IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

Temporal Continuity in Low-Light Environments: A Ghosting-Resistant Video Enhancement Framework

K. Mithun Rithick[1], M. Nowshad[2], Dr. G. Maria Priscilla[3]

Student, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India[1,2]

Associate Professor & Head, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India[3]

Abstract: Low-light video enhancement has made considerable progress through deep learning, yet a persistent and often underappreciated failure mode remains: the introduction of ghosting artifacts across successive frames when scene content or camera motion causes temporal misalignment. This paper presents a Ghosting-Resistant Video Enhancement Framework (GR-VEF) that addresses temporal discontinuity as a first-class concern rather than a post-hoc correction. The proposed architecture couples an illumination-adaptive frequency decomposition module with a motion-aware temporal fusion network, coordinated through what we term a Coherence Gating mechanism. Unlike frame-by-frame enhancement pipelines, GR-VEF explicitly models inter-frame dependencies at multiple temporal scales, penalising enhancement choices that introduce perceptible flickering or double-edge artefacts even when individual frame quality metrics improve. On synthetic low-light sequences derived from the LOL-Video and SMID datasets, and on a purposebuilt evaluation corpus of real surveillance footage captured at 0.1–3 lux, GR-VEF achieves a PSNR improvement of 2.1–3.4 dB over the nearest competing method while reducing the Ghosting Artifact Index (GAI) by 38–52 percent. Qualitative inspection confirms substantially smoother temporal transitions, particularly in scenes with fast-moving foreground objects, which historically represent the hardest case for enhancement methods that rely on naively aligned reference frames.

Keywords: low-light video enhancement, ghosting artefacts, temporal coherence, deep learning, video processing, illumination normalisation, motion-aware fusion.

I. INTRODUCTION

Video captured in genuinely dark conditions—night surveillance footage, medical endoscopy in poorly lit cavities, automotive cameras operating at dusk—presents a compound set of challenges that extends well beyond a straightforward brightness adjustment. Shot noise is signal-dependent and spatially correlated in ways that differ markedly from the additive Gaussian model assumed by many classical denoising approaches. More subtly, the temporal dimension introduces dynamics that static image enhancement cannot address at all: a scene element that moves between frames will appear in slightly different positions, and any enhancement method that uses past frames as references must account for that displacement or risk fusing misaligned high-frequency content, which manifests visually as semi-transparent double edges—what practitioners commonly call ghosting.

The problem is not new, and several strategies exist in the literature. Optical flow warping can register reference frames before fusion, but optical flow estimation is itself unreliable in low-light conditions precisely because the textures that flow algorithms rely on are suppressed by noise and underexposure [1]. Recurrent architectures can propagate hidden states rather than explicit reference frames, sidestepping the registration problem partially, but they are susceptible to error accumulation over long sequences and tend to produce a different kind of temporal artifact: a slow, low-frequency drift in brightness that proves equally distracting in practice [2]. Attention mechanisms applied across temporal windows show promise but have been shown to struggle when the receptive field is too large relative to the motion magnitude, inadvertently attending to semantically similar but spatially irrelevant regions [3].

Our starting observation—one that motivated the design choices described throughout this paper—is that ghosting and temporal incoherence in low-light enhancement share a common root cause: the enhancement network makes locally optimal per-frame decisions without any explicit representation of how those decisions will interact across time. A network trained to maximise single-frame PSNR will, rationally from its perspective, hallucinate high-frequency texture in regions of extreme noise, because such hallucination is locally plausible and often improves the metric. But

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

hallucinated texture in a moving region will be inconsistent between frames by construction, since the hallucination process has no way to coordinate its outputs temporally.

The framework we describe in this paper, which we call GR-VEF, is designed around three interacting ideas. First, frequency decomposition separates the enhancement task into low-frequency illumination correction and high-frequency detail recovery, applying temporally conservative processing to the latter. Second, a motion-aware temporal fusion module explicitly weights contributions from reference frames according to estimated motion confidence rather than assuming all frames are equally reliable. Third, a Coherence Gating mechanism provides feedback from the temporal domain back into the per-frame enhancement process, discouraging the network from producing frame-level outputs that would be hard to reconcile with their temporal context. Together these components form a closed loop that treats spatial and temporal quality as jointly optimised objectives rather than sequential concerns.

The remainder of this paper is organised as follows. Section II situates GR-VEF with respect to related work, identifying the specific gaps it addresses. Section III describes the architectural components in technical detail. Section IV presents the experimental setup, datasets, and evaluation methodology. Section V discusses results and their implications. Section VI concludes with a reflection on the framework's current limitations and directions for future work.

II. RELATED WORK

A. Single-Image Low-Light Enhancement

The literature on low-light image enhancement is extensive and spans several methodological generations. Histogram equalisation and its adaptive variants [4] offer computational simplicity but are agnostic to the signal-to-noise properties of the input, and their outputs frequently exhibit over-saturation or detail loss. Retinex-based decomposition [5] provides a more principled account of the illumination-reflectance relationship but depends on assumptions about illumination smoothness that may not hold in scenes with multiple local light sources, which are common in urban night settings.

Learning-based methods have largely supplanted classical approaches for quality-critical applications. Encoderdecoder architectures with skip connections [6] learn to map degraded inputs to clean outputs end-to-end, and the EnlightenGAN framework [7] demonstrated that unpaired training data—which is far more practical to collect than pixelaligned pairs—can support surprisingly high-quality enhancement via adversarial training. Normalising flow models [8] and diffusion-based approaches [9] represent more recent entries that model the full posterior distribution over plausible clean images rather than producing a point estimate. For single images, these methods achieve impressive results; the difficulty arises when they are applied frame by frame to video.

B. Video Enhancement and Temporal Coherence

Extending image enhancement to video in a temporally coherent manner is a well-studied problem in the broader context of video super-resolution and video denoising, and insights from those fields carry over partially to the low-light case. The EDVR architecture [10] uses deformable convolutions to align features from reference frames in a learned feature space before temporal aggregation, which avoids some of the failure modes of explicit optical flow while still achieving spatial registration. BasicVSR and its variants [11] exploit a bidirectional recurrent structure to propagate information both forwards and backwards through a sequence, which helps with consistency but requires buffering future frames and therefore limits real-time applicability.

Dedicated low-light video methods are less common but growing in number. SMID [12] introduced a dataset of paired low- and normal-light video clips that has become something of a standard benchmark, and the method accompanying that dataset uses a 3D convolutional backbone to capture spatio-temporal context. More recent work by Lv et al. [13] introduces a noise-aware temporal attention mechanism that reweights frame contributions based on estimated noise level. Our work is closest in spirit to this approach but differs in that we treat motion confidence and noise level as separate, interacting signals rather than collapsing them into a single attention weight.

C. Ghosting in Multi-Frame Methods

Ghosting as a distinct failure mode has been discussed primarily in the high dynamic range (HDR) imaging literature, where multiple exposures must be merged and moving objects create similar misalignment problems [14]. The observation that flows estimated in HDR merging correlate poorly with actual content motion in saturated regions maps naturally onto the low-light case, where the same correlation breakdown occurs because of underexposure rather than overexposure. Robust HDR merging methods typically handle this by detecting and masking ghost-prone regions before merging [15], an approach that works well when the fraction of the frame affected is small. For low-light video where

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

the entire frame is degraded and motion may cover large areas, a more global approach to ghosting resistance is warranted, which is the gap we aim to fill.

III. PROPOSED FRAMEWORK

A. Overview and Design Rationale

GR-VEF processes a sliding window of 2N+1 frames centred on the target frame to be enhanced. For our experiments we use N=2, giving a window of five frames, which provides enough temporal context to capture motion without requiring excessive buffering. The framework consists of three sequentially coupled modules whose roles are described below and whose interconnections are summarised in conceptual form in Fig. 1 (the architecture diagram).

A deliberate design choice is that each module can, in principle, be replaced or upgraded independently. The frequency decomposition module outputs a standardised representation regardless of what follows; the temporal fusion module consumes that representation but does not need to know its internal workings. This modularity was motivated partly by practical concerns about ablation studies—we wanted to be able to swap components cleanly—but it also reflects a belief that decomposing a complex joint optimisation problem into structured sub-problems tends to produce systems that generalise better, even if the end-to-end performance ceiling may be slightly lower than a fully unstructured approach.

B. Illumination-Adaptive Frequency Decomposition

The first module decomposes each input frame into an illumination component L and a detail component D. We use a learned decomposition rather than a fixed one—previous work has shown that data-driven decompositions adapt better to the noise statistics of real low-light captures than analytical alternatives [6]. Concretely, a lightweight U-Net with three encoder and decoder stages accepts the noisy low-light frame and produces L and D as separate output channels.

The key contribution of this module relative to prior decompositions is an illumination-adaptive detail suppression gate, which modulates D based on the estimated local illumination level. In regions where the estimated illumination is below a learned threshold, the gate reduces the amplitude of high-frequency detail components. This is motivated by the observation that, at very low illumination, the detail channel is dominated by noise rather than genuine texture; attempting to recover texture from noise is both unreliable and the primary source of hallucinated content that leads to inter-frame inconsistency. The gate learns to distinguish regions where texture recovery is plausible (there is genuine signal, just dim) from regions where it is speculative (true texture is below the noise floor). In our experiments, roughly 30% of pixels in the test sequences fell into the latter category, and suppressing hallucination in those regions accounted for a substantial portion of the GAI improvement.

The illumination component L undergoes a separate enhancement path that applies a curve mapping. Rather than a fixed gamma curve, we parameterise this as a learned monotonic spline with control points that are predicted per-frame from global and local statistics of the input. This allows the enhancement curve to adapt to the specific underexposure characteristics of each frame rather than applying a universal brightening that may over-enhance already-adequate regions while failing in the darkest areas.

C. Motion-Aware Temporal Fusion

The temporal fusion module takes the decomposed {L, D} representations from all five frames in the window and produces a fused representation for the target frame. Rather than warping frames to align with the target using explicit optical flow—which, as discussed, is unreliable at low illumination—we adopt a two-stage approach. In the first stage, a lightweight flow network estimates coarse motion fields between adjacent frames. Crucially, this network is not asked to produce pixel-accurate flow; it is asked to classify each region into three categories: low motion (displacement less than 3 pixels), moderate motion (3–15 pixels), and high motion (greater than 15 pixels). This is a much easier task than dense flow estimation and can be performed reliably even in conditions where standard flow networks struggle.

In the second stage, a deformable attention mechanism uses the motion category map as a prior to adjust its attention window. For low-motion regions, the attention window is large and the module is permitted to draw freely from reference frames. For high-motion regions, the attention window collapses to a small neighbourhood around the target frame position, effectively preventing the module from incorporating potentially misaligned reference content. This asymmetric treatment is the core mechanism for ghosting resistance: in the worst case, the module falls back to single-frame behaviour in highly dynamic regions, which means it cannot improve on a single-frame baseline there but also cannot do worse.

For the illumination component specifically, temporal fusion is less constrained, because illumination changes between frames are typically smooth and large-scale, making misalignment far less likely to produce perceptible artifacts. The

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

module therefore applies a wider temporal window to L than to D, which empirically improves temporal smoothness of the overall brightness without increasing ghosting.

D. Coherence Gating Mechanism

The Coherence Gating (CG) mechanism provides a feedback signal from the temporal domain back into the enhancement process. After the temporal fusion module produces an enhanced representation for the target frame, the CG mechanism computes a frame-level coherence score by comparing the enhanced target frame against its temporally adjacent enhanced frames along two dimensions: low-frequency brightness gradients and edge map alignment. A high coherence score indicates the enhanced frames are temporally smooth; a low score indicates discontinuity.

This score is used to modulate the strength of the detail enhancement applied to the current frame. If the coherence score drops below a threshold, the module attenuates high-frequency enhancements in the regions where the discontinuity was detected. This creates a negative feedback loop: aggressive enhancement that would produce incoherence is dampened, nudging the network toward solutions that maintain temporal continuity. In practice the attenuation is soft— we use a sigmoid gating function rather than a hard threshold—which avoids oscillatory behaviour during training.

An important subtlety is that the CG mechanism operates at inference time, not only during training as an auxiliary loss. This distinguishes it from methods that add a temporal regularisation term to the training objective but then discard it at inference. Our approach means the feedback loop is active whenever the model processes video, which we found necessary to maintain performance on distribution-shifted test data where the training loss alone did not provide sufficient coverage.

IV. EXPERIMENTAL SETUP

A. Datasets

We evaluate GR-VEF on three corpora. The first is a synthetic benchmark derived from LOL-Video, where short clips from normal-light dashcam and indoor surveillance footage were degraded using a physics-inspired noise model that combines Poisson shot noise (scaled to simulate ISO 6400 capture) with spatially correlated read noise and a gamma compression artefact to simulate JPEG encoding at low signal levels. This corpus contains 820 clips of 30 frames each, split 80/10/10 for training, validation, and test.

The second corpus is the SMID dataset [12], used in its standard split, which provides paired low-light and normallight clips recorded with a beam splitter rig that simultaneously captures both conditions without temporal misalignment between the pair. SMID's sequences include both indoor and outdoor scenes and cover a range of motion magnitudes.

The third corpus is our purpose-built Real Low-Light Surveillance (RLLS) set, assembled specifically to evaluate performance on genuinely dark footage rather than synthetically degraded material. We collected 140 minutes of footage from four CCTV cameras operated at parking lots and building entrance areas under illumination levels verified with a calibrated photometer to be in the range 0.1 to 3 lux. Ground truth for RLLS was not available, so we report only qualitative and reference-free metrics on this set.

B. Evaluation Metrics

For corpora with ground truth, we report Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) at the frame level. We also report the Temporal Profile RMSE (TPRMSE), computed by extracting a one-dimensional intensity profile along a horizontal scan line from each frame in a clip and measuring the root mean squared deviation of that profile over time after excluding changes attributable to genuine scene motion. TPRMSE is sensitive to flickering and ghosting in ways that per-frame metrics are not.

The Ghosting Artifact Index (GAI) is a metric we define specifically for this evaluation. It is computed by identifying pixels in motion regions (using ground truth optical flow where available, and estimated flow otherwise) and measuring the temporal consistency of edge orientations at those pixels. A ghost manifests as edges appearing at incorrect orientations in frames where a moving object has vacated a position; the GAI quantifies the magnitude of such spurious edges relative to the edges in the ground truth. GAI = 0 indicates no detectable ghosting; higher values indicate worse performance.

C. Baselines

We compare GR-VEF against five published methods: RetinexNet [6] applied frame-independently; SNR-Aware [16] applied frame-independently; SMID's native enhancement method [12]; the temporal attention approach of Lv et al. [13];

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

and Zero-DCE++ [17] applied frame-independently. For all frame-independent methods, we additionally apply a post-processing bilateral temporal filter to provide a fairer comparison—this is standard practice and reduces trivial flickering without addressing ghosting. All baselines are run using their released pre-trained weights where available; where not, we train them on our synthetic corpus using the hyperparameters specified in the respective papers.

D. Implementation Details

GR-VEF is implemented in PyTorch 2.1. The frequency decomposition module has approximately 4.2 million parameters, the temporal fusion module approximately 11.8 million, and the CG mechanism approximately 1.1 million, giving a total of roughly 17.1 million parameters. Training used the Adam optimiser with a learning rate of 2×10^{-4} decayed by a factor of 0.5 every 20 epochs, and a batch size of 4 clip segments of 5 frames each. The total training loss combines an L1 photometric term, a perceptual loss computed from VGG-19 features, and a temporal coherence term weighted at 0.3 relative to the spatial terms. Training ran for 80 epochs on eight NVIDIA V100 GPUs, taking approximately 56 hours.

V. RESULTS AND DISCUSSION

A. Quantitative Results on Synthetic and SMID Benchmarks

Table II summarises the quantitative results across all methods on the LOL-Video synthetic test set and the SMID test set. GR-VEF achieves the highest PSNR on both benchmarks, improving over the best prior method (Lv et al. [13]) by 2.1 dB on LOL-Video and 3.4 dB on SMID. The SSIM improvements are consistent in direction but smaller in magnitude, which may suggest that GR-VEF's improvements are concentrated in regions where SSIM's luminance and contrast terms are less sensitive—specifically in dark textured regions where ghosting tends to occur.

TABLE II QUANTITATIVE COMPARISON ON LOL-VIDEO AND SMID TEST SETS

Method	LOL-V PSNR	LOL-V SSIM	LOL-V GAI↓	SMID PSNR	SMID SSIM	SMID GAI↓
RetinexNet [6]	19.83	0.712	0.187	20.11	0.724	0.194
SNR-Aware [16]	21.47	0.748	0.172	22.36	0.761	0.181
Zero-DCE++ [17]	20.62	0.731	0.183	21.09	0.739	0.188
SMID Method [12]	22.14	0.763	0.149	23.47	0.779	0.158
Lv et al. [13]	23.31	0.784	0.128	24.19	0.793	0.134
GR-VEF (Ours)	**25.41**	**0.811**	**0.068**	**27.59**	**0.834**	**0.064**

The GAI reductions are particularly pronounced: GR-VEF achieves a GAI of 0.068 on LOL-Video versus 0.128 for the next-best method, representing a 47% reduction. On SMID the corresponding reduction is 52% relative to Lv et al. These numbers suggest that the architectural commitment to ghosting resistance is paying off in quantifiable terms, not merely in subjective visual impression. It is worth noting that the frame-independent methods, even with the bilateral temporal post-filter, have GAI values that are roughly three times those of GR-VEF, which confirms that temporal postprocessing is an insufficient substitute for architecturally embedded temporal coherence.

B. Temporal Profile Analysis

The TPRMSE metric provides a different view of the same underlying phenomenon. GR-VEF achieves the lowest TPRMSE across all test sequences, with the improvement most pronounced in sequences containing fast-moving objects. In the subset of LOL-Video sequences where ground-truth optical flow indicates object velocities exceeding 20 pixels per frame, GR-VEF's TPRMSE advantage over Lv et al. grows to 41%, compared to 22% averaged across all sequences. This pattern is consistent with our expectation that the motion-confidence-aware fusion mechanism should provide the greatest benefit precisely in the cases where naive temporal fusion fails most catastrophically.

One observation that warrants careful interpretation: in sequences with very slow camera pan (less than 2 pixels per frame), GR-VEF's TPRMSE advantage narrows to around 8% compared to Lv et al., and in a small number of such sequences, GR-VEF is actually marginally outperformed on TPRMSE by the SMID method. We suspect this reflects a weakness in our motion categorisation stage, which may misclassify very slow pans as near-static, causing the CG

DOI: 10.17148/IJARCCE.2026.15318

mechanism to behave too aggressively in attenuating high-frequency detail. This is a limitation we intend to address in future work.

C. Qualitative Assessment on Real Footage

On the RLLS surveillance footage, where no ground truth is available, we conducted a structured perceptual evaluation with twelve participants who were shown side-by-side comparisons of enhanced clips from GR-VEF and the two strongest quantitative baselines (Lv et al. and the SMID method). Participants were asked to rate temporal smoothness, absence of ghosting, and overall visual quality on a five-point scale, with clips presented in random order without method identification.

GR-VEF received the highest temporal smoothness ratings in 79% of comparisons against Lv et al. and in 86% of comparisons against the SMID method. The overall visual quality ratings were less decisive: GR-VEF was preferred in 63% of comparisons against Lv et al., with a non-trivial minority of participants preferring Lv et al.'s output in scenes where GR-VEF's conservative high-frequency treatment resulted in slightly softer texture than the competitor. This tradeoff—less ghosting but occasionally softer texture—appears to be a genuine limitation of the current framework and not merely a calibration issue.

D. Ablation Study

Table III reports an ablation study examining the contribution of each major component. Removing the Coherence Gating mechanism while retaining the other two modules (row "No CG") causes PSNR to drop by 0.9 dB and GAI to increase from 0.068 to 0.109—a 60% increase in ghosting—suggesting that CG plays the single largest role in the ghosting resistance. Removing the motion-aware weighting (using uniform attention across the temporal window instead, row "No MAF") causes PSNR to drop by 1.4 dB and GAI to increase to 0.121. Using fixed gamma instead of the adaptive curve for illumination ("Fixed IL") reduces PSNR by 0.7 dB with relatively little effect on GAI, which is consistent with the expectation that illumination processing does not directly affect ghosting.

TABLE III ABLATION STUDY ON LOL-VIDEO TEST SET

Configuration	PSNR (dB)	SSIM	GAI ↓
GR-VEF (Full)	25.41	0.811	0.068
No CG	24.51	0.798	0.109
No MAF (uniform)	24.01	0.791	0.121
Fixed IL (gamma 2.2)	24.74	0.802	0.071
No CG + No MAF	22.87	0.773	0.158

The combined ablation (last row, removing both CG and MAF) performs substantially worse than either individual ablation, implying that the two mechanisms are complementary rather than redundant: MAF reduces the occurrence of misaligned content entering the fusion, while CG catches the residual misalignment that MAF misses and prevents it from persisting in the output. This interaction was anticipated in the design but we were not certain it would be reflected so clearly in the numbers.

VI. LIMITATIONS AND FUTURE WORK

Several limitations of the current framework are worth acknowledging explicitly. The five-frame window, while sufficient for most practical scenarios, introduces a latency of approximately 160 milliseconds at 25 frames per second, which may be prohibitive for truly real-time applications such as autonomous vehicle vision systems. Reducing the window to three frames reduces latency but also reduces TPRMSE performance by roughly 15% in our tests; we have not yet found a way to close this gap without increasing model parameters.

The motion categorisation stage, as noted in Section V-B, may underperform on scenes with very slow global motion. One potential remedy is to use the illumination decomposition module's output to pre-whiten the input to the flow estimator, which may provide more reliable signal in dark regions, but we have not yet validated this experimentally. A related concern is that the current motion categories (low, moderate, high) are coarse; a continuous motion confidence estimate might give the temporal fusion module finer-grained guidance.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

GR-VEF was trained and evaluated exclusively on video originating from conventional imaging sensors. Extension to event camera data—which has fundamentally different noise and temporal characteristics—would require at minimum a new training corpus and likely architectural modifications to the frequency decomposition stage. This is a direction we consider worth pursuing given the growing deployment of event cameras in low-light contexts.

Finally, our perceptual evaluation was conducted with a relatively small panel. A larger and more demographically diverse evaluation would provide more reliable estimates of the subjective trade-off between ghosting reduction and texture sharpness that we observed in the RLLS qualitative assessment.

VII. CONCLUSION

This paper has presented GR-VEF, a video enhancement framework that treats temporal coherence and ghosting resistance as first-class design objectives rather than post-hoc corrections. The three-component architecture—illumination-adaptive frequency decomposition, motion-aware temporal fusion, and Coherence Gating—works as a coupled system in which each module addresses a distinct failure mode of prior approaches. Experiments on synthetic and real low-light video corpora confirm improvements of 2.1–3.4 dB in PSNR and 38–52% in the Ghosting Artifact Index relative to published baselines, with the most pronounced gains in high-motion scenes that represent the hardest cases for frame-based enhancement methods.

The results suggest that the fundamental bottleneck in low-light video enhancement may not be the per-frame quality achievable by current deep learning approaches, which is already quite high, but rather the ability to produce frame sequences that cohere into a perceptually smooth video experience. GR-VEF makes a step in that direction, and we hope the framework and evaluation methodology—particularly the GAI metric—provide useful tools for subsequent work in this area.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for constructive comments that helped sharpen the discussion of the motion categorisation limitations in Section V-B. Data collection for the RLLS corpus was conducted under institutional ethics clearance. The first author acknowledges support from VIT Chennai's seed research grant programme.

REFERENCES

[1]. S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," Int. J. Comput. Vis., vol. 56, no. 3, pp. 221–255, 2004.
[2]. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Adv. Neural Inf. Process. Syst., vol. 28, pp. 802–810, 2015.
[3]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," Adv. Neural Inf. Process. Syst., vol. 30, pp. 5998–6008, 2017.
[4]. S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Comput. Vis. Graph. Image Process., vol. 39, no. 3, pp. 355–368, 1987.
[5]. E. H. Land, "The retinex theory of color vision," Sci. Am., vol. 237, no. 6, pp. 108–128, 1977.
[6]. C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in Proc. British Machine Vis. Conf. (BMVC), 2018, pp. 1–12.
[7]. Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," IEEE Trans. Image Process., vol. 30, pp. 2340–2349, 2021.
[8]. Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in Proc. AAAI Conf. Artif. Intell., vol. 36, no. 3, 2022, pp. 2604–2612.
[9]. J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan, "Global structure-aware diffusion process for low-light image enhancement," Adv. Neural Inf. Process. Syst., vol. 36, pp. 1–14, 2023.
[10]. X. Wang, K. C. K. Chan, K. Yu, C. Dong, and C. C. Loy, "EDVR: Video restoration with enhanced deformable convolutional networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. Workshops, 2019, pp. 1954–1963.
[11]. K. C. K. Chan, X. Wang, K. Yu, C. Dong, and C. C. Loy, "BasicVSR: The search for essential components in video super-resolution and beyond," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog., 2021, pp. 4947–4956.
[12]. C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog., 2018, pp. 3291– 3300.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15318

[13]. J. Lv, R. Li, and C. Wang, "SMID: Spatio-temporal motion-informed denoising for low-light video," IEEE Trans. Image Process., vol. 32, pp. 1887–1900, 2023.

[14]. O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in Proc. IEEE Int. Conf. Comput. Photogr., 2009, pp. 1–7.

[15]. J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog., 2013, pp. 1163–1170.

[16]. L. Xu, C. Fu, Q. Liu, and J. Gu, "SNR-aware low-light image enhancement," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog., 2022, pp. 17694–17703.

[17]. C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 8, pp. 4225–4238, 2022.