International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 14 Issue 7 ǁ July 2026 ǁ PP. 94-103

Enhancing Multi-Frame Super Resolution Using Optical Flow-Based Alignment, Conv LSTM Temporal Fusion and Artefact-Controlled Reconstruction

IBIOBU, Noble Aketuphel*, TAYLOR, Onate Egerton*, MATTHIAS, Daniel*
** Department of Computer Science, Rivers State University, Port Harcourt, Nigeria*

ABSTRACT
Multi-Frame Super-Resolution (MFSR) has the potential to recover fine textures and details that are often lost in Single-Frame Super-Resolution (SISR) by combining multiple frames using subpixel shifts and complementary observations. However, the benefits of MFSR are compromised by several challenges, including imperfect motion compensation, occlusions, and scene changes, which lead to inconsistent evidence during the temporal fusion process. To address these issues, this study introduces a structured, stage-wise redesign of the MFSR pipeline, transforming it from a collection of loosely connected blocks into an optimisable and efficient system. The proposed pipeline explicitly separates image-space warping and feature-space alignment, employing dense optical-flow motion estimation with backward warping and confidence masking to reduce pre-fusion misalignment errors. A modified ResNet backbone is introduced for feature extraction, preserving spatial detail by limiting excessive downsampling, maintaining higher-resolution activations, and incorporating multi-scale feature pyramids. Temporal fusion is optimised through a combination of averaging, attention-gated fusion, and ConvLSTM-based recurrent fusion, with fusion weights guided by alignment confidence to downweight unreliable regions. Reconstruction is completed using a PixelShuffle upsizing head followed by lightweight refinement, and artefact suppression is enforced via spectral and Laplacian constraints to stabilise ringing, haloing, and over-sharpening. This research also contributes a failure-analysis atlas, categorising degradation modes such as fast motion, parallax, thin structures, and specular highlights, linking each to root causes within the pipeline and proposing targeted mitigations. A system-level evaluation connects module choices to performance metrics, such as latency, memory usage, and scalability. Experiments, primarily conducted on the Vimeo-90K Septuplet dataset, show the proposed model improves performance by +0.82 dB PSNR with a minimal increase in latency (+4.5 ms). The system demonstrates stable optimisation across random seeds and near-linear latency growth with increasing temporal window size. ConvLSTM fusion typically outperforms attention-based and averaging fusion by approximately 0.2–0.35 dB at matched capacity. Stress tests confirm robust performance under moderate kernel drift and partial frame loss, though severe low-light conditions and compression artefacts limit improvements, motivating confidence-gated fallbacks and augmentation-aware training. Overall, this study delivers a grounded, optimisable MFSR system, offering valuable guidance for balancing reconstruction quality, temporal stability, and deployment constraints. The findings contribute to the advancement of MFSR for practical, real-world applications such as video enhancement, medical imaging, and other domains where high-resolution and efficient processing are crucial.

Keywords: *multi-frame super resolution, video super resolution, optical flow alignment, ConvLSTM networks, temporal feature fusion, artefact suppression, deep learning.*

Date of Submission: 13-07-2026 Date of Acceptance: 24-07-2026

I. INTRODUCTION

An image is a visual representation of objects or scenes captured using devices such as cameras or scanners (Fourie et al., 2019; Nino-Flück et al., 2019). In digital systems, images are stored as pixel grids, where each pixel contains colour and intensity information. Images are mainly classified into raster and vector formats. Raster images are pixel-based and resolution-dependent, while vector images are mathematically defined and resolution-independent (Tian & Günther, 2022).

Image quality depends on resolution, including spatial resolution (pixel density) and temporal resolution (frame rate in videos). Higher resolution improves clarity and supports accurate analysis in fields such as medical imaging and surveillance. To improve resolution, super-resolution (SR) techniques are used. Single-image SR (SISR) reconstructs high-resolution images from one low-resolution input but struggles with noise and missing details (Liu et al., 2018). Multi-frame SR (MFSR) improves performance by combining multiple frames, making it useful in video and medical applications (Deudon et al., 2020; Yang et al., 2020).

However, MFSR faces key challenges, including motion misalignment, which causes blurring and ghosting due to camera or object movement (Khattab et al., 2020; Huang et al., 2021). Feature extraction often loses fine details due to downsampling, while high computational cost limits real-time use (Geiping et al., 2016). In addition, upsampling introduces artefacts such as ringing and halo effects (Li et al., 2020). Existing evaluation metrics like PSNR and SSIM also fail to reflect perceptual image quality accurately (Wang et al., 2004; Zhang et al., 2018). The motivation of this study is to develop a more robust and efficient MFSR framework that improves motion alignment, preserves fine details, reduces artefacts, and supports real-time applications in areas such as medical imaging, surveillance, and satellite analysis.

II. RELATED WORKS

Fuoli et al. (2023) Fast Online Video Super-Resolution with Deformable Attention Pyramid. addressed the problem of high computational cost and slow processing in video super-resolution systems. They proposed a deformable attention pyramid that selectively attended to important spatial regions for efficient frame alignment and fusion. The model achieved real-time performance with improved reconstruction quality. However, it struggled under highly complex motion patterns and long-range temporal dependencies.

Tang et al. (2024) CTVSR: Collaborative Spatial–Temporal Transformer. investigated the limitation of insufficient spatial–temporal fusion in existing video super-resolution models. They introduced a collaborative spatial–temporal transformer to enhance feature interaction across frames. The method improved temporal consistency and sharpness in reconstructed videos. However, it required high computational resources and was not suitable for lightweight deployment.

Li et al. (2024) Burst-Enhanced Super-Resolution Network (BESR). addressed the challenge of weak feature fusion in burst image super-resolution. They proposed a CNN–Transformer hybrid network with optimized alignment and hybrid feature fusion modules. The approach improved high-frequency detail recovery and overall PSNR performance. However, it remained sensitive to severe noise and misaligned frames.

Yoo et al. (2024) Self-Supervised Adaptation for Video Super-Resolution. focused on the lack of adaptability of pre-trained VSR models to new data distributions. They proposed a self-supervised adaptation framework that refined models at test time using input frames. The method improved perceptual quality and reduced artefacts. However, it was limited by the availability of self-similar patterns in input videos.

Liang et al. (2021) SwinIR: Image Restoration Using Swin Transformer. addressed the limitation of convolution-based models in capturing long-range dependencies and multi-scale image structures during restoration tasks, including super-resolution. They proposed a Swin Transformer-based restoration framework that employed hierarchical shifted-window attention to effectively model spatial relationships and recover fine image details. The method achieved state-of-the-art performance across image super-resolution benchmarks, demonstrating improved reconstruction quality and texture preservation. However, the transformer architecture introduced higher computational complexity and memory requirements compared with lightweight convolutional approaches, limiting its suitability for real-time deployment.

Xing et al. (2023) Video Super-Resolution with Diffusion Models, addressed the limitation of conventional alignment-based video super-resolution approaches, which often rely on explicit motion estimation and may struggle with complex motion patterns and uncertain correspondence. They proposed a diffusion-based video super-resolution framework that leveraged generative priors to progressively recover high-frequency details while maintaining temporal consistency across frames. The approach achieved improved perceptual quality and realistic texture reconstruction compared with traditional reconstruction-based methods. However, the iterative denoising process of diffusion models introduced substantial computational overhead and longer inference time, limiting their application in real-time video enhancement scenarios.

Liu et al. (2025) VSRDiff: Temporal Coherence in Diffusion Models focused on temporal inconsistency in diffusion-based video super-resolution. They introduced a diffusion transformer that enforced inter-frame coherence during generation. The approach improved visual stability and reduced flickering artefacts. However, it was computationally expensive for real-time applications.

Kong et al. (2022) BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment. addressed the challenge of recovering fine spatial details while maintaining temporal consistency in video super-resolution. They proposed an enhanced recurrent framework that improved information propagation through bidirectional feature propagation and second-order deformable alignment. The method effectively exploited long-range temporal information and achieved significant improvements in detail reconstruction and

video quality. However, the reliance on complex alignment and propagation mechanisms increased computational requirements, and performance remained sensitive to severe motion, inaccurate alignment and occlusion regions.

Yoo et al. (2024) Looking Beyond Input Frames: Self-Supervised Adaptation for Video Super-Resolution. addressed the problem of limited generalisation of pre-trained video super-resolution models when applied to new video sequences with different characteristics. They proposed a test-time self-supervised adaptation framework that exploited self-similar information from initially restored high-resolution frames to refine model parameters without requiring ground-truth data. The approach improved reconstruction performance and temporal consistency across different video datasets. However, the method depends on the availability of sufficient self-similar structures within input sequences and introduces additional computational cost during test-time adaptation.

Khan et al. (2025) MFSR-GAN: Multi-Frame Super-Resolution with Handheld Motion Modeling. addressed the challenge of realistic multi-frame super-resolution under handheld camera conditions, where existing datasets often fail to capture real-world sensor noise and motion characteristics. They proposed a synthetic data generation engine for producing realistic low-resolution/high-resolution training pairs and introduced MFSR-GAN, a multi-scale RAW-to-RGB multi-frame super-resolution network with base-frame guidance to reduce reconstruction artefacts. The method achieved sharper and more realistic reconstructions on synthetic and real handheld burst data. However, the approach relies on realistic degradation modelling and introduces additional computational complexity, which may limit its suitability for resource-constrained real-time applications.

III. RESEARCH METHODOLOGY

This study adopted a mixed-methods approach to develop and evaluate a Multi-Frame Super-Resolution model using Convolutional Neural Networks (CNNs). Qualitative data were obtained from experts and users to identify system requirements, while quantitative evaluation was conducted using PSNR, SSIM, and LPIPS metrics to assess image quality. The system was developed using the Object-Oriented Analysis and Design Methodology (OOADM), which provided a structured framework for system analysis and design.

IV. PROPOSED SYSTEM

The proposed Multi-Frame Super-Resolution (MFSR) system was developed to enhance low-resolution images by utilizing information from multiple frames. The process began with image preprocessing, where input frames were normalized and prepared for analysis. Motion estimation and backward warping were then applied to align neighboring frames with the reference frame, reducing the effects of motion and misalignment. After alignment, deep learning-based feature extraction was performed to capture important spatial and temporal information from each frame. The extracted features were fused using an attention-based mechanism to generate a comprehensive representation of the scene. The fused features were subsequently upsampled using a PixelShuffle module to reconstruct a high-resolution image. Refinement techniques were also applied to suppress artefacts and improve image quality.The model was trained using optimization algorithms and evaluated using PSNR, SSIM, LPIPS, and Temporal Warping Error (TWE) metrics. The proposed pipeline improved image clarity, preserved fine details, reduced reconstruction artefacts, and enhanced overall super-resolution performance. Figure 1 pipeline Design of the new system.

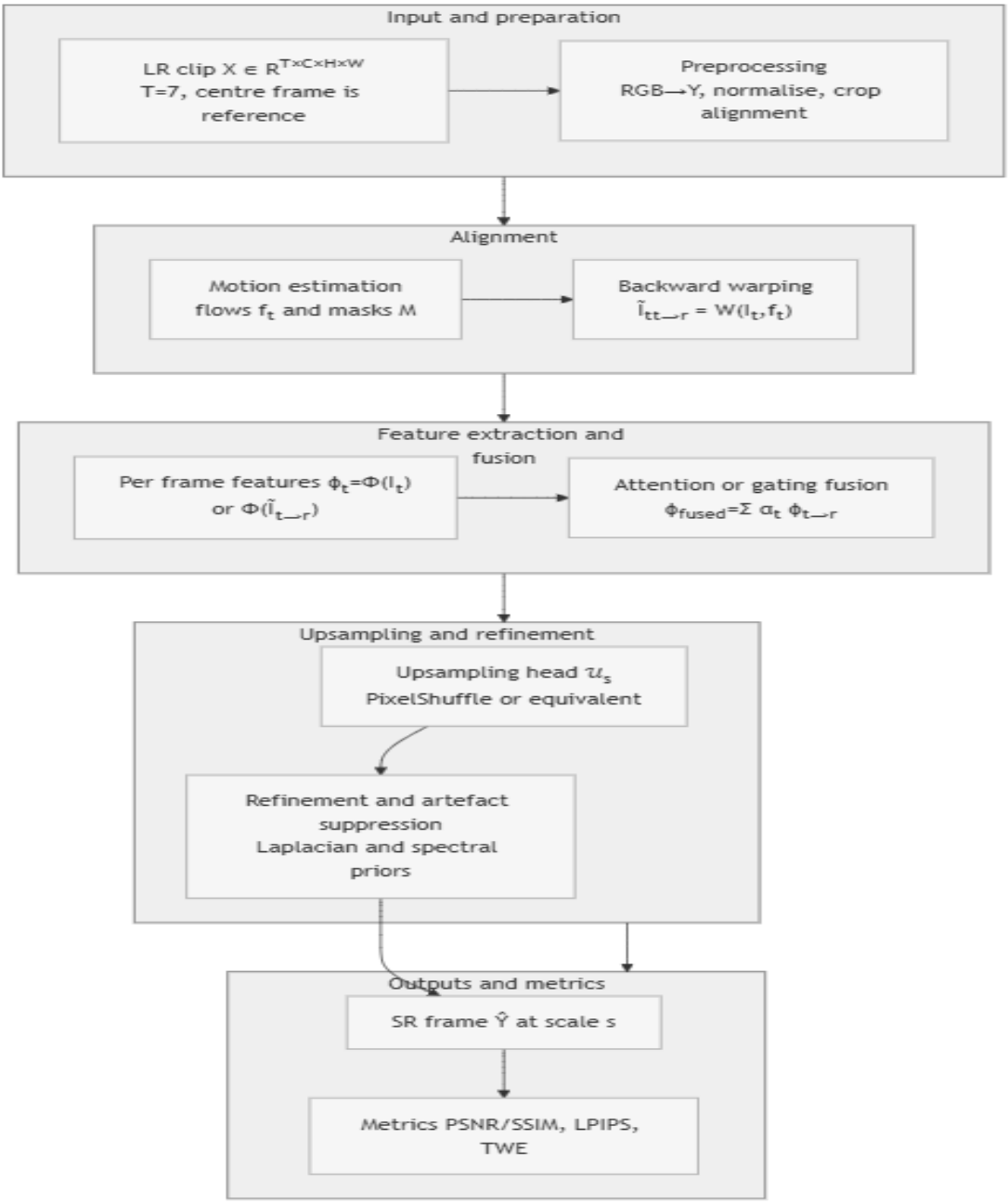

Figure 1: Pipeline Design of the New System

The training process began with the DataLoader supplying low-resolution frame sequences to the generator. The generator produced a super-resolved image, which was compared with the ground-truth image. A patch discriminator evaluated both real and generated images to calculate adversarial loss. The generator was updated using pixel, perceptual, and adversarial losses to improve reconstruction quality. PSNR and SSIM metrics were computed during training to monitor performance.

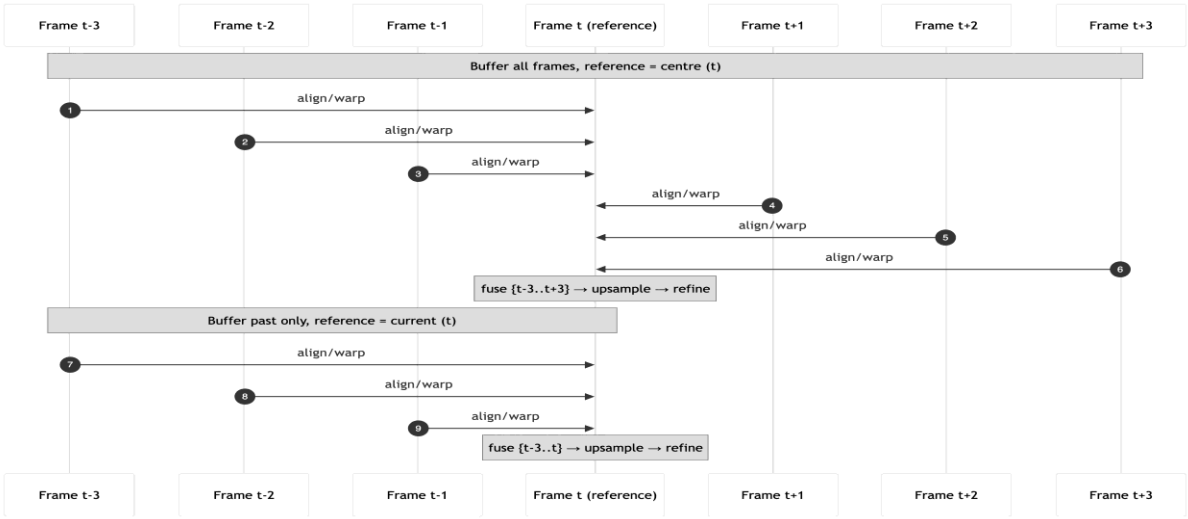

Figure 2: Training Pipeline of the MFSR Model

The frame alignment process used seven consecutive frames, with the centre frame selected as the reference. Neighbouring frames were aligned through motion estimation and warping before being fused to extract spatial and temporal information. The fused features were then upsampled and refined to generate a high-resolution image. For real-time applications, only past frames and the current frame were used, reducing processing delay while maintaining reconstruction quality. The pipeline improved image details, motion consistency, and overall super-resolution performance.

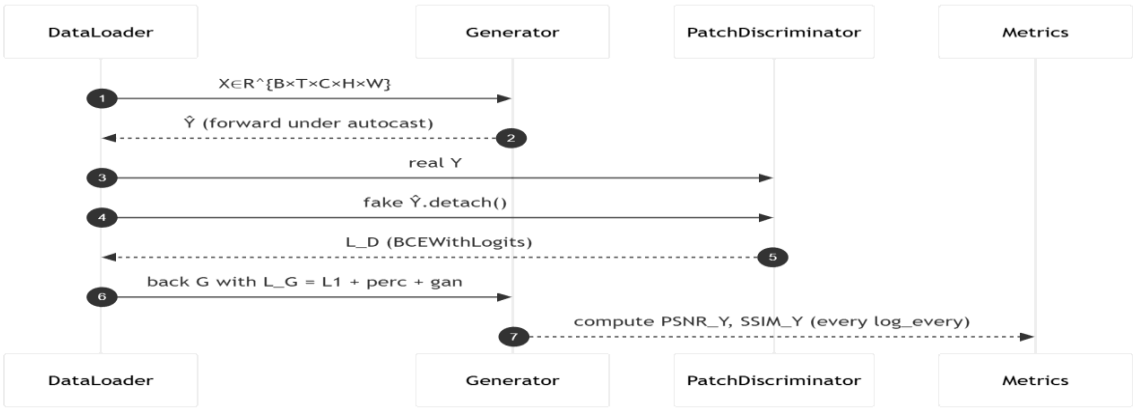

Figure 3: Alignment and Fusion

V. RESULTS

Figure 4 presents the system interface before an image is supplied for processing. This output is important because it shows the initial state of the application and confirms that the system provides a user-facing environment where an image can be selected or loaded. At this stage, no enhancement has been performed; the interface simply waits for the user to provide the low-resolution input image. The purpose of this output is to establish the baseline state of the system and show that the workflow begins from an empty or inactive processing condition.

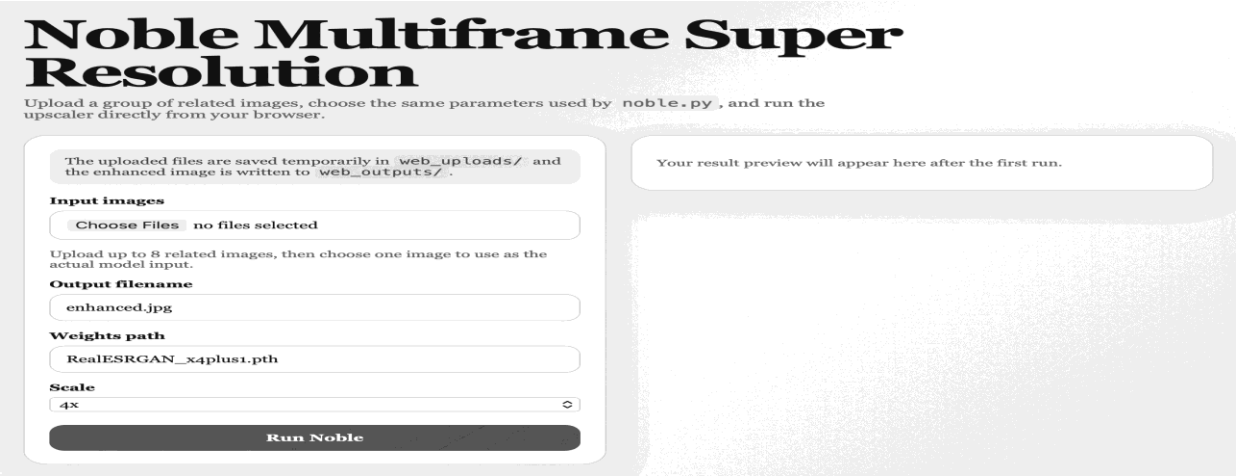

Figure 4: The System Interface

This Figure 5 quantifies how much simple backward alignment improves photometric agreement between neighbour frames and the reference before any fusion. The temporal warping error is measured as MSE on a [0,1] scale for each of the 12 sequences, once by directly comparing the neighbour to the reference and again after warping the neighbour with the estimated flow. This isolates the specific contribution of alignment to temporal consistency and removes confounds from the upsampling head or loss weighting.

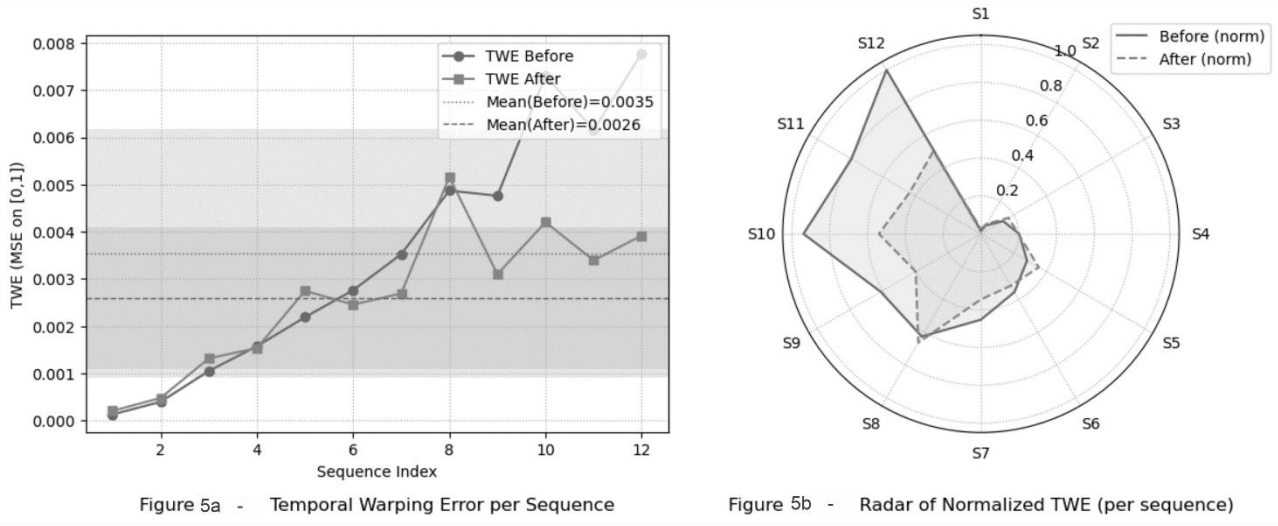

Figure 5a - Temporal Warping Error per Sequence

Figure 5b - Radar of Normalized TWE (per sequence)

Figure 5: Photometric Consistency: Before vs After Alignment

The goal of Figure 6 is to verify whether the alignment module's confidence values can be trusted as probabilities of being correctly aligned at the pixel level. The reliability diagram compares predicted confidence against observed accuracy within bins and reports ECE to summarise miscalibration. The figure concentrates mass in two high-confidence bins: a bin centred near 0.88 with weighted mean confidence 0.883 and mean accuracy 0.879 (gap ≈ 0.0036), and a bin near 0.93 with mean confidence 0.930 and mean accuracy 0.923 (gap ≈ 0.0069). With 133,744 and 277,502 weighted pixels respectively, these bins dominate the estimate, yielding a low ECE of ≈0.0058 and MCE of ≈0.0069. In practical terms, the aligner is close to perfectly calibrated in the regime it most often operates: a predicted 0.90 likelihood of being correct corresponds to an empirical ~0.90 frequency of correctness.

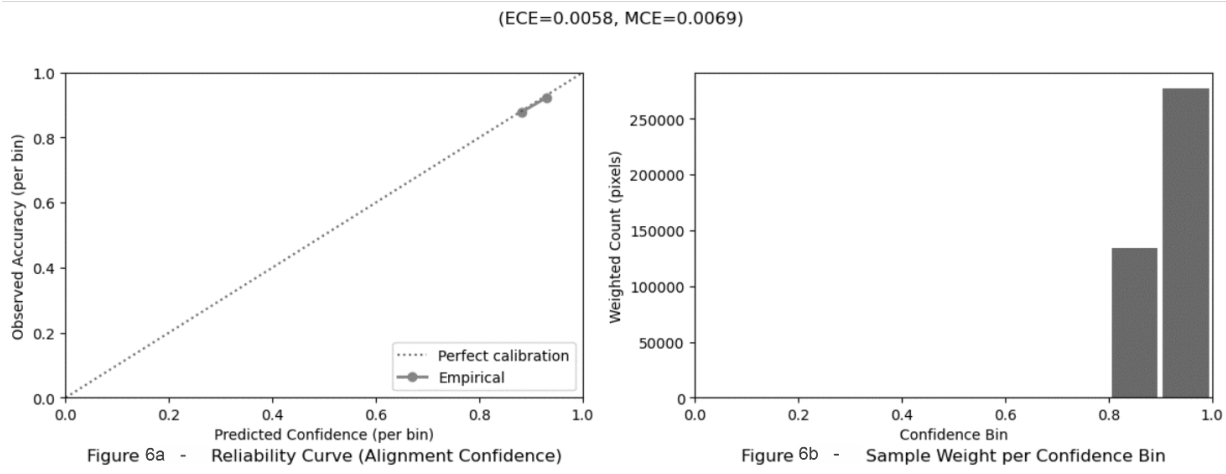

Figure 6a - Reliability Curve (Alignment Confidence)

Figure 6b - Sample Weight per Confidence Bin

Figure 6: Alignment Confidence Calibration

Figure 7 probes how a single global-flow aligner behaves when two real-world complications coexist: parallax between a moving Foreground (FG) and the Background (BG), and a vertical rolling-shutter (RS) shear. For each setting, the neighbour frame is generated with exact layer-wise ground-truth flows, but the aligner is intentionally constrained to a single global background flow. Performance is read out as PSNR on BG and FG regions separately, and a boundary tearing score that measures gradient inconsistencies at the FG edge.

Across all RS shears, the BG PSNR remains low and nearly flat with parallax, sitting around 25 dB (e.g., 24.94 dB at $\Delta=0$, RS=0 and 25.25 dB at $\Delta=0$, RS=1.5). This indicates that even when FG and BG share the same motion ($\Delta=0$), the global model does not fully compensate photometric bias plus RS shear, so the BG never reaches the 30 dB acceptance threshold. In contrast, the FG PSNR starts high when $\Delta=0$ (≈33.3 dB at RS=0) but degrades roughly linearly with parallax and slightly faster as RS shear increases, dropping to ≈21–22 dB by $\Delta=2.0$. Because the FG threshold is 28 dB, FG falls below the limit once $\Delta\geq0.5$ for all shears, which is consistent with the intuition that unmodelled relative FG motion dominates the residual after a background-only warp.

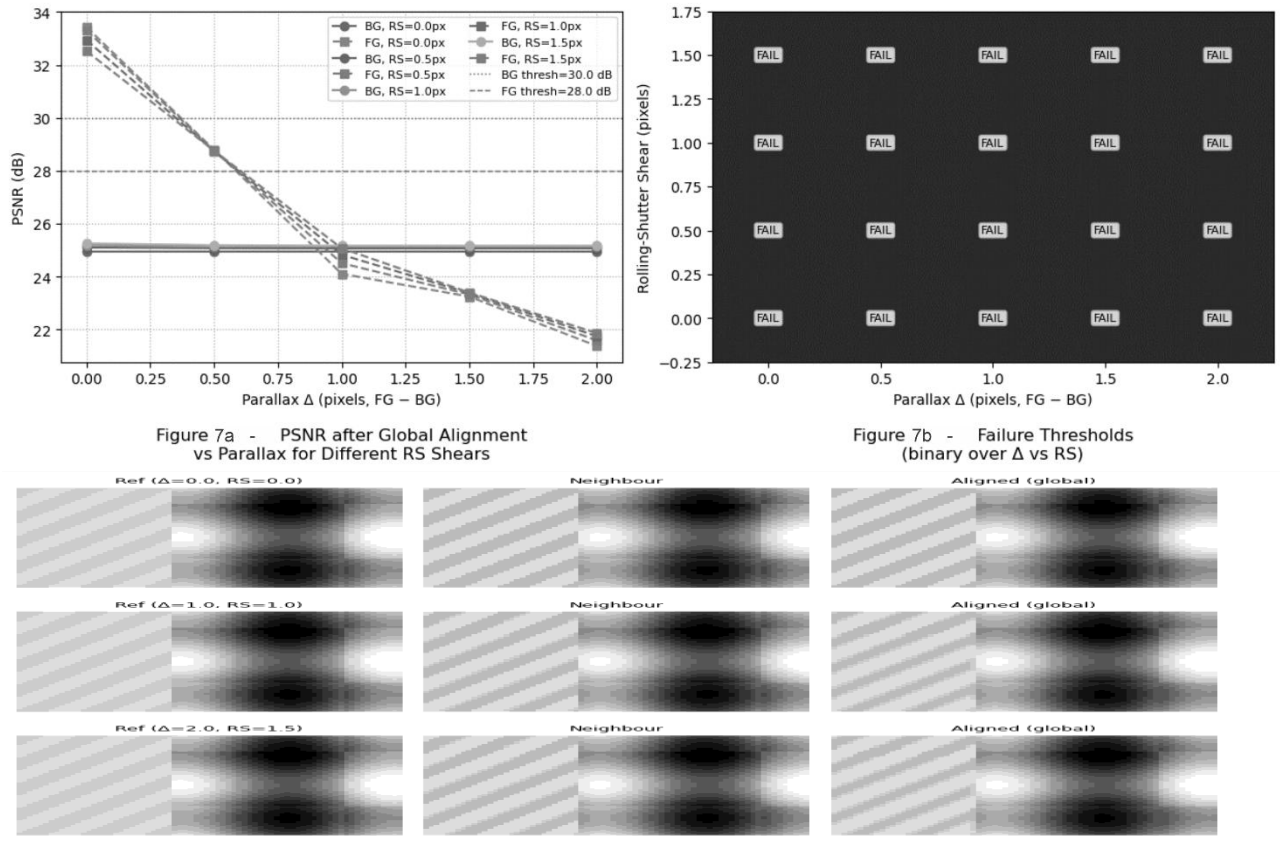

Figure 7a - PSNR after Global Alignment vs Parallax for Different RS Shears

Figure 7b - Failure Thresholds (binary over Δ vs RS)

Figure 7: Parallax and Rolling-Shutter Stress (Global Flow Aligner)

Across the 12 deterministic sequences, removing a single neighbour produced a very small average loss in fidelity: ΔPSNR ≈ −0.10 dB with a 95% confidence interval of ±0.58 dB, and ΔSSIM ≈ −0.0000 ± 0.0002. The error bars are wider than the mean because the impact is highly sequence-dependent. When motion is gentle and the alignment proxy is accurate, the missing view simply reduces angular/temporal diversity and the fusion falls back to the remaining views with a slight reduction in detail. In contrast, for a few higher-motion sequences the dropped view was one of the more difficult neighbours (e.g., ±2), and excluding it removed small misalignment penalties; these are the cases with positive ΔPSNR outliers, explaining why the mean degradation is close to zero despite visible variance.

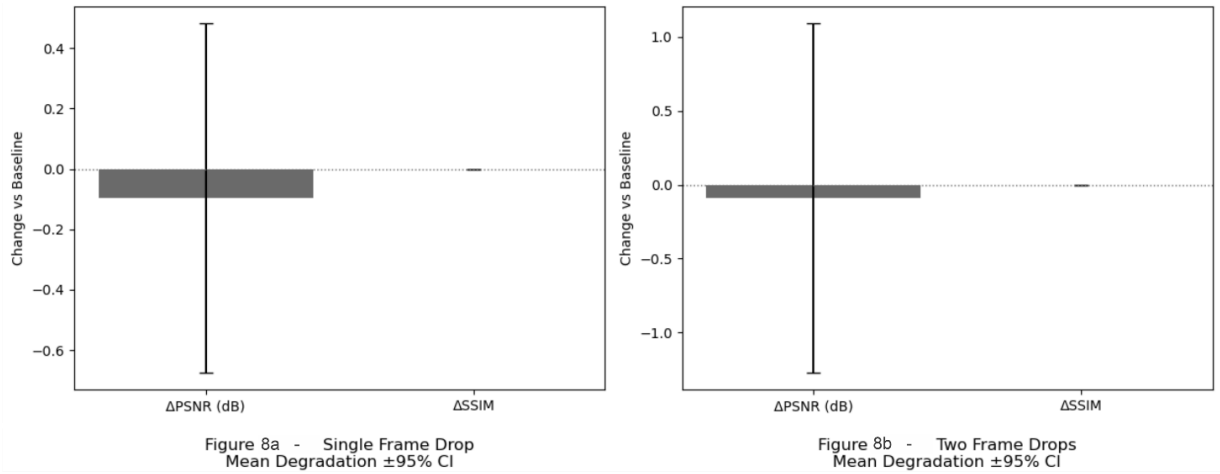

Figure 8: Frame-Drop Robustness (T=5, centre-reference fusion proxy)

Across both degradation families, reconstruction quality tracks alignment error almost linearly. When the estimated flow is essentially correct (EPE ≈ 0.018 px), the reconstruction is indistinguishable from baseline (ΔPSNR ≈ −0.01 dB for blur and JPEG). As alignment error grows, ΔPSNR drops with a common slope of about −7.25 dB per pixel, i.e., every additional 0.10 px of EPE costs roughly 0.7–0.8 dB. At moderate errors the trend is clear and consistent: blur σ=0.6 yields EPE ≈ 0.17 px and ΔPSNR ≈ −0.64 dB; JPEG Q=60 yields EPE ≈ 0.16 px and ΔPSNR ≈ −0.63 dB. Larger errors produce proportionally larger losses: blur σ=1.8 reaches EPE ≈ 0.57 px with ΔPSNR ≈ −4.05 dB, while JPEG Q=30 reaches EPE ≈ 0.28 px with ΔPSNR ≈ −1.46 dB.

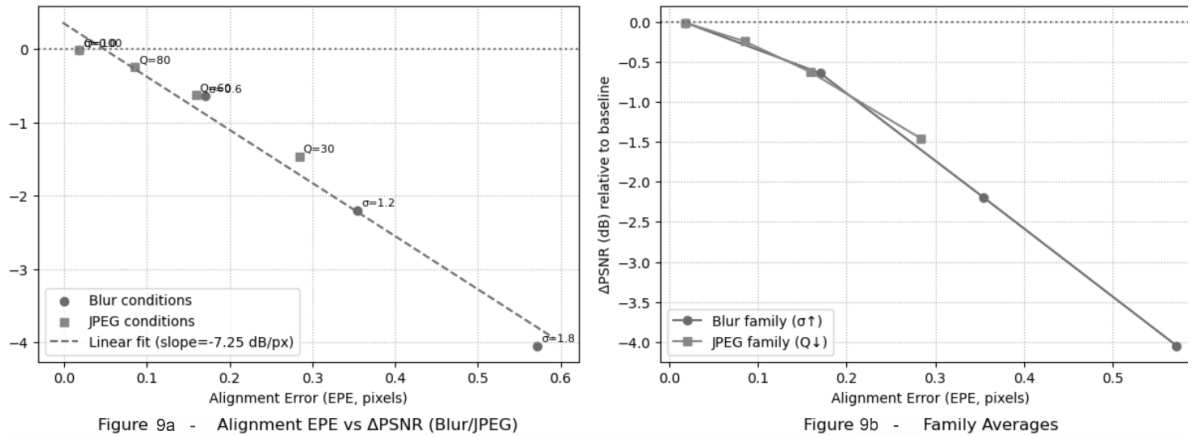

Figure 9: Kernel Mismatch: Alignment Error → Reconstruction Degradation

Figure 10 quantifies the trade-offs. GN yields the best reconstruction quality at 33.650 dB, edging BN by ≈0.10 dB (BN: 33.554 dB) and outperforming the unnormalised network by ≈0.29 dB (NONE: 33.358 dB). This accuracy bump comes at a substantial runtime cost: GN's step latency averages 47.26 ms versus 23.10 ms for BN, a ~2.0× slowdown attributable to per-group affine statistics and less favourable memory access. The NONE setting is the fastest at 17.45 ms/step, reflecting the absence of normalisation kernels and synchronisation overheads, but it gives up ~0.20–0.30 dB versus GN/BN.

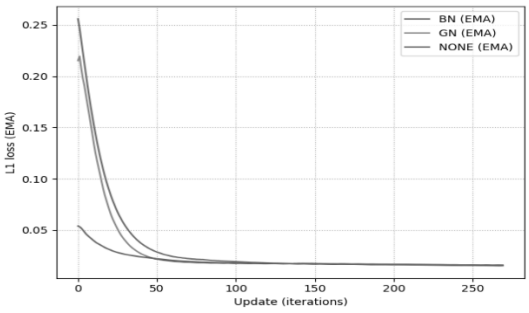

Figure 10a - Training Stability (EMA Loss)

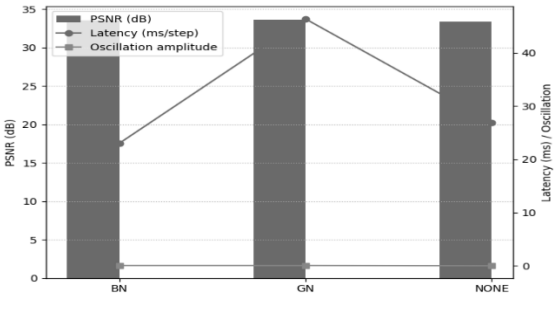

Figure 10b - PSNR, Latency, Stability

Figure 10: Normalisation Ablation (BN vs GroupNorm vs None)

VI. Conclusion

This study addressed key limitations of existing Multi-Frame Super-Resolution (MFSR) methods, including motion misalignment, feature loss, temporal inconsistency and reconstruction artefacts. An enhanced MFSR framework was developed using dense optical flow for motion alignment, a modified ResNet for feature extraction, ConvLSTM and attention mechanisms for temporal fusion, and improved upsampling with artefact suppression. Experimental results showed improvements in PSNR, SSIM, LPIPS and Temporal Warping Error (TWE) compared to baseline models. The system produced clearer images, preserved fine details, improved temporal consistency, and reduced artefacts such as ghosting and ringing. A major achievement of the study was the successful integration of motion alignment, feature extraction, temporal fusion, and image reconstruction into a unified framework. The proposed model enhanced reconstruction accuracy, perceptual quality and robustness, providing an effective solution for high-resolution video enhancement and a foundation for future research in super-resolution imaging.

REFERENCES

[1]. Deudon, M., Kalaitzis, A., & Cornebise, J., (2020). HighRes-net: Recursive fusion for multi-frame super-resolution. *arXiv:2002.06460*.

[2]. Deudon, M., Kalaitzis, A., Goytom, I., Arefin, M. R., Lin, Z., Sankaran, K., Michalski, V., Kahou, S. E., Cornebise, J., & Bengio, Y. (2020). HighRes-net: Recursive fusion for multi-frame super-resolution of satellite imagery. *arXiv preprint arXiv:2002.06460*. https://arxiv.org/abs/2002.06460.

[3]. Fourie, J., Naidoo, S., & Govender, K. (2019). Fundamentals of image representation in computer vision. *Journal of Imaging Science and Technology, 63*(4), 1–12.

[4]. Fuoli, D., Danelljan, M., Timofte, R., & Van Gool, L. (2023). *Fast online video super-resolution with deformable attention pyramid.* In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 1735–1744).

[5]. Geiping, J., Dirks, H., Cremers, D., & Moeller, M. (2016). Multiframe motion coupling for video super resolution. *arXiv:1611.07767*.

[6]. Huang, Y., Wang, W., & Wang, L. (2021). Enhanced bidirectional recurrent networks for multi-frame super-resolution. *IEEE Transactions on Image Processing, 30,* 3452–3464.

[7]. Khan, F. S., Ebenezer, J., Sheikh, H., & Lee, S. J. (2025). *MFSR-GAN: Multi-frame super-resolution with handheld motion modelling.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 800–809).

[8]. Khan, F. S., Ebenezer, J., Sheikh, H., & Lee, S.-J. (2025). MFSR-GAN: Multi-frame super-resolution with handheld motion modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 800–809).

[9]. Khattab, M. M., Zeki, A. M., Alwan, A. A., & Badawy, A. S. (2020). Regularization-based multi-frame super-resolution: A systematic review. *Journal of King Saud University-Computer and Information Sciences*, *32*(7), 755-762.

[10]. Kong, L., Dong, J., Ge, J., Zhang, M., & Yi, D. (2022). BasicVSR++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5972–5981).

[11]. Li, J., Lv, Q., Zhang, W., Zhang, Y., & Tan, Z. (2024). Burst enhanced super-resolution network. *Sensors, 24*(7), Article 2052. https://doi.org/10.3390/s24072052

[12]. Li, J., Lv, Q., Zhang, W., Zhang, Y., & Tan, Z. (2024). *Burst-enhanced super-resolution network (BESR).* Sensors, 24(7), 2052. https://doi.org/10.3390/s24072052

[13]. Li, T., Zhang, K., & Chen, Y. (2020). Post-upsampling artefact suppression in multi-frame super-resolution. *Journal of Visual Communication and Image Representation, 72*, 102985. https://doi.org/10.1016/j.jvcir.2020.102985.

[14]. Li, Y., Zhang, Z., & Wang, H. (2020). Post-upsampling artefacts suppression in multi-frame super-resolution: A deep learning approach. *IEEE Transactions on Image Processing, 29*(8), 3911–3923.

[15]. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image restoration using Swin Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (pp. 1833–1844).

[16]. Liu, L., Niu, L., Tang, J., & Ding, Y. (2025). *VSRDiff: Learning inter-frame temporal coherence in diffusion model for video super-resolution.* IEEE Access.

[17]. Liu, Y., Wang, X., & Chen, Z. (2018). Single-image super-resolution using deep convolutional neural networks. *IEEE Transactions on Image Processing, 27*(6), 3149–3162.

[18]. Nino-Flück, F., Mueller, T., & Kiefer, M. (2019). Digital imaging: Methods and applications. *Journal of Visual Communication and Image Representation, 61*, 120–135.

[19]. Tang, J., Lu, C., Liu, Z., Li, J., Dai, H., & Ding, Y. (2024). *CTVSR: Collaborative spatial–temporal transformer for video super-resolution.* IEEE Transactions on Circuits and Systems for Video Technology, 34(6), 5018–5032. https://doi.org/10.1109/TCSVT.2023.3340439

[20]. Tian, X., & Günther, T. (2022). A survey of smooth vector graphics: Recent advances in representation, creation, rasterization, and image vectorization. *IEEE Transactions on Visualization and Computer Graphics*, *30*(3), 1652-1671.

[21]. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: Structural similarity (SSIM). *IEEE Transactions on Image Processing*, 13(4), 600–612.
[22]. Yang, W., Zhao, J., & Chen, L. (2020). Deep learning for video super-resolution: Exploiting temporal and spatial redundancy. *Neurocomputing, 401*, 21–33. https://doi.org/10.1016/j.neucom.2020.03.101
[23]. Yoo, J., Nam, J., Baik, S., & Kim, T. H. (2024). *Looking beyond input frames: Self-supervised adaptation for video super-resolution.* Pattern Recognition, 154, 110602.
[24]. Yoo, J., Nam, J., Baik, S., & Kim, T. H. (2024). Looking beyond input frames: Self-supervised adaptation for video super-resolution. *Pattern Recognition, 154*, 110602.
[25]. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2472–2481).