudo targets for data augmentation, helping the model to better learn and recognize small targets. Then a Lightweight Stable Encoder is designed to guarantee reliable feature extraction from diverse receptive fields to improve the discrimination of small targets and reduce memory consumption. In addition, we introduce a Coarse-to-fine Hybrid Upsampling Decoder that combines a dual upsampling fusion method and a coarse to fine alignment mechanism to integrate multi-scale features while preserving critical information. Extensive experiments demonstrate that Limi-Net achieves state-of-the-art (SOTA) performance while maintaining a lightweight architecture, making it well-suited for practical deployment. Our code is available at https://github.com/Arrosw/LimiNet.

**Keywords:** Infrared small target, Deep learning, Lightweight, Coarse to fine training.

## 1 Introduction

Infrared small target detection (IRSTD) is critical in various fields such as maritime surveillance, remote sensing, and military defense that require real-time and precise response [14, 15]. In these situations, the infrared small target detector often needs to be deployed on edge devices. Therefore, it is particularly important to develop a high-precision detector with lightweight architecture.

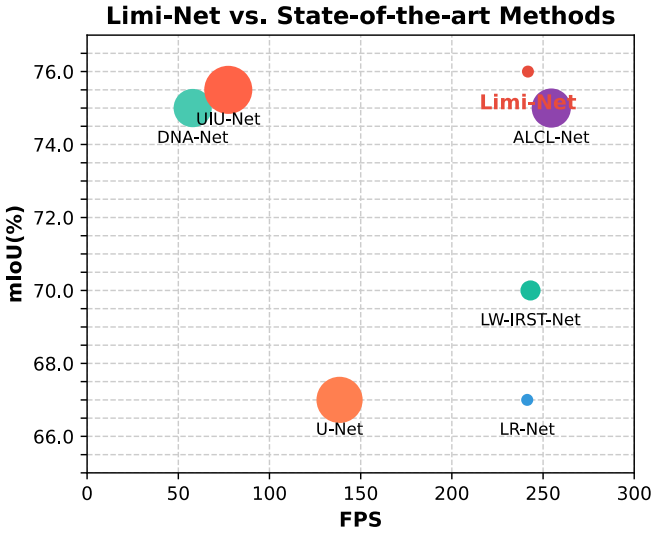


**Fig. 1.** Comparison of mean MIoU, FPS and model size on the IRSTD-1k [12] dataset and NUDT-SIRST [6] dataset. The size of circle corresponds to the number of various model parameters. Frames Per Second (FPS) presents the real-time respond of model, which is measured on an RTX 4090 24 GB GPU.

Infrared sensors excel in low-visibility environments and capture solely thermal radiation signals, which unlike human visual system, making it difficult to maintain details of small targets. To detect infrared small targets, numerous methods have been proposed. Early research on infrared small target detection focused mainly on traditional image processing methods based on human prior knowledge, including filtering-based [1], local-contrast-based [2], and low-rank-based methods [3]. However, these traditional methods rely on manually designed features, which means poor generalization and sensitivity to the scene.

In recent years, with the development of deep learning, especially the powerful feature extraction capability presented by Convolutional neural networks (CNNs), more and more research on infrared small target detection has focused on CNN-based infrared small target detection methods, which can automatically learn intrinsic features from the infrared image in a data-driven manner, making them more robust to various scenarios. Recent studies have shown that it is more effective to treat infrared small target detection as a semantic segmentation task and use U-Net [5] shaped models to process it [4-8]. However, despite their effectiveness, these methods are often computationally expensive and memory-intensive, which limits their practical deployment in real-time or resource-constrained applications. As a result, researchers pay more attention to the lightweight of the network [9-10]. Nevertheless, due to limited feature extraction capabilities, these lightweight models still struggle to accurately detect small targets.

To address these limitations, we aim to construct a lightweight infrared small target detector with both high-precision and real-time response. In this paper, a novel light but mighty network (Limi-Net) is proposed. First, based on the characteristics of infrared small targets, we propose an Infrared Target Simulator that generates pseudo targets for data augmentation, helping the model to learn better and recognize small targets. Given that the instability of initial feature extraction in lightweight architectures will lead to small targets hard to recognize, we explore it in depth and design a Lightweight Stable Encoder that contains a Stability-Enhanced Stem (SES) module and four Lightweight Multi-branch Bottleneck (LMB) modules. Specially, the SES module guarantees initial stable feature extraction and improves the discrimination of small targets. LMB modules fuse features from different receptive fields while reducing memory consumption. To further enhance cross-level feature representation and accuracy detection, we introduce a Coarse-to-fine Hybrid Upsampling Decoder, which contains a Hybrid Upsampling fusion (HUF) module that exploits dual upsampling to integrate cross-layer features and a Coarse-to-fine Multi-scale Head to capture multi-scale targets. The performance comparison shown in Fig. 1 and the experimental results demonstrate that Limi-Net achieves state-of-the-art (SOTA) performance while maintaining lightweight.

## 2 Related Work

Early investigations in this field mainly relied on traditional image processing techniques that incorporate human insights to separate small targets from cluttered backgrounds. Common approaches include filtering-based [1], local-contrast-based [2], and low-rank-based methods [3]. Although these techniques can yield satisfactory results under certain conditions, their reliance on hand-crafted features often restricts their adaptability to more diverse and complex scenarios.

With the rapid advancement of deep learning, an increasing number of studies have focused on applying deep learning techniques to infrared small target detection, particularly the convolutional neural networks (CNNs). CNN-based methods can automatically learn intrinsic representations from infrared images in a data-driven manner, improving their robustness across diverse scenarios. Recent research suggests that formulating infrared small target detection as a semantic segmentation task is more effective [4], leading to the adoption of U-Net shaped architectures [5]. For instance, DNA-Net [6] integrates a dense nested interaction module into the U-Net framework to mitigate the issue of small target loss. UIU-Net [7] enhances multi-scale representation learning by embedding a small U-Net within a larger U-Net backbone. ALCL-Net [8] uses attention-based local contrast learning to intrinsic features of infrared small targets. While these models achieve promising detection performance, their complex architectures result in high computational and memory demands, making them impractical for real-time applications due to hardware constraints.

To address this issue, recent efforts have shifted towards designing lightweight networks. LW-IRST-Net [9] achieves computational efficiency by combining standard convolutions, depth-wise separable convolutions, dilated convolutions, and asymmetric convolution modules, forming an optimized encoding-decoding structure. Similarly,

LR-Net [10] refines ALCL-Net by simplifying its architecture, thereby balancing detection accuracy and computational efficiency for real-world deployment. However, these lightweight models still struggle to accurately detect small targets due to limited feature extraction capabilities.

# 3 Method

## 3.1 Overall Architecture

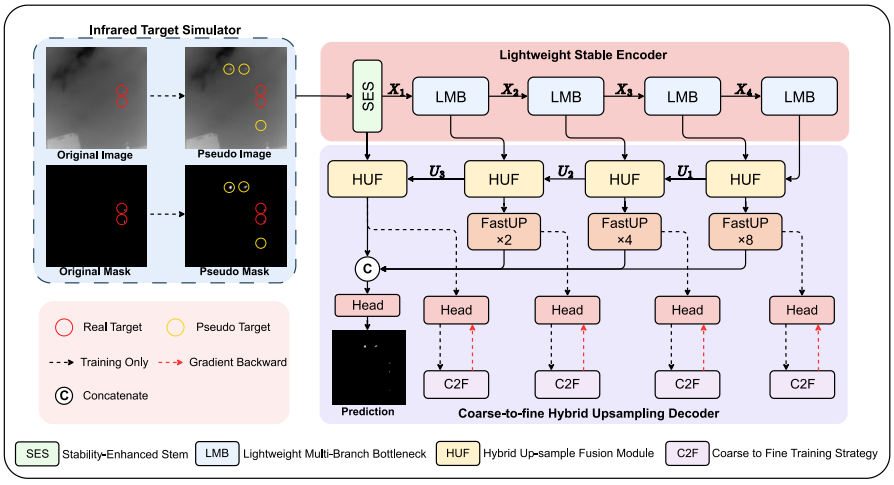


**Fig. 2.** The architecture of Limi-Net. Dashed lines indicate use only in the training phase, and ×k in the FastUp module means upsampling k times. (a) FastUp Module. (b) Segment Head.

To meet the requirements of high accuracy, real-time response, and lightweight design in infrared small target detection, we propose a novel Light but Mighty Network (Limi-Net). The overall architecture of Limi-Net consists of three main components: the Infrared Target Simulator (ITS), the Lightweight Stable Encoder, and the Coarse-to-Fine Hybrid Upsampling Decoder, as illustrated in Fig. 2.

Firstly, the ITS is employed to generate pseudo labels that enhance the training process by generating pseudo targets. Next, the Lightweight Stable Encoder is designed to perform efficient and stable feature extraction while maintaining a lightweight structure. Finally, the Coarse-to-Fine Hybrid Upsampling Decoder is responsible for accurate target segmentation by progressively integrating multi-scale features. The detailed designs of each module are described in the following subsections.

## 3.2 Infrared target simulator

Infrared small targets are generally rare in images, occupy minimal space, and often blend into noisy backgrounds. As a result, negative samples dominate in the training

phase, and it becomes difficult for the model to learn meaningful target features. To address this, an infrared target simulator (ITS) is employed to generate pseudo labels that enrich the training set. Guided by typical infrared target characteristics [19] that bright centers with gradually dimmer peripheries and shapes are mostly round or irregular polygons, ITS places random points, circles, and polygons on the image to simulate real targets. Each pseudo target is randomly assigned a shape, size, and position, whose intensity is determined by the local background and must be brighter than the average of the surrounding points to ensure visibility. This integration of synthetic examples alleviates the low signal-to-noise ratio problem, allowing the model to learn robust representations for small infrared targets.

### 3.3 Lightweight Stable Encoder

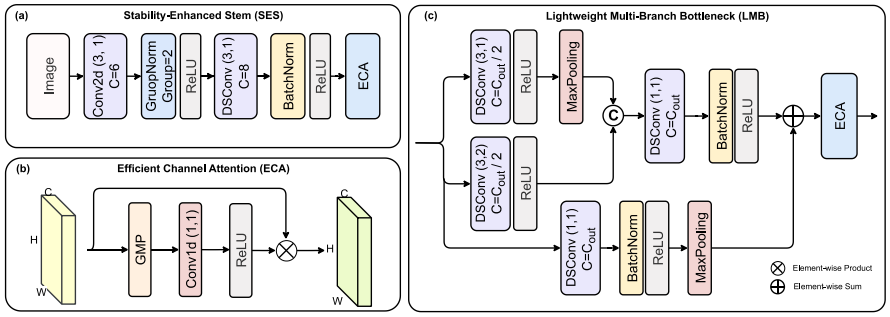


**Fig. 3.** Framework of main modules of lightweight stable feature encoder. (a) Stability-Enhanced Stem (SES). (b) Efficient Channel Attention (ECA). (c) Lightweight Multi-branch Bottleneck (LMB).

Infrared small targets typically exhibit low contrast, making them prone to omission during early processing stages. Once lost in the background, these small targets will be hard to detect. To prevent these subtle features from being lost or misclassified as noise, we propose a Stability-Enhanced Stem (SES) module to extract stable initial features, as illustrated in Fig. 3(a). Rather than employing a depth-wise separable convolution paired with batch normalization, a standard convolution combined with group normalization is adopted at the beginning of SES. This design choice yields richer and more reliable feature representations while maintaining stable gradients, especially when training with the small batch sizes preferred in infrared target detection (large batch size might lead to a very low signal-to-noise ratio in the training phase). A depth-wise separable convolution, supplemented by an Efficient Channel Attention module (ECA, depicted in Fig. 3(b)) [11], is then used to enhance cross-channel interactions while increasing the number of channels to 8.

To further broaden the receptive field and integrate global contextual information, a Lightweight Multi-Branch Bottleneck (LMB) module is introduced. As shown in Fig. 3(c), LMB uses parallel branches that apply depth-wise separable convolutions with different kernel sizes to capture features from various receptive fields and down-

sampling. Through a bottleneck structure, LMB reduces the number of channels before concatenating and fusing the branch outputs, thus controlling the parameter count. Here, we also use the ECA module to achieve feature interaction. Given that the number of feature channels tends to increase with the depth of the model to extract complex features, but infrared small target detection does not require extensive semantic complexity, there will be unnecessary redundancy in the deeper layers. Therefore, we choose a narrower model architectural design over wider alternatives, stacking four LMB modules with output channels set to 8, 16, 32, and 32 respectively.

### 3.4 Coarse-to-fine Hybrid Upsampling Decoder

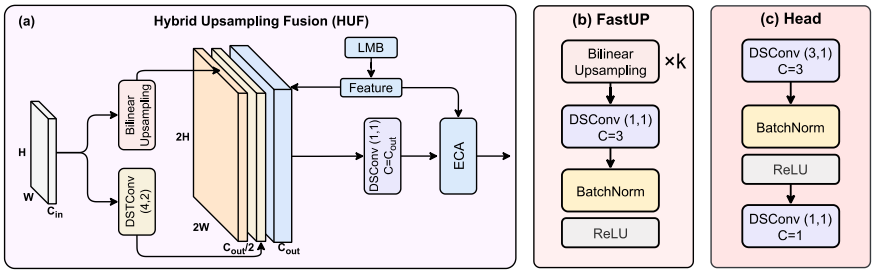


**Fig. 4.** Structure of Hybrid Upsampling Fusion module. (a) Hybrid Upsampling Fusion (HUF) module. (b) FastUP module. (c) Segment Head.

A critical challenge in infrared small target detection networks is maintaining the integrity of target information during feature reconstruction and upsampling stages. Therefore, to solve this problem, we propose a Coarse-to-fine Hybrid Upsampling Decoder as follows.

**Hybrid Upsampling Fusion.** As shown in Fig. 4(a), the hybrid upsampling fusion (HUF) module is designed to blend multi-scale information through a dual-stream upsampling approach. The first stream applies depth-separable transposed convolution (DSTC) to capture intricate details, while the second stream employs bilinear interpolation to retain feature integrity and reduce checkerboard artifacts. It fuses the output with features from the LMB module, which curtails the risk of losing small targets during the integration process. The result is a balanced representation that preserves subtle cues from lower layers and contextual information from deeper layers, improving overall detection performance.

**Coarse-to-fine Multi-scale Head.** Relying solely on the final layer's output for segmentation can overlook details for accurately identifying small infrared targets. Instead, inspired by Feature Pyramid Networks (FPN) [13], we integrate a multi-scale segmentation head that draws on features from various layers. For aligning multi-scale features without lots of computational overhead, we propose a FastUp module (see Fig. 4(b)) to rapidly upsample low-scale features from the HUF module, which uses a combination

of radical bilinear interpolation and depth-separable convolution. These features are then concatenated with the final output to help detection, ensuring that subtle details captured in earlier layers are preserved. To better differentiate small targets from a similar background, a coarse-to-fine strategy (C2F) is adopted, which is encouraged by [17]. Outputs from each HUF module are taken into independent segment heads (see Fig. 4(c)) to conduct soft supervision learning (only during the training phase) by label smoothing technology [16], tempering the confidence of shallower features. The smoothing factor decreases with network depth, and the final prediction is not set smooth. This gradual refinement process helps avoid premature decisions, resulting in a segmentation output that evolves from a broad initial assessment to a precise, fine-grained prediction.

### 3.5 Model Optimization

To enhance the detection of subtle infrared targets, we employ a composite loss function combined with a coarse-to-fine training scheme. At each scale, we introduce a soft binary cross entropy (SoftBCE) loss by leveraging label smoothing that decreases as the network deepens to achieve coarse-to-fine training. Given the pixel label $y_i$ (0 for background and 1 for target) and smoothing factor $\alpha$, the smoothed label $y_i'$ is defined as:

$$y_i' = \alpha^{1-y_i}(1-\alpha)^{y_i} \tag{1}$$

Based on it, the SoftBCE loss is formulated as:

$$L_{SoftBCE}(p, y, \alpha) = -\frac{1}{N}\sum_{i=1}^{N} y_i' \ln\big(\sigma(p_i)\big) + (1-y_i')\ln\big(1-\sigma(p_i)\big) \tag{2}$$

Where $p_i$ denotes the predicted probability that the point belongs to the target, $\sigma(\cdot)$ denotes the Sigmoid function. We then incorporate Dice loss [18] to maintain a balance between foreground and background coverage which as follows:

$$L_{Dice}(p, y) = 1 - \frac{2\sum_{i=1}^{N}\sigma(p_i)y_i + \varepsilon}{\sum_{i=1}^{N}\sigma(p_i) + \sum_{i=1}^{N} y_i + \varepsilon} \tag{3}$$

where $\varepsilon$ is a small constant to prevent division by zero. Finally, following the coarse-to-fine training strategy, we combine these loss terms across multiple scales, resulting in the overall loss, which is formulated as:

$$L_{C2F}(p, y, \alpha) = \sum_{s=1}^{S} \omega_s \left(\lambda_{sce} L_{SoftBCE}(p_i, y_i, \alpha_s) + \lambda_{dice} L_{Dice}(p_i, y_i)\right) \tag{4}$$

Where $\omega_s$ represents the weight assigned to each scale $s$, and there are S scales. $\lambda_{sce}$ and $\lambda_{dice}$ control the contributions of $L_{SoftBCE}$ and $L_{Dice}$, respectively. This hierarchical optimization progressively refines the model's predictions, ensuring both accurate localization and robust generalization.

# 4 Experiments

## 4.1 Datasets and Metrics

**Datasets.** Our experiments are conducted on two public datasets: IRSTD-1k [12] and NUDT-SIRST [6]. The IRSD-1k dataset comprises 1001 images with a resolution of 512×512 and has been divided into training and testing splits with a ratio of 8:2 in previous works. The NUDT-SIRST dataset contains 1327 images with a resolution of 256×256, which we randomly divide into training and testing splits using the same 8:2 ratio. For both datasets, we reserve 20 percent of the training set as our validation dataset and select the best-performing model on the validation set for evaluation on the test sets.

**Evaluation metrics.** To evaluate the performance of our method, we employ both pixel-level and target-level metrics. For pixel-level evaluation, we use Mean Intersection over Union (mIoU), Normalized Intersection over Union (nIoU), and False Alarm Rate ($F_a$) to assess segmentation accuracy and false detections. Additionally, we use the Probability of Detection ($P_d$) to quantify target-level performance.

For multi-class settings, mIoU is defined as follows:

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP_i}{FN_i+FP_i+TP_i} \tag{5}$$

where $TP_i$ , $FN_i$ and $FP_i$ denote the true positive, false negative, and false positive counts for the $i$-th class, respectively.

The normalized IoU (nIoU) is given by

$$nIoU = \frac{\sum_{i=1}^{N} min(Pred_i, Label_i)}{\sum_{i=1}^{N} max(Pred_i, Label_i)} \tag{6}$$

where $Pred_i$ and $Label_i$ represent the predicted and ground truth regions for pixel $i$.

The false alarm rate ($F_a$), which measures the proportion of falsely detected pixels relative to the total number of pixels, is defined as:

$$F_a = \frac{P_{false}}{P_{all}} \tag{7}$$

Where $P_{false}$ is the number of false positive pixels and $P_{all}$ is the number of all pixels in the image.

The probability of detection ($P_d$), which measures the ratio of correctly predicted targets to the total number of targets, is formulated as:

$$P_d = \frac{N_{pred}}{N_{all}} \tag{8}$$

where $N_{pred}$ is the number of correctly predicted targets and $N_{all}$ is the number of all targets.

### 4.2 Implementation Details

All experiments are conducted using PyTorch 2.3.0 and Python 3.12 on Ubuntu 20.04, with CUDA 12.1, on an RTX 4090 24GB GPU. Models were trained for 300 epochs with a fixed random seed. We employed the AdamW optimizer with an initial learning rate of 5e-3. To facilitate convergence, we implemented a warm-up strategy with 30 iterations, the starting factor and the ending factor are 0.05 and 1, respectively. We incorporated Cosine Annealing for learning rate decay with a period of 10 and a minimum learning rate of 5e-4. The batch size was set to 10 during training and 1 during testing.

We set the $L_{SoftBCE}$ and $L_{Dice}$ to 0.6 and 0.4, respectively. For the coarse to fine training strategy, the label smoothing factors ($\alpha$) are set to 0.1, 0.05, 0.01, 0.001 for upsampling 8x, 4x, 2x, and no upsampling respectively. The loss weight $\omega_s$ were assigned to $2000^{-\alpha}$. For the final output, label smoothing was not applied to maintain the sharpness and precision of the predictions.

### 4.3 Performance Comparison

To validate the effectiveness of Limi-Net, we conducted a comprehensive comparison with several leading methods. Our evaluation includes heavyweight models such as U-Net [5], DNA-Net [6], UIU-Net [7], and ALCL-Net [8], as well as lightweight networks like LW-IRST-Net [9] and LR-Net [10].

**Quantitative Comparison.** The quantitative results of different methods are presented in Tab. 1, which summarizes the quantitative results across key metrics.

**Table 1.** Comparison between the Limi-Net and state-of-the-art methods ( ↑ indicates higher is better, ↓ indicates lower is better). Results for the metrics of mIoU (%), nIoU (%), $F_a$ ($10^{-6}$), $P_d$ (%), Parameters(M) and FLOPs(G) are presented. The best values are highlighted with **bold,** and the second are marked with underline.

| | IRSTD-1k | | | | NUDT-SIRST | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | mIoU ↑ | nIoU ↑ | $F_a$ ↓ | $P_d$ ↑ | mIoU ↑ | nIoU ↑ | $F_a$ ↓ | $P_d$ ↑ | Params | FLOPs |
| U-Net | 0.59 | 0.57 | 38.75 | <u>0.88</u> | 0.75 | 0.74 | 12.39 | 0.93 | 31.037 | 54.62 |
| DNA-Net | 0.60 | 0.60 | 19.94 | 0.83 | **0.90** | **0.89** | **0.28** | 0.97 | 4.697 | 13.96 |
| UIU-Net | **0.63** | **0.62** | 28.21 | **0.89** | 0.88 | 0.87 | <u>0.57</u> | 0.97 | 50.541 | 54.35 |
| ALCL-Net | 0.61 | <u>0.61</u> | <u>19.31</u> | <u>0.88</u> | <u>0.89</u> | **0.89** | 3.41 | **0.98** | 5.668 | 6.70 |
| LW-IRST-Net | 0.60 | 0.60 | 20.56 | 0.82 | 0.80 | 0.81 | 15.14 | 0.95 | 0.166 | 0.279 |
| LR-Net | 0.56 | 0.53 | 36.41 | 0.84 | 0.78 | 0.79 | 6.20 | **0.98** | <u>0.020</u> | **0.055** |
| **Limi-Net** | **0.63** | <u>0.61</u> | **17.57** | 0.84 | <u>0.89</u> | **0.89** | 4.65 | **0.98** | **0.016** | <u>0.136</u> |

From the quantitative results, Limi-Net demonstrates superior performance compared to lightweight models while achieving competitive results with significantly larger models, and maintains the lowest parameters and the second lowest computation. Specifically, on the IRSTD-1k dataset, Limi-Net outperforms all other models in pixel-

level evaluation metrics with the highest mIoU and lowest $F_a$, except for the largest model, UIU-Net. On the NUDT-SIRST dataset, the performance of the lightweight models LW-IRST-Net and LR-Net deteriorates sharply, but Limi-Net still maintains superior performance, especially in the mIoU and nIoU metrics, demonstrating its robustness and efficiency. Limi-Net exhibits state-of-the-art performance with minimal computational complexity, achieving high accuracy and real-time response (as shown in Fig. 1). We observed that our model exhibits little slower inference on the RTX 4090 compared to ALCL-Net, which attributed to insufficient utilization of the modern GPU's parallel computing capacity due to our model's compact size. However, the significant reduction in model size and computational requirements makes our approach particularly advantageous for edge device deployment.

**Visual Comparison.** To get an intuitively visual comparison, we select some representative infrared small target images and the prediction mask from above models, as shown in Fig. 5. Limi-Net demonstrates more accurate target detection with fewer missed targets, particularly visible in rows XD219 and XD932, lower false detection such as in XD762 and XD319, and cleaner output with sharper boundaries and less background noise compared to methods like ALCL-Net, LR-Net, and LW-IRST-Net. Thus, Limi-Net can not only achieve high accuracy detection but also precisely segment out the boundaries of small targets.

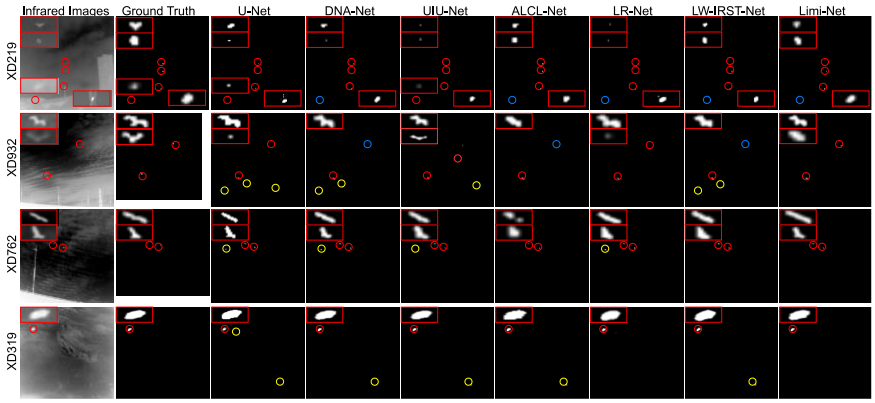


**Fig. 5.** Visual Comparison covering U-Net, DNA-Net, UIU-Net, ALCL-Net, LW-IRST-Net, LR-Net on the IRSTD-1K. Correctly detected targets are magnified and highlighted in red. Targets in blue and yellow denote missed detection and false detection, respectively.

### 4.4 Ablation Studies

We provide experiments on the IRSTD-1k dataset to show the effectiveness of our method. The Performance verification of each component of the Limi-Net is shown in Table 2. Here, "Ours-w/o ITS" trains the model without the Infrared Target Simulator (ITS). "Ours-w/o SES" removes the Stability-Enhanced Stem (SES) module and uses depth-separable convolution with batch normalization instead. "Ours-w/o LMB"

replaces the multi-branch bottleneck (LMB) structure with residual blocks. "Ours-w/o HUF" removes the depth-separable transposed convolution and upsampling using bi-linear interpolation upsampling with depth-separable convolution only. "Ours-w/o C2F" eliminates the coarse-to-fine strategy and uses a plain FPN structure.

**Table 2.** Performance verification of each component of the Limi-Net on the IRSTD-1k dataset.

| Schemes | Components | | | | | Performance | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ITS | SES | LMB | HUF | C2F | mIoU ↑ | nIoU ↑ | $F_a$ ↓ | $P_d$ ↑ |
| Ours-w/o ITS | ✗ | ✓ | ✓ | ✓ | ✓ | 0.60 | 0.59 | 31.26 | 0.82 |
| Ours-w/o SES | ✓ | ✗ | ✓ | ✓ | ✓ | 0.60 | 0.60 | 21.37 | 0.79 |
| Ours-w/o LMB | ✓ | ✓ | ✗ | ✓ | ✓ | 0.62 | 0.60 | 18.14 | 0.81 |
| Ours-w/o HUF | ✓ | ✓ | ✓ | ✗ | ✓ | 0.57 | 0.52 | 32.51 | 0.77 |
| Ours-w/o C2F | ✓ | ✓ | ✓ | ✓ | ✗ | 0.61 | 0.59 | 24.78 | 0.82 |
| Limi-Net | ✓ | ✓ | ✓ | ✓ | ✓ | **0.63** | **0.61** | **17.57** | **0.84** |

From break-down ablation experiments, we demonstrate that each component of Limi-Net plays a role in infrared small target detection, and all of them assembled achieve the best results. Specifically, from Table 2, the Infrared Target Simulator (ITS) does assist the training of Limi-Net by generating diverse pseudo labels, which increases IoU metrics (mIoU from 0.60 to 0.63 and nIoU from 0.59 to 0.61) while significantly reducing the $F_a$ (31.26 to 17.57). By stabilizing feature extraction, Stability-Enhanced Stem (SES) makes the model obtain higher precision ($P_d$ increases from 0.79 to 0.84) and IoU. To demonstrate the importance of the Multi-Branch Bottleneck (LMB), we replace the LMB modules with normal residual blocks, which leads to a dip in mIoU (0.63 to 0.62) and a notable rise in $F_a$ (to 18.14), indicating its effectiveness in multi-scale feature fusion. Hybrid Upsampling Fusion (HUF) modules fuse multi-scale features and curtail the risk of losing small targets during the integration process, as mIoU decreases to 0.57, nIoU drops to 0.52, and Fa dramatically increases to 32.51. Finally, compared to plain FPN structure (mIoU 0.61, Pd 0.82), Coarse-to-fine strategy (C2F) better differentiates small targets, confirming that C2F enhances overall detection. Together, these findings demonstrate that each component is indispensable for achieving the best performance in Limi-Net.

## 5 Conclusion

In this work, we introduced Limi-Net, a novel lightweight network specifically tailored for infrared small target detection, which consists of the Lightweight Stable Encoder, the Coarse-to-fine Hybrid Upsampling Decoder, and the Infrared Target Simulator for data augmentation in the training phase. Our experimental results demonstrate that Limi-Net achieves state-of-the-art performance on infrared small target detection tasks, outperforming existing models in terms of both detection accuracy and computational efficiency. The lightweight design ensures that Limi-Net is suitable for real-time deployment in practical resource-constrained environments.

## References


1. Bai, X., Zhou, F.: Analysis of New Top-Hat Transformation and the Application for Infrared Dim Small Target Detection. Pattern Recognition 43(6), 2145–2156 (2010)
2. Chen, C., Li, H., Wei, Y., Xia, T., Tang, Y.: A local contrast method for small infrared target detection. IEEE Transactions on Geoscience and Remote Sensing 52(1), 574-581 (2014)
3. Gao, C., Meng, D., Yang, Y., Wang, Y., Zhou, X., Hauptmann A.: Infrared patch-image model for small target detection in a single image. IEEE Transactions on Image Processing 22(12), 4996-5009 (2013)
4. Wang, H., Zhou, L., Wang, L.: Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 8509–8518 (2019)
5. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234-241 (2015)
6. Li, B., Xiao, C., Wang, L., Wang, Y., Lin, Z., Li, M., An, W., Guo, Y.: Dense nested attention network for infrared small target detection. IEEE Transactions on Image Processing 32, 1745-1758 (2023)
7. Wu, X., Hong, D., Chanussot, J.: UIU-Net: U-Net in U-Net for infrared small object detection. IEEE Transactions on Image Processing, 32, 364-376 (2023)
8. Yu, C., Liu, Y., Wu, S., Xia, X., Hu, Z., Lan, D., Liu, X.: Pay Attention to Local Contrast Learning Networks for Infrared Small Target Detection. IEEE Geoscience and Remote Sensing Letters, 19, 1-5 (2022)
9. Kou, X., et al.: LW-IRSTNet: Lightweight infrared small target segmentation network and application deployment. IEEE Transactions on Geoscience and Remote Sensing 61, Art. no. 5621313 (2023)
10. Yu, C., et al.: LR-Net: A lightweight and robust network for infrared small target detection. arXiv:2408.02780 (2024)
11. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In: CVPR. pp. 11531-11539 (2020)
12. Zhang, M., Zhang, R., Yang, Y., Bai, H., Zhang, J., Guo, J.: ISNet: Shape matters for infrared small target detection. In: CVPR. pp. 877-886 (2022)
13. Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2117–2125 (2017)
14. Teutsch, M., Kruger, W.: Classification of small boats in infrared images for maritime surveillance. In: Proc. Int. WaterSide Security Conf., pp. 1–7. IEEE (2010)
15. Rawat, S. S., Verma, S. K., Kumar, Y.: Review on recent developments in infrared small target detection algorithms. Proc. Comput. Sci. 167, 2496–2505 (2020)
16. Müller, R., Kornblith, S., Hinton, G. E.: When does label smoothing help? Advances in Neural Information Processing Systems 32 (2019)
17. Du, R., et al.: Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. In: European Conference on Computer Vision, pp. 390–406. Springer, Cham (2020)
18. Sudre, C. H., et al.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 240–248. Springer, Heidelberg (2017)
19. Yuan, S., et al.: Beyond full label: Single-point prompt for infrared small target label generation. arXiv:2408.08191 (2024)