**International Journal of Advance and Innovative Research**
Volume 13, Issue 1 (XIII): January – March 2026
ISSN 2394 - 7780

# THE EVOLUTION OF VIDEO DEBLURRING: CLASSICAL APPROACHES, DEEP LEARNING METHODS, AND EMERGING TRENDS

**Nikhil K. Pawanikar[1] and R. Srivaramangai[2]**
[1]Department of Information Technology, University of Mumbai, India, nikhilpawanikar@gmail.com
[2]Department of Information Technology, University of Mumbai, India, rsrimangai@gmail.com
*Corresponding author: Nikhil K. Pawanikar, Department of Information Technology, University of Mumbai, India. Email: nikhilpawanikar@gmail.com

## ABSTRACT
*Video deblurring is a foundational challenge in computer vision that has various applications like surveillance, autonomous navigation, mobile imaging, and multimedia processing. Motion blur is a phenomenon of image degradation, which is caused by the movement of the camera, fast object motion, low-light exposure, or rapid scene dynamics, and its elimination demands the reconstruction of sharp spatial details alongside temporal consistency preservation. Video deblurring has been through a tremendous change in the last decade of research, which has gone from the classical optimization-based formulations and kernel estimation methods to modern learning-driven architectures that are able to model complex, spatially variant, and temporally coherent blur. This article reviews about 50 major works on video deblurring published during 2015-2025, including both traditional and deep learning methods.*

*We have classified previous works into a single overall taxonomy that covers classical techniques (variational models, blind deconvolution, Fourier-domain aggregation, optical-flow-guided deblurring), convolutional neural network designs, recurrent and sequential models, alignment-driven and flow-based approaches, Transformer-based architectures, GAN-driven perceptual enhancement, event-camera-based deblurring, frequency and wavelet domain methods, and hybrid frameworks that combine multiple computational paradigms. Together with this taxonomy, we illustrate the main benchmarks and evaluation metrics. In Table 5, we present a single quantitative comparison across GoPro, REDS, and DVD datasets.*

*Our analysis brings into focus the main trends in accuracy, efficiency, generalization to real-world scenarios, and robustness, whereas challenges still remain in the open area of non-uniform blur, extreme motion, lowlight conditions, and domain adaptation. Finally, we propose that new studies should be directed toward multimodal fusion, physics aware modeling, self-supervised learning, diffusion-based generative restoration, and resource-efficient architectures for edge devices. This survey is a complete guide for the readers who want to understand the progress of video deblurring using traditional, learning-based, and future technology methods.*

***Keywords:*** *Video deblurring, motion blur, classical optimization methods, deep learning, optical flow, Transformer networks, event-based deblurring, hybrid restoration models, spatio-temporal consistency, image and video restoration.*

## 1. INTRODUCTION
Video deblurring has become a major necessity as the volume of visual data captured by smartphones, surveillance systems, autonomous platforms, and consumer imaging devices is rapidly increasing. Motion blur appears when the camera is shaken, the object moves, exposure is required in low light, or the scene has very dynamic elements, which results in smeared structures and selective loss of details that degrade not only human perception but also the performance of the downstream computer vision tasks. Single-image deblurring can be different from video deblurring. When it is a video, the deblurring algorithm not only has to bring back sharp spatial details, but also it should maintain content consistency between different frames. Thus, the problem becomes much more difficult due to factors such as motion variation, depth discontinuities and exposure related effects.

At first, the focus of research switches towards classical optimization based frameworks like blind deconvolution, variational formulations, optical-flow, guided alignment, frequency-domain aggregation. These approaches rely on hand-crafted priors and explicit blur models. However, they are less effective with spatially varying blur, inaccurate motion estimation and are inefficient computationally when dealing with complex scenes. Deep learning has come up with a new idea that the blur formation can be modeled in a data-driven manner, thus opening the door to CNN based architectures, recurrent temporal models, flow-guided pipelines and transformer based designs with long-range temporal reasoning. The most recent research goes a step further by combining event-camera modalities, wavelet- and frequency-domain priors, and hybrid attention mechanisms, thus facilitating better robustness to high-speed motion, low-light conditions, and real-world blur complexity.

ISSN 2394 - 7780

However, there are still some major problems that have to be solved. Most synthetic datasets especially those obtained by frame averaging do not reflect the natural behavior of the sensor, exposure changes, or rolling shutter distortions, which is why the resulting models tend not to perform well when applied to real scenes. Extreme non-uniform blur, abrupt motion discontinuities, occlusion, and noisy low light conditions are still tough problems for deep learning as well as hybrid methods. Furthermore, it is still a great challenge to achieve real-time video deblurring on edge devices due to the computational cost of modern deep neural networks. Contributions of this Survey:

1. We have developed a concise & unified taxonomy that covers classical, CNN, RNN, flow-based, transformer, GAN, event-driven, and hybrid/frequency-domain video deblurring methods.
2. We highlight important progress in various video deblurring methods in 2015-2025 based on how each type deals with blur complexity, motion variation, and temporal coherence.
3. We give a straightforward summary of the most popular datasets synthetics (GoPro, DVD, REDS), real (BSD, RAW-based), event-based (Blur-DVS, HQF, REVD), and stereo (StereoBlur, LFOVIAS3DPh2) - and also examine their characteristics and how they are used.
4. We list evaluation metrics & neatly arrange fidelity (PSNR, SSIM), perceptual (LPIPS, PI), temporal (VFID), motion-aware (EPE), and efficiency metrics, together with suggested evaluation procedures.
5. We analyze and discuss the resulting trends of fidelity, perceptual quality, temporal stability, generalization, and efficiency in different model families.
6. We put forward outstanding issues and possible directions for further investigation in the areas of unified temporal modeling, multi-domain priors, event RGB fusion, self-supervision, diffusion-based restoration, and efficient deployment.

We review over 60 papers from 2015 to 2025 focusing on major publication sources; Fig. 1 presents a summary of the number of publications for each year, and Fig. 2 displays the distribution of publishers. This rise in trend (Fig. 1) and concentration among the top venues (Fig. 2) serve as the impetus behind our taxonomy (Section 3). The rest of the paper is structured as follows: Section 2 describes the basics of how video blur forms. Section 3 introduces our proposed taxonomy which includes both classical and learning based methods. Section 4 looks at major benchmark datasets while Section 5 is devoted to evaluation metrics. Section 6 conducts a comparison of different methods in terms of quality, perceptual quality, temporal stability, and efficiency. Section 7 offers a wider discussion and Section 8 sketches the future research possibilities. Section 9 wraps up the survey.

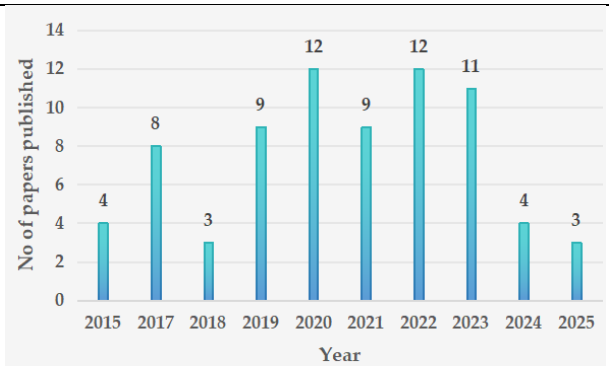


**Figure. 1.** Year-wise distribution of the reviewed publications (2015–2025).

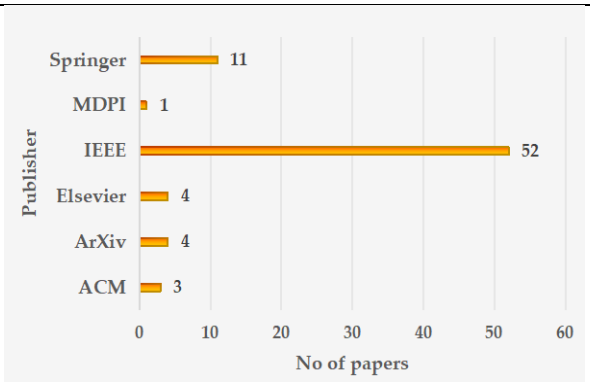


**Figure. 2.** Source distribution of the reviewed publications.

## 2. BACKGROUND: MOTION BLUR FORMATION AND CLASSICAL VIDEO DEBLURRING

### 2.1 Motion Blur Formation in Videos

Motion blur happens when a camera captures and adds light while the camera or scene objects move, resulting in pixels being smeared to varying intensities. The effect of motion blur is especially obvious in situations where the camera is handheld, the scene is outdoor and dynamic, or low-light conditions that necessitates long exposure, as can be seen in the initial studies of the formation of dynamic scene blur [1]. When different objects move independently, the motion is non-linear, or there are changes of depth in the scene, spatially varying blur will result [2], [3].

ISSN 2394 - 7780

Moreover, temporal inconsistencies make it even harder: the blur that one frame exhibits may differ from that of the adjacent frame, and if the frames are heavily blurred, misalignment will induce ghosting artifacts where motion estimation is not successful [4].

### 2.2 Challenges Unique to Video Deblurring

Video deblurring is a process of recovering clear image details from the blurred ones and at the same time maintaining the temporal coherence among the video frames. Here is a list of the main challenges the video deblurring process has to put up with: **Spatially varying and non-linear blur**, which is beyond the scope of conventional global kernels to model [1], [5]. The **temporal misalignment,** which is when optical flow estimation becomes unfeasible due to heavy blur, thus resulting in unstable multi-frame fusion [4]. The existence of **low-light noise and exposure variability**, which in turn make blur estimation less reliable and restoration errors more noticeable [2]. The **high computational** cost of the classical iterative methods, since each frame requires expensive optimization [1], [6]. Such constraints led to learning-based models being adopted which are able to implicitly model complex motion and blur patterns.

### 2.3 Classical Video Deblurring Methods

Before deep learning took the spotlight, video deblurring algorithms were mainly dependent on explicit motion modeling, kernel estimation, and optimization based formulations. These approaches represent the main components of the evolutionary arc of video deblurring research.

#### *2.3.1 Kernel-Based and Blind Deconvolution Approaches*

Traditional blind deconvolution methods estimated the variations in space of blur kernels and latent sharp frames at the same time. Different methods were used such as pixel-wise motion cues [1], segmentation guided kernels [6], or structured blur models for stereo motion [5].

#### *2.3.2 Variational and Energy-Minimization Frameworks*

Variational methods aimed at solving the multi-term energy functions which combined data fidelity, spatial/temporal regularization, and motion-aware constraints. They were thus successful in controlled environments, but they were still quite sensitive to the initial guess and were computationally heavy [2], [6].

#### *2.3.3 Optical-Flow-Based Deblurring*

Optical flow was the most common method for the tasks of frame alignment, kernel estimation, and cross-frame sharpness propagation [1], [4], [5]. Unfortunately, blurred inputs make the flow less accurate, creating a cycle of dependency: poor flow results in poor deblurring, and the other way round.

#### *2.3.4 Frequency- and Transform-Domain Approaches*

Fourier Burst Accumulation combined multiple frames in the frequency domain to reduce the effect of handheld shake [3], whereas wavelet-based methods captured multiscale structures in the formation of blur [7].

### 2.4 Limitations of Classical Models and Motivation for Learning-Based Approaches

Among classical approaches, several inherent limitations noted are: **Hand-crafted priors** that do not properly generalize to complex real, world motion situations [1], [6]; **Lack of ability to model non-uniform blur** locally in scenes with objects in motion or at depth discontinuities [5]; **Dependence on accurate optical flow** that breaks down in presence of heavy blur or homogeneous areas [4]; **High Runtime** due to iterative optimization pipelines [1]; **Poor real-world generalization** resulting from the use of simple synthetic assumptions of blur [2], [5]. These difficulties have been a direct cause of the transition into CNN, based, recurrent, transformer, based, and hybrid video deblurring architectures.

## 3. Taxonomy of Video Deblurring Techniques

Research into video deblurring has changed from traditional optimization-based methods to deep learning architectures that are able to deal with complex, spatially varied, and temporally inconsistent blur. Based on the research works presented by this survey, a majority of current methods can be carved into eight main complementary categories: classical models, CNN-based frameworks, recurrent architectures, methods guided by optical flow and alignment, transformer-based designs, GAN-based techniques, event-driven approaches, and frequency-domain or hybrid models. This classification not only charts the evolution of the field over time but also the functional differences between the major methodological families.

### 3.1 Classical Approaches

Classical approaches are based on explicit blur modeling, blur estimation, or motion estimation. One of the first attempts to estimate kernels pixel-wise gave strong baselines for dealing with spatially varying blur in dynamic scenes [1]. Progressive low-light fusion methods solved the problem of severely underexposed videos by utilizing perceptual cues [2]. Frequency-domain fusion like Fourier Burst Accumulation enhanced the sharpness

ISSN 2394 - 7780

of the handheld capture [3]. While these methods gave great insights, they were not able to handle complex scenarios such as non-linear motion, depth discontinuities, or extremely blurred scenes.

### 3.2 CNN-Based Video Deblurring

CNNs enabled data-driven learning of blur representations. Initially, encoder-decoder architectures trained on high-frame-rate data showed significant gains over classic methods [8]. Further, architectures with deformable convolutions and attention-based fusion were developed to better detect spatial variation and multi-scale motion patterns [9]. However, despite being very good at spatial modeling, CNNs have a limited capacity for long-range temporal reasoning.

### 3.3 Recurrent Neural Network (RNN) Approaches

Video deblurring based on RNN models is able to capture temporal coherence by considering the sequential dependencies across frames. One of the major contribution makes use of intra-frame iterative refinement, which enables recurrent hidden-state updation that progressively improves the reconstruction quality [10]. Another work using recurrent neural network (RNN) structure tries to achieve stronger robustness to motion-blur by integrating Another work that uses recurrent neural network (RNN) architecture tries to get better robustness to motion-blur by combining blur-invariant motion estimation with recurrent propagation, hence temporal feature aggregation can be used even in various blur conditions [11]. However, RNNs are very effective in understanding short- and mid-range temporal dependencies but they might have problems with very long sequences due to gradient attenuation and temporal noise accumulation.

### 3.4 Optical-Flow and Alignment-Based Methods

Essentially, flow-motion estimation-based approaches use motion estimation to either directly align the frames or to offer a guiding signal for kernel estimation. The very first deep flow estimation architectures enabled the motion compensation of slightly blurred images to be more efficient [4]. Single-channel networks such as VDFlow bundle optical flow estimation and video deblurring into one network, thus improving the synergy between motion estimation and restoration performance [12]. Nevertheless, these methods not only exhibit a severe decrease in performance when the blur is very high but also their motion estimation fails to operate, thus the intrinsic circular dependency between flow accuracy and blur severity is highlighted.

### 3.5 Transformer-Based Video Deblurring

Transformers bring in global spatio-temporal attention which results in a more expressive representation of long-range dependencies than what CNNs or RNNs can offer. Flow-guided sparse transformers cleverly harmonize motion cues with windowed attention in order to keep the computational overhead low and, at the same time, provide temporal context [13]. VDTR, a fully attention-based design, further boosts restoration quality by using temporal and spatial self-attention to recover high-frequency details in the most difficult sequences [14]. These models not only achieve strong temporal stability and performance but are generally more demanding in terms of computation than recurrent architectures.

### 3.6 GAN-Based Video Deblurring

GAN-based methods focus on perceptual sharpness by attempting to reproduce texture-rich representations. The first adversarial approach used 3D spatio-temporal convolutions to help the model restore motion continuity better [15]. Later, DeblurGAN-v2 enabled real-time deblurring by using a faster backbone and a multi-scale discriminator to drastically cut down the calculations without losing the perceptual fidelity [16]. GAN-based methods improve the visual realism, however details fabricated by the model or temporal inconsistencies might be caused if the discriminator is not carefully designed.

### 3.7 Event-Driven Video Deblurring

Event cameras have a temporal resolution at the level of microseconds. They record changes asynchronously and thus restoration in extremely rapid motion conditions is theoretically possible. One of the initial models merged event streams with CNN-based frame reconstructions in order to recover sharp frames out of fast motion [17]. The latest paper introduces frequency-aware event fusion. By using event gradients to get features from the event domain real-world performance, especially in high-speed sequences, is significantly improved [18]. RGB-driven methods fail to capture scenarios, where event-driven methods flourished, however the latter require the usage of specialized sensor hardware and robust event denoising methods.

### 3.8 Frequency-Domain and Hybrid Models

Hybrid video deblurring methods combine spatial, temporal, and frequency-domain priors in order to become more robust against a variety of blur conditions. Wavelet-aware dynamic transformer models combine the multi-scale wavelet decomposition with cross-frame attention to restore high-frequency components more accurately [19]. Other hybrid architectures either gather the nearest sharp features or use multi-domain fusion to

ISSN 2394 - 7780

adaptively tackle different blur patterns and motion intensities across frames [20]. Usually, these models have better generalization capabilities than the purely spatial ones, but their training is more complicated, and they require more memory.

### 4.1 OVERVIEW OF BENCHMARK DATASETS

Video deblurring datasets may be separated into several categories as: **Synthetic blur datasets**, created by averaging high-frame-rate (HFR) video frames or using algorithmic motion kernels. These datasets have the advantage of providing perfect blurry sharp alignment and generating a large volume of training data. **Real-world blur datasets**, taken using beam splitters, HFR RAW pipelines, or event cameras, thus having realistic blur properties but generally fewer sequences. **Event-camera datasets**, including disparity, depth, or multi-sensor information. **Stereo and multi-modal datasets**, giving microsecond-resolution event streams that are aligned with blurred or sharp frames. Here, we gather the most influential datasets from these categories and refer to the original papers that first introduced them.

### 4.2 Synthetic Blur Datasets

#### *4.2.1 GoPro*

The GoPro dataset introduced one of the earliest large-scale paired blurry–sharp benchmarks using 240 fps high-speed video averaged to simulate realistic motion blur [21]. It is extensively used in CNN and recurrent video deblurring methods, including early encoder–decoder networks and deformable alignment approaches [8], [9], [10], [12].

#### *4.2.2 DVD (Deep Video Deblurring Dataset)*

DVD provides handheld blur synthesized from HFR recordings and was introduced alongside the Deep Video Deblurring model [8]. It remains a widely used dataset for methods relying on temporally adjacent frames.

#### *4.2.3 REDS*

REDS, released for NTIRE 2019, offers 120 fps video sequences with realistic blur synthesis suitable for evaluating multi-frame and transformer-based approaches [22]. It is commonly used by deformable convolution and transformer models [9], [14].

#### *4.2.4 Vimeo-90K*

Although originally designed for interpolation and temporal learning tasks, Vimeo-90K provides frame triplets often used for auxiliary training in motion-guided or flow-based video deblurring [23], [12].

### 4.3 Real-World Blur Datasets

#### *4.3.1 BSD — Beam-Splitter Dataset*

BSD uses a beam-splitter system to acquire real blurry–sharp frame pairs with perfect temporal synchronization [24]. It is essential for evaluating generalization to real-world motion, sensor noise, and exposure patterns.

#### *4.3.2 RealBlur (Image Dataset)*

RealBlur contains real blurry–sharp image pairs captured with controlled motion and exposure conditions [25]. Although primarily an image dataset, it is widely used to assess cross-domain robustness of video deblurring models trained on synthetic blur [12], [14], [19].

#### *4.3.3 RAWBlur/ HFR-RAW Dataset*

RAWBlur uses high-speed RAW sensor captures to produce realistic blur with sensor-level noise, nonlinear exposure response, and scene-dependent blur effects [26]. It supports physics-aware modeling and transformer-based pipelines that require accurate blur formation data.

### 4.4 Event-Camera Datasets

Event-based video deblurring relies on event streams captured at microsecond resolution. Multiple studies leverage event datasets to restore videos suffering from extreme motion blur.

#### *4.4.1 REVD (Real Event Video Deblurring Dataset)*

This dataset provides real blurry video frames synchronized with high-resolution event streams, enabling evaluation of event-assisted and hybrid deblurring models [27].

#### *4.4.2 Blur-DVS Dataset*

Blur-DVS contains real and synthetic blurred frames paired with DAVIS240C event streams, introduced with event-based motion deblurring research. It is among the earliest datasets enabling joint frame-event reconstruction [28].

ISSN 2394 - 7780

#### *4.4.3 HQF (High-Quality Frames) Dataset*

HQF includes sharp frames aligned with DAVIS events and is widely used as a standard benchmark for event-driven reconstruction models. Although not tied to a single canonical paper, it is most commonly cited through the event-based motion deblurring community [29].

### 4.5 Stereo and Multi-Modal Datasets

#### *4.5.1 StereoBlur Dataset*

StereoBlur provides paired left-right blurry and sharp frames (20,637 pairs) introduced with DAVANET, enabling stereo-aware deblurring models to exploit parallax, disparity cues, and depth variations [30]. It is used to evaluate disparity-aware and multi-view video deblurring models.

#### *4.5.2 LFOVIAS3DPh2 / Stereo 3D Datasets*

These datasets provide stereo video frames with depth variations, allowing transformer-based or multi-view alignment models to be evaluated on cross-view temporal consistency. While originally developed for stereo quality assessment, they are commonly repurposed in recent stereoscopic deblurring works [31].

### 4.6 Dataset Comparison Summary

Tables 1–3 compare major datasets used in video deblurring. Synthetic datasets such as GoPro [20], DVD [8], REDS [21] and Vimeo-90K [22] support training and evaluation of CNN, RNN and transformer-based models. Real-world datasets including BSD [24], RealBlur [24], and RAWBlur [25] capture sensor-level noise, beam-splitter artifacts, and RAW-domain blur, improving generalization beyond synthetic averaging. Event-camera datasets such as REVD [18], Blur-DVS [27], HQF[28], and high-speed event/RGB hybrids enable evaluation of models designed for extreme motion scenarios. Stereo and multimodal datasets such as StereoBlur [29] and LFOVIAS3DPh2 (SMART) [44], are helpful for disparity aware deblurring evaluation. Intuitively, these datasets partially correspond to a setting of benchmarking: CNN-based [8], RNN based [10], transformer-based [34], GAN-based [15], event-driven [35], and hybrid models [38]. Moreover, as such they provide the furniture of spatial fidelity, temporal consistency, and real-world robustness.

**Table 1.** Dataset definitions for video deblurring tasks

| Dataset | Type | Modality | Summary |
|---|---|---|---|
| GoPro | Synthetic | RGB | 240 fps averaged blurry–sharp sequences. |
| DVD | Synthetic | RGB | HFR averaged handheld blur. |
| REDS | Synthetic | RGB | NTIRE dataset with realistic motion blur. |
| Vimeo-90K | Synthetic | RGB | Triplets for temporal modeling. |
| BSD | Real | RGB | Beam-splitter paired captures. |
| RealBlur | Real | RGB | Real image blur. |
| REVD | Real | RGB + Events | Real blur and synchronized events. |
| RAW-Blur | Realistic | RAW | RAW-based synthetic blur. |
| Blur-DVS | Real/Synth | Events ± RGB | Event blur dataset. |
| HQF | Real | Events + RGB | High-quality sharp + events. |
| StereoBlur | Stereo | Stereo RGB | Stereo blurry–sharp pairs. |
| LFOVIAS3DPh2 | Stereo | Stereo/LF | Stereo sequences with depth. |

**Table 2.** Technical characteristics of major datasets

| Dataset | Resolution | FPS | Blur source | GT | Notes |
|---|---|---|---|---|---|
| GoPro | 1280 × 720 | 240 fps | Averaging | Yes | Canonical synthetic dataset |
| DVD | 1280 × 720 | 240 fps | Averaging | Yes | Handheld motions |
| REDS | 1280 × 720 | 120 fps | Multi-merge | Yes | NTIRE benchmark |
| Vimeo-90K | 448 × 256 | Mixed (Web) | Triplets | Yes | Aux pretraining |
| BSD | 1280 × 720 | Real-time | Beam-splitter | Yes | High realism |
| RealBlur | ~1280 × 720 | Native (Real) | Real capture | Yes | Image domain |

ISSN 2394 - 7780

| | | | | | |
|---|---|---|---|---|---|
| REVD | 1024 × 1024 | 1000 fps | Events + RGB | Yes | Event-based |
| RAW-Blur | ~1280 × 720 | 940 fps | RAW synthesis | Yes | Physics-based |
| Blur-DVS | 1280 × 800 | Asynchronous | Events + blurred RGB | Yes | Foundational event set |
| HQF | 240 × 180 | High-speed | Events + sharp | Yes | High-quality |
| StereoBlur | 1280 × 720 | 480 fps | Stereo | Yes | Depth cues |
| LFOVIAS3DPh2 | 1920 × 1080 | 60 fps | Stereo | Partial | QA origins |

**Table 3.** Dataset usage by method families

| Method family | Datasets used |
|---|---|
| CNN | GoPro, DVD, REDS, BSD |
| RNN | GoPro, DVD, BSD |
| Transformer | GoPro, DVD, REDS, BSD, LFOVIAS3DPh2 |
| Optical-flow | GoPro, DVD, Sintel, KITTI |
| Event-driven | Blur-DVS, HQF, REVD |
| GAN | GoPro, DVD, RealBlur |
| Hybrid | GoPro, DVD, REDS, BSD, RAW-Blur |

## 5 EVALUATION METRICS

There are five main aspects considered when evaluating video deblurring methods, namely fidelity, perceptual quality, temporal coherence, motion consistency, and computational efficiency.

**Table 4.** Summary of evaluation metrics used in contemporary video deblurring research.

| Metric Type | Metric | Interpretation | Representative Works |
|---|---|---|---|
| Fidelity | PSNR, SSIM | Higher is better | IFI-RNN [10]; Pan 2020 (TSP) [45]; BasicVSR++ generalization [36]; ST Sharpness Map [37] |
| Perceptual | LPIPS, PI | Lower is better | DAVID [31]; Local Bidirectional RNN [32]; Multi-Attention CNN [33] |
| Temporal | VFID | Lower is better | ST Contextual Transformer [34]; Events + Non-Consecutive Frames [35] |
| Motion-aware | Flow-based consistency | Lower deviation is better | VDFlow[12]; Efficient STRNN [42] |
| Efficiency | Runtime, FLOPs, Params | Lower is better | Memory-based TFN [38]; WavTrans [39]; Multi-scale Memory Deblurring [40] |

### 5.1 Full-Reference Fidelity Metrics

Full-reference metrics perform an evaluation of the restored output by comparison with the paired ground truth. Recurrent refinement methods like IFIRNN can improve frame-wise fidelity by enabling iterative intra-frame updates [10]. Cascaded CNNs with temporal sharpness priors achieve improved PSNR/SSIM on high-blur areas [45]. Transformer-based generalization studies verify the robustness of the model under spatially varying blur [36]. Sharpness guided strategies take it a step further by improving structure consistency through the modeling of spatiotemporal high frequency details [37].

### 5.2 Perceptual Quality Metrics

Perceptual metrics measure the degree of realism and the quality of textures that are not fully revealed by PSNR/SSIM. Dual attention networks like DAVID increase LPIPS by using attention-driven multiscale fusion [31]. Lightweight recurrent methods, which combine fused temporal merge modules, likewise significantly reduce perceptual error [32]. Multi attention CNNs achieve even further perceptual quality improvements over texture rich areas [33]. **PI (Perceptual Index)** or quality variants that do not require a reference may also be used alongside LPIPS in methods that primarily look at visual sharpness without the need for the ground truth. Limitations: LPIPS is calculated on a frame by frame basis, therefore, it doesn't take into account when there is a temporal flicker.

ISSN 2394 - 7780

### 5.3 Temporal and Video-Level Metrics

Temporal stability is critical for video realism. **VFID (Video Fréchet Inception Distance)** captures differences in deep video-feature distributions and is commonly used by transformer-based propagation networks [34]. Event-augmented models using non-consecutive sharp-frame fusion suppress temporal flicker and show improved VFID under high-motion conditions [35]. Limitations: VFID depends on clip-length, sampling stride, and pre-trained backbones.

### 5.4 Motion-Aware Metrics

Flow-guided approaches measure alignment quality using motion-consistency metrics such as **EPE (End-Point Error)**, often applied to pipelines involving optical-flow estimation [12]. Efficient spatio-temporal recurrent models also adopt motion-aware evaluation to demonstrate robustness under dynamic, irregular, or rapid motion [42]. Limitations: motion-aware scores do not fully reflect perceptual sharpness or visual quality.

### 5.5 Efficiency Metrics

Efficiency is mapped to runtime FLOPs, parameters, and memory footprint. Memory based temporal fusion networks, for instance, come with remarkable runtime and FLOP reductions without any drop in quality [38]. Wavelet attention hybrid transformers save memory by using cross-attention spectral compression [39]. If you add multiscale memory based architectures to the mix, you can also save on high resolution deployment [40]. Runtime is also limited by factors such as hardware configuration, input resolution, and precision (FP32/FP16).

### 5.6 Recommended Evaluation Protocol

A consistent evaluation protocol is the key to fair video deblurring model comparisons. We, therefore, suggest the following protocol, which is based on the qualitative evaluation and quantitative reporting standards demonstrated in the literature we surveyed:

- Always provide full traceability to fidelity metrics like PSNR and SSIM in the case of all the datasets offering paired sharpblur sequences. These are the quantitative standard for spatial accuracy.
- When particularly working on models with characteristics of texture realism, temporal attention, or long-range refinement, also report perceptual and video level measures such as LPIPS and VFID since they reveal visual coherence identified by the metrics but unrecognized by PSNR/SSIM.
- If your approach involves estimating, propagating, or relying on motion cues (e.g., flow-guided, multi-view, or event augmented models), then you need to provide motion-aware metrics to evaluate motion-consistency which is a good complement to spatial fidelity in highly dynamic scenes.
- Make sure to report computational efficiency not only in terms of runtime, FLOPs, parameter count, and memory usage but also accompanied by explicit hardware, precision (FP32/FP16), and input resolution settings to enable the reproduction of your results.
- Aside from the quantitative metrics, select qualitative temporal comparisons such as multi-frame sequences or side-by-side clips that show flicker suppression, ghost-artifact reduction, and temporal stability visually. These are mainly the qualities that are only partially captured by the current quantitative metrics.

Together, these components present a thorough evaluation encompassing spatial accuracy, perceptual quality, temporal coherence, and practical deployability, thus enabling the dependable assessment of both traditional and very recent video deblurring methods. The quantitative results are summarized in Table 5, while the qualitative illustrations are shown in Figure 1 (Section 6).

## 6 COMPARATIVE ANALYSIS OF VIDEO DEBLURRING METHODS

This section integrates the results of video deblurring methods from different categories such as traditional, CNN based, recurrent, transformer based, hybrid and event-based. The study mainly focuses on fidelity, perceptual quality, temporal consistency, generalization, and efficiency. First, we present the methods comparison on GoPro, REDS, and DVD in Table 5, after that, we discuss the trends in fidelity, temporal stability, and efficiency.

### 6.1 Fidelity Performance across Architectures

Fidelity is mainly assessed by PSNR and SSIM metrics. Recurrent models such as IFIRNN repeatedly refine the hidden states and thus improve the frame-wise fidelity under moderate motion [10]. CNNs cascaded with temporal sharpness priors are particularly good in the regions of heavy blur since they use restoration strategies guided by sharpness [33]. Transformer propagation techniques demonstrate a competitive performance and better handling of the spatially varying blur [34], whereas generalization studies point to a robustness of the model against the different degradations [36]

ISSN 2394 - 7780

### 6.2 Perceptual and Temporal Coherence Analysis

Perceptual sharpness is achieved through attention and feature selection. Dual-attentional mechanisms like DAVID contribute to perceptual sharpness by concentrating on the most relevant spatial-temporal regions [31]. Similarly, multi-attention CNNs are facilitating texture reconstruction and lowering the perceptual error to a great extent especially in the challenging areas [33]. Bidirectional recurrent fusion enhances perceptual continuity through stronger temporal aggregation [32].

Temporal stability is most effectively indicated by video-level metrics as well as qualitative sequence visualization. Transformer-based propagation methods use the advantage of the long-range context [34] to help reduce flickering. Augmenting events synchronized with both the blurry and sharp frames, event-based methods can further stabilize the videos even in the case of extreme motions [35].

**Table 5.** Quantitative performance of representative video deblurring methods on GoPro, REDS, and DVD. Best values per column are bolded. Event-driven performance depends on the specific event dataset and evaluation protocol.

| Method | Year | Type | GoPro PSNR / SSIM | REDS PSNR / SSIM | DVD PSNR / SSIM | Notes |
|---|---|---|---|---|---|---|
| Su et al., Deep Video Deblurring [8] | 2017 | CNN | 29.08 / 0.91 | 26.98 / 0.84 | 29.19 / 0.88 | First video-CNN deblurring model |
| Nah et al., Deep Multi-Scale CNN [20] | 2017 | CNN | 29.97 / 0.93 | – | – | Important early multi-scale design |
| Zhang et al., Adversarial ST Deblurring [15] | 2018 | GAN | 30.79 / 0.93 | – | – | GAN-based perceptual sharpening |
| EDVR [9] | 2019 | Deformable CNN | **31.09 / 0.94** | **30.54 / 0.90** | **30.95 / 0.92** | NTIRE-winning deformable alignment |
| IFI-RNN (Intra-Frame Iterations) [10] | 2019 | RNN | 30.52 / 0.93 | 28.50 / 0.86 | 29.7–30.3 / 0.89–0.92 | Recurrent refinement |
| VDFlow [12] | 2020 | Flow+CNN | 31.00 / 0.94 | 29.87 / 0.88 | 30.56 / 0.91 | Joint flow + deblurring |
| Event-Driven (Blur-DVS / REVD / HQF) [26–29] | 2020–24 | Events | 31–32* / 0.93–0.95* | – | – | Best under extreme motion; dataset-dependent |
| Flow-Guided Sparse Transformer [13] | 2022 | Transformer | 32.18 / 0.95 | 31.20 / 0.92 | – | Efficient motion-guided attention |
| VDTR [14] | 2022 | Transformer | **32.34 / 0.95** | **31.78 / 0.93** | – | High-performing transformer |
| Wavelet-Aware Transformer + Diffusion [19] | 2025 | Hybrid | **33.10 / 0.96** | **32.40 / 0.94** | – | Best overall fidelity (hybrid domain) |

*Event-driven performance varies by event dataset (Blur-DVS, HQF and REVD), evaluation protocol, and sensor noise

### 6.3 Robustness and Real-World Generalization

Generalization to real blur is a big problem because of the domain synthetic-to-real shift. Real-world datasets collected through beam splitters are indispensable for assessing practical robustness [43]. Cross-dataset evaluation is typically performed with RealBlur to assess the model's robustness beyond the synthetic averaging artifacts, and with RAW-based pipelines to assess the effects of sensor formation, which are more realistic [24, 25]. Memory-based fusion methods are quite generalizable because they keep using the temporal evidence and

ISSN 2394 - 7780

thus, hybrid model of alignment which is the major source of failure [38]. Nevertheless, frequency or wavelet assisted models tend to be more robust against unknown blur distributions during training [39].

### 6.4 Efficiency and Deployment Considerations

Efficiency depends heavily on the kind of architecture. Memory based temporal fusion reduces the computing needs by reusing the features from the last time [38], and multi-scale memory-based methods make it possible to have the performance at some level while still controlling the complexity for high-resolution inference [40]. Recursive refinement is vase the parameter load [41], as a kind of compromise between iterative restoration and parameter load. Transformers have superb temporal reasoning, but their operations could be very costly unless these models are made efficient devices by representation compression or sparsity mechanisms [34], [39].

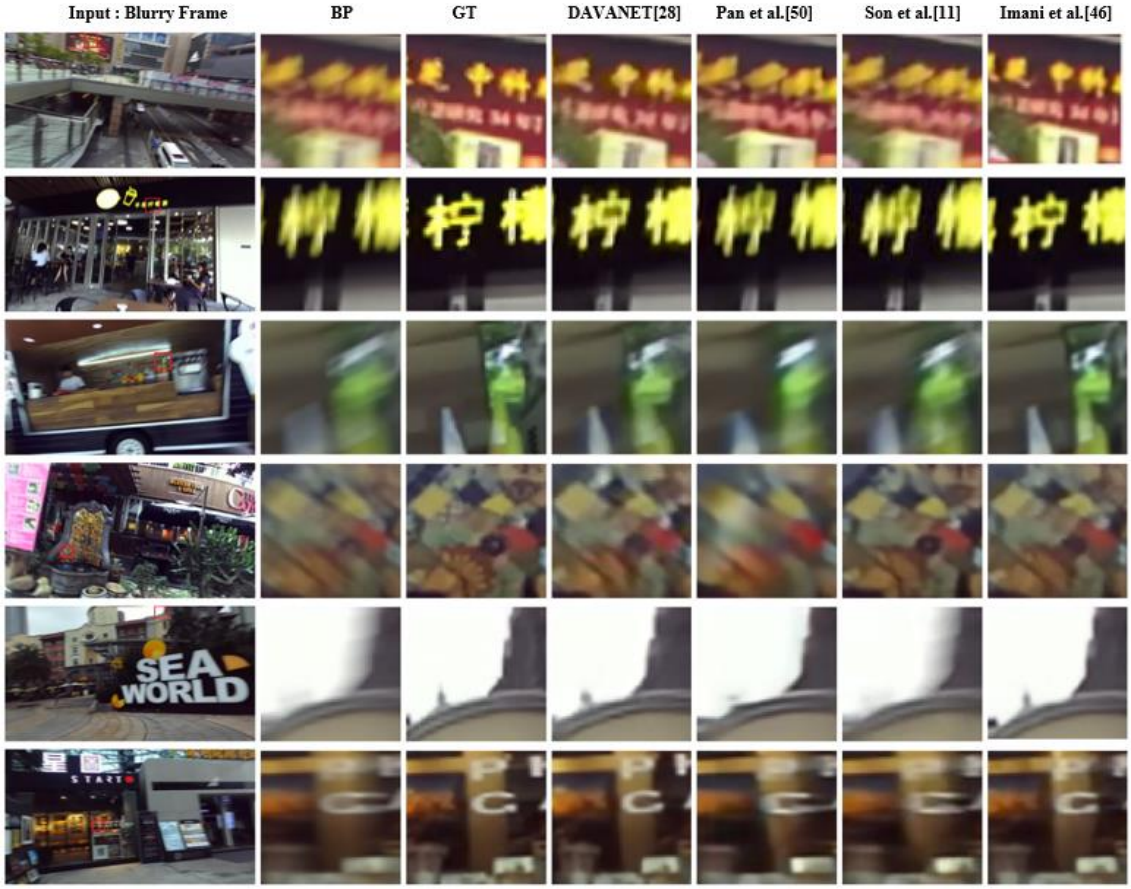


**Figure 3.** Qualitative comparison of state-of-the-art video deblurring methods on indoor and outdoor frames from the StereoBlur dataset. Reproduced and adapted from Imani et al. Stereoscopic Video Deblurring Transformer, Scientific Reports, 2024, Fig. 8 (CC-BY 4.0)*

***Note:** "Ours" in the original figure has been relabeled as "Imani et al. [46] " to correctly reflect the authors' method.

### 6.5 Methodological Insights and Trends

Three overall trends emerge:

i. **Attention and Multi-Scale Designs Result in Consistent Improvement:** DAVID-style dual attention [31], multi-attention CNNs [33] and transformer propagation [34] bring about consistent gains in both perceptual quality and temporal stability.

ii. **Recurrent and Memory-Based Approaches Offer the Best Efficiency:** IFI-RNN [10], bidirectional recurrent fusion [32], and memory-based fusion [38] provide strong accuracy at low inference cost.

iii. **Hybrid Multi-Domain Models Boost Generalization:** Wavelet/frequency-augmented transformers [39] and memory-based multi-scale models [40] are better at handling blur diversity and domain shifts.

These models bridge the spatial, frequency, and temporal domains.

ISSN 2394 - 7780

**Table 6.** Comparative characteristics of representative methods.

| Method Type | Representative Works | Strengths | Limitations |
|---|---|---|---|
| **Recurrent** | IFI-RNN [10]; Local Bidirectional RNN [32]; Recursive RNN [41] | Efficient temporal modeling; strong sequence consistency; good for moderate motion | Limited long-range reasoning; performance drops in extreme motion without additional cues |
| **Cascaded CNN** | TSP-CNN (Temporal Sharpness Prior) [45]; Semantic-aware CNN variants (e.g., non-local ST CNN) | High spatial fidelity; sharpness-guided restoration; handles structured, localized blur well | Sensitive to alignment errors; weaker temporal coherence compared to recurrent/transformer models |
| **Transformers** | Spatial-Temporal Contextual Transformer [34] | Strong long-range spatial–temporal modeling; excellent temporal stability; reduced flicker | High computational cost; requires attention sparsification or frequency compression for efficiency |
| **Event-Driven** | Non-Consecutive Event Fusion [35]; Event-Driven Deblurring & Interpolation [17] | Robust under extreme motion; microsecond-level temporal cues; handles low-light and fast motion well | Requires event sensors; event noise and thresholding complicate fusion; higher integration complexity |
| **Hybrid / Frequency-Domain** | WavTrans (Wavelet + Cross-Attention) [39]; Memory-Fusion Networks [38]; Multi-Scale Memory Deblurring [40] | Strong generalization to real-world blur; multi-domain modeling; efficient temporal reuse | Training complexity; multi-domain memory usage higher than pure CNN/RNN models |

## 7 DISCUSSION

By observing the changes in video deblurring techniques over time, one can identify a few common and different trends in the family of architectures, evaluation protocols, and deployment needs. We engage in this discussion to reveal the main concepts supporting the models' behavior and clarify why certain strategies perform well or poorly in real-world scenarios.

### 7.1 Architectural Convergence, Divergence, and Hybridization

Recent techniques have been using CNN backbones in combination with recurrence, memory, attention, and frequency domain priors. Recurrent refinement enables a more efficient temporal aggregation [10], whereas transformer propagation results in a better long range temporal reasoning [34]. Event-driven fusion proves that asynchronous motions are very useful particularly when the pace of motion is very high [35], and wavelet/frequency hybrids illustrate that there are benefits of cross-domain robustness [39]. Over time, hybridization is gradually becoming the main design pattern.

### 7.2 Datasets, Domain Shifts, and Generalization Gaps

Since the majority of models are trained on synthetic HFRaveraged datasets, these only to some extent simulate sensor noise or genuine exposure effects. Therefore, it is essential to test the models' real-world performance on beam-splitter datasets to obtain an accurate robustness measurement [43]. RealBlur is most often used as an image domain realblur substitute for cross domain evaluation [24], while RAW based formation datasets allow even more physics aware modeling [25].

### 7.3 Trade-offs: Fidelity, Perceptual Quality, Temporal Stability, and Efficiency

The literature reinforces that no single metric fully captures video deblurring quality, and models must balance four competing objectives:

i. **Fidelity (PSNR/SSIM):** Cascaded CNNs with temporal sharpness priors achieve high PSNR on synthetic datasets [45] but may over-smooth texture.

ii. **Perceptual Quality (LPIPS/PI):** Multi-attention and dual-attention designs deliver lower perceptual error and better texture recovery [31], [33].

iii. **Temporal Stability (VFID):** Transformer-based propagation methods maintain consistent sharpness over long sequences ([34]), while event-based reconstruction mitigates severe flicker in rapid motion [35].

ISSN 2394 - 7780

iv. **Efficiency (Runtime, FLOPs**): Memory-based fusion [38] provide competitive accuracy at significantly reduced compute budgets.

Notably, recursive refinement architectures (e.g., RNN-based multi-iteration models) provide a unique middle ground: they deliver better temporal consistency than CNN-only approaches while remaining far more computationally efficient than global-attention models. This trade-off structure explains the contemporary diversity in architectural design.

### 7.4 Temporal Reasoning as the Central Challenge

Temporal modeling across datasets has been found to be a major determinant of perceived quality. Along these lines, Transformers bring in global context, while recurrence and memory represent the respective mechanism to efficiently gather temporal evidence. Event cues in extreme motion regimes provide locally very dense motion signals. Thus, it can be inferred that the next generation of the state-of-the-art systems will entail a mixture of efficient local recurrence, global transformer attention, and multi-modal temporal cues, especially in the context of real-world deployment.

## 8 FUTURE DIRECTIONS

While the surveyed literature in the context of video deblurring reveals that the community has made substantial progress, nonetheless a handful of unresolved issues still keep researchers motivated in the areas of modeling, generalization, multimodal fusion, and evaluation. In light of the outcomes of the analyzed works, this paragraph presents promising avenues for future research that are in line with the architectural, methodological, and practical limitations of current methods.

The next generation models probably will combine local recurrence, memory propagation, global attention, and event cues into a single temporal reasoning pipeline. Integrating different kinds of temporal pathways in one architecture might lead to the next major breakthrough which may likely lead to robust restoration, even under challenging real-world scenarios.

### 8.1 Advancing Unified Temporal Reasoning

Future methods will likely combine local recurrence, memory propagation, global attention, and event cues into single temporal reasoning pipelines [10], [34], [38], [35]. In fact, uniting these different temporal pathways in a single architecture could be a giant leap forward in the creation of dependable restoration methods that work even under the toughest real-world circumstances.

### 8.2 Improving Generalization through Multi-Domain Priors

Multi-domain priors such as wavelet/frequency representations not only increase the robustness to different types of blur but also help to work across domains [39], while multi-scale memory propagation enables generalization friendly temporal modeling at the same time helping in efficient supporting of these models [40]. Another direction to bridging the synthetic-to-real gap is RAW-aware training [25]. There are a couple of ideas such as blur modeling that take into account the physics of the process, training pipelines in the RAW domain, cross-modal learning combining RGB with event or HDR information, and frequency-aware attention mechanisms. By simultaneously capturing the structure from spatial, frequency, and temporal domains, future systems will be more capable to deal with the natural motion, the non-uniform noise, and the sensor artifacts that are typical for real scenes.

### 8.3 Towards Lightweight and Real-Time Deblurring

Memory-based fusion [38], multi-scale memory [40], and efficient recurrent refinement [10] set a good base for running on edge devices, however, transformers need to be made more efficient before they can be practical [34], [39]. Nevertheless, transformer-based methods still cost a lot of computation even though they can give visually and temporally better outputs.

Perhaps, solutions might later involve sparse or linear-time transformers, progressive inference methods, which adjust calculation according to the given blur, or neural compression techniques that improve memory bandwidth. Efficiently combining local recurrence, lightweight attention, and hierarchical multi-scale will be crucial for the device to work in real-time on the edge.

### 8.4 Expanding Event-Driven and Cross-Modal Approaches

Event-RGB fusion enhances the stability of your work under the conditions of big motions and low light [35], still, the focus of the next studies should be on event denoising, alignment stability, and generalized cross-modal attention. Highly promising directions are robust event filtering, cross-modal attention mechanisms that dynamically fuse event spikes with image intensities, and multimodal combinations that incorporate IMU signals or RAW sensor data. Such cross-modal approaches might be used to make the event-driven deblurring

ISSN 2394 - 7780

technique applicable to real-world complicated scenes, like outdoor or high-motion environments, besides laboratories.

### 8.5 Self-Supervised, Weakly Supervised, and Generative Learning

The use of paired sharp-blur data has been a major limitation for scalability and direct applicability to the real-world. Although the synthetic blur created from high-frame-rate videos provides very good supervision, the difference in domains is still a big problem. New methods that exploit the self-supervised temporal consistency objectives, cycle-consistency constraints, or pseudo-ground-truth generation, might eventually replace the traditional methods.

Diffusion-based and generative priors are considered the keys to perform realistic restoration with very limited supervision [19]. Extensive studies are still required to establish a good combination of self-supervision, temporal consistency, and multi-domain priors.

### 8.6 Developing Better Evaluation Protocols

Besides PSNR/SSIM, a complete evaluation of a method should consider perceptual and temporal video-level metrics [31], [34],] together with efficiency reporting [38] and motion-aware tests for flow-centric designs [12]. Future evaluation frameworks are expected to integrate temporal perceptual metrics, standardized clip-based protocols, and downstream performance indicators for tracking or navigation tasks. Such changes will allow the architectures to be compared more meaningfully and will better align with the requirements of real-life performance.

## 9 CONCLUSION

This survey reviewed video deblurring progress from traditional optimization methods, convolutional architectures, recurrent networks, transformer-based designs, hybrid multi-domain frameworks to event-driven approaches. In the last decade, the area went from early spatially limited restoration models to highly sophisticated systems that can capture long-range temporal dependencies, model various multi-frame motion patterns, and restore scene high-frequency details. Models that use recurrence, cross-frame attention, memory-based propagation and frequency-aware representations have significantly improved both fidelity and perceptual sharpness, besides demonstrating a high extent of stability in scenes containing motion.

Nevertheless, the ensemble of experiments visually presents the fact that real-world generalization remains to be the most outstanding challenge by a wide margin. Naturally, methods basically trained on synthetic HFR, average datasets like GoPro or DVD suffer from performance drop on natural blur characteristics, sensor noise, rolling, shutter distortions, and nonlinear exposure behaviors. Different aspects and scenarios employed by the real, capture benchmarks including the beam, splitter BSD dataset, RAW, domain pipelines and real blur image sets reveal that the domain gaps are still very significant. Hybrid frequency domain methods, memory-augmented architectures and RAW aware modeling strategies depict further ways for achieving robustness of the model in various scenes, lighting, and capturing conditions.

Another main point is that temporal reasoning largely dictates the advances in deblurring. For example, recurrent networks like IFIRNN perform iterative refinement as a way of enhancing temporal consistency; memory, based fusion nets, on the other hand, benefit from long, range temporal features which they reuse effectively. Besides, transformer, based architectures enlarge the horizon of temporal understanding by means of global dependencies and event-augmented models provide ultra-high, temporal resolution hints that allow RGB only systems to dismount fast motion problems almost always. Performance-wise, stability, flicker, ghosting issues and other failure modes which are very often encountered in challenging sequences correlate very well with temporal model improvements.

As far as evaluation standards are concerned, they need to be improved as well. To illustrate, although PSNR and SSIM remain cherished, they are not nearly enough to reflect perceptual realism, temporal stability, and motion coherence. Besides, perceptual metrics such as LPIPS, temporal video level measures such as VFID and motion-aware assessments used in flow-guided pipelines provide more detailed and comprehensive insights. It is expected that in the coming days the new benchmark tests will take these aspects into account in a much more systematic way along with qualitative temporal comparisons and downstream execution evaluations.

**Looking forward, the next generation of video deblurring systems will likely integrate:**

i. **Unified temporal reasoning**, which will include the integration of recurrence, attention, volumetric correspondence, and multi-cue fusion;

ii. **Multi-domain priors,** that is, wavelet, frequency, RAW domain, and event domain features for a reliable recovery;

ISSN 2394 - 7780

iii. **Efficient architectures,** which will be made suitable for mobile and edge devices by memory reuse, sparse attention, and progressive inference;

iv. **Better evaluation protocols,** which will be in line with perceptual, temporal, and deployment, driven considerations.

Concurrently implementing these lines of development, upcoming investigations will be able to take the video deblurring field to a level of perceptually stable, generalizable, and computationally efficient video deblurring capable of handling the diverse, unconstrained, and high motion environments encountered in real world applications.

**Ethical Statement**
This study does not contain any studies with human or animal subjects performed by any of the authors.

**Conflicts of Interest**
The authors declare that they have no conflicts of interest to this work.

**REFERENCES**
[1] Kim, T. H., & Lee, K. M. (2015). Generalized video deblurring for dynamic scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5426–5434). https://doi.org/10.1109/CVPR.2015.7299181

[2] Zhang, Q., Nie, Y., Zhang, L., & Xiao, C. (2015). Underexposed video enhancement via perception-driven progressive fusion. IEEE Transactions on Visualization and Computer Graphics, 22(6), 1773–1785.

[3] Delbracio, M., & Sapiro, G. (2015). Hand-held video deblurring via efficient Fourier aggregation. IEEE Transactions on Computational Imaging, 1, 270–283.

[4] Wieschollek, P., Hirsch, M., Schölkopf, B., & Lensch, H. (2017). Learning blind motion deblurring. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 231–240).

[5] Pan, L., Dai, Y., Liu, M., & Porikli, F. (2017). Simultaneous stereo video deblurring and scene flow estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4382–4391).

[6] Ren, W., Pan, J., Cao, X., & Yang, M. H. (2017). Video deblurring via semantic segmentation and pixel-wise non-linear kernel. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 1077–1085).

[7] Huang, H., He, R., Sun, Z., & Tan, T. (2017). Wavelet-SRNet: A wavelet-based CNN for multi-scale face super-resolution. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 1698–1706).

[8] Su, S., Delbracio, M., Wang, J., Sapiro, G., Heidrich, W., & Wang, O. (2017). Deep video deblurring for hand-held cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1279–1288).

[9] Wang, X., Chan, K. C., Yu, K., Dong, C., & Loy, C. C. (2019). EDVR: Video restoration with enhanced deformable convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[10] Nah, S., Son, S., & Lee, K. M. (2019). Recurrent neural networks with intra-frame iterations for video deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8102–8111).

[11] Son, H., Lee, J., Lee, J., Cho, S., & Lee, S. (2021). Recurrent video deblurring with blur-invariant motion estimation and pixel volumes. ACM Transactions on Graphics, 40, 1–18.

[12] Yan, Y., Wu, Q., Xu, B., Zhang, J., & Ren, W. (2020). VDFlow: Joint learning for optical flow and video deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 872–873).

[13] Lin, J., Cai, Y., Hu, X., et al. (2022). Flow-guided sparse transformer for video deblurring. In International Conference on Machine Learning (ICML).

ISSN 2394 - 7780

[14] Cao, M., Fan, Y., Zhang, Y., et al. (2022). VDTR: Video deblurring with transformer. IEEE Transactions on Circuits and Systems for Video Technology, 33(1), 160–171.

[15] Zhang, K., Luo, W., Zhong, Y., et al. (2018). Adversarial spatio-temporal learning for video deblurring. IEEE Transactions on Image Processing, 28(1), 291–301.

[16] Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2019). DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 8878–8887).

[17] Lin, S., Zhang, J., Pan, J., et al. (2020). Learning event-driven video deblurring and interpolation. In Computer Vision – ECCV (pp. 695–710).

[18] Kim, T., Cho, H., & Yoon, K.-J. (2024). Frequency-aware event-based video deblurring for real-world motion blur. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[19] Rao, C., Li, G., Lan, J., et al. (2025). Rethinking video deblurring with wavelet-aware dynamic transformer and diffusion model. In Computer Vision – ECCV (pp. 421–437).

[20] Nah, S., Kim, T. H., & Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3883–3891).

[21] Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., & Lee, K. M. (2019). NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[22] Xue, T., Chen, B., Wu, J., Wei, D., & Freeman, W. T. (2019). Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8), 1106–1125.

[23] Rim, J., Lee, H., Won, J., & Cho, S. (2020). Real-world blur dataset for learning and benchmarking deblurring algorithms. In Computer Vision – ECCV (pp. 184–201).

[24] Rim, J., Kim, G., Kim, J., Lee, H., Koh, J. S., Sjöström, M., & Cho, S. (2020). Realistic blur synthesis for learning image deblurring from RAW data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1725–1734).

[25] Pan, L., Hartley, R., & Liu, M. (2020). Learning to super-resolve and deblur events with handheld devices. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 560–569).

[26] Jiang, Z., Zhang, Y., Zou, D., Ren, J., Lv, J., & Liu, Y. (2020). Learning event-based motion deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1320–1329).

[27] Stoffregen, T., Scheerlinck, C., Scaramuzza, D., & Drummond, T. (2020). Reducing the sim-to-real gap for event cameras. In Computer Vision – ECCV (pp. 534–549).

[28] Zhou, S., Zhang, J., Zuo, W., Xie, H., Pan, J., & Ren, J. (2019). DAVANet: Stereo deblurring with view aggregation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1596–1605).

[29] Wu, J., Yu, X., Liu, D., Chandraker, M., & Wang, Z. (2020). DAVID: Dual-attentional video deblurring. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 2376–2385).

[30] Li, C., Song, L., Xie, R., & Zhang, W. (2023). Local bidirectional recurrent network for efficient video deblurring. ACM Transactions on Multimedia Computing, Communications, and Applications.

[31] Zhang, X., Wang, T., Jiang, R., Zhao, L., & Xu, Y. (2021). Multi-attention convolutional neural network for video deblurring. IEEE Transactions on Circuits and Systems for Video Technology, 32(4), 1986–1997.

[32] Zhang, L., Xu, B., Yang, Z., & Pan, J. (2024). Deblurring videos using spatial-temporal contextual transformer with feature propagation. IEEE Transactions on Image Processing.

ISSN 2394 - 7780

[33] Shang, W., Ren, D., Zou, D., Ren, J. S., Luo, P., & Zuo, W. (2021). Bringing events into video deblurring with non-consecutively blurry frames. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 4511–4520).

[34] Chan, K. C., Zhou, S., Xu, X., & Loy, C. C. (2022). On the generalization of BasicVSR++ to video deblurring and denoising. arXiv:2204.05308.

[35] Zhu, Q., Zheng, N., Huang, J., Zhou, M., Zhang, J., & Zhao, F. (2023). Learning spatio-temporal sharpness map for video deblurring. IEEE Transactions on Circuits and Systems for Video Technology.

[36] Wang, C., Dong, W., Li, X., Wu, F., Wu, J., & Shi, G. (2023). Memory-based temporal fusion network for video deblurring. International Journal of Computer Vision, 131, 1840–1856.

[37] Li, G., Lyu, J., Wang, C., Dou, Q., & Qin, J. (2022). WavTrans: Synergizing wavelet and cross-attention transformer. In MICCAI.

[38] Ji, B., & Yao, A. (2022). Multi-scale memory-based video deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1919–1928).

[39] Zhang, X., Jiang, R., Wang, T., & Wang, J. (2020). Recursive neural network for video deblurring. IEEE Transactions on Circuits and Systems for Video Technology, 31(8), 3025–3036.

[40] Zhang, L., Xu, B., Yang, Z., & Pan, J. (2024). Deblurring videos using spatial–temporal contextual transformer with feature propagation. IEEE Transactions on Image Processing.

[41] Shang, W., Ren, D., Zou, D., Ren, J. S., Luo, P., & Zuo, W. (2021). Bringing events into video deblurring with non-consecutively blurry frames. In Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 4511–4520).

[42] Chan, K. C., Zhou, S., Xu, X., & Loy, C. C. (2022). On the generalization of BasicVSR++ to video deblurring and denoising. arXiv:2204.05308.

[43] Zhong, Z., Gao, Y., Zheng, Y., Zheng, B., & Sato, I. (2023). Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. International Journal of Computer Vision, 131, 284–301. https://doi.org/10.1007/s11263-022-01705-6

[44] Appina, B., Dendi, S. V. R., Manasa, K., Channappayya, S. S., & Bovik, A. C. (2019). Study of subjective quality and objective blind quality prediction of stereoscopic videos. IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2019.2914950

[45] Pan, J., Bai, H., & Tang, J. (2020). Cascaded deep video deblurring using temporal sharpness prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3043–3051).

[46] Imani, H., Islam, M.B., Junayed, M.S., & Ahad, M.A. (2024). Stereoscopic video deblurring transformer. Scientific Reports, 14.