SLIDE: Selective Latent Intervention for Structure-Aware Single-Image Dehazing

Program Riset Unggulan 2025 & PPMI 2026

Nugraha Priya Utama [1,2], Muhammad A.N. Kemas [1,2], Muhammad Ghifary [3]

[1] School of Electrical Engineering and Informatics, ITB, [2] Center for Artificial Intelligence, ITB, [3] INA Digital Edu, Jakarta

Abstract

Single-image dehazing must recover visibility without unnecessarily altering useful scene information. We present SLIDE, a framework that formulates restoration as selective intervention in an invertible latent representation. A Glow backbone retains a reversible encoding of the input. Self-conditioned modulation predicts channel-wise statistical corrections, while attention-guided soft gating blends corrected and original features. This design makes the balance between correction and retention explicit within the architecture. Training uses paired hazy and haze-free images of the same scene. At inference, only the hazy image is required; no clear reference or external style code is needed. Under the reported primary protocols, SLIDE achieves 24.83 dB peak signal-to-noise ratio and 0.880 structural similarity on O-HAZE, and 30.13 dB and 0.950 on RESIDE-SOTS. A separate backbone experiment yields 24.82 dB and 0.884, compared with 21.76 dB and 0.835 for U-Net. Dataset-specific experiments demonstrate adaptability to non-homogeneous and dense physical haze, while transfer to unseen distributions remains limited. Downstream detection gains are model-dependent. SLIDE-Lite provides a real-time operating point under the reported L40 profiling conditions, with 25.84 ms latency and 32.5 images/s throughput. These results support controlled latent correction as a practical restoration design. Backbone invertibility does not guarantee fidelity after editing; structural retention is assessed through the restored outputs.

attention mechanism · image restoration · selective latent alignment · single-image dehazing · normalizing flow · structural preservation

Introduction

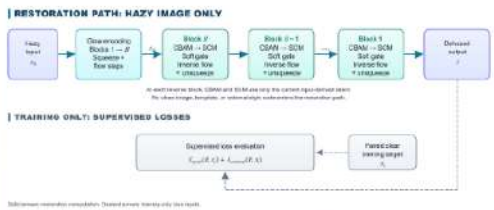
- Atmospheric scattering degrades image quality by reducing contrast, distorting colors, shifting luminance, and obscuring fine structural details, potentially impairing downstream computer vision tasks such as object detection, semantic segmentation, and depth estimation.
- Traditional prior-based methods rely on handcrafted assumptions that may fail under dense or non-uniform haze, bright surfaces, and varying illumination. Although deep learning approaches have improved restoration performance, preserving structural fidelity while effectively removing haze remains a challenge.
- Haze degradation varies spatially within an image. Applying uniform restoration may unnecessarily alter regions that already retain useful contrast and texture, motivating adaptive correction that balances enhancement and information retention.
- SLIDE (Selective Latent Invertible framework for single-image DEhazing) integrates a Glow-based invertible backbone, Self-Conditioned Modulation (SCM), and CBAM-guided soft gating to selectively adjust latent feature statistics based on the hazy input.
- Selective latent modulation aims to enhance visibility while limiting unnecessary structural and content modifications. Its effectiveness is evaluated through image quality metrics, edge agreement, human preference, downstream detection performance, and computational efficiency.

Proposed Method

1. **Invertible Encoding**
   The hazy image is encoded into a multi-scale latent representation using a Glow backbone (3 blocks, 8 flow steps per block).
2. **Attention-Guided Modulation**
   CCBAM highlights informative channel and spatial features, while SCM predicts latent statistical corrections.
3. **Selective Feature Correction**
   A soft gate blends corrected and original latent features, enabling targeted haze removal while retaining useful information.
4. **Inverse Reconstruction**
   The edited latent representation is decoded through the inverse Glow transformations to generate the restored image.

Training Strategy

Supervised training uses paired hazy–clear images with pixel reconstruction (L1) and VGG-19 content losses. During inference, only the hazy image is required.



Picture 1. Overall architecture of SLIDE

Results & Discussion

O-HAZE	RESIDE-SOTS	SLIDE-LITE
24.83	30.13	25.84
PSNR (dB) SSIM 0.880	PSNR (dB) SSIM 0.950	Latency (ms) 32.5 image/s

Track I: Reconstruction and Quality

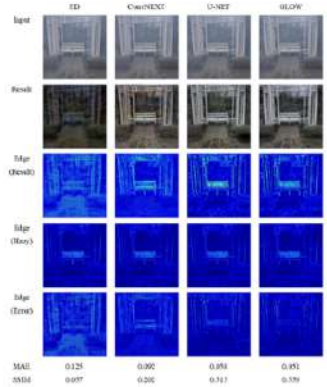
- SLIDE achieved 24.83 dB PSNR and 0.880 SSIM on O-HAZE, outperforming the listed methods in the reported comparison.
- On RESIDE-SOTS, SLIDE achieved 30.13 dB PSNR and 0.950 SSIM; dataset-specific training also demonstrated adaptability to NH-HAZE and Dense-HAZE.



Picture 2. Qualitative dehazing results on RESIDE-SOTS and O-HAZE

Track II: Structure and Perception

- SLIDE achieved the best reported edge preservation scores (Edge SSIM: 0.459; Edge MAE: 0.051).
- Human evaluators significantly preferred SLIDE over DCP (62.3%) and GridDehazeNet (58.7%), while comparisons with FFA-Net and DeHamer were not statistically significant.



Picture 3. Structural: hazy input, proposed output, Canny edge map of the output, and error map relative to ground truth for a representative O-HAZE scene

Track III: Utility and Efficiency

- SLIDE improved YOLOv8-S AP@50:95 by +0.0064, while performance effects varied across larger detector configurations.
- SLIDE-Lite achieved 25.84 ms latency (32.5 images/s) with 1.28M parameters and 41.6 GFLOPs under the reported L40 profiling conditions.



Picture 4. Selected side-by-side detection examples on RTTS at a confidence threshold of 0.25. Left: original hazy input. Right: dehazed output

Architectural Diagnostic and Ablations

- SCM improves latent-statistic alignment and visibility, while CBAM without soft gating can introduce over-correction in clearer regions. Soft gating mitigates these artifacts by adaptively blending corrected and original latent representations.
- The backbone comparison demonstrates the effectiveness of the proposed Glow-based architecture under matched training and leave-one-out cross-validation conditions. On O-HAZE, it achieves 24.82 dB PSNR and 0.884 SSIM, outperforming U-Net (21.76 dB and 0.835 SSIM), ConvNeXt (13.06 dB and 0.211 SSIM), and the standard encoder–decoder (10.11 dB and 0.035 SSIM).
- Ablation results demonstrate the complementary contributions of SCM, CBAM, and soft gating, highlighting the importance of combining feature refinement with controlled latent modification.
- By integrating Glow-based invertible representation, SCM-driven latent-statistic refinement, CBAM attention, and soft gating, SLIDE enables selective restoration rather than uniform image enhancement. This design aims to recover degraded information while preserving useful details in clearer regions.

Table 1. Quantitative comparison of backbone architectures under matched O-HAZE training conditions.

Backbone	PSNR (dB)	SSIM
Standard encoder-decoder	10.11	0.035
ConvNeXt [1]	13.06	0.211
U-Net [2]	21.76	0.835
SLIDE (Glow)	**24.82**	**0.884**



Picture 5. Ablation of SCM, CBAM, and soft gating on O-HAZE.



Picture 6. Qualitative comparison of alternative backbone outputs on O-HAZE.

Conclusion

- **Innovative Framework:** Developed SLIDE, a Glow-based dehazing framework with selective latent correction and reference-free inference.
- **Restoration Performance:** Achieved 24.83 dB PSNR / 0.880 SSIM (O-HAZE) and 30.13 dB / 0.950 SSIM (RESIDE-SOTS).
- **Computational Efficiency:** SLIDE-Lite achieved 25.84 ms latency and 32.5 images/s throughput on NVIDIA L40.

Output

1. **Paper:** Manuscript submitted for publication (2026).
2. **Patent (Application No. P00202507564):** Method and System for Restoration and Clarity Enhancement of Images and/or Videos Degraded by Fog, Smoke, Rain, and Fire.

References

1. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in Proc. IEEE/CVF CVPR, pp. 11976–11986, 2022.
2. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, pp. 234–241, 2015.
3. X. Li, J. Hou, and L. Zhang, "A comprehensive survey and taxonomy on single image dehazing based on deep learning," ACM Comput. Surv., vol. 55, no. 13s, pp. 1–40, 2023.