The 34th IEEE International Symposium on Industrial Electronics
June 20-23, 2025 | Toronto, Canada

Enhancing Low-Light Image Reconstruction via Non-Autoregressive Transformers: A Mask-Aware Latent Integration Framework

Qianyue Wang[1]

[1]Mcmaster University

Paper ID: ISIE25-000258

OUTLINE

Motivation & Introduction

Motivation

- ✓ Low-light images suffer from low contrast, noise, and color distortion.
- ✓ Traditional enhancement methods rely heavily on local context.
- ✓ Lack of global reasoning limits structural and perceptual recovery.
- ✓ Real-world applications demand reliable enhancement under poor illumination.
- ✓ We propose using Non-Autoregressive Transformers to globally refine uncertain regions.

Motivation & Introduction

contribution

✓ We propose a dual-path enhancement framework that integrates Non-Autoregressive Transformers (NAT) into the GLARE model.

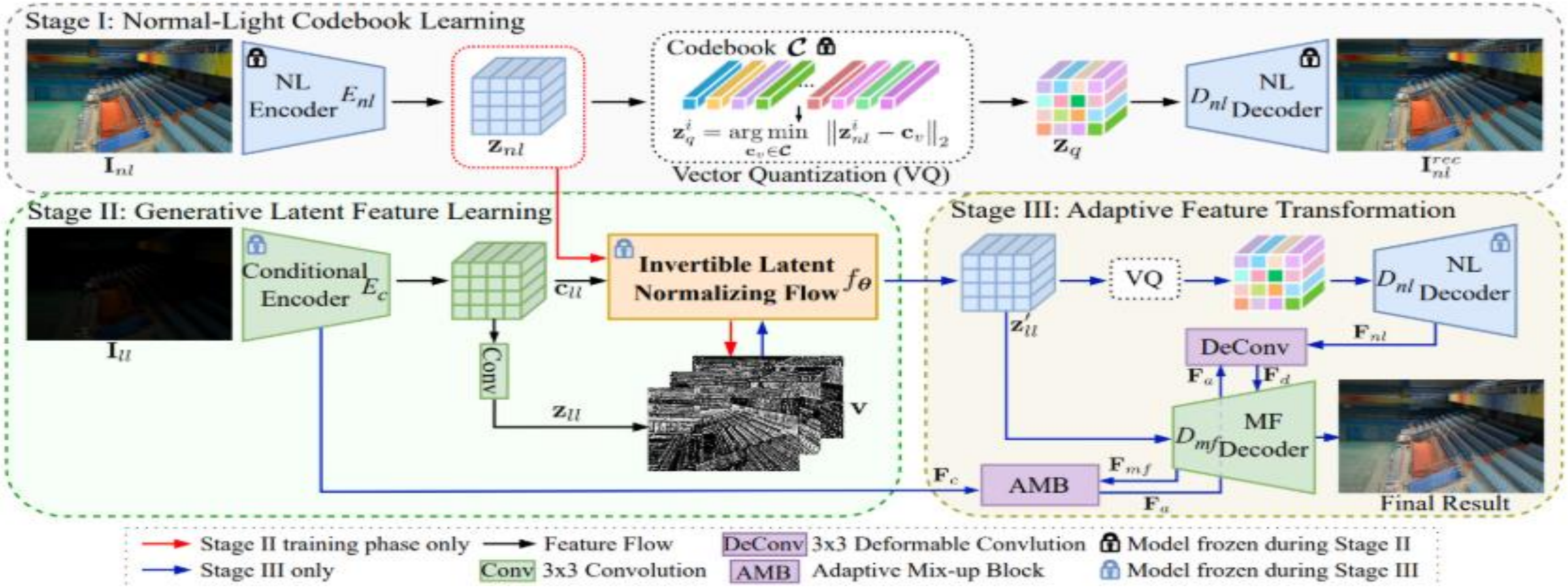
✓ Our design introduces mask-aware refinement at both the latent representation stage and the decoder feature level, targeting uncertain or degraded regions.

✓ Extensive experiments on GLARE and LOL datasets demonstrate consistent improvements in PSNR, SSIM, and LPIPS, validating the effectiveness of multi-level NAT integration.

BACKGROUND AND METHOD

Glare Framework

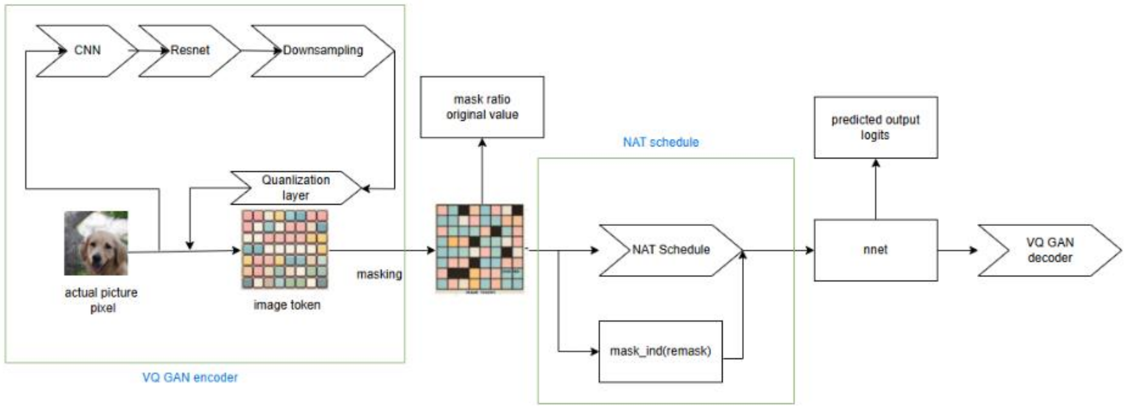
IEEE ISIE 2025

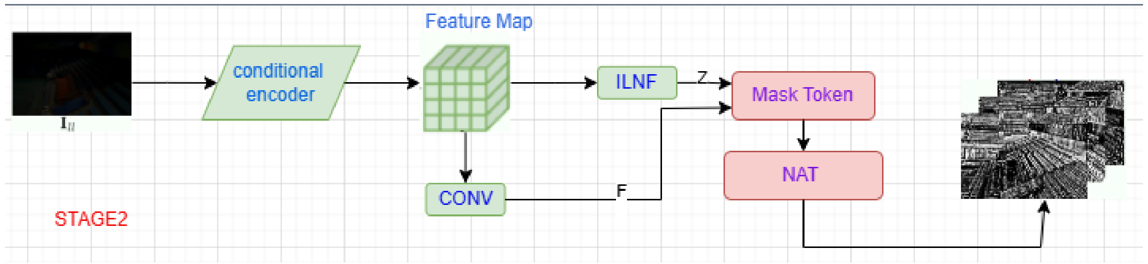


Original GLARE structure

BACKGROUND AND METHOD

Our Method Overview (Dual NAT Integration)





✓ We embed a Non-Autoregressive Transformer (NAT) after latent alignment in Stage 2 of GLARE.

✓ The NAT module refines uncertain latent tokens using mask-aware iterative decoding.

✓ The masking schedule selects low-confidence tokens based on entropy and re-masks them across decoding rounds.

✓ We also apply a second NAT module in Stage 3 (decoder branch) to enhance spatial features, not shown here.

IEEE ISIE 2025

NAT Strategy – Masking and Iteration

- ✓ We identify uncertain tokens using entropy computed from predicted distributions.
- ✓ Tokens with entropy above a threshold τ are replaced with a learnable [MASK] token.
- ✓ The NAT module predicts all masked tokens in parallel at each iteration.
- ✓ After each round, tokens with low confidence are re-masked and refined again.
- ✓ This strategy enables progressive correction of degraded regions using global context.

$$H(V_i) = -\sum_k \hat{p}(V_i = k) \log \hat{p}(V_i = k)$$

$$\text{if } H(V_i) > \tau, \quad \text{then } V_i \leftarrow \texttt{[MASK]}$$

Experiments and Evaluation

Training Strategy and Efficiency

- ✓ Datasets: GLARE and LOL (paired low-/normal-light samples)
- ✓ Metrics: PSNR, SSIM, LPIPS, FID (for generalization test)
- ✓ Optimizer: Adam, learning rate 4×10^{-4}, batch size 8
- ✓ Training steps: 60,150 (≈ 300 epochs), NAT runs 4 iterations
- ✓ Hardware: NVIDIA RTX 3090, 18 hours for full training
- ✓ NAT modules introduce only ~12% extra memory and <60 ms inference overhead

EXPERIMENTS AND EVALUATION

Quantitative Results on GLARE

Model	PSNR	SSIM	LPIPS
GLARE (baseline)	25.52	0.8462	0.1299
LLFlow [8]	26.03	0.8518	0.1275
Restormer [13]	31.00	0.9620	0.0880
+ NAT on latent	26.60	0.8573	0.1247
+ NAT on latent & decoder	**27.08**	**0.8678**	**0.1162**

✓ Our method improves PSNR by +1.56 dB over GLARE baseline.

✓ Both structural (SSIM) and perceptual (LPIPS) metrics are improved.

✓ Dual-path NAT outperforms LLFlow with fewer architectural changes.

✓ Performance is close to Restormer while using a simpler, faster model.

EXPERIMENTS AND EVALUATION

Ablation Study: NAT Iterations

TABLE II: Effect of NAT decoding iterations on GLARE dataset.

Iterations	PSNR	SSIM	LPIPS
2	26.33	0.8551	0.1261
4	**26.60**	**0.8573**	**0.1247**
6	26.62	0.8576	0.1246

✓ Increasing the number of NAT decoding iterations improves performance up to 4 rounds.

✓ Gains saturate beyond 4 iterations, indicating stable convergence.

✓ 4 iterations offer a good trade-off between quality and inference time.

✓ We use 4 iterations during both training and inference in all experiments.

EXPERIMENTS AND EVALUATION

Cross-Dataset Generalization: LOL

LOL	FID	LOL graph	final result
0.3	9.89		
0.7	9.38		
1.0	8.47		

Fig. 2: Qualitative results on the LOL dataset under different brightness levels.

✓ Robust enhancement under brightness ×0.3, ×0.7, and ×1.0

✓ FID scores remain low: 9.89 → 9.38 → 8.47

✓ NAT-enhanced model maintains perceptual quality even on unseen data

✓ Demonstrates strong generalization without retraining or fine-tuning

EXPERIMENTS AND EVALUATION

IEEE ISIE 2025

Visual Results: Qualitative Comparison

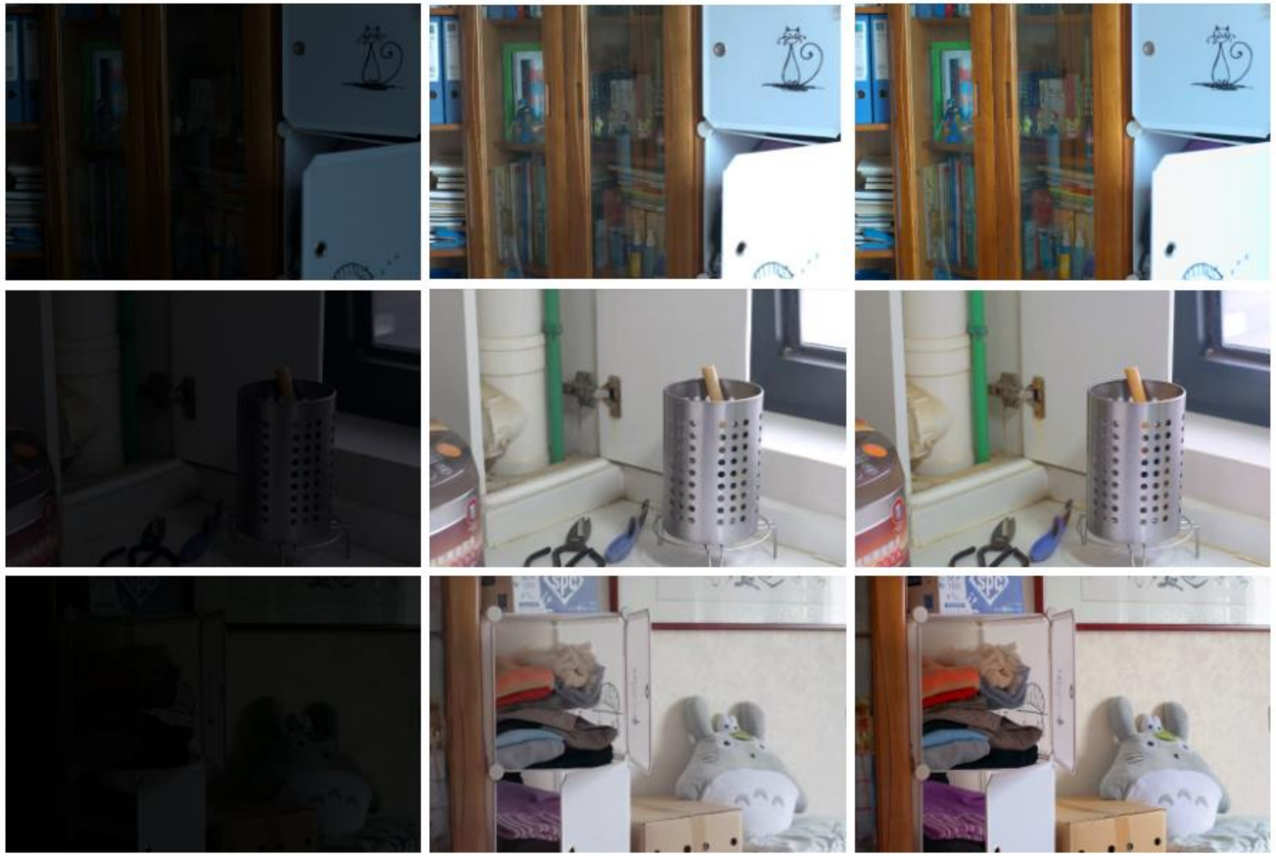


Fig. 3: Visual comparisons on representative scenes. Each row shows one test sample. Left: low-light input. Middle: GLARE baseline output. Right: NAT-enhanced result. NAT yields sharper details, higher contrast, and more accurate color restoration.

CONCLUSION AND FUTURE WORK

✓ We propose a dual-path NAT-enhanced framework for low-light image reconstruction.

✓ Two mask-aware NAT modules are integrated into GLARE at latent and decoder levels.

✓ Our method improves both structural fidelity (SSIM) and perceptual quality (LPIPS/FID).

✓ NAT enables efficient global reasoning with minimal computational overhead.

✓ Future work: explore diffusion-based priors, unsupervised training, and deployment in mobile vision systems.

Thank you for your attention.
I'm happy to take any questions.