PROCEEDINGS OF THE YEREVAN STATE UNIVERSITY

Physical and Mathematical Sciences 2026, **60**(1), p. 63–82

Informatics

KNOWLEDGE-DISTILLED VARIATIONAL BAYESIAN FRAMEWORK FOR EFFICIENT LARGE-SCALE IMAGE DEHAZING

S. V. BABAYAN *

National Polytechnic University of Armenia (NPUA), Armenia

Real-time processing of large-scale image databases with state-of-the-art dehazing methods presents a significant computational challenge: methods that achieve superior generalization typically require substantial inference time, limiting their deployment in real-time applications. High-performing methods employ complex multi-scale processing and deep architectures, typically achievin less than 5 FPS on high-resolution images.

Building on the preliminary multi-scale variational Bayesian framework [1, 2], which achieves strong synthetic-to-real generalization, this paper proposes knowledge distillation to transfer the generalization capabilities of high-performance models to a lightweight Vision Transformer-based student network. The student leverages patch-based processing and reduced architectural complexity to achieve over 150× speedup, while maintaining competitive performance through a theoretically-grounded distillation framework integrated into the variational Bayesian objective.

Additionally, the atmospheric scattering model is extended to estimate space-variant atmospheric light, improving performance on varying haze regions. Trained solely on synthetic Haze4K data, the proposed method stays competitive on synthetic-to-real generalization and downstream object detection (on the augmented KITTI dataset) tasks, while achieving superior inference speed for large-scale real-world applications.

https://doi.org/10.46991/PYSUA.2026.60.1.063

MSC2020: Primary: 68T45; Secondary: 68T07.

Keywords: dehazing, knowledge distillation, deep variational Bayes, atmospheric scattering model, large-scale image processing.

Introduction. The increasing demand for processing large-scale image databases and videos in applications such as aerial imaging, surveillance, and autonomous systems has created a need for efficient real-time dehazing methods. Such applications impose strict requirements: video and streaming workloads need fast inference and parallelization to maintain throughput, so efficiency must be evaluated on

* E-mail: stepanbabayan.h755-3@polytechnic.am

high-resolution inputs. The resolution 3840×2160 (4K UHD) is the standard for high-resolution and 4K video in these domains; we therefore use high-resolution and 4K-scale images to assess inference speed and to motivate efficient, parallelization-friendly architectures. State-of-the-art (SOTA) dehazing methods that achieve superior generalization (including the preliminary multi-scale variational Bayesian framework [1], CasDyF-Net [3], GridFormer [4], and others) employ complex multi-scale processing, deep architectures, and extensive feature interactions to achieve robust generalization. For instance, processing large images at 3840×2160 with these methods typically requires 0.2–10 s per frame (0.1–5 FPS). While this performance may be acceptable for offline processing, real-time applications often require 30 + FPS, creating a significant need for more efficient architectures that maintain the generalization capabilities of high-performance methods.

The computational challenge stems from fundamental architectural characteristics shared by SOTA methods:

1) multi-scale feature processing that operates at multiple resolutions simultaneously;

2) complex network architectures with deep feature hierarchies,

3) extensive feature interactions through attention mechanisms and adaptive filtering.

Existing approaches addressing efficiency – such as model pruning, quantization, or architectural simplifications – typically sacrifice generalization performance [5, 6]. This, along with the absence of large-scale real dehazing datasets, makes synthetically trained methods impractical for real-world haze removal.

This work addresses the large-scale inference problem through knowledge distillation, a theoretically-grounded approach that transfers the generalization capabilities of a high-performance teacher model to a lightweight student network. The teacher network extends the multi-scale variational Bayesian framework [1] by adding a Gaussian output layer that enables probabilistic sampling, providing uncertainty-aware predictions that facilitate better generalization. The key insight is that the student can benefit from the teacher's probabilistic outputs, encoded in the variational posterior, to improve generalization, rather than merely matching deterministic outputs. This also serves as a dataset augmentation step for the student network, while directing its predictions around the mean of the Gaussian distribution (the output of the variational framework). Knowledge distillation is formulated within the variational Bayesian framework by extending the Evidence Lower Bound (ELBO) to include a distillation term that minimizes the Kullback-Leibler divergence between student and teacher variational distributions. Unlike prior Bayesian perspectives on KD [7, 8], which operate without physical constraints, the proposed framework transfers distributional knowledge from a probabilistic teacher, whose uncertainty arises from a principled variational treatment of haze physics. The student simultaneously minimizes the distillation loss and the full physical ELBO, coupling ASM consistency with probabilistic supervision.

The atmospheric scattering model (ASM) [9] formulates the observed hazy image I as:

$$I = J \odot t + A \odot (1 - t), \tag{1}$$

where J is the clean image, t is the transmission map, A is the atmospheric light, and $\odot$ is the element-wise multiplication operator. Traditional formulations assume A to be uniform across the scene, which limits effectiveness in real-world scenarios with spatial variations, particularly in natural conditions with varying illumination. This work also extends the ASM to estimate per-pixel atmospheric light $A(x)$, further improving performance on dense haze regions.

The contributions of this work are:

1. A first-of-its-kind knowledge distillation framework that is integrated into a physics-grounded variational Bayesian dehazing objective, enabling efficient large-scale inference, while preserving the generalization capabilities of strong probabilistic teacher models.

2. The proposed work further extends the ASM to estimate space-variant atmospheric light $A(x)$ on a per-pixel basis, improving performance on dense haze regions seen in real-world conditions.

3. Comprehensive evaluation is performed, demonstrating competitive synthetic-to-real generalization on multiple benchmarks while achieving real-time inference capability, along with significant improvements in downstream object detection, establishing practical utility for large-scale applications.

The rest of the paper is organized as follows. “Revated Work” Section reviews related work. “Variational Bayesian Framework” Section presents the initial variational Bayesian framework and its network architecture. “Knowledge Distillation for Large-Scale Efficiency” Section explains the knowledge distillation for efficient large-scale inference and space-variant atmospheric light estimation. The proposed method is evaluated on both synthetic and real-world datasets in “Experiments” Section. Finally, “Conclusion” Section concludes the paper.

Related Work.

Traditional Prior-Based Methods. Classical dehazing approaches exploit statistical properties commonly observed in natural images to solve the underconstrained inverse problem of recovering clean images from hazy observations. The dark channel prior (DCP) [10] observes that outdoor haze-free patches contain near-zero intensities in at least one color channel, enabling intermediate transmission map estimation. Alternative formulations include the color attenuation prior [11], which models saturation-brightness relationships in hazy regions, and the haze-lines method [12], which exploits color clustering properties. These optimization-based methods offer interpretability, but require per-image iterative solutions and rely on assumptions that often fail under dense haze, non-uniform illumination, or sensor noise – conditions prevalent in real-world large-scale imagery.

Deep Learning-Based Methods. Deep learning has transformed image dehazing through data-driven feature learning (see [13] for a comprehensive survey). Early physics-informed networks estimate ASM parameters: DehazeNet [14] predicts transmission maps, while AOD-Net [15] jointly estimates transmission and atmospheric light in a unified architecture. Recent end-to-end approaches directly learn $I \to J$ mappings: FFA-Net [16] employs feature fusion attention for non-uniform haze, ConvIR [17] and FSNet [18] leverage multi-scale CNN architectures, DEA-Net [19] enhances convolution with detail priors and content-guided attention, GridFormer [4] uses residual dense transformers in a grid structure for multi-weather tasks including dehazing, and CasDyF-Net [3] uses cascaded dynamic filters for adaptive frequency processing. While these methods achieve strong synthetic benchmark performance, the absence of physical constraints often results in degraded synthetic-to-real generalization – a critical limitation for practical deployment in large-scale real-world applications.

Knowledge Distillation for Dehazing. Knowledge distillation has emerged as an effective technique for model compression and acceleration in computer vision tasks. In dehazing, recent works have explored distillation to reduce computational costs while maintaining performance. Hong et al. [20] distil image dehazing with heterogeneous task imitation, where the student learns from a teacher trained on a related task. KTDN [21] proposes knowledge transfer for non-homogeneous dehazing through feature imitation from a pre-trained teacher network trained on clear images. Yet, it focuses on addressing non-homogeneous haze rather than computational efficiency. SDDN [22] develops lightweight dehazing networks based on soft knowledge distillation, achieving compact models with only 2.1M parameters. LOS-DehazeNet [23] introduces omni-scale feature learning for lightweight dehazing with extremely low parameter counts. Beyond dehazing, Bayesian perspectives on KD have been studied by BKD [7] and VBGS-KD [8], mainly in classification and sparsification settings without a physical image-formation model. While existing efficiency-focused distillation methods achieve fast inference, they do not benefit from distilling knowledge from high-performing probabilistic frameworks, limiting their generalization capabilities. This work extends knowledge distillation to variational Bayesian dehazing, enabling efficient processing of large-scale images through a lightweight Vision Transformer-based student network, while maintaining the generalization capabilities of the probabilistic teacher model.

Space-Variant Atmospheric Light. Traditional dehazing methods assume uniform atmospheric light across the scene, which limits their effectiveness in real-world scenarios with varying illumination conditions. Recent works have explored space-variant formulations to address this limitation. Multi-stage approaches [24, 25] improve the atmospheric light estimation and decompose the dehazing process into stages that handle dehazing iteratively. However, these methods often require complex optimization procedures for each frame and may not integrate seamlessly with end-to-end deep learning frameworks. The proposed approach extends the ASM to estimate per-pixel atmospheric light $A(x)$ within a unified variational Bayesian framework, enabling spatial adaptation while maintaining physical consistency with the scattering model.

Efficiency for Large-Scale Processing. The computational challenge of processing large-scale images has received limited attention in dehazing literature. Most existing methods are evaluated on popular resolutions widely used in the literature for ease of processing (e.g., 640×480, 512×512, or 1024×768), where computational cost is manageable. However, for aerial imaging, surveillance, and autonomous driving applications, processing large images (3840×2160 or larger) requires $16\times$ more pixels than standard-resolution images, significantly increasing computational demands. Recent state-of-the-art works in ground image and aerial dehazing [1, 3, 26] have addressed domain-specific challenges, but often lack the efficiency required for real-time deployment. Model compression techniques such as pruning and quantization typically sacrifice generalization performance. Knowledge distillation offers a principled alternative by transferring knowledge from a powerful teacher to an efficient student. Vision Transformers have shown promise for large-scale image processing due to their patch-based architecture, which processes fixed-size patches regardless of input resolution, and their unified design that avoids multi-scale processing overhead. Recent efficient ViT architectures, such as EfficientViT [27], address memory inefficiency through sandwich layouts and cascaded group attention, achieving 40.4% higher GPU throughput and 45.2% higher CPU throughput compared to MobileNetV3-Large [28], with up to $5.8\times$ speedup over MobileViT [29] variants on V100 GPUs. The proposed work is the first to combine knowledge distillation with variational Bayesian frameworks for dehazing, enabling efficient large-scale processing, while preserving the teacher's generalization capabilities.

Probabilistic and Variational Approaches. Probabilistic methods model haze parameters as distributions rather than point estimates, enabling more robust generalization. UDN [30] decomposes uncertainty into aleatoric (data) and epistemic (model) components for adaptive feature refinement. The approach in [31] extends uncertainty modeling to spatially-varying transmission maps for non-homogeneous haze. Variational Bayesian frameworks [32] derive principled objectives from the ASM, accounting for transmission and radiance variability. Prior work [1] established a multi-scale variational formulation with input-conditioned convolutions achieving state-of-the-art synthetic-to-real generalization. This paper extends that framework through knowledge distillation, enabling an efficient student to benefit from the teacher's probabilistic outputs, and space-variant atmospheric light estimation, enabling real-time large-scale processing, while preserving generalization capabilities.

Variational Bayesian Framework. This work builds upon the multi-scale variational Bayesian framework [1], which formulates the dehazing problem within a Bayesian context using the ASM. The joint probability distribution is defined as:

$$\mathrm{p}(J,I,t,A) = \mathrm{p}(J)\mathrm{p}(t)\mathrm{p}(A)\mathrm{p}(I|J,t,A), \tag{2}$$

where J is the haze-free image, t is the transmission map, A is the atmospheric light, and I is the observed hazy image. Improper flat priors $\mathrm{p}(t) \propto \mathit{const}$ and $\mathrm{p}(A) \propto \mathit{const}$ are used for t and A, while the prior on the haze-free image $\mathrm{p}(J)$ combines three multi-scale components:

$$p(J) \propto \exp\left\{-\lambda_q \mathcal{L}_q(J) - \lambda_l \mathcal{L}_l(J) - \lambda_F \mathcal{L}_F(J)\right\}, \tag{3}$$

where $\mathcal{L}_q(J) = \sum_{s=0}^{2} \|J^{(s)} - J_{gt}^{(s)}\|^2$ is a multi-scale quadratic term for noise awareness, $\mathcal{L}_l(J) = \sum_{s=0}^{2} |J^{(s)} - J_{gt}^{(s)}|_1$ enforces spatial consistency, and $\mathcal{L}_F(J) = \sum_{s=0}^{2} |\mathcal{F}(J^{(s)}) - \mathcal{F}(J_{gt}^{(s)})|_1$ preserves frequency-domain fidelity with $J^{(s)}$ denoting the image downsampled by 2^s and $\mathcal{F}(\cdot)$ the Fast Fourier Transform.

The likelihood term ensures consistency with the ASM:

$$p(I|J,t,A) = \mathcal{N}(I|J \odot t + A \odot (1-t), \lambda_\eta^{-1}\mathbf{I}), \tag{4}$$

where λ_η is the precision parameter and $\mathbf{I}$ is the identity matrix.

The posterior $p(J,t,A|I)$ is approximated using variational inference with deterministic (Dirac delta) variational distributions $q^{\theta_J}(J|I) = \delta(J - J^{\theta_J}(I))$, $q^{\theta_t}(t|I) = \delta(t - t^{\theta_t}(I))$, and $q^{\theta_A}(A|I) = \delta(A - A^{\theta_A}(I))$, where $J^{\theta_J}(I)$, $t^{\theta_t}(I)$, and $A^{\theta_A}(I)$ are the outputs of three networks JNet, TNet, and ANet, respectively. Note that while the base framework [1] employs deterministic variational distributions, the teacher network for knowledge distillation extends JNet with a Gaussian output layer to enable probabilistic sampling, as detailed in "Knowledge Distillation for Large-Scale Efficiency" Section.

Maximizing the Evidence Lower Bound (ELBO) yields the training objective:

$$\min_\theta \sum_{I \in \mathcal{I}} \left[\mathcal{L}_{LL}(I^\theta(I)) + \lambda_q \mathcal{L}_q(J^{\theta_J}(I)) + \lambda_l \mathcal{L}_l(J^{\theta_J}(I)) + \lambda_F \mathcal{L}_F(J^{\theta_J}(I))\right], \tag{5}$$

where $\mathcal{L}_{LL}(I^\theta(I)) = \lambda_\eta \|I - I^\theta(I)\|^2/2$ enforces ASM consistency with $I^\theta(I) = J^{\theta_J}(I) \odot t^{\theta_t}(I) + A^{\theta_A}(I) \odot (1 - t^{\theta_t}(I))$, and λ_η, λ_q, λ_l, λ_F are hyperparameters. For a detailed explanation of the variational modeling and inference (see [1]).

Network Architecture. The base variational framework employs three network modules: JNet estimates the haze-free image J, TNet estimates the transmission map t, and ANet estimates the atmospheric light A. JNet is based on CasDyF-Net [3], enhanced with a Dynamic Filter Network (DFN) [33] for adaptive refinement, TNet follows the GCANet architecture [34] for transmission map estimation, and ANet initially applies global average pooling followed by a linear layer for uniform atmospheric light estimation. The proposed approach extends ANet to estimate space-variant atmospheric light $A(x)$ on a per-pixel basis through a lightweight four-layer convolutional module that outputs a spatial map (see "Space-Variant Atmospheric Light Estimation" Section). During training, all three networks are optimized jointly using (5). At inference, only JNet is required for efficient dehazing.

Knowledge Distillation for Large-Scale Efficiency. This section describes the two components of the distilled system: the **teacher model**, the high-capacity variational framework that serves as the source of knowledge, and the **student model**, an efficient network trained via distillation to approximate the teacher. The teacher is introduced first; the student network, which is the main contribution for efficient deployment, is then trained on top of the frozen teacher. The computational requirements for processing large-scale images with high-performance dehazing methods are substantial. For high-resolution images, such as of size 3840×2160 or larger, state-of-the-art methods typically require 4,000–33,000 GFLOPs and 9–62 *GB* memory, resulting in 0.1–1 FPS on modern GPUs. The computational cost comes from:

1) multi-scale processing that operates at multiple resolutions simultaneously;

2) deep network architectures with extensive feature hierarchies;

3) complex feature interactions through attention and adaptive filtering.

To enable real-time processing while maintaining generalization capabilities, knowledge distillation is employed to transfer the generalization capabilities of high-accuracy models to a lightweight student network.

Teacher Network. The teacher is constructed in two steps. First, the full multi-scale variational Bayesian framework [1] (JNet based on CasDyF-Net with DFN, TNet based on GCANet, and ANet) is trained jointly using the objective in (5), achieving strong synthetic-to-real generalization. The trained framework is then frozen, and a Gaussian output layer is added to JNet that enables probabilistic sampling. The Gaussian output layer transforms JNet's deterministic predictions into a probabilistic distribution, where the network output serves as the mean and samples are drawn from a Gaussian with empirically selected variance. This enables the teacher's variational posterior $q^{\theta_J^T}(J|I)$ to encode uncertainty-aware predictions that facilitate robust synthetic-to-real generalization. While the teacher requires $4,890$ GFLOPs and achieves 0.5 FPS on large images, comparable to other high-performance methods, its superior generalization makes it a suitable candidate for knowledge distillation.

Student Network The student employs a lightweight Vision Transformer (ViT) architecture based on EfficientViT [27], designed for memory-efficient large-scale processing. The key efficiency advantages stem from:

1) a sandwich layout that uses a single memory-bound multi-head self-attention (MHSA) layer between efficient feed-forward network (FFN) layers, reducing memory access overhead from frequent tensor reshaping and element-wise operations;

2) cascaded group attention (CGA) that feeds each attention head with different feature splits, reducing computational redundancy, while improving attention diversity;

3) a three-stage sequential architecture with overlapping patch embedding that efficiently processes images at multiple scales;

4) optimized parameter allocation that expands critical components (e.g., value projections), while reducing redundant dimensions.

The architecture consists of:

1) an overlapping patch embedding layer that embeds input patches into tokens;

2) three stages of EfficientViT blocks with sandwich layout and CGA, progressively reducing spatial resolution, while increasing channel dimensions;

3) a lightweight decoder that reconstructs the dehazed image.

This design achieves significant efficiency gains: the student requires only 54 GFLOPs (99% reduction) and 1.7 *GB* memory (92% reduction) compared to the teacher, enabling real-time large-scale processing at 76 FPS – over 150× speedup that reflects EfficientViT's demonstrated efficiency advantages.

Distillation Framework. The key insight enabling the student to maintain competitive performance despite 99% computational reduction is that standard knowledge distillation transfers only point estimates, whereas the proposed probabilistic distillation transfers distributional knowledge that captures the teacher's learned uncertainty structure. Knowledge distillation is formulated within the variational Bayesian framework by extending the ELBO. The student's variational posterior $q^{\theta_J^S}(J|I)$ should approximate the teacher's $q^{\theta_J^T}(J|I)$, enforced through a KL divergence term:

$$\mathcal{L}_{KD}(I) = \mathrm{KL}\left(q^{\theta_J^S}(J|I) \parallel q^{\theta_J^T}(J|I)\right). \tag{6}$$

Since the teacher employs a Gaussian distribution and the student uses a deterministic (Dirac delta) distribution, this KL divergence can be approximated following standard variational inference principles by matching the student's output to both samples drawn from the teacher's Gaussian distribution and ground truth images, with an additional regularization term that encourages parameter efficiency:

$$\begin{aligned}\mathcal{L}_{KD}(I) =& \lambda_{MSE} \|J^{\theta_J^S}(I) - \tilde{J}^{\theta_J^T}(I)\|^2 \\ &+ \lambda_{GT} \|J^{\theta_J^S}(I) - J_{gt}\|^2 + \lambda_{reg} \mathcal{R}(\theta_J^S),\end{aligned} \tag{7}$$

where $\tilde{J}^{\theta_J^T}(I) \sim \mathcal{N}(J^{\theta_J^T}(I), \sigma^2)$ denotes a sample drawn from the teacher's Gaussian distribution (with mean $J^{\theta_J^T}(I)$ being the deterministic output of the teacher's JNet and variance $\sigma^2 = 0.05$), $\mathcal{R}(\theta_J^S)$ is a regularization term encouraging parameter efficiency to avoid overfitting, and λ_{MSE}, λ_{GT}, and λ_{reg} are balancing hyperparameters. The theoretical objective for the student network, formulated within the variational Bayesian framework, is:

$$\begin{aligned}\min_{\theta_J^S} \sum_{I \in \mathcal{I}} \Big[& \mathcal{L}_{LL}(I^{\theta^S}(I)) + \lambda_q \mathcal{L}_q(J^{\theta_J^S}(I)) + \lambda_l \mathcal{L}_l(J^{\theta_J^S}(I)) \\ & + \lambda_F \mathcal{L}_F(J^{\theta_J^S}(I)) + \lambda_{KD} \mathcal{L}_{KD}(I) \Big],\end{aligned} \tag{8}$$

where λ_{KD} controls the distillation strength. In practice, only the student parameters θ_J^S are optimized; the teacher parameters θ_J^T remain frozen throughout student training. The ELBO terms ($\mathcal{L}_{LL}$, $\mathcal{L}_q$, $\mathcal{L}_l$, $\mathcal{L}_F$) are computed using the student's outputs $J^{\theta_J^S}(I)$, while the distillation loss $\mathcal{L}_{KD}$ uses samples $\tilde{J}^{\theta_J^T}(I)$ drawn from the frozen teacher's Gaussian distribution.

Probabilistic Distillation for Efficient Knowledge Transfer. Four complementary mechanisms explain how the student maintains 98% of teacher PSNR on Haze4k despite 99% computational reduction:

1) *Implicit data augmentation*: sampling from the teacher's Gaussian distribution generates diverse training targets around the teacher's prediction, effectively multiplying the training set and exposing the student to variations that improve robustness to domain shift – critical for synthetic-to-real generalization.

2) *Soft label smoothing*: unlike hard ground truth labels, the probabilistic targets provide smoother supervision that prevents overconfident predictions, enabling better generalization from the compact student architecture.

3) *Physical consistency preservation*: crucially, the student is not trained solely on distillation loss, but also on the full ELBO objective including $\mathcal{L}_{LL}$ (ASM consistency), $\mathcal{L}_q$, $\mathcal{L}_l$, and $\mathcal{L}_F$ (multi-scale priors). This ensures the student learns physically-grounded representations rather than merely mimicking teacher outputs, preserving the teacher's generalization capabilities even with drastically reduced capacity.

4) *Efficient architecture design*: EfficientViT's architectural innovations are specifically designed to maximize representational capacity per FLOP. The sandwich layout eliminates memory-bound bottlenecks by reducing redundant tensor reshaping operations in self-attention, while cascaded group attention (CGA) addresses attention head redundancy – a critical inefficiency, where standard multi-head attention computes highly similar attention maps across heads. By feeding each head with different feature splits and cascading outputs, CGA increases effective network depth without additional parameters, enabling the compact student to learn diverse, complementary feature representations. Combined with optimized parameter allocation that expands critical components (value projections), while shrinking redundant ones (Q/K dimensions, FFN expansion), the architecture achieves high representational efficiency that complements the rich supervision from probabilistic distillation. The ablation study validates this design: removing probabilistic sampling degrades O-Haze performance from 18.83 *dB* to 18.6 *dB*, confirming that distributional knowledge transfer is essential for maintaining generalization under extreme compression.

Space-Variant Atmospheric Light Estimation. The ASM is enhanced to estimate per-pixel atmospheric light $A(x)$ instead of a uniform A, better addressing spatial variations in real-world scenarios. The ASM becomes

$$I(x) = J(x) \odot t(x) + A(x) \odot (1 - t(x)).$$

ANet is enhanced by replacing global average pooling with a lightweight spatial feature extraction module that directly processes the input hazy image I to produce a per-pixel atmospheric light map $A(x)$. The module consists of four layers: a dilated convolutional layer with ReLU activation, a standard convolutional layer with ReLU activation, a batch normalization layer with ReLU activation, and a final convolutional layer with sigmoid activation that outputs a three-channel smooth RGB map with the same spatial dimensions as the input. This simple architecture enables the network to capture local atmospheric light variations, while maintaining computational efficiency. The reconstruction term in (5) is updated to use $A^{\theta_A}(x)$ instead of the scalar A^{θ_A}. This formulation improves performance on varying haze regions, particularly on NH-Haze.

Experiments. A comprehensive set of experiments was conducted to evaluate the proposed knowledge distillation framework with emphasis on large-scale inference efficiency. The following describes the experimental setup and the configurations used; the reported numbers were obtained under this setup. Experiments assess:

1) synthetic-to-real generalization;

2) knowledge transfer effectiveness from teacher to student;

3) computational performance on large images (measured on O-Haze and I-Haze, with resolutions exceeding 4K);

4) component contributions via ablation.

Datasets. Training leverages the Haze4K dataset [35] (3,000 training / 1,000 validation pairs), with generalization evaluated on three real-world benchmarks: O-Haze [36] (outdoor scenes, 2833×4657 pixels), I-Haze [37] (indoor scenes, 2833×4657 pixels), and NH-Haze [38] (non-homogeneous haze, 1600×1200 pixels). These datasets provide a wide variety of scenes and conditions for evaluating both generalization capability and computational efficiency on large-scale images. The NH-Haze dataset presents spatially varying haze densities, also making it well-suited to evaluate the proposed space-variant atmospheric light estimation.

Training Protocol. A two-stage knowledge distillation training strategy is employed, implemented in PyTorch on a single NVIDIA A100 GPU.

Stage 1 (Teacher). The full multi-scale variational Bayesian framework (JNet, TNet, ANet) is trained jointly for 1000 epochs using Adam (betas 0.9, 0.999) with cosine annealing from $4 \cdot 10^{-4}$ to $1 \cdot 10^{-6}$, batch size 8, and loss weights $\lambda_\eta = 0.02$, $\lambda_q = \lambda_l = 1$, $\lambda_F = 0.1$. Data augmentation includes random horizontal flips and 256×256 crops. The trained framework is then frozen, and a Gaussian output layer is added to JNet with variance parameter empirically selected as 0.05.

Stage 2 (Student). The EfficientViT-based student is trained for 1500 epochs using the frozen teacher's probabilistic outputs, with distillation weight $\lambda_{KD} = 1$, distillation loss weights $\lambda_{MSE} = 1$ and $\lambda_{GT} = 0.2$, and regularization $\lambda_{reg} = 0.01$. Other hyperparameters for training the student network are the same as those of the variational framework. The student learns from both ground truth supervision and teacher predictions sampled from the Gaussian posterior, effectively combining standard training with uncertainty-aware distillation.

Quantitative Comparison. Evaluation uses PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index), standard full-reference metrics for image quality in restoration, comparing against high-performance methods (Base [1], CasDyF-Net [3], ConvIR [17], FSNet [18], GridFormer [4], DEA-Net [19], FFA-Net [16]). In all quantitative comparison tables except the efficiency comparison, all methods use author-provided weights trained on Haze4K to ensure fair comparison and avoid a reproduction bias. Tab. 1 presents results across all benchmarks. The teacher network degrades performance on the synthetic Haze4k dataset, compared to the Base network [1]. However, it enhances the results on real-world outdoor datasets (O-Haze and NH-Haze), where the space-variant atmospheric light estimation becomes decisive. The student network achieves competitive synthetic performance, while demonstrating strong generalization to real-world datasets. Notably, the student achieves the best performance on NH-Haze ($12.35\ dB$), consistent with space-variant atmospheric light estimation being effective for non-homogeneous haze. The key distinction is that the proposed method maintains this generalization while operating at 76 FPS – over $150\times$ faster than comparable high-performance methods. Efficiency-focused knowledge distillation approaches (SDDN [22], LOS-DehazeNet [23]) are included in the computational efficiency analysis (Tab. 2), but excluded from the quantitative and qualitative comparisons that focus on the synthetic-to-real generalization due to the unavailability of Haze4K-trained weights, which would compromise the fairness of the evaluation.

Table 1

*Comparison of the proposed and state-of-the-art methods trained on synthetic and real-world datasets. Teacher and Student denote the source model and the distilled (efficient) model, respectively. Each cell reports PSNR / SSIM. The highest value is in **bold**; the second-highest value is underscored*

Method	Haze4k (synthetic)	O-Haze (real)	I-Haze (real)	NH-Haze (real)
Teacher	34.93 / 0.991	**18.93** / 0.818	16.74 / 0.773	12.29 / 0.582
Student	34.39 / 0.989	18.83 / **0.822**	16.67 / 0.785	**12.35** / **0.588**
Base [1]	35.28 / 0.991	18.91 / 0.815	16.80 / 0.779	12.26 / 0.576
CasDyF-Net	**35.71** / **0.992**	17.79 / 0.744	15.44 / 0.706	11.43 / 0.531
ConvIR	34.50 / 0.990	17.73 / 0.810	**16.89** / **0.791**	11.75 / 0.565
DEA-Net	34.26 / 0.989	17.72 / 0.694	16.37 / 0.743	11.73 / 0.540
FSNet	34.12 / 0.990	18.45 / 0.806	15.88 / 0.755	12.02 / 0.556
GridFormer	33.27 / 0.986	18.51 / 0.684	15.36 / 0.720	11.79 / 0.518
FFA-Net	26.96 / 0.950	17.06 / 0.763	14.48 / 0.687	11.66 / 0.514

Knowledge Transfer Effectiveness. The distillation results merit detailed analysis: the student retains 98% of teacher PSNR on Haze4k, while achieving over 150× throughput improvement. Compared to existing SOTA methods and efficiency-focused KD dehazing methods – SDDN (112 GFLOPs, 41 FPS) and LOS-DehazeNet (72 GFLOPs, 63 FPS) – the proposed approach achieves competitive accuracy while operating at a higher speed (54 GFLOPs, 76 FPS). This advantage stems from the combination of probabilistic distillation and physics-informed training: while conventional KD methods transfer only deterministic mappings, the proposed framework transfers distributional knowledge through Gaussian sampling, while simultaneously enforcing ASM consistency and multi-scale priors on the student. This dual supervision – probabilistic teacher guidance combined with physical constraints – enables the compact student to learn robust, generalizable representations rather than overfitting to teacher outputs, explaining the strong synthetic-to-real transfer despite extreme compression.

Qualitative Analysis. Figure presents visual comparisons on real-world images from O-Haze and I-Haze, alongside synthetic Haze4K samples. All methods have been trained exclusively on the synthetic Haze4K dataset, making this comparison a direct evaluation of synthetic-to-real generalization capability. Several failure modes emerge among competing methods: FFA-Net and DEA-Net introduce color artifacts; FSNet over-enhances contrast, causing clipping in highlights and shadows; CasDyF-Net struggles with dense haze regions, leaving residual haze or producing minimal improvement (rows 2, 5). ConvIR and GridFormer perform adequately on moderate haze but underperform in challenging dense-haze scenarios. The teacher network produces outputs with sharper edge preservation and more visually coherent reconstructions. The distilled student network approaches the performance of the teacher network and preserves fine structural details while effectively separating atmospheric light from scene content – a capability inherited from the teacher's probabilistic framework. Despite the 99% reduction in computational cost, the student maintains visual quality comparable to the teacher, demonstrating successful knowledge transfer for real-world dehazing.

Qualitative comparison of dehazing results on real-world images from O-Haze and I-Haze datasets and Haze4k synthetic dataset. All competing methods have been trained on the synthetic Haze4k dataset. The IN and GT columns correspond to the observed and ground-truth images, respectively.

Efficiency Analysis for Large-Scale Inference. The primary motivation for this work is enabling real-time processing of large-scale images. Efficiency is evaluated by measuring inference speed and computational cost on high-resolution images. Tab. 2 reports these metrics on images from the O-Haze and I-Haze benchmarks, which have native resolutions up to 2833×4657 pixels – larger than 4K UHD (3840×2160) – thus stressing inference under large-scale conditions. The table compares FPS (frames per second), FLOPs (floating point operations), memory usage, and parameter counts across methods. The key finding is that the student network achieves over $150\times$ speedup (76 FPS vs. 0.5 FPS) compared to the full teacher model, reducing computational cost from $4,890$ GFLOPs to 54 GFLOPs (99% reduction) and memory from 20.84 *GB* to 1.7 *GB* (92% reduction). This enables real-time processing: at 76 FPS, the student can process high-resolution video streams in real-time, exceeding the requirements for real-time applications (30+ FPS). Note that during inference, only JNet is utilized. Thus, the teacher and the base model [1] have the same measures, and only the teacher model is presented. The student also outperforms all other methods in terms of speed while maintaining accuracy that approaches the teacher's within an efficient architecture.

Table 2

Efficiency comparison on O-Haze and I-Haze images (resolutions up to 2833×4657). FPS measured on NVIDIA A100. Methods marked with † are KD/efficiency-focused

Method	FPS	FLOPs, *G*	Mem., *GB*	Param., *M*
Teacher	0.5	4,890	20.84	6.25
CasDyF-Net	0.55	4,730	20.33	6.05
ConvIR	0.5	15,900	13.28	14.83
DEA-Net	0.8	4,077	9.47	3.65
GridFormer	0.1	31,730	61.9	30.2
FSNet	0.6	13,500	29.8	13.28
FFA-Net	0.12	33,400	18.3	4.46
Student	**76**	**54**	<u>1.7</u>	2.3
SDDN†	41	112	2.4	<u>0.56</u>
LOS†	<u>63</u>	<u>72</u>	**1.3**	**0.11**

Downstream Task Evaluation. To demonstrate the practical utility of the dehazing method, its impact on object detection is evaluated using the KITTI Object Detection Benchmark [39]. A random sample of 200 images from the testing set is used, with synthetic haze generation following to create hazy conditions that simulate real-world atmospheric degradation. To dehaze the generated hazy images, all methods have been employed with their weights, trained on the synthetic Haze4k dataset. The pre-trained YOLOv8n [40] object detection model is used for inference without fine-tuning, comparing detection performance on hazy input images versus dehazed outputs from different methods. This evaluation protocol directly assesses whether dehazing improves detection accuracy in degraded conditions, which is critical for autonomous driving and surveillance applications, where haze can significantly impair object detection.

Tab. 3 reports the average number of detections per image (Avg Det), the percentage improvement over hazy input (Improv. vs IN), and the total number of detections across all 200 images. The results reveal that several methods (FSNet, CasDyF-Net) actually degrade detection performance compared to hazy input, likely due to artifacts or over-enhancement that confuse the detection model. In contrast, both the teacher (+13.9% improvement, 5.58 avg detections) and student (+12.4% improvement, 5.51 avg detections) achieve the highest detection improvements among all dehazing approaches, with the teacher ranking first and the student ranking second. Notably, the teacher network outperforms the base model [1] in this task, by better adjusting to real-world scenarios thanks to the space-variant atmospheric light. The student maintains 89% of the teacher's detection improvement while operating at 76 FPS, demonstrating that the efficiency gains do not compromise practical utility. The performance gap to clean ground truth images (5.79 avg detections, +18.2%) indicates room for further improvement, but the substantial gains over hazy input suggest practical value of the proposed approach for real-world applications where haze degradation is prevalent.

Table 3

Object detection performance comparison on KITTI dataset

Method	Avg Det	Improv. vs IN (%)	Total Det
IN (Hazy)	4.90	0.0	980
FSNet	3.86	−21.3	771
CasDyF-Net	4.77	−2.7	954
ConvIR	4.92	0.3	983
FFA-Net	5.40	10.1	1079
GridFormer	5.42	10.6	1084
DEA-Net	5.44	10.9	1087
Base [1]	5.49	12.0	1098
Teacher	**5.58**	**13.9**	**1116**
Student	<u>5.51</u>	<u>12.4</u>	<u>1101</u>
GT (Clean)	5.79	18.2	1158

Ablation Study. To assess the impact of individual components in the proposed dehazing framework, an ablation study with multiple configurations is performed. The following variants are trained:

1) the full student model with all components (KD + space-variant A + Gaussian output layer);

2) the teacher network (full variational framework without knowledge distillation);

3) the student trained with teacher without Gaussian output layer (St. w/o Prob.);

4) the student model with constant atmospheric light (space-invariant A).

These setups are evaluated on all four datasets, with PSNR/SSIM reported in Tab. 4.

Tab. 4 quantifies each component's contribution to the full system. The teacher and the base networks achieve overall higher accuracy, but at 0.5 FPS, they are unsuitable for real-time large-scale processing. Knowledge distillation allows the student to capture much of the teacher's capability with minimal degradation (within 0.1–0.6 dB on synthetic/real datasets), while achieving over 150$\times$ speedup to 76 FPS. The Gaussian output layer in the teacher proves beneficial: training the student without probabilistic sampling results in lower performance (e.g., 18.6 dB vs. 18.83 dB on O-Haze), consistent with the probabilistic outputs improving student generalization. The space-variant atmospheric light estimation proves important for non-homogeneous conditions: using constant A degrades NH-Haze performance (12.35 $\rightarrow$ 12.07 PSNR), consistent with per-pixel $A(x)$ estimation benefiting spatially-varying haze scenarios. Furthermore, by direct comparison with the base network, the teacher model generalizes better in outdoor real-world scenarios (O-Haze and NH-Haze), where the natural atmospheric light is space-variant. These results suggest that KD comes close to preserving the teacher's generalization, while enabling practical large-scale deployment.

Table 4

Ablation study results on Haze4K, O-Haze, I-Haze, and NH-Haze datasets. Each cell reports PSNR / SSIM

Setup	Haze4K	O-Haze	I-Haze	NH-Haze
Teacher	34.93 / 0.991	**18.93** / 0.818	16.74 / 0.773	12.29 / 0.582
Student	34.39 / 0.989	18.83 / **0.822**	16.67 / **0.785**	**12.35 / 0.588**
St. w/o Prob.	34.78 / 0.990	18.6 / 0.810	16.2 / 0.775	12.1 / 0.570
Constant A	34.85 / 0.991	18.4 / 0.808	16.3 / 0.782	12.07 / 0.561
Base [1]	**35.28 / 0.991**	18.91 / 0.815	**16.80** / 0.779	12.26 / 0.576

Conclusion. This paper addresses the computational challenge of real-time large-scale image dehazing by proposing a knowledge distillation framework integrated into the variational Bayesian objective. High-performance dehazing methods achieve superior generalization but require substantial computational resources, limiting their deployment in real-time applications. The proposed approach allows a lightweight Vision Transformer-based student network to benefit from the generalization capabilities of a high-performance teacher model through probabilistic distillation, where the student learns to approximate the teacher's variational posterior within the ELBO framework. The teacher employs Gaussian sampling to enable uncertainty-aware predictions, allowing the student to benefit from distributional knowledge rather than merely point estimates. Additionally, the atmospheric scattering model is extended to estimate space-variant atmospheric light on a per-pixel basis, improving performance on dense haze regions with varying illumination conditions.

Experimental evaluation demonstrates that the student network achieves significant speedup, while maintaining competitive synthetic-to-real generalization performance and improving downstream object detection tasks. The method establishes practical utility for large-scale applications, including aerial imaging, surveillance, and video processing. Future work will explore further optimization for edge devices and extensions to handle extreme atmospheric conditions, such as clouds and heavy precipitation, in real-world environments.

Received 23.02.2026
Reviewed 17.04.2026
Accepted 20.04.2026

REFERENCES

1. Babayan S., Sáez-Maldonado F.J., et al. Bridging the Synthetic-to-real Gap in Single Image Dehazing. *Digit. Signal Process* (2026), 106097. https://doi.org/10.1016/j.dsp.2026.106097
2. Babayan S., Sáez-Maldonado F.J., et al. Variational Bayesian Dehazing with Atmospheric Scattering Model for Synthetic-to-real Generalization. *2025 IEEE ICIPW* (2025), 380–385.
3. Wang Y., He B. CasDyF-Net: Image Dehazing via Cascaded Dynamic Filters. *Proc. IEEE ICVISP* (2024), 1–8. https://doi.org/10.1109/ICVISP64524.2024.10959384
4. Wang T., Zang K. et al. Gridformer: Residual Dense Transformer with Grid Structure for Image Restoration in Adverse Weather Conditions. *IJCV* **132** (2024), 4541–4563. https://doi.org/10.48550/arXiv.2305.17863
5. Park B., Lee H., et al. Progressive Pruning of Light Dehaze Networks for Static Scenes. *Applied Sciences* **14** (2024), 10820. https://doi.org/10.3390/app142310820
6. Zheng X. Research on Hardware Acceleration of Dehazing Network Based on Deep Learning. *International Journal of Computer Science and Information Technology* **5** (2025), 69–87. https://doi.org/10.62051/ijcsit.v5n1.07
7. Fang L., Chen Y. , et al. Bayesian Knowledge Distillation: a Bayesian Perspective of Distillation with Uncertainty Quantification. *Forty-first International Conference on Machine Learning* (2024).
8. Ming Y., Fu H., et al. Variational Bayesian Group-level Sparsification for Knowledge Distillation. *IEEE Access* **8** (2020), 126628–126636. https://doi.org/10.1109/ACCESS.2020.3008854
9. McCartney E.J. *Optics of the Atmosphere: Scattering by Molecules and Particles.* New York (1976).
10. He K., Sun J., Tang X. Single Image Haze Removal Using Dark Channel Prior. *IEEE TPAMI* **33** (2011), 2341–2353. https://doi.org/10.1109/CVPRW.2009.5206515
11. Zhu Q., Mai J., Shao L. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE TIP* **24** (2015), 3522–3533. http://dx.doi.org/10.1109/TIP.2015.2446191
12. Fattal R. Dehazing Using Color-Lines. *ACM Trans. Graph.* **34** (2015), 1–14. https://doi.org/10.1145/2651362

13. Zheng C., Ying W., Hu Q. Comparative Analysis of Dehazing Algorithms on Real-world Hazy Images. *Scientific Reports* **15** (2025), 10822.
https://doi.org/10.1038/s41598-025-95510-z
14. Cai B., Xu X., et al. DehazeNet: An End-to-end System for Single Image Haze Removal. *IEEE TIP* **25** (2016), 5187–5198.
https://doi.org/10.48550/arXiv.1601.07661
15. Li B., Peng X., et al. AOD-Net: All-in-one Dehazing Network. *Proc. IEEE ICCV* (2017), 4780–4788.
16. Qin X., Wang Z., et al. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. *Proc. AAAI Conf. on AI* **34** (2020), 11908–11915.
17. Cui Y., Ren W., et al. Revitalizing Convolutional Network for Image Restoration. *IEEE TPAMI* **46** (2024), 9423–9438.
https://doi.org/10.1109/TPAMI.2024.3419007
18. Cui Y., Ren W., et al. Image Restoration via Frequency Selection. *IEEE TPAMI* **46** (2024), 1093–1108.
https://doi.org/10.1109/TPAMI.2023.3330416
19. Chen Z., He Z., Lu Z.M. Dea-net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-guided Attention. *IEEE TIP* **33** (2024), 1002–1015.
https://doi.org/10.1109/TIP.2024.3354108
20. Hong M., Xie Y., et al. Distilling Image Dehazing with Heterogeneous Task Imitation. *Proc. IEEE/CVF CVPR* (2020), 3462–3471.
21. Wu H., Liu J., et al. Knowledge Transfer Dehazing Network for Non-homogeneous Dehazing. *Proc. IEEE/CVF CVPRW* (2020), 478–479.
22. Tran L.A., Park D.C. Lightweight Image Dehazing Networks Based on Soft Knowledge Distillation. *The Visual Computer* **41** (2025), 4047–4066.
https://doi.org/10.1007/s00371-024-03645-3
23. Chen Z., Bi X., et al. Omni-scale Feature Learning for Lightweight Image Dehazing. *Applied Intelligence* **54** (2024), 10039–10054.
https://doi.org/10.1007/s10489-024-05721-6
24. Lu H., Li Y., et al. Single Image Dehazing through Improved Atmospheric Light Estimation. *Multimed. Tools Appl.* **75** (2016), 17081–17096.
https://doi.org/10.48550/arXiv.1510.01018
25. He Y., Li C., Li X. Remote Sensing Image Dehazing Using Heterogeneous Atmospheric Light Prior. *IEEE Access* **11** (2023), 18805–18820.
https://doi.org/10.1109/ACCESS.2023.3247967
26. Kulkarni A., Murala S. Aerial Image Dehazing with Attentive Deformable Transformers. *Proc. IEEE/CVF WACV* (2023), 6305–6314.
27. Liu X., Peng H., et al. Efficientvit: Memory Efficient Vision Transformer with Cascaded Group Attention. *Proc. IEEE/CVF CVPR* (2023), 14420–14430.
https://doi.org/10.48550/arXiv.2305.07027
28. Koonce B. *Mobilenetv3. Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization.* Springer (2021), 125–144.
29. Mehta S., Rastegari M. *Mobilevit: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer* (2021).
https://doi.org/10.48550/arXiv.2110.02178
30. Hong M., Liu J., et al. Uncertainty-Driven Dehazing Network. *Proc. AAAI Conf. on AI* **36** (2022), 906–913.
https://doi.org/10.1609/aaai.v36i1.19973

31. Wang B., Ning Q., et al. Uncertainty Modeling of the Transmission Map for Single Image Dehazing. *IEEE TCSVT* **34** (2024), 11115–11127.
https://doi.org/10.1109/TCSVT.2024.3412093
32. Im E.W., Shin J., et al. Deep Variational Bayesian Modeling of Haze Degradation Process. *Proc. ACM CIKM* (2023), 895–904.
https://doi.org/10.1145/3583780.3614838
33. Jia X., De Brabandere B., et al. Dynamic Filter Networks. *Proc. NeurIPS* (2016), 667–675.
34. Chen D., He M., et al. Gated Context Aggregation Network for Image Dehazing and Deraining. *Proc. IEEE WACV* (2019), 1375–1383.
https://doi.org/10.48550/arXiv.1811.08747
35. Liu Y., Zhu L. et al. From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data. *Proc. 29th ACM Int. Conf. on Multimedia* (2021), 50–58.
https://doi.org/10.1145/3474085.3475331
36. Ancuti C.O., Ancuti C., et al. O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-free Outdoor Images. *Proc. IEEE CVPRW* (2018), 754–762.
https://doi.org/10.48550/arXiv.1804.05101
37. Ancuti C.O., Ancuti C., et al. I-HAZE: A Dehazing Benchmark with Real Hazy and Haze-free Indoor Images. *Proc. ACIVS* (2018), 620–631.
38. Ancuti C.O., Ancuti C., Timofte R. Nh-haze: An Image Dehazing Benchmark with Non-homogeneous Hazy and Haze-free Images. *Proc. IEEE/CVF CVPRW* (2020), 444–445.
https://doi.org/10.1109/CVPRW50498.2020.00230
39. Geiger A., Lenz P., et al. Vision Meets Robotics: The Kitti Dataset. *IJRR* **32** (2013), 1231–1237.
https://doi.org/10.1177/0278364913491297
40. Sohan M., Sai Ram T., et al. A Review on Yolov8 and Its Advancements. *Proc. ICDICI* (2024), 529–545.
https://doi.org/10.1007/978-981-99-7962-2_39

Ս. Վ. ԲԱԲԱՅԱՆ

ԳԻՏԵԼԻՔՆԵՐԻ ԹՈՐՄԱՄԲ ՎԱՐԻԱՑԻՈՆ ԲԱՅԵՍՅԱՆ ՇՐՋԱՆԱԿ՝ ՄԵԾԱԾԱՎԱԼ ՊԱՏԿԵՐՆԵՐԻ ԱՐԴՅՈՒՆԱՎԵՏ ՄԱՌԱԽՈՒՂԱԶԵՐԾՄԱՆ ՀԱՄԱՐ

Մեծածավալ պատկերների փվյալների բազաների իրական ժամանակի մշակումը առաջադեմ մշուշահեռացման (dehazing) մեթոդներով ներկայացնում է էական հաշվարկային մարտահրավերներ: Մեթոդները, որոնք ապահովում են գերազանց ընդհանրացում, որպես կանոն պահանջում են զգալի ժամանակ եզրակացության (inference) փուլում, ինչը սահմանափակում է նրանց կիրառումը իրական ժամանակում աշխատող հավելվածներում: Բարձր արդյունավետությամբ մեթոդները օգտագործում են բարդ բազմամակարդակ մշակում և խորը ճարտարապետություններ՝ սովորաբար հասնելով 5 ՖՊՍ-ից ցածր ցուցանիշի բարձր որակի պատկերների պրոցեսինգի համար:

Հիմնվելով նախնական բազմամակարդակ վարիացիոն Բայեսյան մոտեցման վրա [1,2], որը ապահովում է հզոր սինթետիկից-իրական ընդհանրացում,

այս աշխատանքը առաջարկում է գիտելիքների թորում (knowledge distillation)՝ բարձր արդյունավետությամբ մոդելների ընդհանրացման կարողությունները փոխանցելու թեթև Vision Transformer-ի վրա հիմնված ուսանող ցանցին: Ուսանող ցանցը օգտագործում է բլոկ առ բլոկ մշակում և ունի նվազեցված ճարտարապետական բարդություն՝ հասնելով ավելի քան 150 անգամ արագացման՝ միաժամանակ պահպանելով մրցունակ արդյունավետություն փեսականորեն հիմնավորված թորման մոտեցման միջոցով, որը ինտեգրված է վարիացիոն Բայեսյան օբյեկտիվի մեջ:

Բացի այդ, մթնոլորտային ցրումի մոդելը (Ատմոսփերից Սցատտերինգ Մոդել) ընդլայնվել է՝ գնահատելու տարածական-փոփոխական մթնոլորտային լույսը, ինչը բարելավում է մառախուղի հեռացումը տարբեր մշուշային խտությամբ տարածքներում: Ուսուցանված լինելով բացառապես սինթետիկ Haze4K տվյալների վրա, առաջարկվող մեթոդը մրցունակ է մնում սինթետիկից-իրական ընդհանրացման և օբյեկտների հայտնաբերման առաջադրանքներում (KITTI տվյալների բազայի հիման վրա), միաժամանակ ապահովելով գերազանց եզրակացության արագություն իրական աշխարհի մեծածավալ մշակման հավելվածների համար:

С. В. БАБАЯН

ВАРИАЦИОННАЯ БАЙЕСОВСКАЯ СТРУКТУРА
С ДИСТИЛЛЯЦИЕЙ ЗНАНИЙ ДЛЯ ЭФФЕКТИВНОГО УСТРАНЕНИЯ
ДЫМКИ НА ИЗОБРАЖЕНИЯХ БОЛЬШОГО МАСШТАБА

Обработка баз данных крупномасштабных изображений в реальном времени с использованием современных методов удаления дымки (dehazing) представляет собой существенную вычислительную проблему: методы, обеспечивающие превосходное обобщение, как правило, требуют значительного времени на этапе инференса (inference), что ограничивает их применение в приложениях реального времени. Методы с высокой производительностью используют сложную многоуровневую обработку и глубокие архитектуры, обычно достигая менее 5 FPS на изображениях высокого качества.

Основываясь на предварительном многоуровневом вариационном байесовском фреймворке [1, 2], который обеспечивает сильное обобщение от синтетических данных к реальным, данная работа предлагает дистилляцию знаний (knowledge distillation) для передачи обобщающих способностей высокопроизводительных моделей на облегченную студенческую сеть на основе Vision Transformer. Студенческая сеть использует обработку по блокам и обладает сниженной архитектурной сложностью, достигая более чем 150-кратного ускорения, одновременно сохраняя конкурентоспособную производительность благодаря теоретически обоснованному фреймворку дистилляции, интегрированному в вариационную байесовскую целевую функцию.

Кроме того, модель атмосферного рассеяния была расширена для оценки пространственно-вариабельной атмосферной подсветки, что улучшает удаление дымки в регионах с различной плотностью тумана. Обученная исключительно на синтетических данных Haze4K, предложенная методика остается конкурентоспособной в задачах обобщения от синтетических данных к реальным и обнаружения объектов (на основе датасета KITTI), одновременно обеспечивая превосходную скорость инференса для крупномасштабных приложений реального мира.