# Motion Reconstruction from a Single Blurry Noisy Image using Deep Image Deblurring and Sequence Modeling

Mahika Annie Verghese, Shreya Shree Srikanth and Tetiana Ufimtseva

**Abstract**—Motion blur decomposition aims to recover sharp frame sequences from a single blurry image, but existing methods degrade noticeably in the presence of sensor noise. We address this limitation by extending the Multi-model Blur Decomposition (MBD) framework with two components: a lightweight noise estimation network, and a FiLM-based feature modulation step. Together, these additions allow the model to adapt to varying noise levels and stabilise frame-to-frame transitions. To further analyse our model, we run a Restormer denoiser in the inference step to understand how the use of Restormer aids our model. We also introduce an improved loss function with a temporal consistency term. On clean inputs ($\sigma = 0$), this enhanced loss configuration improves PSNR from 22.87 dB to 28.68 dB, an increase of 5.8 dB (≈25%), with corresponding gains in SSIM and LPIPS. Under moderate noise ($\sigma = 20$), NAMBD improves PSNR from 20.42 dB to 24.10 dB (≈18%), and the Restormer with NAMBD pipeline maintains this performance while reducing degradation at higher noise levels ($\sigma \geq 30$). Evaluated on the GoPro dataset across $\sigma$ = 0 to 50, the proposed approach consistently outperforms the baseline in PSNR, SSIM, and LPIPS, and the noise estimator predicts $\sigma$ with low mean absolute error. This work provides a practical solution for motion deblurring in real-world scenarios where noise corruption is inevitable. Our work is available at Github: **https://github.com/tufimtseva/animation_from_blur_and_noise**.

**Index Terms**—Computational Photography, Denoising, Deblurring, Motion Blur Decomposition, Sequence Reconstruction

## 1 Introduction

In many real-world settings, one of the most common and challenging issues that arises during the image capturing process is motion blur. The ability to capture live scenes accurately under diverse conditions is critical across multiple domains. Surveillance cameras record scenes for security purposes, athletes are captured to analyze and improve their techniques, and high-speed vehicles are monitored by roadside cameras for traffic management. Motion blur also occurs frequently in dynamic environments captured by autonomous vehicles or drones, in mobile photography, and in medical imaging. When only a single blurred frame is available, valuable temporal information about movement is lost, making the task of reconstructing the original motion particularly difficult. Extensive research has been conducted to tackle the problem of motion reconstruction from a single blurry image. The first research goal of this project is to improve the motion reconstruction quality of an existing state-of-the-art approach, which we select as the baseline for this study.

In real-world scenarios, image blur is often accompanied by sensor noise. For example, in low-light surveillance settings, cameras increase ISO or exposure to amplify the signal, which also amplifies noise. Despite this, little attention has been given to motion reconstruction under the combined effects of blur and noise. Existing motion blur decomposition methods often fail when the input images contain sensor noise, resulting in temporally inconsistent and visually degraded frame sequences. Notably, one of the most recent and strong algorithms, introduced in the Zhong et al. [1] paper, experiences a 7.7% drop in image reconstruction quality on its native dataset [2] when only a moderate level of Gaussian noise ($\sigma = 20$) is present.

The second and main research focus of this project is to reconstruct short motion sequences from a single blurred and noisy image using a combination of image processing techniques and sequence modeling. We simulate sensor noise with additive white Gaussian noise and evaluate our algorithm under different noise intensities ($\sigma = 0, 5, 10, 20, 30, 40, 50$), enabling a thorough assessment of performance from noise-free to highly degraded inputs.

The key contributions of this work are as follows:

- We develop a noise-aware extension of the MBD framework that incorporates a lightweight noise estimator with FiLM-based feature modulation.
- We introduce a temporal consistency loss that reduces frame-to-frame flicker and improves the stability of reconstructed motion sequences.
- We evaluate a two-stage Restormer with NAMBD pipeline that improves performance under high-noise conditions.
- We conduct a systematic study across noise levels ($\sigma$ = 0 to 50) and analyse how noise affects motion decomposition quality.

## 2 Related Work

For our project, we primarily looked at five papers that deal with different approaches to recovering motion sequences from blurry images. Although each method introduces strong ideas, they share two major limitations: they assume

• *M.A. Verghese, S.S. Srikanth and T. Ufimtseva are with the Department of Computer Science, University of Toronto*

TABLE 1
PSNR (in dB) under Gaussian noise levels and the corresponding drop in quality at $\sigma = 20$.

| **Method** | $\sigma = 0$ | $\sigma = 5$ | $\sigma = 10$ | $\sigma = 20$ | **Quality drop, %** |
|---|---|---|---|---|---|
| Zhong et al. [1] | 22.87 | 22.74 | 22.28 | 20.42 | 10.71% |
| Jin et al. [3] | 30.02 | 28.71 | 25.59 | 21.32 | 29.0% |
| Zhang et al. [5] | 30.22 | 24.73 | 23.14 | 18.49 | 38.8% |

clean, noise-free input and often struggle with severe or complex motion blur. These shared limitations are what we want to address in our work by working with images in noisy conditions. We categorize these works into deterministic CNN methods, physics-based modeling, multi-modal frameworks, and transformer-based approaches.

**Deterministic CNN Approaches:** Jin et al. [3] introduces a supervised deep learning approach to recover a sequence of 7 sharp frames from a single motion-blurred image. Using the GoPro dataset [4], which averages high-frame-rate video frames to create realistic motion blur, the authors train a set of seven CNNs (one per frame), which have the same architecture but are trained with different temporal order-invariant losses to reconstruct a plausible motion sequence. To address the temporal order ambiguity, since both forward and backward sequences can explain the blur, they use an order-invariant loss between symmetric frame pairs, combined with adversarial training for perceptual sharpness. The method successfully generates temporally consistent and visually sharp sequences and outperforms prior deblurring models that only restore a single frame. However, the method remains deterministic, produces a fixed number of frames, and cannot represent multiple plausible motion trajectories.

**Physics-based Trajectory Modeling:** The authors of the paper [5] introduce a novel framework to recover exposure trajectories, which are dense, continuous, and non-linear motion paths, from a single blurry image using a physics-driven blur formation model. The researchers are also using the GoPro dataset [4] and training a motion offset estimation network with a differentiable re-blurring module to estimate object trajectories and reconstruct plausible motion paths during exposure time. Their method addresses the problem of the loss of temporal consistency and the ill-posed nature of inferring motion without ground-truth dynamic information. The approach achieves strong results on synthetic datasets, but its quadratic motion model struggles with highly complex real-world motions, and the reliance on synthetic blur introduces domain-shift issues.

**Multi-modal and Generative Frameworks:** The study [1] introduces a motion guidance representation and a two-stage network to generate multi-modal sequences. The main focus and the novelty of the research was in dealing with the temporal inconsistency problem among the reconstructed sharp images. The authors introduce a motion guidance network that quantises the optical flow into 5 directional bins that immediately converts the motion reconstruction problem into a deterministic one. Their proposed architecture involves a two stage network where the first network generates a coarse motion sequence and the second network refines it for high-quality details. A core innovation lies in this multi-modal motion guidance prediction network, built using a conditional VAE-GAN framework, which enables the model to learn a distribution of plausible motions rather than a single deterministic output and reconstruct the motion relying on each directional guidance separately. This allows them to generate multiple realistic motion sequences for the same image. They suggest 3 ways to compute the motion guidance: using GAN-based prediction, adjacent video frames, or user annotations. However, this work struggles with very severe blur, complex texture and does not handle varying noise intensity.

The authors of [6] addresses the ambiguity in frame ordering in a different way and proposes a self-supervised ordering scheme that explicitly assigns a temporal direction to each training video. They introduce the ordering of the sequence of blurry images (preceding image, current image and current image, succeeding image) and train two separate networks based on the order of the images. Each sequence is mapped into a high-dimensional latent space after which a hyperplane that separates a sequence from its reversed version is built. This strategy allows the model to learn sharper, more realistic motion sequences with unambiguous temporal ordering. In addition, the authors also introduce a real-image motion blur dataset for the image-to-video deblurring problem called real blur2vid (RB2V). This dataset covers three main categories of blurry images: face, hand and street. While HyperCUT shows impressive results under controlled lighting and consistent motion, it does not address more challenging scenarios like noise, or blur caused by longer exposures. These limitations create a gap that we aim to address.

**Transformer-based Approaches:** The authors of the paper [7] design a transformer-based framework that can handle complex, real-world blur and do not rely on simplified motion assumptions. Instead of just predicting optical flow or a sharp image, the authors proposed a Blur Interpolation Transformer (BIT) that learns to interpolate intermediate sharp frames directly. They combine multi-scale Swin Transformer blocks together with dual-end temporal supervision and temporally symmetric ensembling and achieve state-of-the-art results on synthetic datasets such as Adobe240. They argue that blur inherently encodes temporal information, and with the right architecture, it's possible to extract this motion accurately. A key contribution is the introduction of a hybrid camera system and the first real-world dataset of one-to-many blur-sharp video pairs, which significantly improves generalization to real-world blur. The model still struggles with extremely strong or highly non-uniform blur as well as fast motions.

**Baseline Selection:** We selected three baseline works: Zhong et al. [1], Jin et al. [3] and Zhang et al. [5] as they represent the main methodological directions in single-image

motion recovery: early CNN-based video extraction, physically grounded exposure-trajectory modeling and multi-modal motion decomposition. The paper [6] is strongly dependent on their dataset and their main focus on unsupervised temporal ordering differs from ours, which is on noise robustness and reconstruction quality. BiT [7] performs blur interpolation across video frames rather than inferring motion from a single blurred input, making it adjacent but not directly comparable to our setting. We then evaluated these 3 baselines on their native datasets to understand their strengths and limitations. Across all three works, the core limitations are shared: they struggle with large or complex blur, and none of the original models explicitly handle noise. Among these baselines, the Animation from Blur by Zhong et al. [1] model showed strong generalisation: even without being trained on GoPro dataset its performance degraded only marginally under increasing Gaussian noise ($\sigma$ = 5, 10, 20), unlike the other models. Therefore, we chose this paper [1] as our baseline for this work and focus on addressing one of its key limitations, its lack of noise modeling, by introducing noise-aware components into the framework.

## 3 Proposed Method

Motion blur decomposition from a single image is inherently challenging due to the loss of temporal information during exposure and the ill-posed nature of the problem. While existing methods achieve promising results on clean images, they often produce temporally inconsistent sequences and degrade significantly under realistic noise conditions. To address these limitations, we propose three independent yet complementary enhancements to the Multi-modal Blur Decomposition (MBD) framework [1] to improve both the visual quality of reconstructed sequences and the robustness of the model under noisy conditions. The first enhancement focuses on improving temporal coherence and perceptual quality by introducing an active perceptual loss module combined with a temporal consistency constraint.

Before introducing our noise-aware components, we first model the discrepancy between synthetic clean blur and real sensor degradation. In real-world scenarios, particularly in low-light environments where longer exposure times cause motion blur, images are invariably corrupted by sensor noise. To mitigate this domain gap, we explicitly model sensor imperfections by injecting gaussian noise into the blurred inputs during training. As illustrated in Figure 1, our framework moves beyond standard clean blur (center) to handle inputs with significant stochastic degradation (right). We evaluate the model's robustness by applying varying noise intensities, with standard deviations ranging from $\sigma = 0$ to $\sigma = 25$, simulating conditions ranging from ideal studio lighting to severe sensor degradation.

The second enhancement targets the sensitivity of the decomposition network to input noise. We incorporate a pre-denoising stage using the Restormer network, which effectively suppresses high-frequency sensor noise prior to motion reconstruction. The third enhancement introduces a Noise-Aware Motion Blur Decomposition (NAMBD) network. This network leverages a noise estimation module

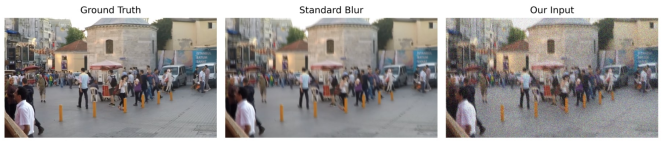



Fig. 1. Comparison of input modalities. From left to right: the clean Ground Truth image, the Standard Blur usually used in training benchmarks, and Our Input, which incorporates gaussian noise to simulate realistic sensor artifacts found in dynamic scene capture.

that predicts the noise level of the input image and conditions the feature maps via Feature-wise Linear Modulation (FiLM) [8].

Finally, we combine the Restormer pre-denoising stage with the original baseline model architecture (motion guidance predictor and MBD block) and later, instead of the original MBD block, we are using our NAMBD block to form a two-stage pipeline that jointly mitigates sensor noise and adaptively modulates features for robust motion reconstruction. The complete proposed methodology, including all three enhancements and their interactions, is illustrated in Fig. 2.

### 3.1 Enhanced Loss Function with Temporal Consistency

The first enhancement focuses on improving the visual quality and smoothness of the reconstructed animation. This allows the network to generate smoother and visually sharper sequences by explicitly penalizing abrupt changes between consecutive frames. The original baseline model included a perceptual loss module. We enable this module and augment it with a temporal consistency loss, which enforces coherent frame evolution across the reconstructed sequence.

Given a predicted sequence of sharp frames $\{\hat{I}_t\}_{t=1}^{T}$ and ground-truth frames $\{I_t\}_{t=1}^{T}$, we define the enhanced loss as:

$$\mathcal{L}_{total} = \mathcal{L}_{reconstruction} + \lambda_{perc}\mathcal{L}_{perceptual} + \lambda_{temp}\mathcal{L}_{temporal}, \quad (1)$$

where $\mathcal{L}_{reconstruction}$ is the standard $\ell_2$ reconstruction loss, $\mathcal{L}_{perceptual}$ is the perceptual loss computed using a pre-trained VGG network, and $\mathcal{L}_{temporal}$ enforces smooth transitions between consecutive frames. $\lambda_{perc}$ and $\lambda_{temp}$ are weighting factors tuned empirically.

This enhancement significantly reduces flicker and unstable motion trajectories, particularly in noisy conditions.

### 3.2 Restormer Pre-Denoising

The second enhancement addresses the sensitivity of existing motion blur decomposition models to sensor noise. This stage ensures that the decomposition network receives cleaner inputs, maintaining accurate motion estimation even under moderate to high noise levels. We introduce a pre-denoising stage using Restormer [9], a state-of-the-art transformer-based denoiser. In this approach, the noisy blurred input image is first passed through the Restormer network to suppress high-frequency noise before decomposition.

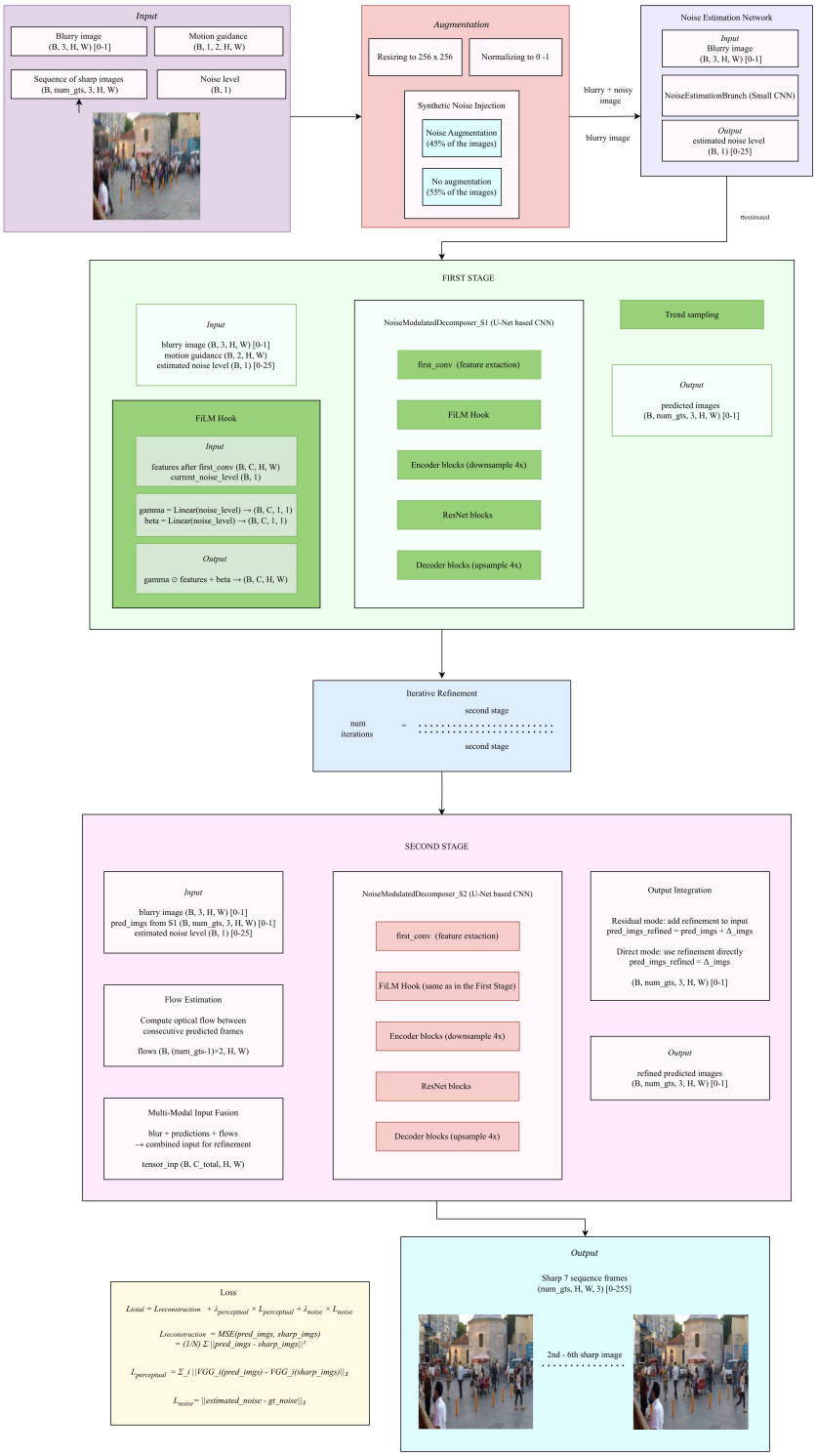



Fig. 2. Overview of the proposed methodology

Let $I_b$ denote the input blurry image with additive Gaussian noise. The denoised output is:

$$I_d = \text{Restormer}(I_b). \tag{2}$$

This denoised image $I_d$ is then fed into the baseline decomposition network. This simple but effective two-stage approach improves model robustness under moderate and high noise levels while preserving motion cues.

### 3.3 Noise-Aware Motion Blur Decomposition (NAMBD)

The third enhancement introduces noise-adaptive feature modulation to the decomposition network. We design a Noise-Aware Motion Blur Decomposition (NAMBD) model

where a lightweight noise estimator predicts the noise level $\hat{\sigma}$ of the input image. These predictions condition the CNN features via Feature-wise Linear Modulation (FiLM) layers [8], allowing dynamic adjustment to the network's processing based on the estimated noise. By dynamically modulating feature representations based on the predicted noise, the network adapts its processing to both clean and noisy images, resulting in stable motion reconstruction across a wide range of real-world scenarios.

For a feature map $F$ in the decomposition network, the FiLM-modulated feature $F'$ is given by:

$$F' = \gamma(\hat{\sigma}) \odot F + \beta(\hat{\sigma}), \tag{3}$$

where $\gamma(\hat{\sigma})$ and $\beta(\hat{\sigma})$ are scale and shift parameters produced by the noise estimator network, and $\odot$ denotes element-wise multiplication. Here, $\gamma(\hat{\sigma})$ and $\beta(\hat{\sigma})$ act as learned modulation parameters that determine how strongly each feature channel should be scaled or shifted based on the estimated noise level. This mechanism enables the network to maintain high-quality motion reconstruction across varying noise intensities.

### 3.4 Two-Stage NAMBD with Restormer

To further strengthen robustness under severe noise, we combine the second and third enhancements into a two-stage inference pipeline. This ordering - denoise first, then decompose - ensures that the motion estimation stages receive features where noise has been suppressed but blur structure remains intact. During inference, the input noisy blurry image $I_b$ is first denoised by Restormer, producing $I_d$, which is then passed through the NAMBD network for motion blur decomposition. This combined approach effectively mitigates high-frequency noise while dynamically adapting to residual artifacts, resulting in more stable and temporally coherent motion sequences.

The overall architecture of our proposed methodology is illustrated in Fig. 2, depicting the noise-aware feature modulation.

### 3.5 Implementation Details

We implemented the proposed framework using PyTorch on the University of Toronto SLURM cluster NVIDIA GPUs. The model was trained on the GoPro dataset for 100 epochs with a batch size of 1. We used the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$ and weight decay of $1 \times 10^{-4}$, decayed using a cosine annealing scheduler with $\eta_{min} = 10^{-6}$. To address sensor noise, we enabled noise-aware training with an augmentation probability of 0.45. During these steps, Gaussian noise was injected dynamically with $\sigma$ sampled uniformly from the range $[5, 25]$. The motion decomposer consists of two stages, each containing 12 bottleneck blocks with an expansion factor of 128. Optical flow was initialized using a pre-trained RAFT checkpoint from the original work [1].

## 4 Experimental Results

In this section, we evaluate the performance of our proposed enhancements to motion blur decomposition under varying noise conditions. We systematically analyze the impact of the improved loss function, the Noise-Aware Motion Blur Decomposition (NAMBD), the Restormer pre-denoising stage, and their combination, using standard metrics including PSNR, SSIM, and LPIPS. Our experiments aim to quantify both the perceptual quality of reconstructed frames and the robustness of the models to sensor noise.

### 4.1 Improved Loss Function

We first isolate the contribution of the modified loss function. The original motion-decomposition model supervises each predicted frame independently using pixel-wise and perceptual losses, with no constraint enforcing cross-frame consistency. This leads to sequence flicker and unstable temporal evolution, especially under noise. By adding a temporal consistency term, we get smoother, more coherent trajectories.

To verify that this additional constraint does not hinder optimization, we analyze the training dynamics over 70 epochs. **Figure 3(a)** illustrates the training and validation loss curves. The steady decline of the validation loss (red line) alongside the training loss (blue line) indicates that the model generalizes well and has stable optimization. This confirms that the spatio-temporal loss stabilises the learning process without inducing overfitting, even when handling the complex task of enforcing temporal coherence.

Quantitatively, the enhanced loss configuration provides the strongest single-stage performance (Table 2). Relative to the baseline MBD model on clean GoPro inputs ($\sigma = 0$), PSNR improves from 22.87 dB to 28.68 dB (≈25%), SSIM increases from 0.753 to 0.880 (≈17%), and LPIPS decreases from 0.217 to 0.148 (≈32%). These gains confirm that the temporal consistency term not only stabilises training but also produces sharper and more coherent reconstructed sequences.

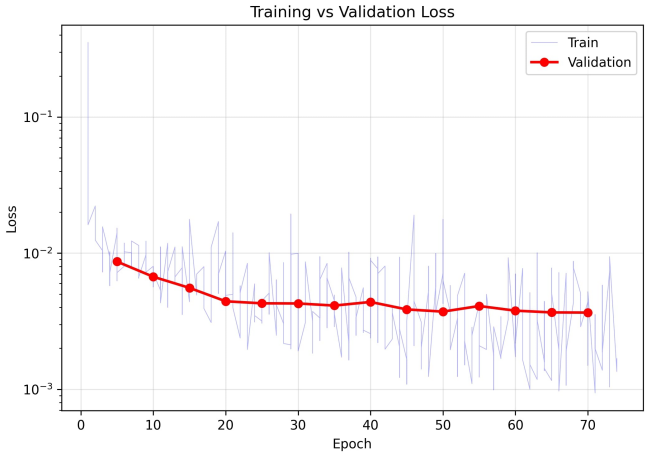
TABLE 2
Spatio-Temporal Perceptual Loss

| Metric | PSNR (dB) | SSIM | LPIPS |
|---|---|---|---|
| Value | 28.6758 | 0.8797 | 0.1484 |

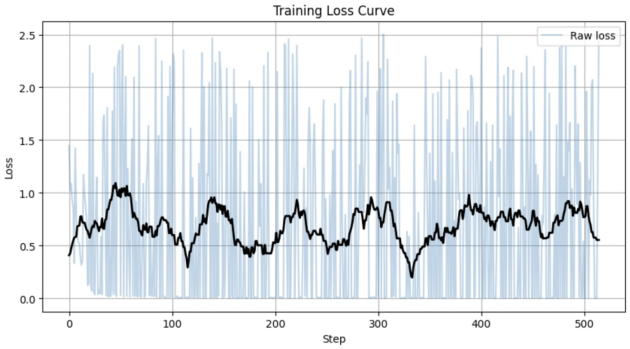
### 4.2 Noise-Aware Motion Blur Decomposition (NAMBD)

NAMBD performs strongly across all noise levels, consistently outperforming the baseline MBD model in PSNR and SSIM (Table 3). The PSNR gain grows from 1.9 dB at $\sigma = 0$ to nearly 6 dB at $\sigma = 50$, and SSIM is higher for all $\sigma$. For LPIPS, NAMBD is slightly worse than the baseline on clean and very low-noise inputs ($\sigma = 0, 5$), but from $\sigma \geq 10$ onward it yields substantially lower LPIPS, indicating better perceptual quality.

This behaviour is expected: LPIPS is highly sensitive to fine texture and local contrast, and the FiLM layers apply mild feature modulation even when the true noise level is close to zero. Because the noise estimator tends to slightly over-predict $\sigma$ on clean images, the resulting modulation smooths subtle high-frequency details that LPIPS penalises more than PSNR or SSIM. Once noise increases, however, this same modulation becomes beneficial, leading to the improved LPIPS scores observed for $\sigma \geq 10$.




(a) Improved Loss Function (Training vs. Validation)




(b) NAMBD Training Loss (Raw vs. Moving Avg)

Fig. 3. Training stability analysis. **(a)** The Improved Loss Function shows consistent convergence between training (blue) and validation (red) loss, indicating good generalization. **(b)** The NAMBD model demonstrates a stable downward trend in the moving average (black) despite the high variance of the raw loss (blue) caused by stochastic noise injection.

TABLE 3
Comparison across models at different noise levels

| $\sigma$ | Baseline [1](on GoPro dataset) | | | NAMBD | | | Restormer + AFB [1] | | | Restormer + NAMBD | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS | PSNR (dB) | SSIM | LPIPS |
| 0 | 22.87 | 0.753 | 0.217 | **24.75** | 0.764 | 0.268 | 20.61 | 0.56 | 0.39 | 22.06 | 0.64 | 0.43 |
| 5 | 22.74 | 0.695 | 0.228 | 24.71 | 0.762 | 0.266 | 21.24 | 0.65 | 0.35 | 22.74 | 0.69 | 0.39 |
| 10 | 22.28 | 0.612 | 0.274 | 24.59 | 0.754 | 0.259 | 22.36 | 0.71 | 0.27 | 24.25 | 0.74 | 0.31 |
| 20 | 20.42 | 0.477 | 0.404 | 24.10 | 0.718 | 0.240 | 22.23 | 0.70 | 0.26 | 24.15 | 0.73 | 0.30 |
| 30 | 18.31 | 0.377 | 0.542 | 23.25 | 0.638 | 0.294 | 21.88 | 0.62 | 0.30 | 23.65 | 0.65 | 0.32 |
| 40 | 16.59 | 0.305 | 0.672 | 22.23 | 0.557 | 0.376 | 20.70 | 0.50 | 0.44 | 23.00 | 0.65 | 0.33 |
| 50 | 15.21 | 0.254 | 0.793 | 21.15 | 0.487 | 0.459 | 19.77 | 0.43 | 0.50 | 22.05 | 0.60 | 0.40 |

To verify that adding these adaptive modulation layers does not introduce training instability, we analyzed the optimisation trajectory. Figure 3(b) plots the training loss for the NAMBD model. Despite the stochastic nature of the noisy inputs, the moving average (black line) shows a smooth and consistent downward trend, confirming that the network effectively learns to disentangle noise from motion blur without divergence.

At very high noise levels, however, feature modulation alone cannot recover motion cues that have already been overwhelmed by noise, and the gains over the baseline begin to saturate. This motivates the use of a stronger pre-denoising stage in the two-stage Restormer with NAMBD pipeline.

### 4.3 Restormer + Animation From Blur (AFB) [1]

We evaluate Restormer purely as a denoiser applied to blurry–noisy inputs before passing them to the main motion reconstruction algorithm from the baseline paper [1]. Across all noise levels, Restormer produces cleaner images visually, but this comes at the cost of oversmoothing edges and motion-dependent structures.

This behaviour is reflected directly in Table 3, Restormer achieves the lowest PSNR and SSIM of all methods, even at low noise. This confirms that Restormer's strong spatial denoising is not well aligned with the type of blur–motion structure required for downstream decomposition. Restormer suppresses the additive noise, but on the low noise level it also can suppress high-frequency motion cues that our decomposition model relies on, leading to degraded reconstruction fidelity when the noise level is low

For all experiments, we use the publicly released Restormer model trained for Gaussian color denoising at $\sigma = 25$. This model is optimized for additive white Gaussian noise with standard deviation up to 25 (in the 0–255 intensity range). Since our decomposition model is also trained on noise levels spanning $\sigma \in [0, 25]$, we ensure that the pre-denoising stage operates under noise conditions consistent with our training setup.

### 4.4 Restormer + NAMBD

Restormer + NAMBD is designed for the regime where noise becomes strong enough that feature modulation alone starts to saturate. This two-stage approach helps stabilise performance under moderate and high noise ($\sigma \geq 20$), even though it does not produce the best scores at low noise.

At $\sigma = 20$, its PSNR (24.15 dB) closely matches NAMBD (24.10 dB) and remains far above the baseline (20.42 dB). As the noise increases, the benefit becomes clearer: at $\sigma = 40$ and $\sigma = 50$, the combined model reaches 23.00 dB and 22.05 dB, outperforming NAMBD by nearly 1 dB and exceeding the baseline by more than 6 dB. SSIM shows a similar trend:

(a) Input Blurry Image

(b) Reconstructed Sequence

Fig. 4. Qualitative results. (a) Input image corrupted by blur and noise. (b) Our reconstructed sequence presented. The proposed method successfully suppresses noise while maintaining temporal coherence across the 7 generated frames.

lower than NAMBD at small $\sigma$, but substantially more stable than the baseline or Restormer alone at higher noise.

Restormer mitigates the high-frequency sensor noise that NAMBD alone cannot suppress, while NAMBD's FiLM modulation adapts to the residual artifacts left after denoising. Together, these components slow the rate of degradation at $\sigma \geq 30$, where NAMBD begins to lose motion-relevant structure. At low noise levels, however, the two-stage model performs slightly worse than NAMBD. When the input is already clean, Restormer introduces mild smoothing that removes fine texture and edges that NAMBD could have used, producing a small but measurable drop in PSNR and SSIM. This behaviour reflects an inherent tension between preserving detail and removing noise, particularly in regimes where the input does not benefit from pre-denoising.

Despite this limitation, the two-stage setup is the most reliable choice in realistic settings, where noise levels are not fixed or predictable. Under strong noise, it consistently outperforms both NAMBD and the baseline across all metrics, confirming that combining denoising with adaptive modulation is more effective than relying on either mechanism alone.

Overall, our experiments demonstrate that each enhancement contributes a complementary benefit: the improved loss reduces flicker, NAMBD strengthens performance under moderate noise, and Restormer provides a necessary buffer when noise becomes dominant. The combined Restormer + NAMBD model delivers the most stable results across all tested noise levels, validating it as a practical solution for motion reconstruction under real-world imaging conditions.

## 5 Limitations

While our enhancements improve model strength under noisy conditions, several limitations remain. First, our experiments rely on additive Gaussian noise as a stand-in for real sensor noise. In practice, noise can vary with exposure, ISO, sensor architecture, and color channels, and often includes non-Gaussian artifacts such as shot noise and compression effects. Because of this gap, additional testing on real noisy images and diverse camera pipelines is needed to fully validate the approach.

Second, although the temporal consistency loss reduces flicker, it can also smooth out motion more than desired in scenes with abrupt or high-frequency movement. When ground-truth motion is limited or ambiguous, the model may prefer overly stable trajectories, which can dampen fine motion details.

Finally, like the baseline MBD framework, our method inherits the constraint of producing a fixed number of output frames. This limits its ability to model longer motion sequences or motions with variable duration, which may be important in real-world applications.

## 6 Future Work

A natural next step is to make the system noise-adaptive. Our results show that different parts of the model behave better at different noise levels: Restormer works well on cleaner inputs, while the two-stage Restormer + NAMBD setup holds up better as noise increases. This suggests adding a small noise estimator during inference that decides which branch to use. Such a mechanism would avoid unnecessary denoising on clean images while still giving the model extra support when the input is heavily degraded.

Another direction is to move beyond the simple Gaussian noise we used in our experiments. Building a more realistic noise model would make the method easier to apply to images captured in everyday settings.

It may also be valuable to bring denoising and motion decomposition closer together. Right now, Restormer acts as a separate preprocessing step, which sometimes removes motion-related details along with noise. A joint, end-to-end model could preserve more of the information needed for motion reconstruction.

Finally, improvements in temporal modeling could help recover longer and more complex motion sequences. Approaches based on transformers, recurrent warping, or diffusion models may offer more flexibility than the fixed-length sequences produced by the current architecture. Testing on real blur-sharp datasets with natural sensor noise would also help bridge the gap between our synthetic experiments and practical use cases.

## 7 Conclusion

In this work, we introduced three enhancements to the Animation from Blur framework to make motion reconstruction more reliable under noisy conditions. Adding a temporal consistency term helped reduce flicker between frames, the NAMBD module let the model adapt to different noise levels, and the Restormer stage improved results when the input was heavily degraded. Across our experiments, NAMBD handled low and moderate noise well, while the Restormer + NAMBD combination worked best once noise increased. Together, these results suggest that including some form of noise awareness, whether through estimation, conditioning, or preprocessing, makes a clear difference in recovering motion from a single blurry image. In terms of PSNR, SSIM, and perceptual quality, our methods consistently outperform the original baseline. Overall, our study points toward motion reconstruction models that respond more naturally to the kinds of noisy images common in real settings. Our github repository is also referenced below [10].

## References


[1] Z. Zhong, T. Sun, W. Wu, Y. Zheng, J. Lin, and I. Sato, "Animation from blur: Multi-modal blur decomposition with motion guidance," in *Proc. ECCV*, 2022, pp. 1–17.

[2] R. Li, S. Yang, D. A. Ross, and A. Kanazawa, "Learn to dance with aist++: Music conditioned 3d dance generation," 2021.

[3] M. Jin, M. Meishvili, and P. Favaro, "Learning to extract a video sequence from a single motion blurred image," in *Proc. CVPR*, 2018, pp. 1–10.

[4] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. CVPR*, 2017, pp. 1–9.

[5] Y. Zhang, L. Wang, S. Maybank, and D. Tao, "Exposure trajectory recovery from motion blur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 1–14, 2022.

[6] B.-D. Pham, P. Tran, A. Tran, C. Pham, R. Nguyen, and M. Hoai, "Hypercut: Video sequence from a single blurry image using unsupervised ordering," in *Proc. CVPR*, 2023, pp. 1–12.

[7] Z. Zhong, M. Cao, X. Ji, Y. Zheng, and I. Sato, "Blur interpolation transformer for real-world motion from blur," in *Proc. CVPR*, 2022, pp. 1–11.

[8] E. Perez, F. Strub, H. D. Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proc. AAAI*, vol. 32, no. 1, 2018.

[9] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. CVPR*, 2022, pp. 5728–5739.

[10] "Codebase github repository link," animation_from_blur_and_noise.