IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 August 2025, accepted 23 September 2025, date of publication 17 October 2025, date of current version 31 October 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3622993

RESEARCH ARTICLE

Camera Sensor Raw Data-Driven Video Blur Effect Prevention: Dataset and Study

ABDELWAHED NAHLI[1], DAN LI[1], RAHIM UDDIN[1], TAHIR RAZA[1], MUHAMMAD IRFAN[2,3], QIYONG LU[1], (Member, IEEE), AND JIAN QIU ZHANG[1], (Senior Member, IEEE)

[1]Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai 200007, China
[2]Institute of Biomedicine, University of Turku, 20741 Turku, Finland
[3]School of Biomedical Engineering, University of Sydney, Camperdown, NSW 2050, Australia

Corresponding authors: Dan Li (lidan@fudan.edu.cn) and Abdelwahed Nahli (21110720143@m.fudan.edu.cn)

This work was supported in part by Fudan University, and in part by the National Natural Science Foundation of China under Grant 12374431.

ABSTRACT Recent advances in machine vision have played an important role in addressing the challenging problem of motion blur. However, most deep learning–based deblurring methods operate in the RGB domain, rely on recursive strategies, and are often trained on unrealistic synthetic data. In this paper, we introduce a preventive solution from a new perspective, leveraging the opportunity to operate directly in the RAW domain on high-bit sensor data. Since no publicly available high–frame rate RAW-based blur prevention dataset exists, we construct Blurry-RAW, a novel dataset containing paired blurry and sharp frames in both RAW and RGB formats. We further propose 3D-ISPNet, a CNN–Transformer hybrid architecture, trained exclusively on RAW sensor data. This model achieves superior quantitative and qualitative performance compared to RGB-based counterparts. Moreover, by fine-tuning on data from different camera sensors, 3D-ISPNet demonstrates strong generalization across diverse hardware. Ultimately, the introduction of RAW-driven blur prevention and the new dataset paves the way for further research in this emerging direction.

INDEX TERMS Sensor raw data, camera ISP, blur effect prevention.

I. INTRODUCTION

Blurry video frames are the result of an accumulation of optical signals acquired by the camera sensor over the course of an exposure period. A shaking camera or fast-moving objects in captured scenes usually result in this problem. It has been appealing to researchers to work on blur removal tasks to restore sharp videos in order to support various computer vision tasks such as tracking moving object and video interpolation. The problem of blur effect remains highly ill-posed to solve. For complicated blur kernel modeling, conventional techniques mostly require some sort of pre-assumptions or/and constraints [1], [2], [3], [4]. In real-world applications, nevertheless, these approaches are not readily applicable due to the inaccurate approximations. In [5], the authors have attempted to reproduce the blur effect phenomenon in recent years. By averaging successive sharp frames, they synthesize blurry frames from high fps videos. They contribute to the flourishing of related research by relieving us of the complicated design of blur kernels.

However, the averaged RGB values are still inadequate for replicating real-world blur effects, as they are influenced by camera ISP processing. The ISP's non-linear processes such as demosaicing, white balance, gamma compression, and color correction can result in RGB frames that are not linear to the underlying sensor data. Additionally, this problem made post-processing necessary for building the previous dataset [5], in order to reduce the domain gap. Learning to recover blurry videos with RGB datasets has significant issues, highlighting the necessity of utilizing RAW datasets.

As a part of this work, Blurry-RAW is created as a novel video dataset so that deep learning-based blur effect

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

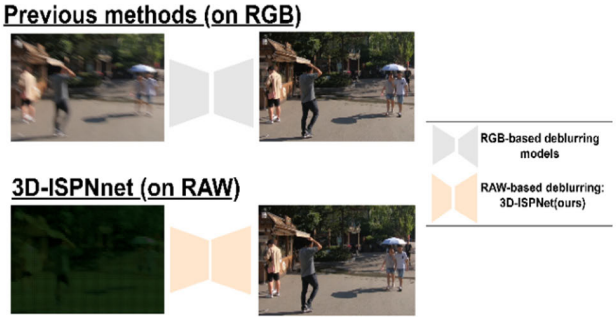

FIGURE 1. To tackle blur effect phenomenon, conventional approaches mainly rely on recursive strategies and operate in the RGB domain, whereas in this study, we promote a novel blur preventive strategy by training on sensor raw data.

prevention can be carried out on raw sensor data, where pairs of blurry and sharp RAW frames are included, along with their RGB counterparts. With the appropriate camera settings, we record high fps sharp RAW videos, split them into successive RAW frames, and as in [5], we also synthesize their corresponding blurry RAW frames by averaging every eight consecutive RAW frames, using a sliding window of one frame stride. The direct manipulation of sensor raw data allows us to create more realistic blurry frames without undergoing any post-processing, for instance, gamma correction.

A RAW video frame, unlike an RGB frame, the information is stored using color filters, such as the Bayer pattern or X-tran, for example. Besides spatial information, each position in an image contains a specific color sensor value. Thus, we also propose 3D-ISPNet as a novel network architecture that takes advantage of RAW images' special characteristics, as shown in Fig. 1. Using spatiotemporal convolutions, 3D-ISPNet is a simple, scalable approach for video blur effect prevention. 3D-ISPNet has the capability to generate end-to-end predictions for multiple frames in a single forward pass, without the need for external flow or depth maps. By learning from large scale video data, it implicitly handles complex motions and occlusions, and it is easier to deploy, while improving inference speed and achieving state-of-the-art blur prevention accuracy. 3D-ISPNet can easily be deployed as a blur-aware camera Image Signal Processor (ISP), as it can perform all the essential subtasks of an ISP system in an end-to-end fashion.

Our experiment section emphasizes the advantage of directly deblurring on RAW data. The great amount of information preserved in RAW frames allows us to restore finer details and structure. The performance gains we get after applying our proposed novel network architecture are further enhanced. Hence, the designed network is better suited for video blur prevention. Our approach surpasses existing RGB data-driven deblurring methods in both quantitative and qualitative performance. Here are some key highlights of our contributions:

1) 3D-ISPNet architecture: We propose 3D-ISPNet, an efficient and scalable CNN–Transformer hybrid designed to learn directly from camera sensor data for single-shot, multi-frame blur prevention. By performing one of the core sub-tasks of an Image Signal Processor (ISP), it can be seamlessly integrated into a blur-aware camera ISP pipeline.
2) Blurry-RAW dataset: We introduce Blurry-RAW, a novel high–frame rate RAW dataset created to overcome the limitations of existing RGB datasets. It provides paired sharp and blurry video frames in both RAW and RGB formats, enabling more realistic and sensor-level blur prevention research.
3) RAW-domain advantage: Through extensive experiments, we demonstrate that operating directly in the RAW domain on high-bit sensor data significantly improves blur prevention while better preserving fine scene details and structures. Our approach consistently outperforms existing RGB-based deblurring methods, both quantitatively and qualitatively.

II. RELATED WORKS

There are several deblurring approaches, but most of them rely on the blur model in [10] and [11], which can be expressed as below:

$$I_B = K\,(M) * I_S + N \tag{1}$$

where I_B denotes the blurred picture, $K\,(M)$ are the blur effect kernels related to the field of motion M, I_S represents the sharp image and N stands for the environment noise. Deblurring problems can be categorized as non-blind or blind, depending on whether information about blur or blur kernel ($K\,(M)$) is available or not. Most conventional methods prioritize non-blind deblurring, while blind blur kernels are typically modeled based on simplistic assumptions and priors. In their work, Kim et al. [3] suggested segmenting blurry images and estimating non-uniform blur kernels within each segment. In [4], Kim and Lee further noted that the blur kernel at the pixel level can be approximated as linear. As a result of patch-based priors, Michaeli and Irani [2] were able to restore sharp images in down-scaled images. Using a patch-wise non-uniform deblurring algorithm, Yu et al. [12] estimated each kernel locally and recovered a latent image by using total variation regularization. Despite this, blur kernels are often unknown in practice.

There has been significant progress in machine vision-based image generation techniques, such as blur reduction [5], [26], [27], [36], denoising [19], [20], object removal [21], [22], style transfer [23], [25] and super-resolution [13], [14], [15], [16], [17], [18]. In certain blur reduction techniques, artificial neural networks are used to approximate the blur kernels, and then clear images are restored [10], [11], [28]. Xu et al. [28] proposed a method that suggests deblurring can be accomplished by utilizing blur kernels that can be decomposed into a small set of filters. In a previous work [11], Sun et al. used convolutional neural networks (CNN) to estimate the probabilities of predefined motion kernels for image patches, and then applied

an optimization technique to reconstruct the latent image. In a previous study [10], a fully convolutional network was employed to estimate motion flow, followed by non-blind deconvolution to restore the sharpness of the image. These methods involved deep learning techniques to estimate blur kernels more accurately. Nevertheless, these models frequently encounter challenges in real-world scenarios due to the difficulty of accurately modeling non-uniform and complex blur kernels using simplistic kernel assumptions. In an effort to address the intricate and uncertain characteristics of blur kernels in real-world images, certain researchers have experimented with blind deconvolution methods for directly restoring sharp images [5], [6], [26], [27], [29], [30], [31], [38], [42]. Nah et al. [5] utilized techniques that do not rely on blur kernels for both generating datasets and estimating latent images. They developed an advanced method for image deblurring using a multi-scale network, and also released a widely-used dataset for subsequent researchers to utilize. Galshetwar et al. [6] introduced a straightforward yet effective approach for multi-frames deblurring by designing an edge-enhanced model branch. Kupyn et al. [26] employed a Generative Adversarial Network (GAN) based model to restore clear images. Tao et al. [29] improved upon the multi-scale approach by incorporating a recurrent structure that enables weight sharing across different scales within the network. Zhang et al. [30] attained cutting-edge performance on GoPro datasets by adopting a strategy of deblurring small patches cropped from complete images, as opposed to deblurring at multiple scales. In a recent work Zhong et al. [31] introduce a frame-oriented neural network structure for learning high motion video deblurring task. As only datasets with RGB images exist, these methods primarily focus on deblurring RGB images. Furthermore, RGB data often lack the abundance of valuable information that can only be obtained from sensor raw data. As compared to RGB data, RAW data contain all the information captured by the image sensor, including color information and other details, such as brightness, contrast.

There have been also some previous attempts to address the blur artifacts by analyzing sensor raw data [32], [33], [39]. Zhen [32] employed inertial sensors to recover RAW images, capturing RAW image data using a digital imaging system and 3-axis acceleration data. An estimate of the blur kernel and camera motion can be made using acceleration data.

In their work [33], Trimeche et al. introduced a multi-channel image restoration technique aimed at mitigating optical blur resulting from the camera's optical system. RAW images were processed using a modified iterative Landweber algorithm, applied separately to each color channel, along with adaptive denoising. They employed an adaptive filter that utilized neighboring pixels' local polynomial approximation from dynamically picked windows to enhance the iterative process resilience. In addition, they suggest a unique saturation control mechanism to reduce the impact of iterative recorvery in near-saturated parts, preventing false colors resulting from independent channel filtering in RGB domain. Despite utilizing RAW data, these methods remain unpractical as they rely on unrealistic simplistic pre-assumptions. In addition, they tend to focus more on blur caused by camera shaking than on object motion-caused blur.

Prior studies have demonstrated that sensor raw data can benefit many computational imaging tasks [7], [8], [9], [34], [35].

In [34], Plötz et al have built a denoising dataset, which was greatly inspiring. In their study [35], Schwartz et al introduced DeepISP, a comprehensive deep neural model that encompasses the entire camera image signal processing unit. By learning from sensor raw data, Chen et al. [7] addressed the extremely low-light image enhancement problem. The authors of a recent study [8] employed RAW data to assist in the restoration of details and structures for super-resolution, aiming for improved performance in real-world scenarios. A RAW image-featured dataset SR-RAW for super-resolution was constructed in [9] by Zhang et al, where camera optical zoom was used to collect the dataset. They showed that operating directly on sensor raw data can indeed be advantageous. The various image processing tasks can be enhanced by using high-bit sensor raw data, as shown in all of these studies.

For video deblurring, there is, however, no high frame rate RAW datasets, which is why most of prior methods have only worked in the RGB domain. Consequently, we constructed Blurry-RAW as a novel RAW video dataset for blur effect prevention. On the other hand, we introduce 3D-ISPNet, a newly designed network structure which can directly learn from sensor raw data. The proposed network can directly operate on sensor raw data, recovering finer details and achieving a superior performance compared to the previous RGB data-driven methods.

III. THE CONSTRUCTED BLURRY-RAW DATASETS

We create a new dataset called blurry-RAW, which consists of sets of blurry and sharp RAW frames along with their RGB counterparts. This dataset is designed to facilitate end-to-end learning for RAW video blur removal. Blurriness in images primarily occurs due to the accumulation of optical signals that are acquired by camera sensors over the course of the time of exposure. The process of accumulating signals can be described in the following way:

$$B_{RAW} = \frac{1}{T}\int_{t=0}^{T} S\,(t)\,dt \cong \frac{1}{M}\sum\nolimits_{i=0}^{M} S\,[i] \qquad (2)$$

T represents the exposure time, while $S\,(t)$ stands for the signals that the camera sensor captures at a specific time t. Similarly, M denotes the total count of recorded frames, while $S\,[i]$ refers to the i-th clear RAW frame in the sequence. We create blurry RAW images by taking an average of consecutive sharp RAW frames, using a sliding window of one frame stride, ensuring that the synthesized blurry RAW video has the same fps value as the origin sharp RAW video. Subsequently, we can generate realistically smooth blurry frames directly using RAW data without undergoing any further

Algorithm 1 3D-ISPNet Training

```
Input: paths to theBLURRY-RAWvideo data subsets
Output: a trained 3D-ISPNet, Deblurred video
Trainet(x, e, b, lr)
    prepare the input datax, as a list of 5D tensors
    e is the number of iteration epochs
    b is the number of batches
    lr is the learning rate
    fit the model to the input datax
    compilethe model
    for 0 to edo
        for 0 to bdo
            use the L1 loss function in (12)
            train 3D-ISPNet model
        end for
        return the trained 3D-ISPNet
    end for
end Trainet()
Testnet(mdl, b, r, n)
    mdl is the pre-trained 3D-ISPNet
    b is a 5D tensor, contains the blurry frames
    r is a list the deblurred frames
    n is the number of the input tensors
    while i < ndo
        r = mdl.predict(b)
        i++
    return r
end Testnet()
```

post-processing stages, avoiding the common parallel noises and effects such as gamma correction, as in [5]. Fig. 2 showcases three different possible ways to construct blurry data.

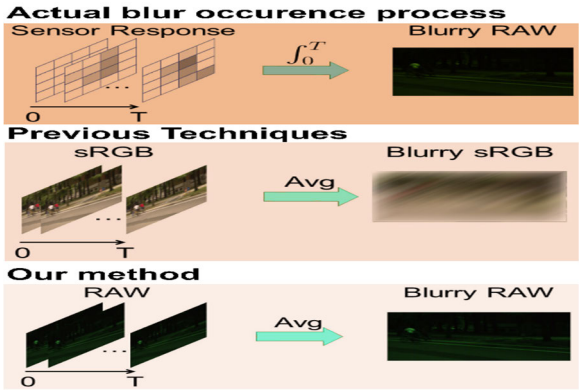

FIGURE 2. Blur effects occurrence by signals accumulation. Directly averaging on high fps RAW data allows a more realistic blur effects synthesize.

We recorded 360 RAW video clips using the Canon EOS-1D X Mark III camera, paying attention to quality, diversity, lighting, and scene dynamics. In order to ensure recording clear frames and contain enough lightness, we fix the aperture of the camera to the largest value f/4.0. The videos we record are taken at various locations, such as gardens, schools, shopping malls, intersections, sports fields, playgrounds, etc.

Using a sliding window of one frame stride, we averaged each 8 successive sharp frames of the recorded RAW videos to synthesize their blurred counterparts. Ending up constructing a set of sharp/blurry pairs of RAW video clips.

A total of 12600 pairs, consisting of RAW format sharp/blurred frames, and their post-processed RGB counterparts are included in the built Blurry-RAW dataset, which is of much greater size and quality compared to the related datasets. Table 1 illustrates a comparison between constructed dataset and other commonly used datasets in the related literature. The introduced Blurry-RAW dataset allows RAW data-driven end-to-end training of blur effect prevention models, and initiates new opportunity to enhance the prior RGB data-powered approaches. The proposed dataset is publicly released, and currently accessible through the following link: https://github.com/nahliabdelwahed/BLURRY-RAW-Dataset

IV. THE PROPOSED DEBLURRING METHOD

The proposed video blur effect prevention approach is described in detail in this section, particularly the mutual relationship between the proposed technique and the newly introduced Blurry-RAW dataset.

TABLE 1. Comparison of deblurring datasets: fps stands for the frame rate used for data recording. The higher the better.

Datasets	RGB	RAW	fps
GoPro [5]	√	×	240
Deblur-RAW [41]	√	√	30
REDs [38]	√	×	120
Blurry-RAW (ours)	√	√	60

A. NETWORK ARCHITECTURE

The proposed 3D-ISPNet model structure is illustrated in Fig. 3 and Table 2. It's an autoencoder structure formed with 3D convolutions (3DConv) in both encoder and decoder levels, and centered by a 3D vision transformer (3DVT) block, Fig. 4. This allows for accurate representation of temporal dynamics between input frames, resulting in enhanced video restoration performance. 3Dconv is a five-dimensional filter of size $C_{in} \times C_{out} \times t \times h \times w$, where t denotes the temporal dimension and $(h \times w)$ denotes its spatial size. C_{in} and C_{out} are the layer's input and output number of channels. Video motion trajectories, actions, and correspondence between frames can be modeled using the additional temporal dimension. In this study, we use 3D-ResNet (3DR-18) with 18 layers as the base backbone. In principle, any 3D CNN structure could be adoptable as a backbone. However, in order to strike a balance between accuracy and speed, we assess various forms of 3D Convolutional Neural Networks (CNN)

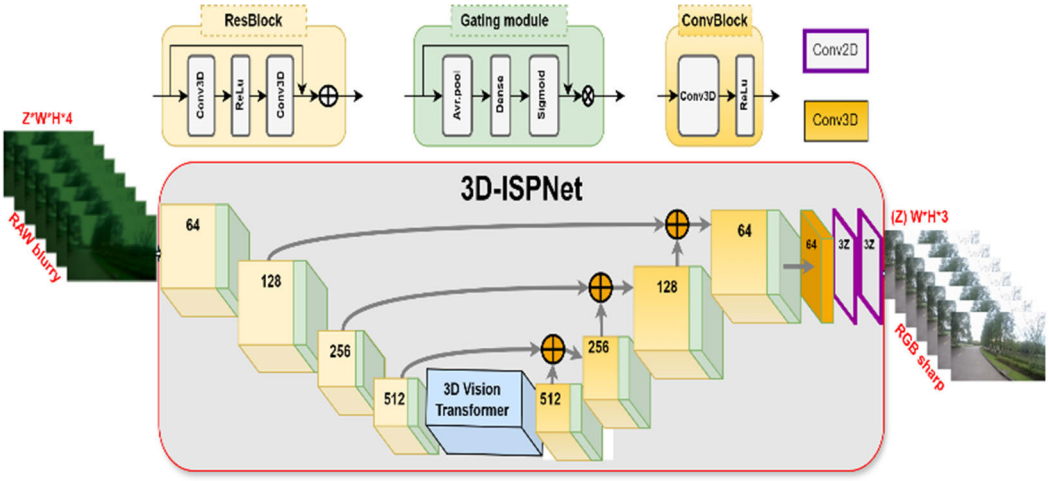

FIGURE 3. 3D-ISPNet Architecture: (a) An autoencoder neural structure, incorporating a feature gating module after each 3D space-time (de-)convolution block, as well as a 3D vision transformer block in the middle.

TABLE 2. 3D-ISPNet Architecture: Z, C, P AND FC denote number of in/output frames, channels, padding and fully connected, respectively.

	Layers		C	Strides	Kernel size	P	Output size
Input layer	5D Tensor	(batch, frames, w, h, channels)	-	-	-	-	Z*512*512*4
3D-ResBlocks	ResBlock * 4	Conv3D, ReLu, skip-connection	512	(1,1,1) and (1, 2,2)	1, 3 and 7	0	Z*64*64*512
Gating Modules	Gating Module * 8	Avg.pooling, FC, sigmoid	-	-	-	-	32
3DtransConvs-Blocks	TransCons-Block * 4	ConvTranspose3D, ReLu	128	(1,1,1) and (1, 2,2)	3	0	Z*512*521*64
3DVT Block	1	-	-	-	-	-	Z*512*521*64
Conv3D layer	3D convolution * 1	Conv3D, ReLu	64	1	3	0	Z*512*521*64
Fusion layer	2D convolution * 1	Conv2D, BatchNorm, ReLu	3Z	1	3	0	512*521*3Z
Output layer	2D convolution * 1	Conv2D, Tanh	3	1	7	0	(Z) 512*521*3

as potential backbones. By removing the final classification layer from 3DRD-18, we are left with five convolutional blocks, each consist of two 3Dconvs with a cross-layers connection.

Additionally, we eliminate temporal striding to preserve important details in each generated frame, as down-sampling operations like striding and pooling can sometimes result in loss of sharpness. However, in order to keep the computation manageable, we do choose to use a spatial stride of 2 in the first, second, and the fourth convolutional blocks. The decoder utilizes a progressive approach to up-sample and fuse features at multiple scales, using the deep latent representation captured by the encoder to construct the output frames.

Up-sampling is performed using transpose convolution (TransConv-3D) layers with a stride of 2. A Conv-3D layer is added after the last TransConv-3D layer in order to mitigate the commonly observed checkerboard artifacts. Additionally, encoder-decoder cross-connections are also used to ensure that accurate and sharp frames are recovered by low and high-level information fusion.

Once the decoder produces a three-dimensional feature map, it goes through a temporal fusion layer. This layer uses a 2D convolution to merge the temporal dimension features along with the channel dimension, resulting in a 2D spatial feature map. This step is for merging and aggregating information contained within multiple consecutive frames for prediction. Finally, this feature map undergoes a convolution using a 7 × 7 kernel in two dimensions (2Dconv), which predicts an output with dimensions of h x w x 3Z.

TABLE 3. Parameters configuration of the 3DVT architecture used in the grid search.

D	L	D	k	Configuration
2048	4	64	4	1
		32	8	2
		16	16	3
	6	64	4	4
		32	8	5
		16	16	6
	8	64	4	7
		32	8	8
		16	16	9
3072	4	64	4	10
		32	8	11
		16	16	12
	6	64	4	13
		32	8	14
		16	16	15
	8	64	4	16
		32	8	17
		16	16	18

This output is then divided into Z individual frames along the channel dimension. Our network has been carefully designed to handle video blur effect with high efficiency, regardless of the value of Z, while making only minimal modifications to the main architecture.

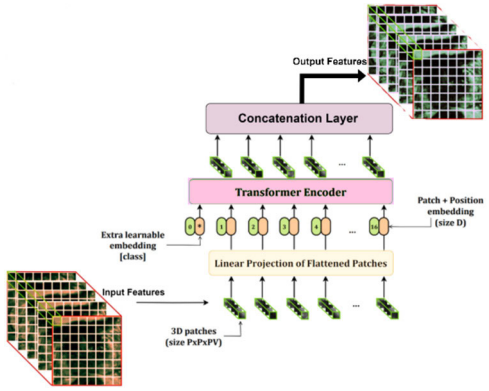

FIGURE 4. 3D Vision transformer (3DVT) block.

B. FEATURE GATING MODULE

As a self-attention mechanism in one of its forms, spatio-temporal feature gating technique is utilized in deep learning models for several tasks, such as image classification [18], video interpolation [9] image restoration [39]. In our architecture, we implement the gating module following every layer.

Considering a feature map of size $f_i = C \times T \times H \times W$, whereas C, T and $H \times W$ are channels number, temporal size and spatial size, respectively. The gating layer will produce the following output:

$$f_0 = \sigma\left(\odot.pool\left(f_i\right) + b\right) \odot f_i \tag{3}$$

where σ, $b \in R^c$ and $\odot \in R^{c \times c}$ are the sigmoid activation function, bias parameters and learnable weight, respectively. Whereas, $\odot$ is the element-wise multiplication over the channel dimension and *pool* is a spatio-temporal pooling layer. A feature gating mechanism of this nature would effectively learn to increase the importance and focus on relevant dimensions of feature maps that capture valuable patterns for preventing frame blur effects.

C. 3DVT BLOCK

The 3DVT block in Fig. 4 and Table. 3, is inspired from ViT model [24], which was initially designed for two-dimensional data, as it has been modified in this study to handle three-dimensional features. Instead of flattening a 2D patch, each embedding is obtained by horizontally tokenizing an extracted 3D feature on its temporal dimension. We define the 3DVT input features as x ∈ R H×W×Z×C, where (H, W, Z) represents the resolution of the volumetric input, and C denotes the number of filters applied to obtain these features. To tackle the high memory consumption of transformers and promote token efficiency, we follow the formalism in [43] to divided the 3D input features into H non-overlapping horizontal strips. Whereas each strip contains Z tokens with dimensions of C/2.

We downscaled the original ViT architecture in [24], significantly reducing the number of learnable parameters. To find the optimal configuration, we conducted a grid search, exploring different values for the multilayer perceptron (MLP) size (d), hidden size (D), number of layers (L), and number of attention heads (k).

In our grid search, we also considered the value of d used in the ViT architecture, which was set to 3072 in the original article [24]. The parameters of each configuration explored in the grid search are summarized in Table 3. Furthermore, Fig. 4 provides a visualization of a generic 3DVT architecture, highlighting the parameters varied during the grid search.

D. LOSS FUNCTION

By utilizing a pixel-level loss, such as L1, to compare the predicted and ground truth frames, we can train the entire network in an end-to-end manner.

$$l\left(\left\{\hat{I}\right\},\{I\}\right) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{Z-1}\left\|\hat{I}_j^{(i)} - I_j^{(i)}\right\|_1 \tag{4}$$

where $I_j^{(i)}$ and $\hat{I}_j^{(i)}$ are the j^{th} ground-truth and the j^{th} predicted frame of the i^{th} training clip, respectively. Whereas Z is the number of predicted frames at a single shot, and N is the mini-batch size used in training.

(a) blurry (b) sharp (c) restored(ours)

FIGURE 5. 3D-ISPNet visual results on blurry-RAW testing subset.

E. IMPLEMENTATION DETAILS

3D-ISPNet has been trained on the training split of the proposed Blurry-RAW dataset and evaluated on its test split.

We choose a crop size of 512 by 512. Random frames order reversal and horizontal flipping have been utilized for data augmentation. A learning rate of $2 \times 10{-}4$ was initially employed, which was reduced by half whenever the training progress stagnated or reached a plateau. The network under consideration undergoes training for a total of 200 epochs, utilizing a mini-batch size of 64 and executed on a GeForce RTX 3080 GPU. Data augmentation is performed during training by randomly selecting a patch of frame sequences and applying temporal order inversion, along with random horizontal flipping. The Inference time measurements were

FIGURE 6. Qualitative comparison on blurry-RAW testing set.

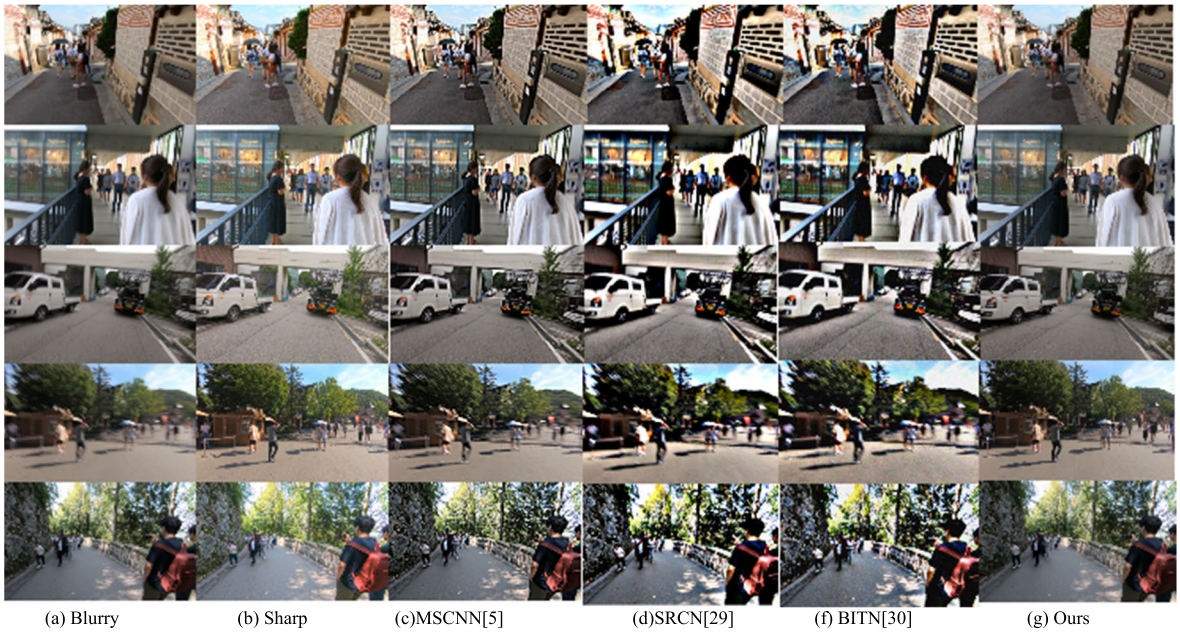

FIGURE 7. Qualitative comparison on RAW-REDs testing set.

conducted on NVIDIA-2080Ti GPU. The measured time excludes the data pre-processing and CPU/GPU transfers time, and only considers the required time for a single forward pass.

The obtained results are found by inferring over 100 samples from the Blurry-RAW testing set using a crop size of 512×512. The total time needed to restore all frames in multi-frame blur prevention is computed as the cumulative time for the restoration process. During the calculation of inference time, GPU warm up time and CUDA operations were also taken into consideration.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section provides evidence of the superiority of the proposed dataset and method through a series of experiments, conducted on a standardized hardware environment to ensure an unbiased evaluation of all methods. The prior related RGB data-driven methods are also fine-tuned using the RGB data from the newly constructed Blurry-RAW dataset.

We employ the structure similarity index (SSIM) and the peak signal-to-noise ratio (PSNR) as the metrics for evaluation. Detailed descriptions of all experiments can be found in the following sections.

FIGURE 8. Qualitative comparison on RAW-GoPro testing subset.

Upon evaluating different types of 3D-CNN as potential backbones, we ultimately opted for the 3DR-18 backbone to serve as the standard encoder for the designed 3D-ISPNet. This decision was made to strike the optimal balance between preventing blur effects and achieving high accuracy and speed in our approach. Besides Blurry-RAW testing dataset, we also utilize RAW-GoPro and RAW-REDs testing datasets for evaluation the proposed model performance, by reversing the GoPro and REDs datasets's RGB content to their RAW format using the CycleISP method [39].

A. QUANTITATIVE RESULTS

Within this research, we have compared our proposed method with a selection of established deblurring approaches that are considered representative in the field.

In Zhang et al.'s work [30], they achieve impressive results on the GoPro dataset with low computational cost by deblurring cropped patches. In Nah et al.'s paper [5], they release the GoPro dataset publicly and employ a multi-scale deep learning model for blur removal. In Kupyn et al's study [26], they propose a GAN-based model to address blur effects, while in Tao et al's article [29], they improve the multi-scale model by incorporating recurrent structure. On the other hand, Zhong et al in a recent work [31] introduce a frame-oriented neural network structure for learning high motion video deblurring task.

All the methods are fine-tuned on the RGB frames from the novel Blurry-RAW dataset, using their publicly released weights and following their setup. Table. 4 and Fig. 9 illustrate the evaluation results, which reveal that the performance of the proposed method surpasses the state-of-the-art image deblurring techniques, while maintaining low computational cost. Directly operating in the RAW domain on high fps sensor raw data for preventing the appearance of blur effects, is what distinguish our approach from the related RGB data-powered deblurring techniques. Since the low-bit RGB frames that have undergone processing do not contain crucial detailed information about the actual scene, which can only be obtained from high-bit sensor raw data. By training on the newly constructed Blurry-RAW dataset, the designed 3D-ISPNet can directly operate on RAW data and learn video blur effect prevention task, which allows the proposed technique in this study, to achieve a superior and favorable performance compared to the prior conventional RGB data-driven approaches.

B. QUALITATIVE RESULTS

We have conducted a qualitative comparison of our visual findings with prior state-of-the-art techniques on various testing datasets, as depicted in Fig. 5, Fig. 6, Fig. 7, and Fig. 8. In addition to the synthetic cases, we have also presented the results of our proposed method on actual blurry frames captured using a camera's long exposure mode (1/10 sec), as depicted in Fig. 12. Our approach has demonstrated superior performance compared to the related methods [5], [29], [30], as evidenced in Table. 4. Unlike the RGB data-based state-of-the-art deblurring techniques, our proposed method is capable of generating frames with enhanced

structural clarity and finer details. This implies that the valuable and abundant information preserved in RAW frames is indeed advantageous for the task of preventing video blur.

C. ABLATION STUDY

To ensure the effectiveness of the key components of the 3D-ISPNet and its training strategy, comprehensive ablation studies were conducted. In order to maintain fairness, all models were trained with the same configuration and settings. The outcomes of these studies are summarized in Table. 5.

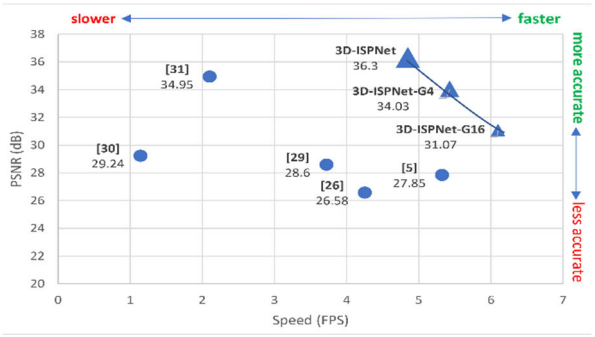

FIGURE 9. 3D-ISPNet accuracy (PSNR) versus inference speed (frames per second).

1) BACKBONE ARCHITECTURE

In our study, we explore the use of 3D convolutions to model frame sequences spatio-temporal relationships and enhance the task of video blur prevention. To test this hypothesis, we also train a video blur prevention model using 2D convolutions instead, and the results are presented in Table 5(a). In the training of the 2D ResNet, the RGB channels of the input are concatenated before being input into the network.

TABLE 4. Quantitative results.

Datasets / Methods	BLURRY-RAW		RAW-GOPRO		RAW-REDS		DEBLUR-RAW	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCN [29]	28.60	0.824	29.10	0.856	28.87	0.837	28.91	0.841
MSCNN [5]	27.85	0.819	28.03	0.839	27.91	0.828	27.96	0.837
BITN[31]	34.93	0.932	34.95	0.944	33.98	0,923	33.92	0.918
Deblur-GAN [26]	26.58	0,811	26.84	0,861	26.63	0,831	26.04	0.801
DSHN [30]	29.24	0.880	29.72	0.887	29.53	0.883	29.13	0.809
Uform[42]	34.09	0.935	33.06	0.921	33.94	0.910	33.62	0.902
RID [41]	29.45	0.919	29.17	0.892	29.30	0.912	29.80	0.928
3D-ISPNet (ours)	**36.38**	0.971	36.22	0.967	36.31	0.969	36.35	**0.976**

2) CHOICE OF FUSION

As shown in Table 5(c), there are a number of different choices for combining features across encoders and decoders using skip connections. The term ''fusion-concat'' indicates the concatenation of corresponding feature maps along the channels dimension. On the other hand, ''fusion-add'' involves adding the features from the encoder to the decoder, while ''No fusion'' denotes the absence of cross-layer skip connections between the encoder and decoder. It is observed that incorporating some form of feature transfer between the encoder and decoder is crucial, as it allows for the aggregation of supplementary information learned from both low-level and high-level features, resulting in more accurate prevention of frame blur effects. In our final model, we have opted for ''fusion-concat'' as it has shown superior performance compared to ''fusion-add''.

3) TEMPORAL STRIDING

Convolutional neural networks (CNNs) commonly use pooling or striding techniques, but these can inadvertently discard important fine details in images, which are critical for generative tasks like deblurring. To verify this experimentally, we applied 2x (1/2x) and 4x (1/4x) temporal striding in the encoder (decoder), and the results in Table 5(b) supported this hypothesis, showing a drop-in performance from 36.3 to 35.2 with larger temporal striding. As a result, we opted to use a temporal stride of 1 in all the 3DConv layers.

4) LOSS FUNCTION

Previous studies [45] have examined the comparison between pixel-based losses and perception-powered losses [24]. Relying solely on L1 or L2 loss can increase the PSNR score, however, such optimization often leads to overly smooth outputs and visually blurry predictions. On the other hand, incorporating VGG-based perception loss could yield visually sharper images. However, based on the results presented in Table 5(d), it appears that adopting additional loss functions such as Huber loss or VGG loss did not lead to improvement in PSNR and SSIM metrics, except for the case of using only L1 loss, which resulted in increased visual sharpness of the image in our study.

5) CHANNEL GATING

Channel gating and 3DVT blocks play a significant role in the designed network, operating in two distinct attention mechanism fashions. Notably, the model trained without the spatiotemporal feature gating demonstrates a considerable decrease in the PSNR value, it drops from 36.37 to 32.1dB, as shown in Table 5(e). Similarly, the PSNR value drops from 36.37 to 31.97dB while omitting the 3DVT block, as illustrated in Table 5(g). This can be interpreted as that the introduced model has successfully learnt to put more attention weight on regions with significant motion, as visualized in Fig. 10, which further validate the effectiveness of the adopted attention mechanisms for this deblurring task.

Ablation results revealed the importance and the influence of each designed component. By taking in consideration the findings of this ablation study, we settle on a final 3D-ISPNet model structure, which enables an effective and robust video blur effect prevention using sensor raw data.

TABLE 5. Ablation results for 3D-ISPNet design on different components; (a) backbones, (b) temporal striding, (c) fusion methods, (d) loss functions, (e) channel gating block, and (g) 3DVT block.

Backbone	PSNR	SSIM
2DR-18	33.96	0.959
2DR-50	34.99	0.964
3DR-18	**36.37**	**0.978**
	(a)	

Stride	PSNR	SSIM
No stride	**36.37**	**0.978**
2x stride	35.4	0.961
4x stride	35.21	0.96
	(b)	

Fusion	PSNR	SSIM
No-fusion	35.11	0.972
Fuse-add	35.73	0.974
Fuse-concat	**36.37**	**0.978**
	(c)	

Loss	PSNR	SSIM
L2+VGG	35.91	0.961
L1-Loss	**36.37**	**0.978**
Huber-Loss	35.34	0.959
	(d)	

C Gating	PSNR	SSIM
With	**36.37**	**0.978**
Without	32.07	0.894
	(e)	

3DVT	PSNR	SSIM
With	**36.37**	**0.978**
Without	31.97	0.887
	(g)	

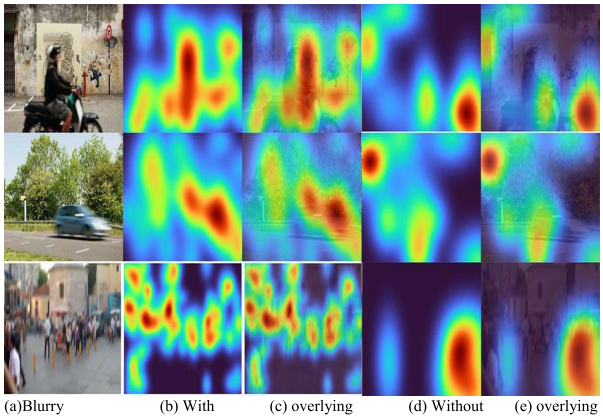

FIGURE 10. Attention map visualization with versus without 3DVT. (a) depicts the input blurry images. (b), and (d), display the resultants attention maps with and without 3DVT, respectively. Whereas (c) and (f) are overlying attention maps on the input blurry images.

6) INPUT CONTEXT Z

The number of input frames, denoted as Z, plays an important role in video blur prevention, as it determines the temporal context available to the model. A larger Z provides richer motion cues and generally improves accuracy, but it also increases computational cost and memory usage, which can hinder real-time performance. In Fig. 11, through empirical evaluation, we found that using 8 consecutive frames strikes the best balance between effectiveness and efficiency. This setting provides sufficient temporal information for reliable blur suppression while maintaining practical inference speed, making it a suitable choice for real-time deployment scenarios.

D. GENERALIZATION TO OTHER SENSORS

In order to verify that the 3D-ISPNET model pre-trained on the novel Blurry-RAW dataset is also able to perform well on different sensors/devices. As an additional resource, we chose HdM-HDR, as a burst photography public dataset [37]. HdM-HDR was constructed originally for low-light imaging and high dynamic range (HDR) on smartphone cameras. The dataset content was captured using various phones running the Android's Camera2 API. Each burst is of 15 or 30 fps.

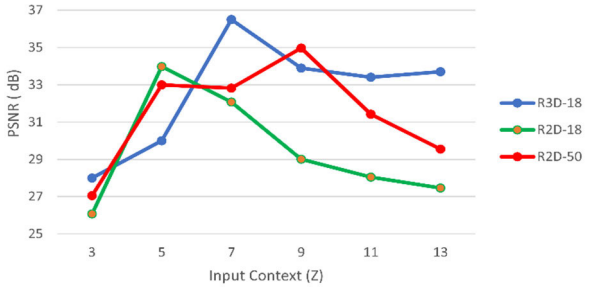

FIGURE 11. The Influence of input context Z, for video blur prevention.

FIGURE 12. 3D-ISPNet real blurry frames restoration. Blur effects are due to objects motion under the long exposure mode.

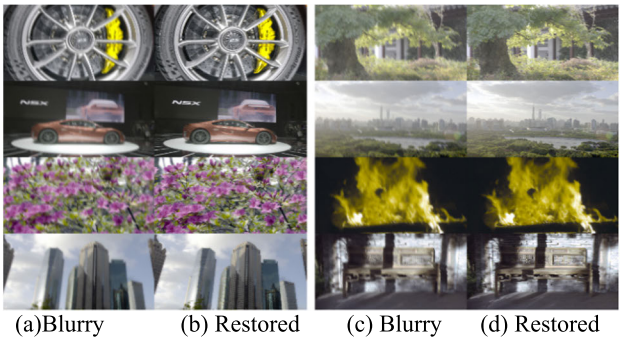

FIGURE 13. 3D-ISPNet generalization capability to various camera sensors.

We select the HdM-HDR subset which consists of 153 bursts. In each burst, the first five frames are utilized to synthesize blurry images, whereas the middle ones are kept as the ground truth sharp images. Less than five frame bursts are simply discarded. We fine-tuned the pre-trained 3D-ISPNet with a learning rate of 1e-4 for 60 K iterations. The outcomes are visualized in Fig. 13, showcasing the model's adaptive blur prevention capability learned through the training process. Thus, the deployment of the proposed pre-trained model is generalizable to various camera sensors.

VI. DISCUSSION AND FUTURE WORK

In our research, we explore the potential and effectiveness of the proposed method for preventing video blur using RAW data. We have developed a unique dataset called Blurry-RAW, which is the first blur prevention-oriented video dataset with high frame rate RAW features. By averaging high fps videos, we synthesized more realistic and smoother blur effects without aliasing artifacts. Through a series of experiments, we verified that learning directly from the RAW domain is indeed beneficial for video blur effect prevention task.

Although there may be more sophisticated cameras which can record higher frame rates under RAW video mode, to the extent of our ability, the 60 fps Canon EOS-1D X Mark III camera, is what we possess at the moment for high frame rate RAW data collection. Nevertheless, the constructed Blurry-RAW dataset is still relatively of an optimal quality, as a RAW-featured high fps video dataset which holds promise for blur effect prevention. It facilitates end-to-end model training using high-bit sensor raw data, enhancing the potential for advancements in this field.

Beyond video blur effect prevention, the proposed 3D-ISPNet model also can easily be deployed as a major component in a blur-aware camera ISP system, as it can perform one a core subtask of an ISP system in an end-to-end fashion. Our future works will investigate the possibility of extending the 3D-ISPNet capabilities, to be a multi-tasks model, able to cover further functions, like, video interpolation and high dynamic range reconstruction.

VII. CONCLUSION

In this study, high-bit and high fps camera sensor data have been leveraged to address the very ill-posed blur effects phenomenon through a newly introduced preventive strategy. As there has been a scarcity of RAW-featured datasets, the majority of the previous methods have been concentrating on the processed low-bit RGB data, missing crucial details that can only be obtained from high-bit RAW data. As a solution, Blurry-RAW has been constructed, as the first high fps RAW-featured blur prevention video dataset. The introduction of the Blurry-RAW dataset has opened up a new possibility to explore models end-to-end training using informative sensor raw data. Subsequently, as an innovative CNN-Transformer hybrid model architecture, 3D-ISPNet has been designed specifically to fit the high-bit camera sensor data-driven video blur prevention task. Extensive and diverse experiments have shown that the valuable and rich information present in RAW data provides significant benefits for blur prevention task. The trained 3D-ISPNet model has exhibited a high effectiveness in restoring finer structures and textural details. Thus, the fact of being able to directly operate in the RAW(high-bit) domain on high fps sensor data for preventing the occurrence of blur effect in the RGB domain, is what has distinguished the proposed method from the prior techniques. The method presented in this study has demonstrated superior performance compared to other related approaches, both in terms of quantitative and qualitative evaluations, resulting in a cutting-edge level of achievement.

ACKNOWLEDGMENT

The authors express appreciation to their entire laboratory team for engaging in discussions and permitting access to the ultra-fast computing resources.

REFERENCES

[1] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2013, pp. 1–8.
[2] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2014, pp. 783–798.
[3] T. H. Kim, B. Ahn, and K. M. Lee, "Dynamic scene deblurring," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3160–3167.
[4] T. H. Kim and K. M. Lee, "Segmentation-free dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2766–2773.
[5] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 257–265.
[6] S. Moriyama and K. Ichige, "An edge-enhanced branch for multi-frame motion deblurring," *IEEE Access*, vol. 12, pp. 156929–156938, 2024.
[7] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
[8] X. Xu, Y. Ma, and W. Sun, "Towards real scene super-resolution with raw images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1723–1731.
[9] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3757–3765.
[10] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3806–3815.
[11] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.
[12] X. Yu, F. Xu, S. Zhang, and L. Zhang, "Efficient patch-wise non-uniform deblurring for a single image," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1510–1524, Oct. 2014.
[13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[14] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843.
[15] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
[16] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.
[17] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, "Real-time video super-resolution with spatio-temporal networks and motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2848–2857.
[18] X. Yang, H. Mei, J. Zhang, K. Xu, B. Yin, Q. Zhang, and X. Wei, "DRFN: Deep recurrent fusion network for single-image super-resolution with large factors," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 328–337, Feb. 2019.
[19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[20] X. Liao and X. Zhang, "Multi-scale mutual feature convolutional neural network for depth image denoise and enhancement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
[21] X. Cai and B. Song, "Semantic object removal with convolutional neural network feature-based inpainting approach," *Multimedia Syst.*, vol. 24, no. 5, pp. 597–609, Oct. 2018.
[22] J. Chen, C.-H. Tan, J. Hou, L.-P. Chau, and H. Li, "Robust video content alignment and compensation for rain removal in a CNN framework," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6286–6295.
[23] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6997–7005.
[24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
[25] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
[26] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.
[27] A. Nahli, S. Cao, Z. Jia, R. Ma, and S. Xu, "Dataset and network structure: Towards frames selection for fast video deblurring," *IEEE Access*, vol. 9, pp. 61369–61382, 2021.
[28] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.
[29] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.
[30] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5971–5979.
[31] Z. Zhong, M. Cao, X. Ji, Y. Zheng, and I. Sato, "Blur interpolation transformer for real-world motion from blur," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 5713–5723.
[32] R. Zhen, "Enhanced raw image capture and deblurring," Ph.D. dissertation, Dept. Elect. Eng., Univ. Notre Dame, Notre Dame, IN, USA, 2013.
[33] M. Trimeche, D. Paliy, M. Vehvilainen, and V. Katkovnic, "Multichannel image deblurring of raw color components," in *Computational Imaging III*, vol. 5674. San Jose, CA, USA: Society of Photo-Optical Instrumentation Engineers (SPIE), 2005, pp. 169–178.
[34] T. Plötz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2750–2759.
[35] E. Schwartz, R. Giryes, and A. M. Bronstein, "DeepISP: Toward learning an end-to-end image processing pipeline," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 912–923, Feb. 2019.
[36] A. Nahli, D. Li, R. Uddin, M. Irfan, M. Oubibi, Q. Lu, and J. Q. Zhang, "ExposureNet: Mobile camera exposure parameters autonomous control for blur effect prevention," *IET Image Process.*, vol. 18, no. 12, pp. 3403–3414, Oct. 2024.
[37] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, Nov. 2016.
[38] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. M. Lee, "NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1996–2005.
[39] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "CycleISP: Real image restoration via improved data synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2693–2702.
[40] S. Nah et al., "NTIRE 2021 challenge on image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 149–165.
[41] C.-H. Liang, Y.-A. Chen, Y.-C. Liu, and W. H. Hsu, "Raw image deblurring," *IEEE Trans. Multimedia*, vol. 24, pp. 61–72, 2022.
[42] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17662–17672.
[43] F. J. Tsai, Y.-T. Peng, Y.-Y. Lin, C.-C. Tsai, and C.-W. Lin, "Stripformer: Strip transformer for fast image deblurring," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, vol. 2022, pp. 146–162.
[44] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 237–246.
[45] X. Wang, K. C. Chan, K. Yu, C. Dong, and C. C. Loy, "EDVR: Video restoration with enhanced deformable convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1954–1963.

ABDELWAHED NAHLI received the B.S. degree in computer engineering from Hassan II University, Casablanca, in 2018, and the M.S. degree in information and communication engineering from Shanghai University, Shanghai, in 2021. He is currently pursuing the Ph.D. degree with Fudan University, Shanghai. He is a Research Fellow at Shanghai Institute for Pattern Recognition and Data Science. His research interests include the Internet of Things, computer vision, artificial intelligence, and data science.

DAN LI received the B.Sc., M.S., and Ph.D. degrees in electrical engineering from Fudan University, Shanghai, China, in 2003, 2006, and 2013, respectively. He is currently an Associate Professor with the Department of Electronic Engineering, Fudan University. His research interests include signal processing and its application to NDT and medicine.

RAHIM UDDIN is currently a dedicated Ph.D. Researcher specializing in advanced digital signal processing (DSP) for next-generation communication technologies, including 6G and beyond. His work focuses on developing robust DSP frameworks encompassing high-throughput data handling, low-latency protocols, and adaptive signal analysis techniques to ensure seamless and efficient wireless communication. By integrating cutting-edge algorithms with emerging wireless hardware, he aims to drive innovation in future connectivity standards and elevate the performance and reliability of next-generation networks.

TAHIR RAZA is currently a dedicated Ph.D. Researcher specializing in laser-induced graphene (LIG) technology, flexible electronics, and multimodal sensing systems for biomedical applications. With a strong background in laser processing and flexible conductive materials. His work focuses on developing next-generation wearable sensors for digital wound monitoring. Additionally, designing a wireless microchip system for remote health monitoring.

MUHAMMAD IRFAN received the bachelor's degree in electronic engineering from UET Taxila, Pakistan, in 2015, and the master's degree in electronics and communication engineering and the Ph.D. degree in biomedical engineering from Fudan University, Shanghai, in 2021 and December 2024, respectively. Currently, he is a Biomedical Researcher with the University of Turku, Finland. He has been awarded prestigious scholarships, including the Finnish National Agency for Education Fellowship, in 2023, the FCFH Grant, in 2024, the University of Turku Grant, in 2024, Chinese Government Scholarship, from 2018 to 2024, and others. His work resulted in publications in top-tier journals, such as IEEE Internet of Things Journal, IEEE Journal of Biomedical and Health Informatics, IEEE Transactions on Biomedical Engineering, IEEE Sensors Journal, and Elsevier platforms. His expertise includes biomedical signal processing (EEG, EMG, EOG, and ECG), sleep analysis, the IoT, edge computing, federated learning, biomedical image analysis, emotion AI, and large language models (LLM).

QIYONG LU (Member, IEEE) received the B.S. and M.S. degrees from Fudan University, in July 1988 and July 1993, respectively. He has been the Vice Dean of the EE Department, from December 1998 to December 2003, and the Vice Dean of the School of Information Science and Engineering, from July 2003 to March 2010. He is currently the Vice Dean of Wuxi Institute of Fudan University. He has authored three books, more than 30 journal publications, and ten Chinese patents. He is a member of IET and Chinese Institute of Electronics.

JIAN QIU ZHANG (Senior Member, IEEE) received the B.Sc. degree from the Electronic Engineering Department, East of China Institute of Engineering, Nanjing, China, in 1982, and the M.S. and Ph.D. degrees from the Department of Electrical Engineering, Harbin Institute of Technology (HIT), Harbin, China, in 1992 and 1996, respectively. From 1999 to 2002, he was a Senior Research Fellow at the School of Engineering, University of Greenwich, Medway Campus, Chatham Maritime, U.K. In 1998, he was a Visiting Research Scientist at the Institute of Intelligent Power Electronics, Helsinki University of Technology, Espoo, Finland. From 1995 to 1997, he was an Associate Professor with the Department of Electrical Engineering, HIT, where he was a Lecturer, from 1989 to 1994. From 1982 to 1987, he was an Assistant Electronic Engineer at the 544th Factory, Hunan, China. He is currently a Professor with the Department of Electronic Engineering, Fudan University, Shanghai, China.