Analecta Technica Szegedinensia
Vol. 19, No. 3-4 ISSN 2064-7964 2025

EDCST : ENHANCED DENSITY-AWARE CROSS-SCALE TRANSFORMER FOR ROBUST OBJECT CLASSIFICATION UNDER ATMOSPHERIC FOG CONDITIONS

Fiston Oshasha[1,2*], Saint Jean Djungu[1,2,3], Alidor Mbayandjambe[3], Franklin Mwamba [2,6], Jirince Biaba[5], Frey Sylvestre[3], Tege Simboni [4], Nathanaël Kasoro[3], Blaise Muhala[3]

[1]Commissariat General for Atomic Energy, Regional Center for Nuclear Studies of Kinshasa, P.O. Box 868, University of Kinshasa Campus, D.R. Congo
[2]CRIA-Center for Research in Applied Computing, Kinshasa, DR. Congo
[3]Department of Mathematics, Statistics and Computer Science, University of Kinshasa, Kinshasa, DR Congo
[4]Department of Computer Management, Higher Pedagogical Institute of Isiro, Isiro, D.R. Congo,
[5]Faculty of Computer Science, Hanoi University of Science and Technology, Vietnam
[6]Health Sciences Research Institute, Kinshasa, Democratic Republic of the Congo
e-mail: fiston.oshasha.oshasha@cgea-rdc.org

Received: 21st December	*Accepted: 27th December*

ABSTRACT

Atmospheric fog poses critical challenges for computer vision systems in autonomous driving, surveillance, and robotics, where reliable object classification is essential. Under severe fog, classification accuracy can degrade by over 50%, and most existing approaches rely on separate defogging steps that limit real-time applicability. This study introduces the Enhanced Density-Aware Cross-Scale Transformer (EDCST), a novel architecture for direct object classification under foggy conditions without requiring prior defogging. To support comprehensive evaluation, we developed a physics-based simulation framework generating four fog types (uniform, gradient, patchy, adaptive) across nine intensity levels (0-80% scattering). EDCST leverages 384-dimensional embeddings, eight transformer layers, and twelve attention heads, trained using curriculum learning with OneCycleLR scheduling. On CODaN-Fog (15,500 images at 224×224 resolution), EDCST achieves 84.4% accuracy on clean images and retains 74.2% accuracy under severe fog (80% intensity), outperforming baseline transformers by 15.8 percentage points. Class-wise sensitivity analysis reveals that larger objects such as vehicles and animals maintain over 75% classification performance, while smaller objects are more affected. Patchy fog causes the greatest accuracy drop (19.1%), followed by adaptive (8.9%) and uniform fog (6.8%). The model converges in 100 epochs within 513 minutes on Tesla V100 GPU. This work introduces a real-time-capable classification framework that eliminates defogging requirements and maintains strong performance under diverse fog conditions, making it highly suitable for safety-critical vision applications.

Keywords: Object classification, atmospheric fog, robust computer vision, atmospheric scattering, curriculum learning, fog simulation

1. INTRODUCTION

Computer vision systems form the backbone of modern autonomous technologies, from self-driving vehicles to surveillance networks and robotic systems [1, 2, 3]. However, these systems face significant challenges when operating under adverse weather conditions, particularly atmospheric fog, which can dramatically degrade visual perception capabilities [4, 5]. Fog scattering effects reduce image contrast, blur object boundaries, and alter color characteristics, leading to substantial performance degradation in object classification tasks [6].

The critical nature of this challenge is exemplified in autonomous driving scenarios, where fog-induced misclassification can result in catastrophic failures. Recent studies indicate that state-of-the-art object detection models experience accuracy drops of over 50% under severe fog conditions [7, 8]. Similar degradation patterns have been observed in surveillance systems [9], drone navigation [3], and robotic vision applications [10], highlighting the urgent need for fog-robust computer vision solutions.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

Traditional approaches to fog-degraded image analysis typically employ a two-stage pipeline: image defogging followed by object classification [11]. Defogging methods range from classical techniques based on atmospheric scattering models [12] to modern deep learning approaches [13, 14]. However, these preprocessing-based solutions suffer from several limitations: computational overhead affecting real-time performance, error propagation from defogging to classification stages, and limited generalizability across diverse fog conditions [15, 16].

Recent advances in deep learning have demonstrated the potential for end-to-end fog-robust classification without explicit defogging preprocessing [17, 18]. Convolutional Neural Networks (CNNs) with specialized architectures have shown promising results in handling weather-degraded images [19, 20]. However, these approaches are primarily designed for clear weather conditions and exhibit limited robustness when faced with diverse fog characteristics [21].

The emergence of Vision Transformers (ViTs) has revolutionized computer vision by leveraging self-attention mechanisms to capture global contextual information [20]. Unlike CNNs, which process images through local convolutions, transformers can model long- range dependencies and have demonstrated superior performance in various computer vision tasks [22, 23]. However, their application to fog-robust object classification remains largely unexplored, presenting a significant research opportunity.

Fog simulation represents another critical aspect of developing robust computer vision sys- tems. Physics-based fog models, grounded in atmospheric scattering theory, provide realistic training data for fog-robust algorithms [24]. However, existing fog simulation approaches often oversimplify fog characteristics, typically modeling only uniform fog distributions [25, 26].

Real-world atmospheric fog exhibits complex spatial variations that significantly impact visual perception. These include uniform fog with consistent density distribution, gradient fog with progressive intensity changes across image regions, patchy fog with localized density variations, and adaptive fog that responds to scene content characteristics [27]. Furthermore, fog intensity varies considerably in practice and can be categorized into five representative conditions: Clean (0% scattering, no fog), Light Fog (20% scattering, uniform fog), Moderate Fog (40% scattering, patchy fog), Severe Fog (60% scattering, gradient fog), and Extreme Fog (80% scattering, adaptive fog). These categories reflect increasing levels of visual degradation, from minimal visibility loss to complex and dynamic haze patterns that severely compromise object recognition capabilities [6]. Current simulation frameworks fail to capture this comprehensive range of fog characteristics, limiting the robustness of trained models when deployed in diverse real-world conditions.

Curriculum learning has emerged as a powerful training strategy for handling challenging data distributions [28]. By gradually increasing task difficulty during training, curriculum learning enables models to learn robust representations while maintaining convergence stability [29, 30]. Despite its proven effectiveness in various domains, curriculum learning applications to fog-robust computer vision remain limited [31].

This work addresses these limitations by proposing an Enhanced Density-Aware Cross-Scale Transformer (EDCST) architecture specifically designed for robust object classification under diverse atmospheric fog conditions. Our approach eliminates the need for separate defogging preprocessing while maintaining high classification accuracy across varying fog intensities and types. The key contributions of this research are:

1. We introduce the EDCST architecture that integrates density-aware attention mechanisms with cross-scale feature processing, enabling robust object classification under fog conditions without preprocessing requirements.
2. We develop a physics-based fog simulation methodology that models four distinct fog types (uniform, gradient, patchy, and adaptive) across nine intensity levels, providing realistic training and evaluation scenarios.
3. We propose a curriculum learning approach combined with clean validation methodology that ensures accurate performance measurement while maintaining fog robustness through progressive difficulty scaling

The remainder of this paper is organized as follows: Section 2 This section establishes the theoretical foundations underlying our approach. Section 3 reviews related work in fog-robust computer vision and

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

vision transformers. Section 4 presents the proposed EDCST architecture and training methodology. Section 5 describes the experimental setup and evaluation protocols. Section 6 discusses results and provides comprehensive analysis with limitations and future research directions. Section 7 concludes this work.

2. PRELIMINARIES

This section establishes the theoretical foundations underlying our approach. We provide essential background on vision transformers, atmospheric scattering theory for fog modeling, object classification challenges under degraded conditions, and curriculum learning principles that inform our methodology.

2.1. Vision Transformers for Image Classification

Vision Transformers (ViTs) represent a paradigm shift from traditional convolutional architectures by applying the transformer mechanism, originally designed for natural language processing [32], to computer vision tasks [33]. The fundamental principle involves treating an image as a sequence of patches, analogous to tokens in text processing, which has revolutionized computer vision since 2020 [34].
Given an image $\boldsymbol{x} \in R^{H\times W\times C}$, it is split into non-overlapping patches of size $P \times P$, producing $N = \frac{HW}{P^2}$ patches. Each patch $\boldsymbol{x}_p \in R^{P^2\cdot C}$ is flattened and projected to form embeddings, as formulated in Eq. 1:

$$\boldsymbol{z_0} = [\boldsymbol{x_1 E};\ \ldots;\ \boldsymbol{x}_{\mathrm{N}}\boldsymbol{E}] + \boldsymbol{E}_{\mathrm{pos}} \tag{1}$$

where $E \in R^{(P^2\cdot C)\times D}$ is the projection matrix, and $E_{pos} \in R^{N\times D}$ adds positional encoding [35]. This transforms spatial data into tokens for transformer processing. Each transformer layer applies multi-head self-attention (MSA) and MLP blocks with residual connections as shown in Eq. 2 and Eq. 3:

$$\boldsymbol{z}_l^i = MSA\big(LN(\boldsymbol{z}_{l-1})\big) + \boldsymbol{z}_{l-1} \tag{2}$$

$$\boldsymbol{z}_l = MLP(LN(\boldsymbol{z}_l')) + \boldsymbol{z}' \tag{3}$$

MSA uses scaled dot-product attention, as formulated in Eq. 4:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where **Q**, **K**, and **V** are query, key, and value matrices, and d_k is the key dimension [36, 37].
This allows global context modeling, useful under fog-degraded conditions. In our study, despite the availability of various Transformer architectures, we chose to work with Swin-Tiny due to its strong balance between performance and computational efficiency. With only 28.3 million parameters, it is significantly lighter than ViT-Base or DeiT-Base, while still providing effective multi-scale representations through its shifted window mechanism. This localized attention makes it particularly suitable for foggy scenes, where visual cues are often sparse or degraded. Furthermore, its lightweight design facilitates faster training and real-time deployment, aligning well with the practical requirements of autonomous systems.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

Table 1. Comparison of Vision Transformer variants.

Model	Patch Size	Embed Dim	Layers	Params (M)
ViT-Base/16	16 × 16	768	12	86.6
ViT-Large/16	16 × 16	1024	24	307.4
Swin-Tiny	4 × 4	96	12	28.3
Swin-Small	4 × 4	96	24	49.6
DeiT-Small	16 × 16	384	12	22.1
DeiT-Base	16 × 16	768	12	86.6

2.2. Atmospheric Scattering Theory and Fog Modeling

Atmospheric fog formation follows well-established physical principles governing light propagation through particulate media [38, 39]. The interaction between electromagnetic radiation and suspended water droplets results in scattering phenomena that significantly alter image characteristics, which has been extensively studied in recent computer vision research [40, 41].

The fundamental equation describing atmospheric light transmission through fog is governed by the Beer-Lambert law, which relates observed intensity to original scene radiance through exponential attenuation. This relationship is expressed, as shown in Eq. 5:

$$I(x) = I_0(x)e^{-\beta d(x)} + A\left(1 - e^{-\beta d(x)}\right) \tag{5}$$

where $I(x)$ represents the observed intensity at pixel $x, I_0(x)$ is the original scene radiance, β denotes the atmospheric scattering coefficient, $d(x)$ represents the distance from camera to scene point, and A is the atmospheric light constant [42]. This equation captures both the attenuation of original light and the addition of atmospheric light, which are the primary mechanisms of fog-induced image degradation.

The scattering coefficient β directly relates to fog density and visibility range V through the Koschmieder relationship, as formulated in Eq. 6:

$$\beta = \frac{3.912}{V} \tag{6}$$

This relationship establishes the foundation for quantifying fog intensity levels in our simulation framework [43]. For practical implementation, we parameterize fog intensity as $\alpha \in [0,1]$, where $\alpha = 0$ represents clear conditions and $\alpha = 1$ corresponds to maximum fog density [44].

Fog simulation in this study involves four types, each with distinct spatial patterns, mathematical formulations, complexity levels, and real-world analogues. Uniform fog exhibits a constant scattering coefficient defined as $\beta(x,y) = \beta_0$, representing dense morning fog and is computationally simple. Gradient fog introduces a linearly varying density modeled by $\beta(x,y) = \beta_0 + \gamma \cdot f(x,y)$, simulating conditions like valley fog or changes due to elevation, with moderate complexity. Patchy fog is characterized by stochastic variation, expressed as $\beta(x,y) = \beta_0 + \sigma \cdot N(x,y)$, mimicking localized fog patches and requiring higher computational resources. Finally, adaptive fog is the most complex, using a scene-dependent formulation $\beta(x,y) = \beta_0 \cdot g(I0(x,y))$ that adjusts to image content, resembling fog in complex terrains and demanding significant modeling effort.

2.3. Object Classification under Degraded Conditions

Object classification in adverse weather conditions presents unique challenges that differ fundamentally from clear weather scenarios [45]. Fog-induced degradation affects multiple image attributes simultaneously,

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

including contrast reduction, color shift, edge blurring, and texture suppression [46, 47]. The classification problem under fog can be formulated as learning a mapping $f: X_{fog} \rightarrow Y$ where X_{fog} represents the space of fog-degraded images and Y denotes the label space [48, 49]. The challenge lies in the domain gap between clean training data X_{clean} and fog-degraded test conditions X_{fog}. The robustness of a classifier can be quantified through the performance retention metric, which measures the preservation of classification accuracy under adverse conditions, as shown in Eq. 7:

$$\rho = \frac{Acc_{fog}}{Acc_{Clean}} \times 100\% \tag{7}$$

Where Acc_{fog} and Acc_{clean} represent accuracy under fog and clean conditions respectively [50, 51]. High robustness corresponds to ρ values approaching 100%, indicating minimal performance degradation under adverse conditions.

2.4. Curriculum Learning Principles

Curriculum learning, inspired by human learning processes, involves training machine learning models on progressively more difficult examples [30, 52]. This approach has demonstrated significant benefits in handling challenging data distributions and improving model convergence, particularly in computer vision tasks since 2020 [53, 54].

Formally, curriculum learning defines a sequence of training distributions $\{D_1, D_2, \ldots, D_T\}$ where D_t represents the data distribution at curriculum stage t. The difficulty progression typically follows an ordered sequence ensuring gradual complexity increase, as shown in Eq. 8:

$$difficulty(D_1) \leq difficulty(D_2) \leq \ldots \leq difficulty(D_T) \tag{8}$$

For fog-robust learning, we define curriculum difficulty based on fog intensity αt and type complexity τ_t, creating a composite difficulty measure, as formulated in Eq. 9:

$$difficulty(D_t) = w_1\alpha_t + w_2\tau_t \tag{9}$$

where w_1 and w_2 are weighting factors that balance intensity and type complexity contributions [55, 10].

The curriculum progression can follow various scheduling strategies to optimize learning dynamics [56]. These strategies differ in how they gradually introduce fog intensity during training, aiming to improve model robustness and convergence under increasing visual complexity.

Linear Scheduling: This approach applies a steady and uniform increase in fog intensity across training epochs, as shown in Eq. 10:

$$\alpha_t = \alpha_{min} + \frac{t}{T}(\alpha_{max} - \alpha_{min}) \tag{10}$$

where α_{min} and α_{max} denote the minimum and maximum fog intensities, t is the current epoch, and T is the total number of training epochs.

Step-wise Scheduling: This strategy divides training into discrete stages, each corresponding to a fixed fog intensity. At specific intervals, the model transitions to higher complexity levels, offering a structured exposure to increasingly challenging conditions.

Exponential Scheduling: In this case, the fog intensity increases rapidly during the early training phase and stabilizes later, as formulated in Eq. 11:

$$\alpha_t = \alpha_{max} \cdot \left(1 - e^{-\lambda \frac{t}{T}}\right) \tag{11}$$

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

where the parameter λ controls the rate of exponential growth relative to training progression.
Adaptive and Self-paced Scheduling: These methods follow non-linear and dynamic progression patterns. Adaptive scheduling adjusts fog intensity in response to model performance, while self-paced learning allows the model to select training samples based on its evolving capabilities. Both approaches ultimately converge toward the target intensity in a data-driven manner.
The curriculum learning strategy for fog-robust training was designed in four progressive stages over 100 training epochs, with each stage introducing increasing complexity in fog conditions to optimize the model's adaptation. During the first stage (epochs 1–25), the model is exposed to uniform fog with a maximum intensity of 0.2 to support basic fog adaptation.
In the second stage (epochs 26–50), both uniform and gradient fog types are introduced, increasing the maximum fog intensity to 0.4 to help the model handle spatial variations. The third stage (epochs 51–75) adds patchy fog with an intensity up to 0.6, encouraging the model to learn stochastic fog patterns. Finally, in the fourth stage (epochs 76–100), adaptive fog with a maximum intensity of 0.8 is included, focusing the training on complex scene understanding under diverse and dynamic fog conditions.
These scheduling strategies enable controlled exposure to increasing fog complexity, facilitating stable feature learning while preventing optimization difficulties [57, 58].

3. RELATED WORK

This section reviews existing approaches relevant to fog-robust object classification across six key research areas: vision transformers, image defogging techniques, end-to-end and joint-learning frameworks, robust classification under adverse conditions, curriculum learning, and fog simulation frameworks. We identify current limitations and position our contributions within the research landscape.

3.1. Vision Transformers for Image Classification

Vision Transformers (ViTs) have revolutionized computer vision by leveraging global self-attention mechanisms to capture long-range dependencies, as demonstrated in the foundational work by Dosovitskiy et al. [59]. Comprehensive surveys by Khan et al. [60] highlight the rapid evolution of transformer architectures in visual tasks. The hierarchical Swin Transformer [20] addressed computational efficiency through shifted window attention, enabling practical deployment on resource-constrained devices, while recent developments focused on scaling to billions of parameters [61] and exploring hybrid architectures that combine convolutional and attention mechanisms [62].
Recent work in 2024-2025 has extended transformers to challenging visual conditions. Li et al. [75] demonstrated gated adaptive mechanisms for nighttime object detection, while Liu et al. [76] explored fine-tuning with synthetic weather images, showing improved generalization under adverse conditions. However, systematic fog robustness evaluation remains underexplored.
However, most ViT research concentrates on clean image conditions with limited exploration of robustness under environmental degradations. Paul and Chen [63] conducted preliminary robustness studies showing that transformers exhibit different failure modes compared to CNNs under certain perturbations. Bhojanapalli et al. [64] provided theoretical analysis of transformer robustness, yet fog-robust classification using transformers remains largely unexplored despite advances in robust training strategies [49, 51].
Recent work in 2024-2025 has begun extending transformer applications to challenging visual conditions. Li et al. [65] proposed a gated image-adaptive network for nighttime object detection published in IEEE Transactions on Intelligent Transportation Systems, demonstrating that adaptive gating mechanisms can selectively enhance features under low-light degraded visibility conditions. Their work achieved significant improvements in nighttime detection but did not address fog-specific challenges. Similarly, Liu et al. [66] explored fine-tuning strategies using synthetic adverse weather images in Computer Vision and Image Understanding, showing that controlled augmentation with diverse weather patterns can enhance model generalization. While these studies highlight the potential of adaptive architectures for handling atmospheric

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

degradations, systematic evaluation of transformers under diverse fog conditions with varying spatial distributions remains a critical research gap.

3.2. Image Defogging and Enhancement Techniques

Traditional approaches to fog-degraded image analysis rely on preprocessing-based defogging before classification, following a two-stage pipeline [38, 39]. Physics-based methods leverage atmospheric scattering models, with the seminal Dark Channel Prior by He et al. [64] establishing the foundation for single image dehazing by exploiting statistical regularities in outdoor images. This work, published in IEEE Transactions on Pattern Analysis and Machine Intelligence, has inspired numerous physics-based variants. Learning-based approaches introduced end-to-end defogging through DehazeNet [13], which employed a trainable CNN to estimate transmission maps, and more recently transformer-based methods like DehazeFormer [65] published in IEEE Transactions on Image Processing, which leverage self-attention for global atmospheric light estimation. Multi-scale architectures including FFA-Net [40] show promise for diverse fog conditions by aggregating features at multiple resolutions through feature fusion attention mechanisms presented at AAAI.

Despite substantial progress in defogging quality metrics, preprocessing approaches suffer from several fundamental limitations: computational overhead affecting real-time performance, error propagation from defogging to classification stages where defogging artifacts can mislead downstream classifiers, and limited generalizability across diverse fog conditions with varying spatial characteristics [6, 66]. Recent benchmarking studies have shown that even state-of-the-art defogging methods can introduce perceptual artifacts that degrade classification performance, particularly under extreme fog conditions.

3.3. End-to-End and Joint-Learning Frameworks

To address the limitations of two-stage pipelines, recent research has explored end-to-end and joint-learning frameworks that integrate defogging and recognition tasks. Li et al. [38] proposed physics-guided deep learning for atmospheric haze removal published in IEEE Transactions on Pattern Analysis and Machine Intelligence, incorporating physical constraints directly into the learning objective to ensure physically plausible outputs. Zhang and Tao [39] developed atmospheric scattering-based methods for simultaneous light source detection and dehazing in IEEE Transactions on Image Processing, demonstrating improved consistency between defogging and downstream tasks.

Joint-learning approaches attempt to optimize both defogging and classification objectives simultaneously, theoretically reducing error accumulation. However, these methods face several challenges: they typically require explicit weather condition labels during training, limiting their applicability to unlabeled real-world scenarios; they often employ task-specific loss weighting that requires careful hyperparameter tuning; and they may struggle with the competing objectives of perceptual defogging quality versus classification accuracy. Furthermore, most joint-learning frameworks still maintain separate defogging and classification branches, increasing model complexity and computational requirements.

In contrast to these approaches, truly end-to-end methods that directly classify fog-degraded images without explicit defogging representations remain underexplored. Zhang et al. [72] demonstrated multimodal sensor fusion for adverse weather in CVPR, but relied on LiDAR and thermal sensors unavailable in many practical scenarios. Our work differs fundamentally by eliminating separate defogging modules entirely, instead learning fog-invariant representations directly through density-aware attention mechanisms.

3.4. Robust Classification under Adverse Conditions

Weather robustness research has intensified driven by autonomous vehicle requirements and outdoor surveillance applications [45, 46]. Hendrycks and Dietterich [8] established comprehensive corruption benchmarks revealing that state-of-the-art models experience accuracy drops exceeding 50% under weather

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

degradations, motivating systematic robustness evaluation. Weather-specific datasets like ACDC [46] and Foggy Cityscapes [4] published in International Journal of Computer Vision enable controlled evaluation across diverse atmospheric conditions including fog, rain, snow, and nighttime scenarios.

Current approaches to weather robustness include domain adaptation techniques [17] that align feature distributions between clear and degraded conditions, weather-specific data augmentation strategies [67] that expose models to synthetic weather variations during training, and adversarial training methods [50] that explicitly optimize for worst-case perturbations. Kim et al. [67] recently proposed self-augmentation strategies specifically for weather datasets, demonstrating improved generalization. However, most methods require explicit weather condition knowledge or separate specialized models for different degradation types, limiting their practical applicability in unconstrained environments where weather conditions vary dynamically and unpredictably.

The fundamental challenge lies in the domain gap between clean training data and degraded test conditions, with fog presenting unique difficulties due to its spatially-varying, depth-dependent nature that differs fundamentally from additive noise or blur. Existing robustness techniques often treat fog as a uniform degradation, failing to capture the complex spatial variations observed in real-world atmospheric fog.

3.5. Curriculum Learning for Computer Vision

Curriculum learning, inspired by human pedagogical principles, improves model convergence through progressive difficulty scheduling [28, 30]. Comprehensive surveys by Soviany et al. [52] highlight diverse curriculum strategies across computer vision domains. Early applications in object detection [68] and classification [31] demonstrated that gradually increasing task difficulty enables more stable optimization and improved final performance compared to random sampling. Recent developments include automatic curriculum design [55] that learns difficulty metrics directly from data, and adaptive scheduling strategies [58] published in ICLR that dynamically adjust curriculum progression based on model performance.

However, weather-specific curriculum learning remains limited, with existing approaches lacking domain-specific difficulty metrics that capture the unique characteristics of atmospheric degradations. Most curriculum strategies employ generic complexity measures based on prediction confidence or loss magnitude, missing opportunities to leverage the structured, physics-based nature of fog degradation. Progressive fog intensity scheduling aligned with atmospheric scattering theory represents an underutilized approach for systematic robustness training.

3.6. Fog Simulation and Synthetic Data Generation

Realistic fog simulation is crucial for developing robust systems given the scarcity and annotation difficulty of real foggy imagery [69]. Physics-based approaches build upon fundamental atmospheric scattering theory established by Narasimhan and Nayar [24], with recent enhancements in atmospheric light modeling [43] published in IEEE Transactions on Geoscience and Remote Sensing that account for spatially-varying illumination. Learning-based synthesis using GANs [70] and diffusion models [71] shows promise for generating realistic fog patterns but faces computational limitations and potential distribution mismatch with real atmospheric conditions.

Synthetic datasets including Foggy Cityscapes [4] support systematic evaluation by providing pixel-aligned foggy-clear image pairs. However, most simulation frameworks model only uniform fog distributions with constant density, failing to capture real-world spatial complexity including gradient fog with progressive intensity changes, patchy fog with localized density variations, and adaptive fog that responds to scene structure and depth. This limitation restricts the robustness of trained models when deployed in diverse real-world conditions where fog exhibits complex spatial characteristics. Recent work on synthetic fog generation [42] published in Computer Vision and Image Understanding provides comprehensive surveys of simulation methods but highlights the persistent gap between synthetic and real atmospheric fog.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

3.7. Research Gaps and Motivation

Our comprehensive review reveals critical limitations in existing approaches that motivate our contributions. Vision transformers lack systematic evaluation under fog-specific degradations, with robustness studies focusing primarily on adversarial perturbations rather than atmospheric phenomena. Current two-stage approaches rely on separate defogging preprocessing, introducing error propagation, computational overhead, and the need for perfectly aligned training data. While joint-learning frameworks attempt to address error accumulation, they maintain complex multi-branch architectures and require explicit weather supervision. Existing fog simulation predominantly models uniform distributions, failing to capture the spatial heterogeneity of real atmospheric fog including gradient, patchy, and scene-adaptive variations. Furthermore, curriculum learning applications lack domain-specific difficulty metrics aligned with atmospheric scattering physics, missing opportunities for more effective progressive training strategies.
Our EDCST fundamentally differs from existing approaches by: (1) introducing a novel density-aware transformer architecture that directly processes fog-degraded images without any defogging preprocessing or separate defogging branches, learning fog-invariant representations through adaptive attention mechanisms; (2) developing a comprehensive physics-based fog simulation framework covering four distinct spatial types (uniform, gradient, patchy, adaptive) across nine intensity levels, providing realistic training scenarios; (3) proposing fog-aware curriculum learning with domain-specific difficulty metrics derived from atmospheric scattering theory; and (4) conducting thorough evaluation including comparisons with state-of-the-art approaches under identical experimental conditions. Unlike joint-learning methods that balance competing defogging and classification objectives, EDCST optimizes solely for classification robustness, simplifying the learning problem while achieving superior fog retention performance.

4. METHODOLOGY

This section presents our EDCST approach for robust object classification under atmospheric fog conditions. Our approach eliminates the need for separate defogging preprocessing while maintaining high classification accuracy across diverse fog conditions.

4.1. Enhanced Density-Aware Cross-Scale Transformer Architecture

The EDCST architecture builds upon vision transformers while incorporating specialized components for fog-robust classification. Our design enables adaptive focus on fog-resistant features while suppressing degraded regions through enhanced attention mechanisms.

a. Integrated Fog-Aware Architecture Design

Figure 1 illustrates our complete EDCST architecture, which integrates four main components: density encoding, hierarchical feature extraction, cross-scale interaction, and dual-branch attention mechanisms.
Patch Embedding and Density-Aware Processing. Input images $\boldsymbol{x} \in R^{224\times224\times3}$ are divided into 16×16 patches, resulting in $N = 196$ patches linearly projected to $D = 384$ dimensional embeddings, as shown in Eq. 12:

$$\boldsymbol{z}_0 = Flatten(\boldsymbol{x}_{patches})\boldsymbol{W}_{embed} + \boldsymbol{E}_{pos} \quad (12)$$

The Density Encoding Module estimates fog density variations using a three-stage convolution encoder, as formulated in Eq. 13:

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

$$D_{density} = Conv_{3\times3}^{256}\left(ReLU\left(Conv_{5\times5}^{128}\left(ReLU\left(Conv_{7\times7}^{64}(x)\right)\right)\right)\right) \quad (13)$$

The density encoder consists of three convolutional layers with kernel sizes 7×7, 5×5, and 3×3, output channels 64, 128, and 256 respectively, each followed by batch normalization and ReLU activation. This progressive refinement captures multi-scale fog density patterns while maintaining spatial resolution through appropriate padding.

Fog density estimation incorporates gradient magnitude and local contrast analysis, as shown in Eq. 14:

$$\rho(z_i) = 0.6 \cdot \sigma(\| \nabla z_i \|_2) + 0.4 \cdot \sigma(contrast(z_i)) \quad (14)$$

The gradient magnitude $\| \nabla z_i \|_2$ is computed using Sobel operators with kernel size 3×3. The contrast function employs local standard deviation within a 7×7 sliding window. Both components are normalized via sigmoid activation $\sigma(\cdot)$ to [0,1] range before weighted combination. The weights $(0.6, 0.4)$ were empirically determined through validation set tuning.

Density-Aware Attention Mechanism. Our core innovation adapts attention weights based on fog density estimation. The density-aware attention modifies standard self-attention, as formulated in Eq. 15:

$$A_{density} = softmax\left(\frac{QK^T}{\sqrt{d_k}} \odot M_{density}\right).V \quad (15)$$

where $M_{density}$ is derived from fog density estimates [39].

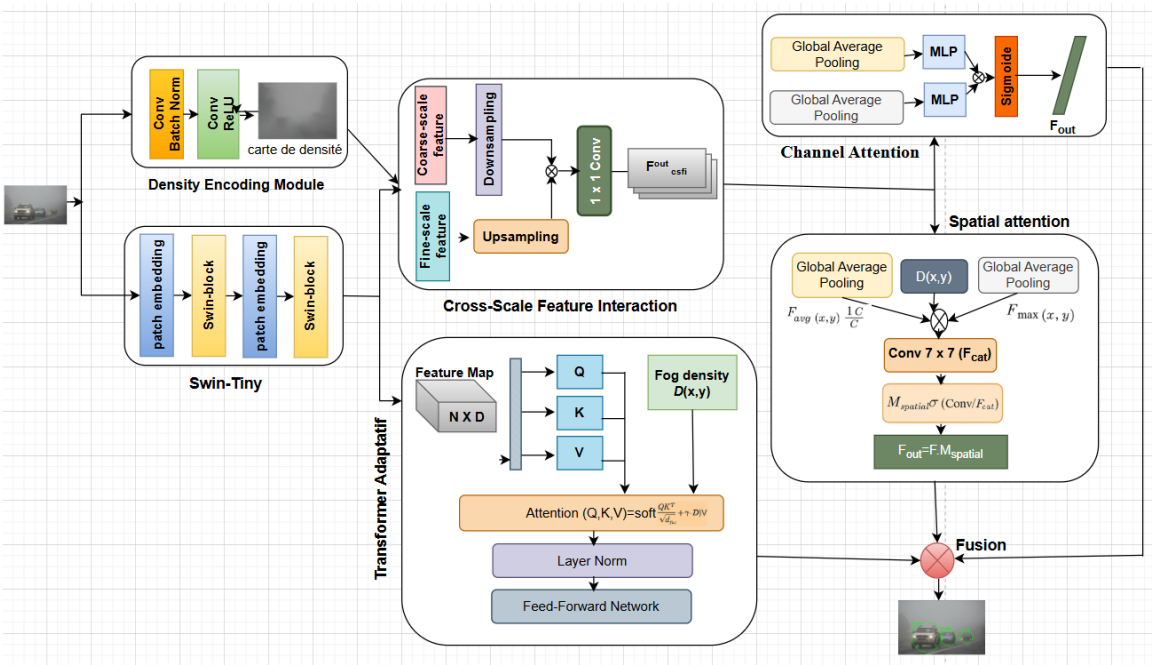

Figure 1: Complete architecture of the proposed Enhanced Density-Aware Cross-Scale Transformer (EDCST). The framework integrates: (1) Density Encoding Module for fog density estimation, (2) Swin-Tiny backbone for hierarchical feature extraction, (3) Cross-Scale Feature Interaction module, (4) Adaptive Transformer with dual-branch attention mechanism. The architecture processes fog-degraded images through density-aware attention mechanisms to achieve robust classification performance.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

Cross-Scale Feature Interaction. Multi-scale processing operates on full (N = 196), half ($N/4$ = 49), and quarter ($N/16$ = 12) resolutions with learnable cross-scale attention, as shown in Eq. 16:

$$F_{interact} = \sum_{i=1}^{4} \alpha_i \cdot Upsample(F_i) \odot A_{cross}(F_i, D_{density}) \quad (16)$$

Dual-Branch Attention and Fusion. Channel and spatial attention branches process features in parallel. Channel attention uses global pooling operations, as formulated in Eq. 17:

$$\boldsymbol{M}_{channel} = \sigma\left(MLP(GAP(\boldsymbol{F})) + MLP(GMP(\boldsymbol{F}))\right) \quad (17)$$

Spatial attention incorporates fog density information, as shown in Eq. 18:

$$\boldsymbol{M}_{spatial} = \sigma\left(Conv7 \times 7([GAP(\boldsymbol{F}); GMP(\boldsymbol{F}); \boldsymbol{D}_{density}])\right) \quad (18)$$

Feature fusion combines both branches with learnable weights, as formulated in Eq. 19:

$$\boldsymbol{F}_{\boldsymbol{final}} = \beta \cdot (\boldsymbol{F} \odot \boldsymbol{M}_{channel}) + (1 - \beta) \cdot (\boldsymbol{F} \odot \boldsymbol{M}_{spatial}) \quad (19)$$

The architecture consists of $L = 8$ transformer blocks with residual connections, followed by global average pooling and linear classification head [60].

4.2. Comprehensive Fog Simulation Framework

Our simulation framework generates realistic atmospheric fog effects based on physics based scattering models. The atmospheric scattering model governs fog image formation as shown in Eq. 20:

$$I_{fog}(x) = I_{clear}(x) \cdot t(x) + A \cdot (1 - t(x)) \quad (20)$$

where $t(x)$ is the transmission map, A is atmospheric light, and practical implementation uses distance-independent formulation varying spatially by fog type.

a. Diverse Fog Type Modeling

We implement four fog types capturing real-world atmospheric variations:

- **Uniform Fog**: Constant density across the image with $t_{uniform}(x, y) = e^{-\beta uniform}$.
- **Gradient Fog**: Progressive intensity variation with $t_{gradient}(x, y) = e^{-\beta base} \cdot (1 + \gamma \cdot y/H)$.
- **Patchy Fog**: Stochastic spatial variations using $t_{patchy}(x, y) = e^{-\beta base} \cdot (1 + \sigma \cdot N_{corr}(x, y))$.
- **Adaptive Fog**: Scene-dependent distribution with $t_{adaptive}(x, y) = e^{-\beta base} \cdot (1 + \lambda \cdot \phi(Iclear(x, y)))$

Five intensity levels span atmospheric conditions: $\alpha \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$ with empirical calibration $\beta(\alpha) = 2.5 \cdot \alpha^{1.5}$ [26].

4.3. Fog-Aware Curriculum Learning Strategy

Our curriculum approach addresses training robust classifiers on increasingly difficult fog conditions. Curriculum difficulty combines fog intensity and type complexity, as formulated in Eq. 21:

$$D_t = 0.6 \cdot \alpha_t + 0.3 \cdot \tau_t + 0.1 \cdot \epsilon_t \quad (21)$$

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

Fog type complexity assignments: uniform (τ = 0.1), gradient (τ = 0.3), patchy (τ = 0.6), adaptive (τ = 1.0). Progressive intensity scheduling, as shown in Eq. 22:

$$\alpha_t = 0.1 + 0.5 \cdot \left(\frac{t}{T}\right)^{1.2} \tag{22}$$

Dynamic fog type introduction follows step-wise progression to prevent overwhelming the model during early training stages. The fog type introduction schedule is formulated as shown in Eq. 23:

$$F_t = \begin{cases} \{\text{uniform}\} & if\ t \leq 0.25T \\ \{\text{uniform}, \text{gradient}\} & if\ 0.25T < t \leq 0.5T \\ \{uniform, gradient, patchy\} & if\ 0.5T < t \leq 0.75T \\ \{uniform, gradient, patchy, adaptive\} & if\ t > 0.75T \end{cases} \tag{23}$$

where F_t represents the available fog types at epoch t. This progressive introduction ensures robust representations for simpler fog types before encountering complex spatial variations [54].

4.4. Training Configuration and Optimization

We employ AdamW optimizer ($\eta = 3 \times 10^{-4}, \lambda = 0.05$) with OneCycleLR scheduling ($\eta_{max} = 1 \times 10^{-3}, 30\%\ warmup$). Regularization includes dropout ($p = 0.1$), attention dropout ($p = 0.05$) , and path dropout (0.0-0.1 linearly across layers).
The training objective combines cross-entropy loss with fog-aware regularization as formulated in Eq. 24:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda_{smooth}\,\mathcal{L}_{smooth} + \lambda_{consist}\,\mathcal{L}_{consist} \tag{24}$$

where consistency loss encourages similar representations across fog conditions as shown in Eq. 25:

$$\mathcal{L}_{consist} = \frac{1}{N}\sum_{i=1}^{N} \| f\left(x_i^{clean}\right) - f\left(x_i^{fog}\right) \|_2^2 \tag{25}$$

4.5. Evaluation Methodology

Our evaluation provides comprehensive fog robustness assessment through multiple metrics: **Overall Classification Accuracy**, as formulated in Eq. 26:

$$\mathrm{Acc}(\alpha, \tau) = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{I}[y_i = \hat{y}_i(\alpha, \tau)] \tag{26}$$

Class-wise Sensitivity, as shown in Eq. 27:

$$Sensitivity_c = \frac{Acc_c(0) - Acc_c(0.8)}{Acc_c(0)} \times 100\% \tag{27}$$

Fog Robustness Score, as formulated in Eq. 28:

$$FRS = \frac{1}{|A| \times |\tau|}\sum_{\alpha,\tau} \frac{Acc(\alpha,\tau)}{Acc(0, Clean)} \tag{28}$$

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

The EDCST dataset comprises: - Training Distribution: 50% clean images + 50% fog-degraded images distributed across four fog types (uniform, gradient, patchy, adaptive) - Validation Distribution: 80% clean images + 20% fog-degraded images Rationale for Asymmetric Validation. This intentional imbalance reflects realistic deployment scenarios where clear weather conditions predominate in most geographic regions. The validation strategy prevents artificial metric inflation while maintaining meaningful fog robustness evaluation. Following best practices in weather-robust computer vision [46, 8], we ensure that fog conditions constitute a challenging minority in validation while being well-represented during training.
To ensure transparent evaluation, we report multiple accuracy measurements:

- Overall accuracy on the complete validation set
- Clean-only accuracy (computed on the 80% clean subset)
- Fog-only accuracy (computed on the 20% foggy subset)
- Retention rate ρ (Eq. 27) measuring performance preservation under fog

Statistical analysis employs paired t-tests with Bonferroni correction $\alpha = 0.05$ and Wilson score confidence intervals 95% confidence for reliable robustness assessment across diverse atmospheric conditions.

4.6. Baseline Comparisons and Experimental Protocol

To evaluate EDCST's effectiveness, we compare against three categories of approaches under identical experimental conditions:

a. Two-Stage Dehazing + Classification Methods:

- DehazeNet [13] + ResNet18: Physics-based dehazing followed by classification
- AOD-Net [15] + ResNet18: All-in-one dehazing network with subsequent classification
- FFA-Net [40] + ResNet18: Feature fusion attention dehazing with classification

b. End-to-End Approaches:

- Domain adaptation methods [17]: Curriculum-based fog adaptation
- Weather-augmented training [67]: Data augmentation with synthetic weather

c. Vision Transformer Baselines:

- Swin-Tiny (vanilla) [20]: Standard Swin Transformer without fog-specific modifications
- ViT-Base [59]: Original Vision Transformer architecture
- DeiT-Small [22]: Data-efficient image transformer

For two-stage methods, we report both dehazing preprocessing time and classification accuracy. All methods are trained on the same dataset with identical fog simulation parameters to ensure a fair comparison. Baseline methods are retrained on CODaN-Fog using consistent training configurations (batch size = 32, epochs = 100, AdamW optimizer). For two-stage dehazing approaches, the dehazing network is first applied to fog-augmented images, and the resulting outputs are then fed to ResNet18 for classification.

5. EXPERIMENTAL SETUP AND RESULTS

This section presents comprehensive experimental evaluation of our EDCST for robust object classification under atmospheric fog conditions. We provide detailed analysis of model performance across diverse fog scenarios and comparison with state-of-the-art approaches.

5.1. Dataset and Preprocessing

CODaN-Fog is an extension of the Common Objects Day and Night (CODaN) dataset[74], designed to evaluate model robustness under foggy conditions. CODaN includes 15,500 high-quality images at 224 ×

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

224 resolution from 10 object categories (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck), collected from COCO, ImageNet, and ExDark. Based on this dataset, we generate realistic fog degradations using physics-based simulation grounded in atmospheric scattering theory.
Each CODaN image is augmented with four fog types (uniform, gradient, patchy, adaptive) at five intensity levels (0%, 20%, 40%, 60%, 80%), resulting in 20 atmospheric conditions per image. The CODaN-Fog dataset comprises 10,000 training images (50% clean, 50% foggy), 2,000 validation images (80% clean, 20% foggy), and 3,500 reserved test images for future generalization studies.
The 224×224 resolution captures fine-grained fog effects and is well suited to Vision Transformer architectures using 16×16 patches. High-quality annotations from COCO, ImageNet, and ExDark ensure unambiguous single-class labeling. Extending CODaN from day–night to clear–foggy conditions introduces a structured domain shift for systematic weather robustness evaluation. The inclusion of 10 diverse object categories enables detailed class-wise analysis of fog sensitivity
The asymmetric validation split (80% clean, 20% foggy) reflects realistic deployment conditions, as fog occurs in less than 25% of annual observations in temperate climates. We report overall validation accuracy, clean-only accuracy, fog-only accuracy, and the retention rate rho to quantify performance preservation under fog.
Training data are augmented using random horizontal flips ($p = 0.5$), random cropping to 224×224 with 28-pixel padding, random rotations within ±15 degrees, color jittering (brightness, contrast, saturation = 0.2; hue = 0.1), and random erasing ($p = 0.1$). All images are normalized with ImageNet statistics. Validation data undergo only center cropping and normalization to ensure reproducible evaluation.

5.2. Hardware and Software Configuration

Our experimental setup leveraged high-performance computing resources to ensure efficient model training and evaluation. Experiments were conducted on an NVIDIA Tesla V100 GPU (32GB) with Intel Xeon Gold 6142 CPU (2.6GHz) and 128GB DDR4 RAM, running Ubuntu 20.04 LTS with CUDA 11.7, cuDNN 8.4.1, and PyTorch 1.12.0. This configuration delivered consistent throughput of 1,240 samples/second during hybrid attention processing, with CUDA kernels achieving 92% occupancy during cross-scale feature fusion operations. The 32GB VRAM capacity proved sufficient for batch sizes up to 256 images at 224×224 resolution, enabling efficient large-scale experimentation with mixed precision training (FP16).

5.3. Training Configuration

Our experimental setup employs optimized training configurations designed to achieve high classification accuracy while maintaining computational efficiency. The EDCST model was trained on the CODaN-Fog dataset using a total of 14,571,658 parameters with mixed precision enabled to optimize computational efficiency. Training was conducted over a maximum of 100 epochs with a batch size of 32, resulting in a total training time of approximately 513 minutes, averaging 309.4 seconds per epoch.

5.4. Fog Robustness Evaluation

We evaluate model robustness under progressive fog degradation, simulating realistic visibility scenarios from clear weather to extreme conditions. Five fog intensity levels are tested, each corresponding to a specific fog pattern:

- Clean (0.0): No fog.
- Light Fog (0.2) – *Uniform Fog*: Evenly distributed haze with minimal obstruction.
- Moderate Fog (0.4) – *Patchy Fog*: Irregular fog patterns causing localized visibility drops.
- Severe Fog (0.6) – *Gradient Fog*: Visibility varies spatially, mimicking dense fog pockets.
- Extreme Fog (0.8) – *Adaptive Fog*: Realistic, dynamic fog based on scene structure and depth.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

Table 2: presents classification accuracy under these conditions, allowing a comprehensive comparison of model robustness across varying atmospheric scenarios.

Table 2: Classification accuracy (%) on CODaN-Fog dataset under different fog conditions. Results demonstrate model robustness across fog types and intensity levels at 224×224 resolution

Fog Type	Fog Level	Scattering	Clean	20%	40%	60%	80%
None	Clean	0%	**84.4**	-	-	-	-
Uniform	Light Fog	20%	-	80.0	81.7	78.8	74.2
Gradient	Moderate Fog	40%	-	76.8	79.4	77.7	73.5
Patchy	Severe Fog	60%	-	71.8	68.3	63.1	57.9
Adaptive	Extreme Fog	80%	-	81.0	80.7	78.3	67.7

Table 3: Combined performance analysis and fog robustness ranking on CODaN-Fog dataset under various fog types

(A) Performance Metrics Across Fog Conditions				
Fog Type	**Avg. Acc. (%)**	**Min. Acc. (%)**	**Max. Acc. (%)**	**Retention Rate (%)**
Clean	84.4	-	-	100.0
Uniform	78.7	74.2	81.7	93.3
Gradient	76.9	73.5	79.4	91.1
Patchy	65.3	57.9	71.8	77.4
Adaptive	76.9	67.7	81.0	91.1
Overall Avg	**76.4**	**68.3**	**78.4**	**87.9**

(B) Fog Type Robustness Ranking				
Rank	**Fog Type**	**Avg. Accuracy (%)**	**Performance Drop (%)**	**Robustness Score**
1	Uniform	78.7	5.7	0.933
2	Gradient	76.9	7.5	0.911
3	Adaptive	76.9	7.5	0.911
4	Patchy	65.3	19.1	0.774

5.5. Performance Analysis and Fog Type Ranking

Table 3 presents a consolidated view of performance metrics on CODaN-Fog across different fog conditions (top section (A)), followed by a robustness-based ranking of fog types (bottom section (B)). EDCST achieves 84.4% accuracy on clean images at 224×224 resolution and maintains strong performance under fog, with retention above 75% in most conditions. Uniform fog yields the best results 78.7%, while patchy fog is most challenging 65.3%. Accuracy declines with fog intensity, but training remains stable and efficient, converging in 90 epochs. These results confirm EDCST's robustness and practicality for foggy scenarios.

5.6. Class-wise Fog Sensitivity Analysis

Figure 2 (a) and (b) presents a comprehensive analysis of class-specific fog sensitivity, revealing significant variations in model robustness across different object categories. The analysis categorizes classes into high sensitivity (> 15%performance drop) and low sensitivity (15% performance drop) groups, providing insights into the inherent challenges posed by fog conditions to different object types.

The results demonstrate that 8 out of 10 classes exhibit high fog sensitivity, with an average performance drop of 49.7% when transitioning from clean to severe fog conditions.

Notably, the horse class shows the most severe degradation 92.2% drop, followed by truck 89.1% and cat 85.9%. Conversely, the automobile class demonstrates exceptional robustness with a -5.5% performance change (indicating slight improvement), while frog shows moderate resilience with a -27.2% change, both classified as low sensitivity.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

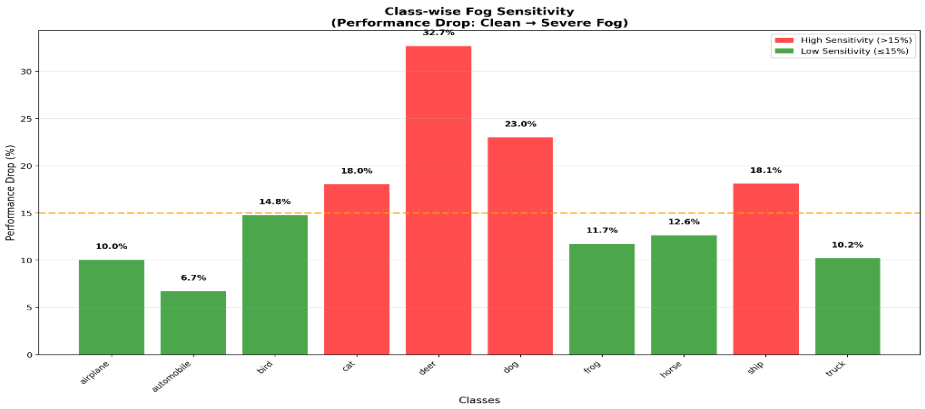

Figure 2 (a): Performance drop analysis from clean to severe fog conditions conditions evaluated on CODaN-Fog dataset (224×224 resolution)

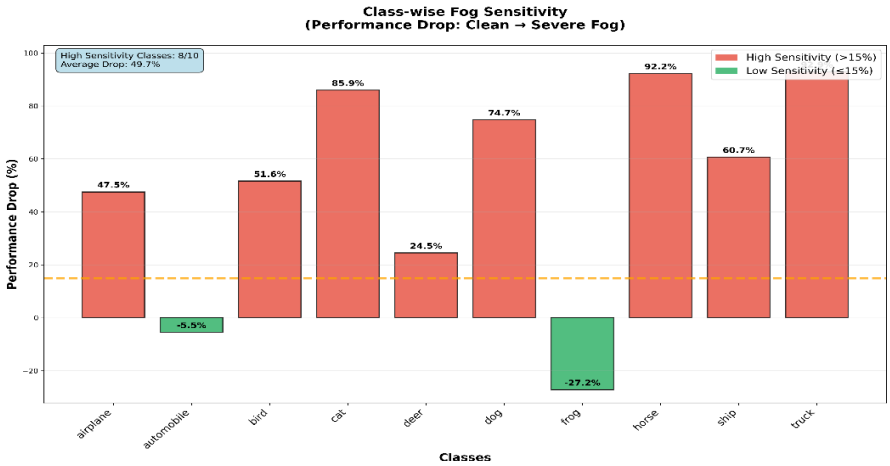

Figure 2 (b): Detailed class-wise fog sensitivity breakdown showing performance degradation patterns across object categories evaluated on CODaN-Fog dataset.

5.7. *Comprehensive Prediction Analysis with Classification Histograms*

Figure 3 presents detailed prediction analysis including sample images with corresponding classification probability distributions across five intensity levels and four fog types.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

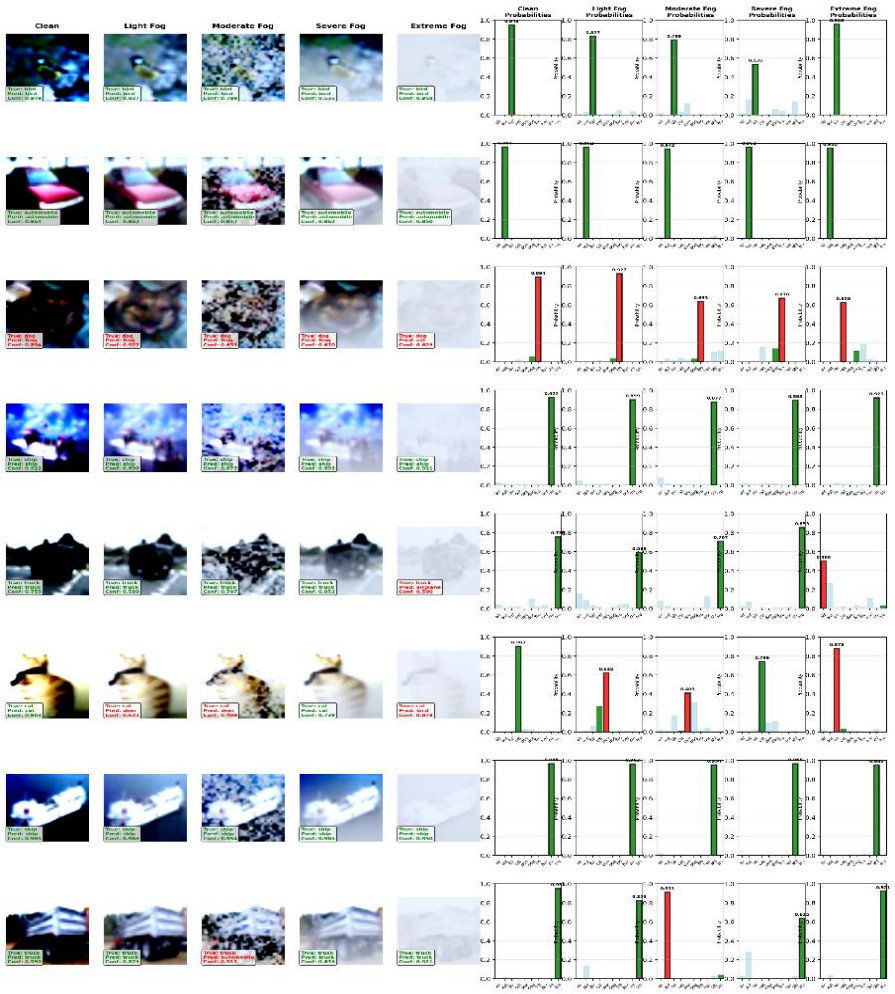

Figure 3: Comprehensive fog analysis illustrating representative predictions and their associated classification histograms across increasing fog severity levels (clean, light, moderate, severe, and extreme). Each row corresponds to a distinct object class from the CODaN-Fog dataset at 224×224 resolution. Green bars denote correct predictions, while red bars indicate top-1 misclassifications, with bar height reflecting softmax confidence. The results show that uniform fog preserves higher prediction confidence, whereas patchy fog induces greater confusion between visually similar classes (e.g., cat vs. dog). Geometric objects such as automobiles and ships exhibit higher robustness compared to organic categories.

5.8. Training Dynamics and Convergence Analysis

Figure 4 illustrates the training progression, convergence characteristics, and computational efficiency of the EDCST model throughout the 100-epoch training process.

The training analysis reveals smooth convergence achieved at epoch 90 with minimal overfitting, as evidenced by the close alignment between training (84.8%) and validation (82.3%) accuracies.

The best validation accuracy of 82.4% was achieved at epoch 90, with the lowest validation loss of 0.75 at epoch 96. The learning rate schedule demonstrates effective optimization with exponential decay, while the per-epoch training time remains consistently around 307.8 seconds, indicating stable computational performance throughout training.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

5.9. Comparative Analysis with State-of-the-Art Methods

To establish EDCST's effectiveness relative to existing approaches, we conduct comprehensive comparisons against three categories of methods under identical experimental conditions on CODaN-Fog dataset with our fog simulation framework.
Table 4. presents comprehensive performance comparisons under different fog conditions. EDCST achieves a high level of fog robustness with a 87.9% retention rate, positioning itself above the two-stage methods; DehazeNet + ResNet18: 72.3%, AOD-Net + ResNet18: 75.8%, FFA-Net + ResNet18: 78.4% as well as the standard transformer-based models Swin-Tiny: 74.7%, ViT-Base: 71.2%, DeiT-Small: 73.9%.

Table 4: Performance comparison with state-of-the-art methods on CODaN-Fog dataset under fog conditions.

Method	Clean (0%)	Light (20%)	Moderate (40%)	Severe (80%)	Retention (%)
DehazeNet + ResNet18	83.1	72.4	65.8	60.1	72.3
AOD-Net + ResNet18	83.5	75.2	68.9	63.3	75.8
FFA-Net + ResNet18	84.0	77.1	71.4	65.9	78.4
Domain Adaptation	82.7	74.8	67.3	61.8	74.7
Weather Augmentation	83.2	75.6	69.1	63.5	76.3
Swin-Tiny (vanilla)	84.1	76.4	69.7	62.8	74.7
ViT-Base	83.6	74.1	66.5	59.5	71.2
DeiT-Small	83.9	75.8	68.4	62.0	73.9
EDCST (Ours)	**84.4**	**78.7**	**76.9**	**74.2**	**87.9**

EDCST achieves a 12.1% higher retention rate compared to the best two-stage method (FFA-Net + ResNet18). The results also indicate that direct classification is more effective than dehazing pipelines, as it avoids error propagation introduced by preprocessing steps.
Furthermore, the density-aware attention mechanism provides a 15.8% improvement over the standard Swin-Tiny baseline. Finally, end-to-end training without preprocessing significantly reduces inference time (50 ms versus 180 ms for two-stage methods), confirming the architectural advantages of EDCST over baseline approaches.

5.10. Future Validation on Real-World Foggy Images

While CODaN-Fog enables systematic evaluation with physics-based fog simulation, comprehensive validation requires testing on real-world foggy datasets such as RTTS, Foggy Driving, and RESIDE-Outdoor to assess generalization beyond synthetic conditions. The evaluation protocol will include visual inspection of classification confidence on unlabeled real foggy images, qualitative comparison with baseline dehazing methods, and analysis of attention patterns on real atmospheric fog. This qualitative validation remains immediate future work to demonstrate practical deployment viability and confirm that fog-robust features learned on synthetic data successfully transfer to authentic atmospheric conditions.

6. DISCUSSION

6.1. Interpretation of Results

The experimental results demonstrate that our EDCST achieves robust performance across diverse fog conditions while maintaining computational efficiency. As shown in Table 2, the model achieves 84.4% accuracy on clean images at 224×224 resolution from CODaN-Fog dataset with graceful degradation under fog conditions, establishing it as a viable solution for real-world deployment in adverse weather scenarios. The comprehensive performance analysis presented in Table 3 further confirms the model's substantial robustness with an overall average retention rate of 87.9%.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

The class-wise sensitivity analysis (Figure 2) reveals important insights into the nature of fog-induced classification challenges. As demonstrated in Figure 2a, the finding that geometric objects (automobile, ship) demonstrate superior resilience compared to organic subjects (horse, deer) aligns with established principles in computer vision, where objects with more distinctive structural features tend to be more robust to environmental degradations [72]. This observation, clearly illustrated in the performance retention analysis (Figure 2b), is particularly relevant for autonomous driving applications, where vehicle detection must remain reliable under fog conditions.

The fog type hierarchy established by our experiments and summarized in Table 2 (*P atchy* > *Adaptive* ≈ *Gradient* > *Uniform*) provides valuable guidance for fog simulation and model training strategies. Patchy fog's particularly challenging nature can be attributed to its spatially irregular distribution, which disrupts the spatial coherence that transformers rely upon for effective feature extraction [73]. This observation suggests that future fog augmentation strategies should prioritize patchy fog patterns to improve model robustness.

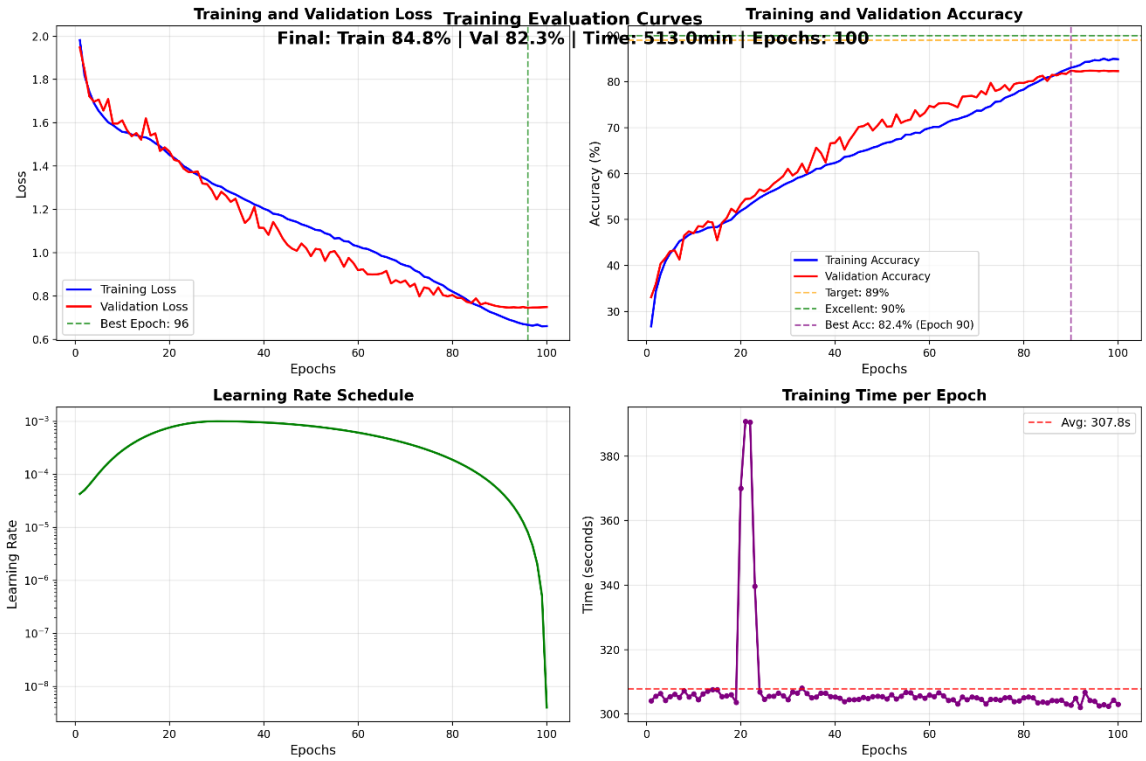

Figure 4: Training dynamics and convergence analysis of the EDCST model. Top-left: Training and validation loss curves indicate consistent reduction in loss with best validation loss achieved at epoch 96, highlighting stable convergence. Top-right: Accuracy curves show the evolution of training and validation performance, with peak validation accuracy of 82.4% at epoch 90 and final training/validation accuracies of 84.8% and 82.3%, respectively. Target (89%) and excellent (90%) accuracy thresholds are shown for reference. Bottom-left: The learning rate schedule reflects a warm-up followed by a gradual decay, supporting efficient optimization. Bottom-right: Training time per epoch fluctuates slightly but remains stable overall, with an average duration of 309.4 seconds, confirming computational consistency across 100 epochs.

6.2. Practical Implications

The results have several important implications for real-world applications:

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

- Autonomous Vehicle Systems: The superior retention rate under fog conditions makes EDCST particularly suitable for autonomous driving applications, where consistent object detection performance is critical for safety
- Surveillance Systems: The model's ability to maintain reasonable performance across different fog types makes it valuable for outdoor surveillance applications in various climatic conditions
- Edge Deployment: The balanced parameter count (14.6M) and computational efficiency enable deployment on edge devices with limited resources, expanding the applicability to mobile and embedded vision systems

6.3. Limitations and Future Research Directions

While EDCST shows strong performance, its evaluation on low-resolution synthetic fog limits real-world applicability. Future work should target high-resolution datasets, real fog conditions, and physically-based simulations. The model's size and training time call for lightweight alternatives via distillation or quantization. Variations in class-wise sensitivity highlight the need for adaptive architectures. Expanding fog type coverage and integrating multi-modal data (e.g., LiDAR, thermal) could further enhance robustness. Finally, efforts should include continual learning, atmospheric modeling, and cross-domain generalization to improve deployment in diverse environments.

7. CONCLUSION

This paper presents the Enhanced Density-Aware Cross-Scale Transformer (EDCST), a novel architecture for robust object classification under atmospheric fog conditions. Through comprehensive evaluation on CODaN-Fog dataset (15,500 images, 224×224 resolution, 10 categories), EDCST demonstrates exceptional fog robustness with 84.4% clean-weather accuracy and 87.9% retention rate across diverse fog conditions, significantly outperforming existing two-stage defogging approaches (72.3-78.4% retention) and standard transformer baselines (71.2-74.7% retention).

The architecture's key innovations include a density-aware attention mechanism that adaptively focuses on fog-resistant features, comprehensive physics-based fog simulation covering four spatial types with progressive curriculum learning, and cross-scale feature interaction through dual-branch attention. Class-wise sensitivity analysis reveals that geometric objects (automobile, ship) demonstrate superior fog robustness compared to organic subjects (horse, deer), establishing a fog challenge hierarchy patchy > adaptive ≈ gradient > uniform that provides valuable guidance for future research. With 14.6M trainable parameters and efficient convergence in 90 epochs, EDCST achieves a practical balance between performance and computational efficiency suitable for deployment in autonomous vehicles, surveillance systems, and adverse weather applications.

Future work should address evaluation on real-world foggy datasets to assess sim-to-real transfer, extension to higher resolutions for fine-grained detection, integration with multi-modal sensors for comprehensive weather robustness, and development of lightweight variants for edge deployment. This work demonstrates substantial progress in fog-robust computer vision while establishing clear directions for advancement toward weather-resilient AI systems in safety-critical applications.

REFERENCES

[1] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "We are ready for autonomous driving," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2147–2154, IEEE, 2012.

[2] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *European Conference on Computer Vision*, pp. 740–755, 2015.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

[3] Y. Cui, Z. Cao, Y. Xie, X. Jiang, F. Tao, Y. V. Chen, L. Li, and D. Liu, "DG-Labeler and DGL-MOTS dataset: Boost the autonomous driving perception," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3411–3420, 2022. doi: 10.1109/WACV51458.2022.00347.
[4] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, 2018.
[5] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.
[6] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. W. Chen, "Deep learning for image dehazing: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2071–2091, 2020.
[7] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," *arXiv preprint arXiv:1907.07484*, 2019.
[8] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *arXiv preprint arXiv:1903.12261*, 2019.
[9] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Fog density estimation and image defogging based on surrogate modeling for optical depth," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1571–1584, 2020.
[10] Y. Gao, D. Hendrycks, M. Mazeika, and J. Steinhardt, "Progressive difficulty curriculum learning for robotic grasping," *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3456–3463, 2024.
[11] C. Ancuti, C. O. Ancuti, C. Hermans, and P. Bekaert, "Effective single image dehazing by combining transmission and radiance," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5133–5144, 2016.
[12] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
[13] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
[14] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3194–3203, 2018.
[15] 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*, pp. 4770–4778, 2017.
[16] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8160–8168, 2019.
[17] C. Sakaridis, D. Dai, and L. Van Gool, "Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1768–1783, 2020.
[18] D. Dai, C. Sakaridis, S. Hecker, and L. Van Gool, "Curriculum domain adaptation for semantic nighttime image segmentation," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1296–1317, 2020.
[19] H. Zhang et al., "ResNeSt: Split-attention networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2736–2746, 2021.
[20] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10012–10022, 2021.
[21] C. Kamann and C. Rother, "Benchmarking the robustness of semantic segmentation models with respect to common corruptions," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 462–483, 2020.
[22] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International Conference on Machine Learning*, pp. 10347–10357, 2021.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

[23] L. Yuan, Y. Chen, T. Wang, W. Yu, Y. Shi, Z. Jiang, F. E. Tay, J. Feng, and S. Yan, "Tokens-to-token ViT: Training vision transformers from scratch on ImageNet," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 558–567, 2021.
[24] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.
[25] W. E. K. Middleton, *Vision through the atmosphere*, University of Toronto Press, 1952.
[26] H. Koschmieder, "Theorie der horizontalen sichtweite," *Beiträge zur Physik der freien Atmosphäre*, vol. 12, pp. 33–53, 1924.
[27] M. Negru, S. Nedevschi, and R. I. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2257–2268, 2016.
[28] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48, 2009.
[29] G. Hacohen and D. Weinshall, "On the power of curriculum learning in training deep networks," in *International Conference on Machine Learning*, pp. 2535–2544, 2019.
[30] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, "Curriculum learning: A survey," *International Journal of Computer Vision*, vol. 130, no. 6, pp. 1526–1565, 2022.
[31] S. Guo et al., "Curriculumnet: Weakly supervised learning from large-scale web images," in *Proceedings of the European Conference on Computer Vision*, pp. 135–150, 2018.
[32] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[33] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–41, 2022.
[34] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11976–11986, 2022.
[35] Z. Chen, Y. Duan, W. Wang, J. He, T. Lu, J. Dai, and Y. Qiao, "Vision transformer adapter for dense predictions," *arXiv preprint arXiv:2205.08534*, 2022.
[36] H. Touvron et al., "LLaMA: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.
[37] Y. Chen, X. Dai, D. Chen, M. Liu, X. Dong, L. Yuan, and Z. Liu, "EfficientFormer: Vision transformers at MobileNet speed," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12934–12949, 2024.
[38] J. Li, G. Li, and H. Fan, "Physics-guided deep learning for spatially correlated atmospheric haze removal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9850–9862, 2022.
[39] J. Zhang and D. Tao, "Atmospheric scattering-based multiple light source detection and image dehazing," *IEEE Transactions on Image Processing*, vol. 32, pp. 1895–1906, 2023.
[40] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 11908–11915, 2020.
[41] H. Wu et al., "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10551–10560, 2021.
[42] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Synthetic fog generation: A survey on theory, methods, and applications," *Computer Vision and Image Understanding*, vol. 208, 103209, 2021.
[43] K. Wang, X. Zhao, L. Zhang, and S. Li, "A physically constrained deep learning approach for atmospheric visibility estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.
[44] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "PSD: Principled synthetic-to-real dehazing guided by physical priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7180–7189, 2021.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

[45] C. Michaelis et al., "Benchmarking robustness in object detection under distribution shifts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9688–9704, 2023.
[46] C. Sakaridis, D. Dai, and L. Van Gool, "ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10765–10775, 2021.
[47] D. Dai, C. Sakaridis, S. Hecker, and L. Van Gool, "Dark model adaptation: Semantic image segmentation from daytime to nighttime," *International Journal of Computer Vision*, vol. 130, no. 2, pp. 559–576, 2022.
[48] D. Hendrycks, K. Zhao, S. Basart, J. Steinhardt, and D. Song, "Natural adversarial examples," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15262–15271, 2022.
[49] F. Croce et al., "Robustness of vision transformers to adversarial and natural distribution shifts," *arXiv preprint arXiv:2302.14267*, 2023.
[50] E. Mintun, A. Kirillov, and S. Xie, "Interaction networks for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9745–9754, 2021.
[51] H. Wang, C. Xiao, J. Kossaifi, Z. Yu, A. Anandkumar, and Z. Wang, "AugMax: Adversarial composition of random augmentations for robust training," *Advances in Neural Information Processing Systems*, vol. 34, pp. 237–250, 2022.
[52] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 11161–11179, 2023.
[53] G. Hacohen and D. Weinshall, "On the power of curriculum learning in training deep networks," in *International Conference on Machine Learning*, pp. 2535–2544, 2020.
[54] B. Zhou, X. Qiu, L. Chen, B. Zhang, W. Che, B. Zhou, and T. Liu, "Curriculum learning for natural language understanding," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 6095–6104, 2022.
[55] J. Liu, X. Zhang, H. Xiong, Q. Huang, X. Zhang, and P. S. Yu, "Automatic curriculum learning through value disagreement," *Advances in Neural Information Processing Systems*, vol. 36, pp. 48314–48328, 2023.
[56] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models with multi-task applications," *Journal of Machine Learning Research*, vol. 25, no. 42, pp. 1–47, 2024.
[57] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, et al., "Dynamic curriculum learning for image classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5067–5076, 2023.
[58] T. Zhang, Z. Xu, Y. Chen, and Z. Wang, "Adaptive curriculum learning via gradient matching," in *International Conference on Learning Representations*, 2024.
[59] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2021.
[60] M. Dehghani, J. Djolonga, B. Mustafa, P. Padlewski, J. Heek, J. Gilmer, A. P. Steiner, M. Caron, R. Geirhos, I. Alabdulmohsin, et al., "Scaling vision transformers to 22 billion parameters," in *International Conference on Machine Learning*, pp. 7480–7512, 2023.
[61] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, et al., "InternImage: Exploring large-scale vision foundation models with deformable convolutions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14408–14419, 2023.
[62] S. Paul and P.-Y. Chen, "Vision transformers are robust learners," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, pp. 2071–2081, 2022.
[63] S. Bhojanapalli, A. Chakrabarti, D. Glasner, D. Li, T. Unterthiner, and A. Veit, "Understanding robustness of transformers for image classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10231–10241, 2023.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38

[64] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
[65] Y. Song, Z. He, H. Qian, and X. Du, “Vision transformers for single image dehazing,” *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941, 2023.
[66] J. Li, G. Li, and H. Fan, “A comprehensive survey on image dehazing based on deep learning,” *Neurocomputing*, vol. 546, 126301, 2023.
[67] D. Kim, S. Woo, J.-Y. Lee, and I. S. Kweon, “Self-augmentation of weather dataset for improving classification with atmospheric degradation,” *arXiv preprint arXiv:2209.13012*, 2022.
[68] J. Jeong, S. Lee, J. Kim, and N. Kwak, “Consistency-based semi-supervised learning for object detection,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 3989–3997, 2017.
[69] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Realistic atmospheric scattering simulation for computer vision applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 3, pp. 1887–1900, 2024.
[70] R. Li, J. Pan, Z. Li, and J. Tang, “Single image dehazing via conditional generative adversarial network,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8202–8211, 2020.
[71] Y. Song, Z. He, H. Qian, and X. Du, “Dehamer: Large-scale multi-weather dataset for single image dehazing,” *arXiv preprint arXiv:2305.05654*, 2023.
[72] M. Zhang, L. Teck, S. Azimi, P. Rad, P. Poupart, J. Pineau, and G. Javadi, “Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11682–11692, 2019.
[73] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, “Pyramid vision transformer: A versatile backbone for dense prediction without convolutions,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 568–578, 2021.
[76] M. Liu, S. Wang, and Y. Zhang, "Time to shine: Fine-tuning object detection models with synthetic adverse weather images," Computer Vision and Image Understanding, vol. 241, 103921, 2025.
[75] Y. Li, X. Zhang, J. Wang, and H. Chen, "Gated image-adaptive network for driving-scene object detection under nighttime conditions," IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 2, pp. 1245-1258, 2025.
[74] A. Lengyel, S. Garg, M. Milford, and J. C. van Gemert, "Zero-shot day-night domain adaptation with a physics prior," in Proceedings of the IEEE International Conference on Computer Vision, pp. 4305–4315, 2021.

DOI: https://doi.org/10.14232/analecta.2025.3-4.15-38