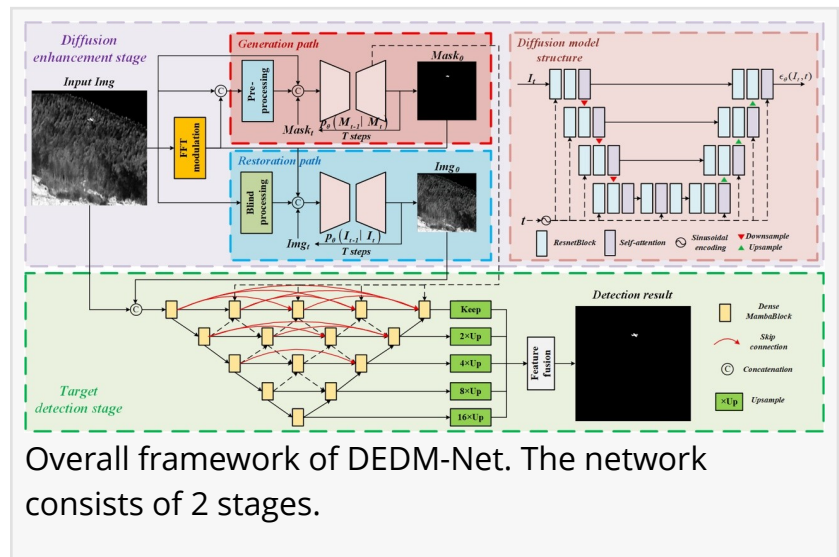


Diffusion-mamba hybrid sets new standard for infrared target detection

GA, UNITED STATES, September 24, 2026 /EINPresswire.com/ -- [Infrared small target detection](https://www.einpresswire.com/news/498888/infrared-small-target-detection-is-critical-for-remote-sensing-fire-prevention-and-surveillance-yet-existing-methods-struggle-with-tiny-low-contrast-targets-that-lack-distinct-shape-or-texture-researchers-have-developed-a-two-stage-deep-learning-network-that-combines-diffusion-based-feature-enhancement-with-state-space-modeling-ssm-to-suppress-background-clutter-while-amplifying-target-signals-the-approach-achieves-state-of-the-art-detection-rates-across-three-public-datasets-significantly-reducing-both-missed-detections-and-false-alarms-in-complex-imaging-environments) is critical for remote sensing, fire prevention, and surveillance, yet existing methods struggle with tiny, low-contrast targets that lack distinct shape or texture. Researchers have developed a two-stage deep learning network that combines diffusion-based feature enhancement with state-space modeling (SSM) to suppress background clutter while amplifying target signals. The approach achieves state-of-the-art detection rates across three public datasets, significantly reducing both missed detections and false alarms in complex imaging environments.



Overall framework of DEDM-Net. The network consists of 2 stages.

Infrared small target detection enables binary segmentation of weak targets within complex backgrounds, serving applications from forest fire early warning to remote sensing threat assessment. However, practical challenges persist: infrared targets often occupy fewer than 81 pixels (typically under 9×9), exhibit extremely low energy with signal-to-noise ratios around 3, and lack prominent shape or texture information. These characteristics cause targets to be easily submerged in background clutter, making feature extraction particularly difficult. Deep learning methods have improved performance, but most focus exclusively on target features while neglecting background information—the vast majority of the image—leading to severe class imbalance between positive and negative samples. Based on these challenges, there is an urgent need for an approach that simultaneously models both targets and backgrounds to achieve robust detection in complex scenes.

On June 30, 2026, researchers from the Research Center for Space Optical Engineering at Harbin Institute of Technology published their findings in the Journal of Remote Sensing. Their proposed Diffusion-Enhanced Dense Mamba Network (DEDM-Net) addresses a critical challenge in remote sensing: detecting infrared small targets that are easily overwhelmed by background clutter. This technology directly impacts forest fire prevention, surveillance early warning systems, and

military threat assessment—applications where missed detections or false alarms can have severe consequences.

The team developed a two-stage network that achieves a synergistic effect greater than the sum of its parts. The first stage employs a dual-path diffusion model with a novel blind processing module that predicts each pixel using only surrounding information—never the pixel itself—preventing extremely small targets from being misclassified as background. The second stage introduces a dense nested Mamba architecture based on the state space model (SSM), which captures long-range correlations across global and local features with linear computational complexity—a significant advantage over conventional Transformers. A cross-stage prediction fusion module further integrates features from both stages, improving contour segmentation accuracy. Together, these innovations deliver superior performance across all evaluation metrics compared to 11 state-of-the-art methods.

The DEDM-Net was evaluated on three public datasets: NUAA-SIRST (427 images), NUDT-SIRST (1,327 images at 256×256), and IRSTD-1k (1,000 images at 512×512). On the NUDT-SIRST dataset, the method achieved 93.40% IoU, 93.28% nIoU, 98.37% detection probability (P_d), and a remarkably low false-alarm rate of just 3.75×10^{-6} —outperforming DNA-Net (92.99% IoU, 93.22% nIoU) and ISTDU-Net (91.69% IoU, 91.84% nIoU). On the IRSTD-1k dataset, DEDM-Net achieved 73.71% IoU and 93.89% P_d with only 11.10×10^{-6} false alarms, surpassing all competitors. The blind processing module uses a dual-window structure with outer radius $R=4$ and inner radius $r=2$, ensuring that target regions are "blindly processed" while surrounding context is captured. Ablation studies confirmed that each component—the generation path, restoration path, dense nested structure, and residual Mamba blocks—contributes meaningfully to overall performance. The network was trained on an NVIDIA RTX 4080 GPU using the Adam optimizer.

"Infrared small targets are extremely challenging because they lack shape and texture—they're essentially just a few bright pixels in a sea of background," said corresponding author Dr. Shikai Jiang of Harbin Institute of Technology. "By modeling both the target-free background and potential target regions simultaneously, our diffusion-enhanced approach effectively amplifies what matters while suppressing what doesn't. The Mamba architecture then provides the global context needed to distinguish true targets from bright clutter."

The method employs a two-stage training framework. In the diffusion enhancement stage, a U-Net backbone estimates noise across 1,000 diffusion steps, with a dual-path scheme modeling both target masks and background images. The blind processing module generates pixel-wise convolution kernels that exclude the center pixel, effectively removing small targets from background reconstruction. In the detection stage, a dense nested network with feature pyramid connections and residual Mamba blocks extracts multiscale features. The loss function combines binary cross-entropy (BCE) and Dice losses to address class imbalance.

While DEDM-Net achieves superior accuracy, the diffusion-based two-stage design increases inference time compared to single-stage networks. Future work will focus on model distillation,

mixed-precision inference, and faster samplers to reduce the required diffusion steps. The approach holds promise for real-time surveillance systems, autonomous drone navigation in low-visibility conditions, and early wildfire detection networks. As the team noted, the framework could also inspire new thinking about how generative models and state-space architectures can be combined for other challenging computer vision tasks where target-background separation is critical.

References

DOI

[10.34133/remotesensing.1046](https://doi.org/10.34133/remotesensing.1046)

Original Source URL

<https://spj.science.org/doi/10.34133/remotesensing.1046>

Funding information

This work was supported by the National Natural Science Foundation of China under Grant 62305086, the China Postdoctoral Science Foundation under Grant 2023M740901, the Natural Science Foundation of Heilongjiang Province of China under Grant LH2024F032, and in part by the National Key Laboratory of Air-Based Information Perception and Fusion under Grant 20220001077001.

Lucy Wang

BioDesign Research

[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/944840723>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.