INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NAS RA

Sargis Hovhannisyan

OBJECT DETECTION IN ADVERSE WEATHER USING NOVEL DEEP LEARNING AND THERMAL-VISIBLE IMAGING

ABSTRACT

of the dissertation for obtaining a Ph.D. degree in Technical Sciences on specialty 05.13.05 “Mathematical modelling, numerical methods and program complexes”

Yerevan 2025

ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Սարգիս Անդրեասի Հովհաննիսյան

Օբյեկտների հայտնաբերումը անբարենպաստ եղանակային պայմաններում՝ օգտագործելով տեսանելի ու ջերմային պատկերներ և նոր խորը ուսուցման մեթոդներ

Ե.13.05 - «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի համար

ՍԵՂՄԱԳԻՐ

Երևան 2025

Relevance of the Research

Object detection is identifying and localizing objects within images or video frames, typically marked by bounding boxes indicating the location and class of detected objects (Figure 1 illustrates an example of object detection output). Accurate object detection is critical across many practical domains, impacting safety, efficiency, and security. In Autonomous Vehicles, object detection supports functionalities including collision avoidance, pedestrian detection, lane identification, and traffic sign recognition [1]. In Unmanned Aerial Vehicles (UAV) systems, object detection enables route planning, obstacle avoidance, and target monitoring [2]. Accurate detection and identification of wildlife species in automated monitoring systems facilitate improved ecological monitoring, conservation efforts, and anti-poaching initiatives [3]. In smart-city surveillance, object detection improves urban safety through better monitoring traffic incidents, crime detection, and public safety management [4]. In industrial inspection systems, object detection supports critical tasks such as defect detection, quality assurance, and safety inspections in industrial environments [5].

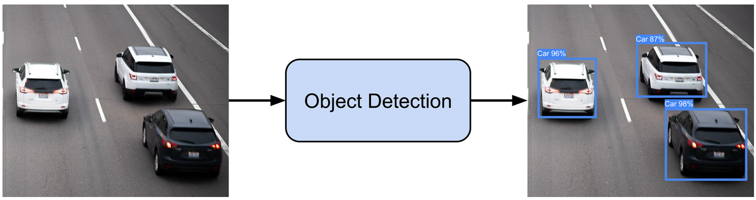

Figure 1: Example of an object detection output.

These systems utilized two types of imaging technologies: Red Green Blue (RGB) visible and Thermal Infrared (TIR) cameras. RGB cameras capture visible light across the red, green, and blue spectrum, providing high-resolution color imagery with excellent detail in well-lit conditions. TIR cameras detect heat signatures by capturing emitted infrared radiation from objects, enabling object detection regardless of lighting conditions and allowing temperature-based differentiation.

While current detection systems achieve human-level precision on benchmark datasets and are widely deployed in commercial products due to deep-learning models that learn visual patterns directly from data, they are typically optimized for clear-weather images with good lighting conditions. Despite these strengths, these systems are significantly degraded under adverse weather conditions. Numerous challenges persist that limit the reliability and effectiveness of these systems in real-world applications. Studies have shown that state-of-the-art (SOTA) detectors' accuracy can drop by around **30-40%** as haze density increases [6].

RGB sensors are inherently dependent on ambient lighting and highly susceptible to visual degradation caused by fog, haze, shadows, and nighttime darkness. Although TIR sensors function effectively in darkness, through light fog, and under headlight glare, they

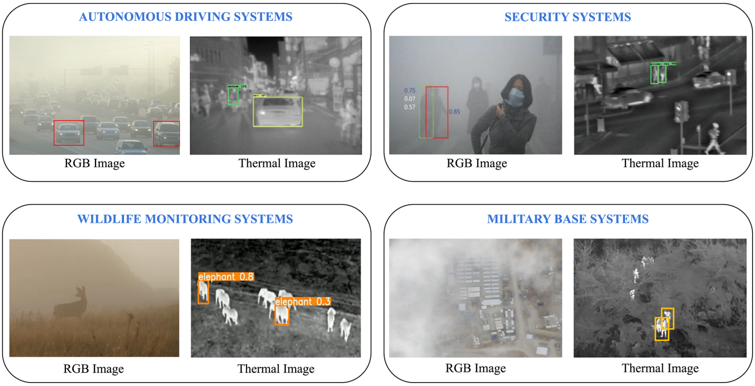

Figure 2: Examples of object detection failures under bad weather conditions across different imaging modalities.

suffer from inherent limitations, including low contrast, weak edge definition, and sensor noise that complicate accurate object detection.

Even thermal imaging, while robust in low-visibility scenarios, can be compromised by severe atmospheric interference such as heavy fog or haze, which reduces contrast and edge sharpness, further impairing detection accuracy [7]. Thermal video introduces additional complications, including motion blur and camera jitter, making object detection even more challenging in dynamic environments where temporal consistency becomes critical for reliable performance (see Figure 2 for examples of such failures).

These detection failures translate directly into real-world harms with significant consequences. When object detection systems fail due to environmental degradations such as fog, haze, poor illumination, or occlusions, critical image features become obscured, resulting in missed detections or misclassifications that can have severe implications.

- Road safety reports by the United States (U.S.) Federal Highway Administration and the National Highway Traffic Safety Administration indicate that low-visibility conditions, including fog, haze, and nighttime driving, contribute to a disproportionate number of fatal crashes. Although these conditions account for a small percentage of total driving exposure, they collectively account for nearly **50%** of all traffic fatalities in the U.S. [8]. The increased risk results from reduced driver awareness, delayed reaction times, and reduced effectiveness of vehicle safety systems. Studies further illustrate that the effectiveness of autonomous braking systems can be reduced by 30% to 80% when visibility is reduced due to severe fog conditions, significantly narrowing reaction times and increasing collision risks [9].

- Surveillance systems at fixed locations also experience severe impairments under low visibility conditions. For instance, camera-based motion detection systems frequently generate false alarms triggered by fog, dust, and insects. According to U.S. policing studies, false alarm rates for burglar alarm dispatches range from **94% to 98%**, placing unnecessary burdens on law enforcement resources [10]. Furthermore, criminals exploit visibility impairments, with analyses indicating that approximately **50%** of residential burglaries occur during nighttime or under low visibility when RGB camera systems demonstrate the lowest reliability [11].

- Wildlife conservation efforts are similarly impacted, as around **80%** of unauthorized wildlife hunting incidents occur during nighttime or under dense atmospheric haze [12]. Although drone surveillance provides crucial monitoring capabilities, aerial operations frequently encounter significant disruptions from fog and dust, limiting operational hours substantially. For example, a U.S. border-security audit showed that visibility restrictions kept unmanned aircraft airborne for about **22%** of their scheduled hours [13].

- Aviation safety and military operations routinely face ”degraded visual environments”, including conditions induced by fog, dust, and smoke. According to U.S. Army safety analyses, disorientation or obstacle collisions account for approximately **24%** of helicopter crashes and **44%** of fatalities. Additionally, commercial airports face operational disruptions due to fog, with major airports operating under instrument flight rules between **15% to 23%** of annual operational hours, incurring significant financial costs due to delays and reinforcing the critical need for effective visual enhancement solutions [14].

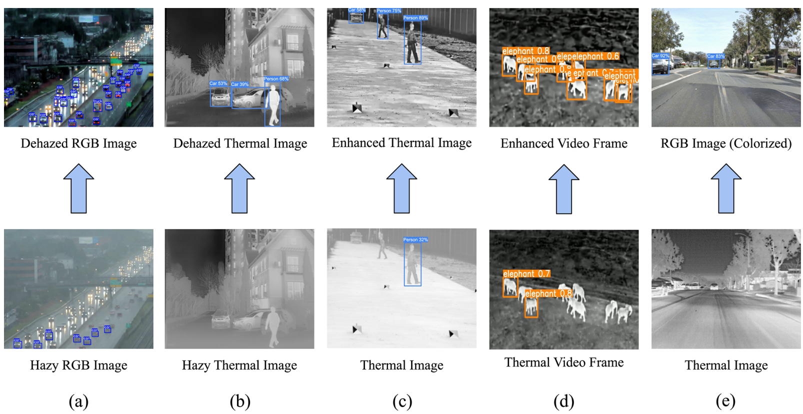

Figure 3: Impact of Enhancement Methods on Object Detection for RGB and TIR Imagery.

Several promising methodological approaches have emerged to address the challenges of reliable object detection under adverse conditions. Image dehazing represents a foundational approach to improving visibility in degraded visual conditions by removing atmospheric interference such as haze or fog from images (see Figure 3 (a)). This technique addresses the significant impairment of computer vision algorithms caused by reduced visibility, contrast, and detail clarity in adverse weather. Recent deep learning-based approaches, such as DMPHN [15] and AOD-Net [16], demonstrate considerable advancement by automatically learning important features from data. However, these methods face limitations in generalization capability across varying real-world conditions and often struggle with complex backgrounds or heavy atmospheric degradation.

Similarly, TIR dehazing addresses the unique degradation patterns affecting thermal imagery in adverse weather conditions. While thermal cameras offer inherent advantages in low-light conditions, they remain susceptible to quality reduction from heavy atmospheric interference. Thermal images experience distinct degradation characteristics, including diminished contrast, edge blurring, and decreased clarity resulting from atmospheric scattering and absorption (see Figure 3 (b)). Contemporary approaches leverage Convolutional Neural Networks (CNN) to enhance thermal image quality [17]. However, significant challenges persist regarding limited dataset availability, the fidelity of synthetic data generation, and model adaptability across diverse thermal imaging systems with varying specifications.

General TIE techniques aim to improve the overall quality and interpretability of thermal imagery by addressing common challenges, including low contrast, detail obscuration, ghosting effects from overlapping thermal radiation, and inconsistent sensor characteristics (see Figure 3 (c)). Recent approaches, including GAN-based approaches and CNN architectures [18], [19], have advanced image quality considerably but continue to face generalization difficulties in complex scenarios, particularly those involving reflective materials and ambiguous thermal patterns.

Thermal video enhancement extends beyond static image processing by addressing the temporal characteristics inherent to thermal video data. This introduces additional complexities, such as motion blur, temporal inconsistencies, rapid scene dynamics, and variability in sensor responses over time (see Figure 3 (d)). Methods that leverage temporal context encounter significant challenges in effectively handling non-rigid motion and complex thermal variations across frames. These temporal artifacts and inconsistencies severely undermine object detection performance, leading to unreliable detections, increased false positives, and missed targets, particularly in dynamic environments where consistent and accurate detection is essential.

Thermal image colorization represents a transformative approach to bridging the gap between TIR and RGB domains by addressing the inherent lack of color information and typically low contrast with unclear object boundaries in thermal imagery. TIR to RGB colorization is the process of transforming a single-channel TIR image into a three-channel color image that corresponds to visible-spectrum images. This process aims to generate realistic textures, colors, and visual details that would be present if the scene were captured by an RGB camera under favorable lighting conditions, as illustrated in Figure 3 (e). Colorizing thermal images enhances both human interpretability and compatibility with RGB-trained models. It provides a more intuitive visual representation and allows existing RGB-based algorithms to be applied to thermal data without extensive retraining. Recent research has explored supervised and unsupervised deep learning frameworks, including CycleGAN [20] and U-GAT-IT [21], to translate thermal imagery into colored RGB representations. These techniques, however, encounter significant challenges including semantic distortions, inconsistent preservation of critical details, temporal instability in video sequences, and suboptimal performance on small objects, making accurate and consistent translation from thermal to visible spectrum while maintaining semantic integrity a complex research challenge, particularly for applications requiring high precision such as autonomous driving systems.

Given these substantial and persistent challenges, there remains a critical need for innovative approaches to enhance object detection systems' reliability under adverse weather and low-visibility conditions. This thesis addresses these challenges by proposing novel deep-learning methodologies tailored to improve image and video quality in visually degraded environments. In particular, we concentrate on advanced image dehazing methods that effectively mitigate atmospheric interference in RGB and TIR modalities. Furthermore, we investigate specialized enhancement techniques for thermal images and videos, employing recent advancements in neural architectures to handle unique degradations such as low contrast, edge ambiguity, and temporal inconsistencies. Finally, this work explores TIR-to-RGB colorization methods, bridging the gap between these imaging modalities to leverage RGB-based algorithms without extensive retraining, thus significantly improving object detection accuracy and reliability across practical, real-world scenarios.

Challenges of Object Detection

Reliable object detection remains challenging due to the inherent limitations and distinct vulnerabilities associated with different imaging modalities when operating in adverse environmental conditions. While detection technologies are increasingly accurate in controlled or optimal conditions, their performance rapidly deteriorates when facing real-world scenarios involving degraded visual environments.

Despite **RGB imaging** widespread use, RGB-based detection methods inherently depend on ambient illumination and visibility conditions. Adverse scenarios such as fog, haze, heavy shadows, nighttime darkness, or noise introduced by bad weather significantly degrade RGB image quality, resulting in reduced contrast and loss of fine details critical for accurate detection. Consequently, detection reliability is severely affected, leading to frequent object mislocalization and misclassification. Furthermore, as most deep-learning detectors are trained predominantly on clear-weather images, their performance in challenging conditions is often compromised, highlighting the need for alternative imaging modalities that are more resilient to environmental impairments.

TIR imaging offers advantages for detection tasks by capturing emitted radiation rather than reflected light, providing resilience in varied lighting conditions. However, adoption is limited by several key challenges: *domain shift from RGB* data requiring specialized model training, as visible-spectrum trained models struggle with the fundamentally different visual features and contrast patterns in thermal data. *Low contrast and blurred edges* significantly reduce the detail visibility needed for accurate object detection, while sensor and spectral variability across different thermal imaging systems complicates model generalization across deployments. *Atmospheric degradation* from fog or haze affects image quality despite TIR's relative robustness, reducing contrast and blurring edges critical for detection accuracy. Additionally, *reflection artifacts* from surfaces like metal and glass create misleading contours that confuse detection algorithms. In video applications, *motion blur, camera instability, and temporal noise* accumulation further compromise detection consistency. These combined challenges significantly affect detection reliability in critical applications, requiring advanced enhancement techniques to ensure consistent performance across challenging operational environments.

The aim of the Work

The aim of this thesis is to address the significant challenges posed by adverse weather conditions on object detection systems, developing accurate deep learning methodologies tailored to overcome these issues. The proposed approaches aim to surpass existing (SOTA) methods, achieving superior performance across multiple benchmark datasets and real-world scenarios. To accomplish these objectives, the thesis focuses on the following technical tasks:

1. Develop **dehazing frameworks** explicitly designed for **RGB** and **TIR** images.
2. Develop **thermal image and video enhancement** networks.
3. Develop a TIR-to-RGB **colorization pipeline** capable of translating thermal images into visually intuitive RGB representations.
4. Conduct meticulous evaluations of the proposed frameworks, assessing their performance using key metrics such as **detection accuracy** and **generalization capability**.

The practical significance of the work

The methodologies proposed in this thesis form a unified enhancement framework that improves visual clarity and cross-modal consistency in RGB and TIR imagery. As a result, the research has broad applicability across a range of safety-critical and high-value domains:

- **Autonomous Driving and Advanced Driver-Assistance Systems** enhance visibility in fog, haze, and nighttime conditions, enabling safer navigation and more reliable detection of pedestrians, vehicles, and road signs.
- **UAV-Based Wildlife Monitoring and Anti-Poaching Systems** facilitate reliable detection of animals and humans in dense forests or nighttime conditions, strengthening conservation efforts and UAV-based patrolling.

- **Search and Rescue Operations in Disaster Environments** improve visibility and consistency across video frames, enhancing detection of survivors and obstacles during UAV-assisted missions.

- **Medical Imaging and Computer-Assisted Diagnostics** enhance contrast and suppress distortions in thermal medical images, facilitating early disease detection in applications like breast thermography and skin diagnostics.

The methods of investigation

In this thesis, we have used a wide range of approaches from different fields, including signal processing, machine learning, deep learning, and related fields. The Python programming language and its associated packages were used to train deep neural networks, process data, and design algorithms. Previous related results also served as a basis for this work.

Publications

All results are new and have been published in international and local journals, and presented at international conferences. The main findings of this thesis have been published in six scientific articles in various journals. The list of these articles is provided at the end of the Synopsis.

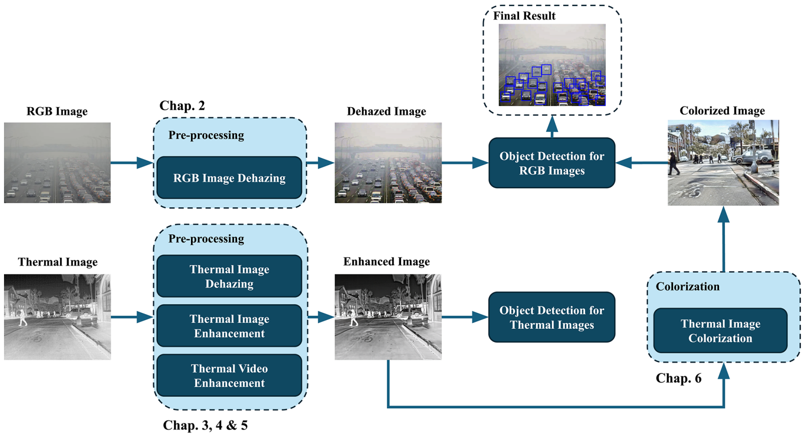

Figure 4: Overall Workflow of the Thesis Framework.

Structure of the Thesis

The dissertation consists of 7 chapters and a list of used literature. The thesis is written in 150 pages and has 247 literature references. The thesis contains 45 figures and 18 tables.

- **Chapter 1** introduces the research context, clearly outlining existing limitations in object detection under adverse weather conditions. It articulates the research questions and objectives and highlights the key novel contributions made throughout the thesis.

- **Chapter 2** aims to develop an innovative deep-learning solution to enhance object detection performance in challenging weather conditions, particularly addressing the significant degradation caused by haze and fog. Despite achieving considerable success, current SOTA approaches often struggle with non-homogeneous haze, fail to preserve natural color properties, and perform poorly on small training datasets. Most critically, these methods are not optimized for downstream computer vision tasks such as object detection, which is essential for real-world applications.

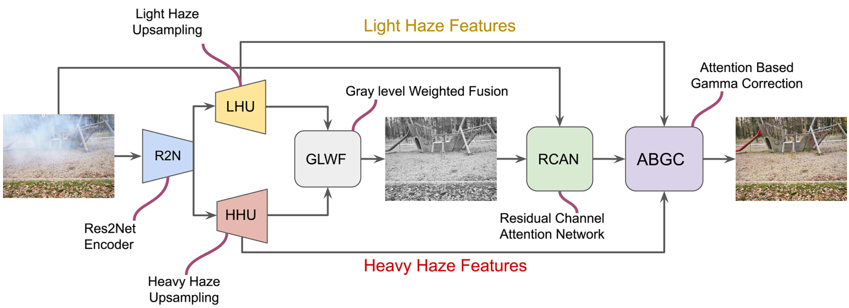

Figure 5: Overall architecture of EOD-Net.

To overcome these challenges, this chapter proposes *EOD-Net*, a novel end-to-end RGB image dehazing architecture designed to improve object detection in hazy environments. Overall pipeline of the EOD-Net is presented in Figure 5. Key innovations include a dual-branch dehazing system with a Gray-Level Weighted Fusion module and a specialized enhancement Attention-Based Gamma Correction module for color restoration, effectively addressing the limitations of previous approaches. Comprehensive evaluations on synthetic (I-Haze [22], O-Haze [23], NH-Haze2 [24]) and real-world hazy datasets demonstrate superior performance over existing SOTA methods across multiple image quality metrics (LPIPS [25], PSNR, SSIM [26], FSIM [27], VIF [28]). Furthermore, practical testing on traffic surveillance footage significantly improves object detection performance as shown in Figure 6. Unprocessed hazy images allow detection of only **3-4%** of vehicles from all present vehicles, while EOD-Net enables detection of approximately **40%** of all vehicles in heavily hazed environments. This substantial improvement demonstrates EOD-Net's effectiveness for real-world applications where visual clarity directly impacts safety and operational decisions.

- **Chapter 3** addresses thermal image dehazing under severe atmospheric degradation such as haze, smoke, and fog, which obscure details, lower contrast, and degrade

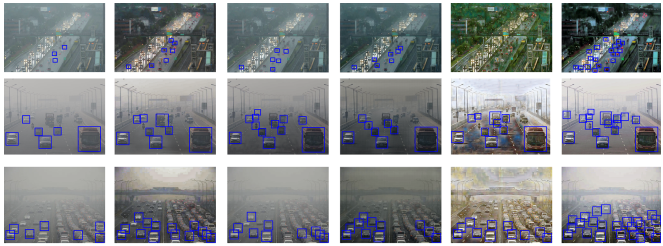

Figure 6: Vehicle detection results compared with other SOTA dehazing methods.

downstream performance. Current SOTA methods work reasonably well on thermal images captured in visible-light environments but still struggle under these adverse conditions. To overcome these limitations, we propose *MTIE-Net*, a Mamba-based thermal image dehazing framework built on the Enhancement and Denoising State Space Model. By integrating CNNs with state-space modeling, the network performs joint denoising and enhancement, restoring visibility while preserving critical edges necessary for reliable object detection. Extensive experiments on the M3DF dataset [29] show that MTIE-Net outperforms both traditional and deep-learning baselines across PSNR and SSIM [26], EME, BDIM, and MDIMTE [30]. Table 1 presents the quantitative results on the M3FD dataset for object detection evaluation. Nearly all enhancement methods significantly improved detection performance compared to using only the original infrared images. Our proposed MTIE-Net outperformed other methods in terms of detection mean Average Precision (AP), achieving up to a 25% improvement in detection accuracy compared to when using artificially generated challenging hazy infrared images, and around 8% improvement over the best competing enhancement method under challenging conditions. Furthermore, it generalizes well to real-world domains, making it practical for surveillance and other safety-critical applications where object-detection performance is essential.

Table 1: Object detection evaluation ($mAP_{0.5}$) under adverse weather conditions.

Measure	**Day**	**Overcast**	**Night**	**Challenge**	$mAP_{0.5}$
Hazy Infrared	0.718	0.721	0.620	0.618	0.710
Original Infrared	0.806	0.798	0.712	0.739	0.786
Visible	0.827	0.789	0.764	0.759	0.758
AGCCPF	0.811	0.799	0.739	0.748	0.789
BBCNN	0.815	0.805	0.743	0.747	0.790
IE-CGAN	0.816	0.808	0.768	0.754	0.791
WTHE	0.818	0.810	0.785	0.783	0.792
MTIE-Net	**0.828**	**0.819**	**0.849**	**0.871**	**0.812**

- **Chapter 4** aims to develop an innovative physics-guided framework for thermal image enhancement that overcomes inherent challenges such as low contrast, ghosting artifacts, blurred edges, and sensor noise, all impairing downstream vision tasks. Existing SOTA techniques can boost contrast but often amplify noise and leave ghost artifacts, limiting their generalization across domains. To address these challenges, this chapter introduces *PB-IID-Net*. Figure 7 illustrates the architecture of *PB-IID-Net*.

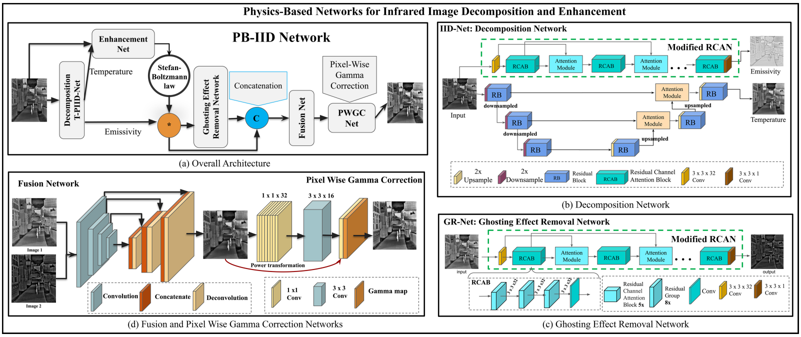

Figure 7: Overall architecture of PB-IID-Net.

The network leverages physics principles specific to thermal imagery by applying the extended Stefan-Boltzmann law in its decomposition module to separate temperature and emissivity components. It then removes ghosting through an artifact-suppression block and applies adaptive fusion with pixel-wise gamma correction, restoring visibility while preserving fine structural details. Extensive tests on five datasets (LTIR [31], CVC-14 [32], Autonomous Vehicles [33], Solar Panel [34], and Breast [35]) show that PB-IID-Net outperforms traditional and learning-based baselines across the non-reference metrics EME, BDIM, MDIMTE, BRISQUE, and NIQE [30]. PB-IID-Net achieves superior enhancement across different types of infrared images (near, mid, and far infrared), consistently outperforming other methods with notable improvements in object detection metrics across all spectra.

Table 2: Detection performance (mAP) on various inputs and models

Measure	Orig. NIR	Orig. MIR	Orig. FIR	WTHE on NIR	WTHE on MIR	WTHE on FIR	PB-IID-Net on NIR ($k=2$)	PB-IID-Net on MIR ($k=2$)	PB-IID-Net on FIR ($k=2$)	PB-IID-Net on NIR ($k=4$)	PB-IID-Net on MIR ($k=4$)	PB-IID-Net on FIR ($k=4$)
$\text{mAP}_{50} \uparrow$	51.9	56.6	48.2	53.1	58.4	49.9	**60.5**	**63.8**	55.9	59.7	63.2	**56.4**
$\text{mAP}_{75} \uparrow$	14.8	20.1	16.3	15.3	22.1	18.5	**19.9**	**27.5**	20.7	18.4	27.1	**21.3**
$\text{mAP}_{50:95} \uparrow$	22.6	25.1	22.5	23.5	26.8	24.1	**28.4**	**31.5**	27.7	26.9	30.9	**29.1**

With spectrum-specific optimal empirical parameters (K=2 for NIR and MIR, K=4 for FIR), the model demonstrated significant performance in the MIR spectrum, representing up to **5.4%** higher $mAP_{0.5}$ and **4.7%** higher $mAP_{0.5:0.95}$ compared to other enhancement techniques, as shown in Table 2.

- **Chapter 5** focuses on designing a Mamba-based framework for thermal video enhancement that addresses core issues such as low contrast, motion blur, sensor noise, and frame-to-frame inconsistencies that hinder object tracking and detection. Existing SOTA methods often enhance contrast at the cost of introducing temporal flicker, misaligned motion, and residual noise, which limits their reliability in real-world scenarios. This chapter proposes **TVEMamba**, which tackles the aforementioned challenges in thermal videos. First, sharpening and denoising the network improve each frame through denoising and sharpening. Next, the blur-resistant motion estimation network generates blur-resistant optical-flow maps from consecutive frames to capture complex motion patterns. Finally, the motion deblurring network combines the aligned frames, creating a temporally coherent, high-clarity video without motion-induced artifacts while preserving fine spatial detail. Extensive evaluations on five datasets (BIRDSAI [36], FLIR [37], CAMEL [38], Autonomous Vehicles [33], and Solar Panels [34]) demonstrate that TVEMamba outperforms SOTA methods across several non-reference quality metrics: EME, BDIM, DMTE, MDIMTE, LGTA [30], and BIE [39]. Additionally, as shown in Table 3, it boosts object detection accuracy by more than **6%** relative to the original degraded wildlife monitoring footage. These improvements highlight the practical advantages of TVEMamba for applications involving severely deteriorated video quality. This makes the method particularly valuable for wildlife monitoring, autonomous driving systems, and UAV-based military operations, where precise object detection is essential.

Table 3: Object detection performance on the BIRDSAI dataset. $YOLO_1$ and Hyper-$YOLO_1$ models are trained on original datasets, and $YOLO_2$ and Hyper-$YOLO_2$ models are trained on enhanced datasets produced by the TVEMamba framework.

Classes	2		3		2	
Architecture	$YOLO_1$	$YOLO_2$	$YOLO_1$	$YOLO_2$	Hyper-$YOLO_1$	Hyper-$YOLO_2$
$mAP_{0.5}$	38.1	**44.2**	25.0	**29.7**	38.0	**43.9**
$mAP_{0.5:0.9}$	13.2	**16.8**	9.3	**10.9**	12.9	**16.4**

- **Chapter 6** aims to bridge the gap between TIR and RGB imaging modalities through a novel colorization framework that transforms thermal imagery into visually realistic RGB representations. This enhances autonomous driving capabilities in challenging low-visibility conditions and enables applicability for RGB-based object detection. Existing thermal-to-visible translation methods face significant limitations, including semantic distortions, temporal inconsistency across video frames, poor performance on small objects, and insufficient detail preservation in complex textures. These challenges become particularly pronounced when translating objects with minimal thermal differences, such as buildings and trees. These challenges substantially reduce their effectiveness for critical downstream applications such as object detection in autonomous driving systems. This chapter proposes FWGAN, a frame-based Weber-law-driven Generative Adversarial Network, as shown in Figure 8.

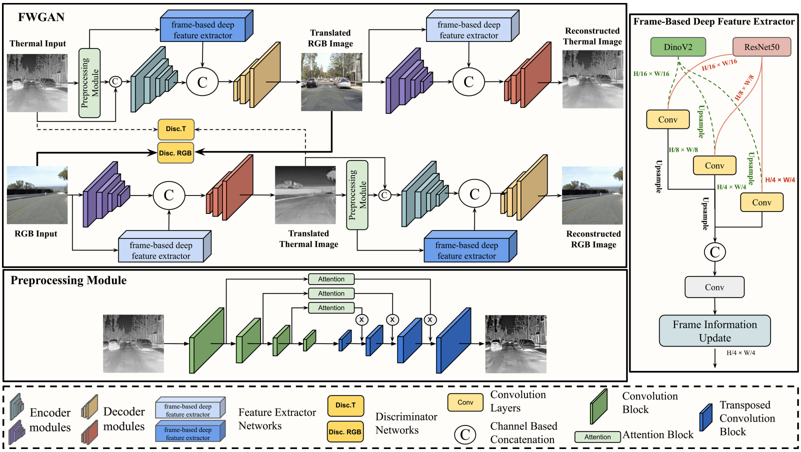

Figure 8: Overall Architecture of FWGAN for Thermal Image Translation and Analysis.

This innovative thermal-to-visible translation network addresses current limitations through specialized components: a Preprocessing Module that enhances sharpness and reduces noise, a Frame-based Deep Feature Extractor Module that integrates complementary features from pre-trained backbones, and a Frame Information Update mechanism that ensures temporal consistency across video sequences. Integrating extended Weber Contrast Enhancement-driven edge detection improves fine detail preservation when minimal temperature differences exist. Evaluations on FLIR [37] and KAIST [40] datasets demonstrate FWGAN's superior performance across multiple image quality metrics, including NIQE, BRISQUE, and PIQE [30] for image naturalness assessment. Most critically, compared to existing SOTA colorization techniques, FWGAN enhances object detection accuracy from **61% to 63%**, confirming its potential for integration into autonomous driving systems operating in adverse visibility conditions.

- **Chapter 7** concludes the thesis by summarizing the main contributions, discussing current limitations, and proposing potential future research on robust multi-modal perception in challenging environmental conditions.

Figure 4 illustrates the logical progression of this thesis, demonstrating how each chapter systematically contributes to a comprehensive framework for robust object detection under challenging visibility conditions.

List of author's publications

The obtained results were reported in several international and local scientific workshops:

1. S. Hovhannisyan et al., "AED-Net: A single image dehazing", IEEE Access, 10, 12465-12474, 2022.

2. S. Hovhannisyan et al., "EOD-Net: enhancing object detection in challenging weather conditions using an innovative end-to-end dehazing network", In 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6), IEEE, October 2023.

3. S. Hovhannisyan, "Mamba-based Thermal Image Dehazing", Mathematical Problems of Computer Science, 62, 126-144, 2024.

4. S. Hovhannisyan et al., "Thermal Video Enhancement Mamba: A Novel Approach to Thermal Video Enhancement for Real-World Applications", Information, 16(2), 125, 2025.

References

[1] R. Wang, Z. Wang, Z. Xu, *et al.*, "A real-time object detector for autonomous vehicles based on yolov4," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 9 218 137, 2021.

[2] Z. Cao, L. Kooistra, W. Wang, L. Guo, and J. Valente, "Real-time object detection based on uav remote sensing: A systematic literature review," *Drones*, vol. 7, no. 10, p. 620, 2023.

[3] L. Chen, G. Li, S. Zhang, W. Mao, and M. Zhang, "Yolo-sag: An improved wildlife object detection algorithm based on yolov8n," *Ecological Informatics*, vol. 83, p. 102 791, 2024.

[4] Y. Li, Y. Huang, and Q. Tao, "Improving real-time object detection in internet-of-things smart city traffic with yolov8-dsaf method," *Scientific reports*, vol. 14, no. 1, p. 17 235, 2024.

[5] H. M. Ahmad and A. Rahimi, "Deep learning methods for object detection in smart manufacturing: A survey," *Journal of Manufacturing Systems*, vol. 64, pp. 181–196, 2022.

[6] Z. Liu, Y. He, C. Wang, and R. Song, "Analysis of the influence of foggy weather environment on the detection effect of machine vision obstacles," *Sensors*, vol. 20, no. 2, p. 349, 2020.

[7] F. Erlenbusch, C. Merkt, B. de Oliveira, *et al.*, "Thermal infrared single image dehazing and blind image quality assessment," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 459–469.

[8] Association for the Advancement of Automotive Medicine. "Nighttime driving." Accessed: 2025-05-07, Association for the Advancement of Automotive Medicine (AAAM). (Jan. 2021), [Online]. Available: `https://www.aaam.org/nighttime-driving/`.

[9] A. Wulfeck. "Vehicle safety systems struggle in bad weather, aaa finds." Accessed: 2025-05-10, FOX Weather. (2021), [Online]. Available: `https://www.foxweather.com/weather-news/vehicle-safety-systems-struggle`.

[10] Fairfax County Police Department. "False alarm security systems." Accessed: 2025-05-09, Fairfax County. (2025), [Online]. Available: `https://www.fairfaxcounty.gov/police/ServicesAH/FalseAlarmSecuritySystems`.

[11] J. Knott. "What time do most burglaries occur?" Accessed: 2025-05-10, Pro CE. (Nov. 2, 2023), [Online]. Available: `https://www.cepro.com/news/what-time-do-most-burglaries-occur/118261/`.

[12] J. Worland. "Drones are helping catch poachers operating under cover of darkness." Accessed: 2025-05-07, TIME. (May 17, 2018), [Online]. Available: `https://time.com/5279322/drones-poaching-air-shepherd/`.

[13] Office of Inspector General, U.S. Department of Homeland Security, "Cbp drones are dubious achievers," U.S. Department of Homeland Security, Washington, DC, Tech. Rep., 2015.

[14] U.S. Army Combat Readiness Center, "Dve: The varying degree," *Flightfax*, May 2022.

[15] S. D. Das and S. Dutta, "Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 482–483.

[16] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.

[17] R. Soundrapandiyan, S. C. Satapathy, C. M. PVSSR, and N. G. Nhu, "A comprehensive survey on image enhancement techniques with special emphasis on infrared images," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9045–9077, 2022.

[18] M. A. Marnissi and A. Fathallah, "Gan-based vision transformer for high-quality thermal image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 817–825.

[19] K. Lee, J. Lee, J. Lee, S. Hwang, and S. Lee, "Brightness-based convolutional neural network for thermal image enhancement," *Ieee Access*, vol. 5, pp. 26 867–26 879, 2017.

[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[21] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *arXiv preprint arXiv:1907.10830*, 2019.

[22] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-haze: A dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*, Springer, 2018, pp. 620–631.

[23] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 754–762.

[24] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 444–445.

[25] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[27] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[28] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[29] J. Liu, X. Fan, Z. Huang, *et al.*, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5802–5811.

[30] S. Agaian, H. Ayunts, T. Trongtirakul, and S. Hovhannisyan, “A new method for judging thermal image quality with applications,” *Signal Processing*, vol. 229, p. 109 769, 2025.

[31] A. Berg, J. Ahlberg, and M. Felsberg, “A thermal object tracking benchmark,” in *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*, 2015.

[32] A. González, Z. Fang, Y. Socarras, *et al.*, “Pedestrian detection at day/night time with visible and fir cameras: A comparison,” *Sensors*, vol. 16, no. 6, p. 820, 2016.

[33] K. Takumi, K. Watanabe, Q. Ha, A. Tejero-De-Pablos, Y. Ushiku, and T. Harada, “Multispectral object detection for autonomous vehicles,” in *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, 2017, pp. 35–43.

[34] E. Alfaro-Mejía, H. Loaiza-Correa, E. Franco-Mejía, A. D. Restrepo-Girón, and S. E. Nope-Rodríguez, “Dataset for recognition of snail trails and hot spot failures in monocrystalline si solar panels,” *Data in brief*, vol. 26, p. 104 441, 2019.

[35] L. Silva, D. Saade, G. Sequeiros, *et al.*, “A new database for breast research with infrared image,” *Journal of Medical Imaging and Health Informatics*, vol. 4, no. 1, pp. 92–100, 2014.

[36] E. Bondi, R. Jain, P. Aggrawal, *et al.*, “Birdsai: A dataset for detection and tracking in aerial thermal infrared videos,” in *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, 2020, pp. 1747–1756.

[37] deepnewbie, *Flir thermal images dataset*, Accessed: 2024-09-21, 2024. [Online]. Available: `https://www.kaggle.com/datasets/deepnewbie/flir-thermal-images-dataset`.

[38] E. Gebhardt and M. Wolf, “Camel dataset for visual and thermal infrared multiple object detection and tracking,” in *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*, IEEE, 2018, pp. 1–6.

[39] H. Ayunts, A. Grigoryan, and S. Agaian, “Novel entropy for enhanced thermal imaging and uncertainty quantification,” *Entropy*, vol. 26, no. 5, p. 374, 2024.

[40] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, “Multispectral pedestrian detection: Benchmark dataset and baseline,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1037–1045.

Ամփոփում

Սարգիս Անդրեասի Հովհաննիսյան

Օբյեկտների հայտնաբերումը անբարենպաստ եղանակային պայմաններում՝ օգտագործելով տեսանելի ու ջերմային պատկերներ և նոր խորը ուսուցման մեթոդներ

Աշխատանքը նվիրված է օբյեկտների հայտնաբերման համակարգերի բարելավմանը և աշխատանքի կայունության բարձրացմանը՝ հատկապես վատ տեսանելիության պայմաններում (մառախուղ, ծուխ, մշուշ, ցածր լուսավորություն): Այս նպատակով մշակվել են խորը ուսուցման նոր մեթոդներ, որոնք օգտագործում են ինչպես տեսանելի (RGB), այնպես էլ ջերմային (TIR) պատկերներ: Կատարված գնահատումները հիմնավորում են մշակված լուծումների արդյունավետությունը և ցույց են տալիս դրանց ներուժը իրական կիրառական խնդիրներում:

Աշխատանքի ***հիմնական նպատակներն*** են՝

1. Մշակել խորը ուսուցման նոր ցանցեր՝ ինչպես տեսանելի (RGB), այնպես էլ ջերմային (TIR) պատկերներից մառախուղը կամ մշուշը հեռացնելու համար:
2. Ստեղծել խորը ուսուցման ցանցեր՝ ջերմային պատկերների և տեսանյութերի որակը բարելավելու համար (բարձրացնել հակադրությունը, նվազեցնել աղմուկը), որպեսզի օբյեկտների հայտնաբերումը դառնա ավելի ճշգրիտ և կայուն:
3. Մշակել նոր խորը ուսուցման ցանց՝ ջերմային պատկերները տեսանելի (RGB) գունավոր պատկերների վերածելու (գունավորման) համար: Սա թույլ կտա օգտագործել արդեն ստեղծված օբյեկտների հայտնաբերման մեթոդները, խուսափելով ջերմային պատկերներով վերաուսուցումից:
4. Մանրամասն գնահատել առաջարկվող լուծումների արդյունավետությունը (հատկապես հայտնաբերման ճշգրտությունը)՝ օգտագործելով տարատեսակ տվյալների հավաքածուներ (ինչպես հենանիշային, այնպես էլ իրական պայմաններում ստացված):

Առաջին գլխում ներկայացվում է հետազոտության ոլորտը, բացատրվում է, թե ինչու է դժվար օբյեկտներ հայտնաբերել անբարենպաստ եղանակային պայմաններում, սահմանվում են աշխատանքի խնդիրները ու նպատակները և ընդգծվում ատենախոսության հիմնական նոր մոտեցումները:

Երկրորդ գլուխը ներկայացնում է խիտ մշուշոտ պայմաններում օբյեկտների հայտնաբերման արդյունավետությունը բարձրացնելու լուծումը՝ նոր խորը ուսուցման EOD-Net ցանցի միջոցով: Արդյունքները ցույց են տվել հայտնաբերման արդյունավետության զգալի բարելավում՝ ***3-4%***-ից հասնելով մոտ ***40%***-ի:

Երրորդ գլխում առաջարկվում է MTIE-Net անունով նոր ցանց (հիմնված «Մամբա» մոդելների վրա)՝ ջերմային պատկերներից մշուշը հեռացնելու համար: Այս մեթոդը բարելավում է հայտնաբերման ճշգրտությունը ***25%***-ով՝ համեմատած չմշակված ջերմային պատկերների հետ:

Չորրորդ գլխում ներկայացվում է ջերմային պատկերների որակը բարելավելու մեթոդ (մասամբ հիմնված ֆիզիկական օրենքի վրա)՝ PB-IID-Net ցանցի միջոցով: Այս լուծումը օգնում է բարձրացնել օբյեկտների հայտնաբերման արդյունավետությունը և կիրառելի է ինֆրակարմիր սպեկտրի ցանկացած տիրույթի համար:

Հինգերորդ գլխում առաջարկվում է TVEMamba ցանց (հիմնված Մամբա մոդելների վրա)՝ ջերմային տեսանյութերի որակը բարելավելու համար: Այն բարձրացնում է օբյեկտների հայտնաբերման ճշգրտությունը մոտ ***6%***-ով:

Վեցերորդ գլխում մշակվել է FWGAN անունով նոր ցանց, որի միջոցով կարելի է ջերմային պատկերներից ստանալ տեսանելի (RGB) գունավոր պատկերներ: Այս մեթոդը բարձրացնում է օբյեկտների հայտնաբերման ճշգրտությունը ***61%***-ից մինչև ***63%***:

Յոթերորդ գլխում ամփոփվում են ատենախոսության հիմնական արդյունքները և նշվում են հնարավոր հետագա հետազոտությունների ուղղությունները՝ կապված բարդ պայմաններում աշխատող տեսողական համակարգերի զարգացման հետ:

Աշխատանքի կիրառական և գիտական նշանակությունը

Ատենախոսական աշխատանքը ունի ինչպես զգալի կիրառական, այնպես էլ գիտական նշանակություն:

- Աշխատանքի արդյունքները թույլ են տալիս զգալիորեն բարձրացնել օբյեկտների հայտնաբերման համակարգերի հուսալիությունը և անվտանգությունը իրական կիրառություններում (ինչպիսիք են՝ ինքնավար տրանսպորտը, հսկողության համակարգերը, որոնողափրկարարական աշխատանքները)՝ հատկապես վատ տեսանելիության և բարդ եղանակային պայմաններում:
- Մշակած մեթոդների արդյունավետությունը գնահատվել է տարատեսակ տվյալների վրա կատարված փորձերով: Այս փորձերը ցույց են տվել, որ դրանք գերազանցում են առկա առաջատար լուծումներին՝ ապահովելով օբյեկտների հայտնաբերման ավելի բարձր ճշգրտություն և կայունություն:
- Ստացված արդյունքները և մշակված մոտեցումները կարող են հիմք հանդիսանալ գիտական և կիրառական մի շարք հետագա ուղղությունների զարգացման համար, մասնավորապես՝ օբյեկտների հայտնաբերումը էլ ավելի բարելավելու՝ տեսանելի (RGB) և ջերմային (TIR) պատկերները միաժամանակ օգտագործելու ուղղությամբ:

Заключение

Оганнисян Саргис Андреасович

Обнаружение объектов в неблагоприятных погодных условиях с использованием видимых и тепловых изображений и новых методов глубокого обучения

Работа посвящена улучшению систем обнаружения объектов и повышению стабильности их работы, особенно в условиях плохой видимости (туман, дым, дымка, низкая освещенность). С этой целью разработаны новые методы глубокого обучения, которые используют как видимые (RGB), так и тепловые (TIR) изображения. Проведенные оценки подтверждают эффективность разработанных решений и показывают их потенциал в реальных прикладных задачах.

Основные цели работы:

1. Разработать новые сети глубокого обучения для удаления тумана или дымки как с видимых (RGB), так и с тепловых (TIR) изображений.
2. Разработать сети глубокого обучения для улучшения качества тепловых изображений и видео (повышение контрастности, снижение шума), с целью сделать обнаружение объектов более точным и стабильным.
3. Разработать новую сеть глубокого обучения для преобразования (колоризации) тепловых изображений в видимые (RGB) цветные изображения. Это позволит использовать уже существующие методы обнаружения объектов, избегая необходимости их переобучения на тепловых изображениях.
4. Провести детальную оценку эффективности предложенных решений (в частности, точности обнаружения) с использованием различных наборов данных (как эталонных, так и полученных в реальных условиях).

В первой главе представляется область исследования, объясняется, почему сложно обнаруживать объекты в плохую погоду, ставятся задачи и цели работы и подчеркиваются основные новые подходы диссертации.

Вторая глава посвящена решению для повышения эффективности обнаружения объектов в условиях плотной дымки посредством новой сети глубокого обучения EOD-Net. Результаты показали значительное улучшение эффективности обнаружения с **3-4%** до примерно **40%**.

В третьей главе предлагается новая сеть MTIE-Net (основанная на моделях "Мамба") для удаления дымки с тепловых изображений. Этот метод улучшает точность обнаружения на **25%** по сравнению с необработанными тепловыми изображениями.

В четвертой главе представляется метод улучшения качества тепловых изображений (частично основанный на физическом законе) посредством сети PB-IID-Net. Это решение способствует повышению эффективности обнаружения объектов и применимо для любого диапазона инфракрасного спектра.

В пятой главе предлагается сеть TVEMamba (основанная на моделях "Мамба") для улучшения качества тепловых видео. Она повышает точность обнаружения объектов примерно на **6%**.

В шестой главе разработана новая сеть FWGAN, посредством которой можно получать с тепловых изображений видимые (RGB) цветные изображения. Этот метод повышает точность обнаружения объектов с **61%** до **63%**.
В седьмой главе подводятся итоги диссертации, отмечаются возможные направления дальнейших исследований, связанные с развитием визуальных систем, работающих в сложных условиях.

Прикладное и научное значение работы
Диссертационная работа имеет как значительное прикладное, так и научное значение.

- Результаты работы позволяют значительно повысить надежность и безопасность систем обнаружения объектов в реальных приложениях (таких как автономный транспорт, системы наблюдения, поисково-спасательные работы), особенно в условиях плохой видимости и сложных погодных условиях.
- Эффективность разработанных методов оценена в экспериментах, проведенных на различных наборах данных. Эти эксперименты показали, что они превосходят существующие передовые решения, обеспечивая более высокую точность и стабильность обнаружения объектов.
- Полученные результаты и разработанные подходы могут послужить основой для развития ряда научных и прикладных дальнейших направлений, в частности, для дальнейшего улучшения обнаружения объектов путем одновременного использования видимых (RGB) и тепловых (TIR) изображений.