PVDD: A Practical Video Denoising Dataset with Real-World Dynamic Scenes

Xiaogang Xu†[1], Yitong Yu†[2], Nianjuan Jiang[2], Jiangbo Lu[2]
Bei Yu[1], and Jiaya Jia[1]
† indicates equal contribution

[1] The Chinese University of Hong Kong
[2] SmartMore
{xgxu, byu, leojia}@cse.cuhk.edu.hk, 2016312358@email.cufe.edu.cn,
jnianjuan@gmail.com, jiangbo.lu@gmail.com

Abstract. To facilitate video denoising research, we construct a compelling dataset, namely, "Practical Video Denoising Dataset" (PVDD), containing 200 noisy-clean dynamic video pairs in both sRGB and RAW format. Compared with existing datasets consisting of limited motion information, PVDD covers dynamic scenes with varying and natural motion. Different from datasets using primary Gaussian or Poisson distributions to synthesize noise in the sRGB domain, PVDD synthesizes realistic noise from the RAW domain with a physically meaningful sensor noise model followed by ISP processing. Moreover, based on this dataset, we propose a shuffle-based practical degradation model to enhance the performance of video denoising networks on real-world sRGB videos. Extensive experiments demonstrate that models trained on PVDD achieve superior denoising performance on many challenging real-world videos than on models trained on other existing datasets.

为了促进视频去噪研究，我们构建了一个引人注目的数据集，即"实用视频去噪数据集"（PVDD），其中包含200个sRGB和RA W格式的噪声清洁动态视频对。

与现有的由有限运动信息组成的数据集相比，PVDD涵盖了具有变化和自然运动的动态场景。

与使用原始高斯或泊松分布来合成sRGB域的噪声的数据集不同，PVDD使用物理意义上的传感器噪声模型和ISP处理来合成RA W域的现实噪声。
此外，基于这个数据集，我们提出了一个基于洗牌的实际退化模型，以提高现实世界sRGB视频的视频去噪网络的性能。广泛的实验证明，在PVDD上训练的模型在许多具有挑战性的真实世界视频上取得了比在其他现有数据集上训练的模型更出色的去噪性能

Keywords: Video Denoising Dataset, Dynamic Scenes, Realistic Noises

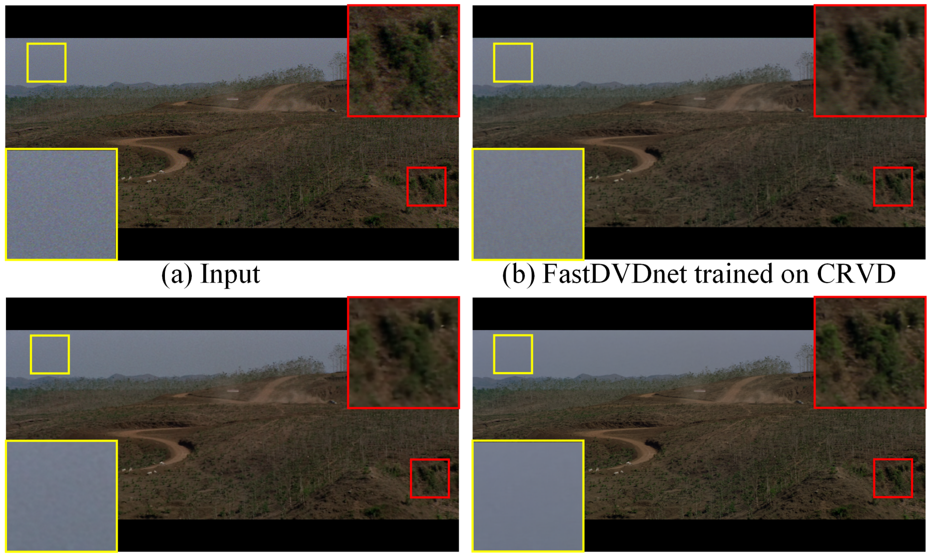
1 Introduction

The signal degradation caused by the sensor noise is common when capturing videos, resulting in poor visual quality. Therefore, video denoising is a task with significant importance. Deep-learning-based denoising techniques have emerged in recent years [36,35,24,10,47]. These methods learn the mapping from the noisy image/video domain to the clean image/video domain, while the corresponding performance heavily depends on the characteristic of the training data.

Although practical image denoising datasets have been introduced to the community, they are still rather limited regarding the diversity in motion. Directly employing them for the video denoising task generally results in the lack of variety in temporal space [47] and thus cannot provide sufficient training features for challenging video processing tasks. More practical and capable video denoising datasets, which fall short currently, are in critical demand.

虽然实用的图像去噪数据集已被引入社区，但它们在运动的多样性方面仍然相当有限。直接将它们用于视频去噪任务通常会导致在时间空间上缺乏多样性[47]，因此不能为具有挑战性的视频处理任务提供足够的训练特征。目前，更多实用的、有能力的视频去噪数据集是迫切需要的，而这些数据集的不足之处在于

There are two categories of video denoising datasets currently. The first kind collects video pairs. It creates clean video frames by averaging neighboring noisy



(a) Input (b) FastDVDnet trained on CRVD

(c) FastDVDnet trained on DAVIS (d) FastDVDnet trained on PVDD

Fig. 1: Models trained on our video denoising dataset yield superior performance on real-world noisy videos. (a) An example frame of a challenging video with the complicated noise pattern that never appears in the training set. (b)-(d) show output of FastDVDnet trained on different datasets. The model trained on our PVDD achieves the best denoising performance (yellow boxes in (a)-(d)) and preserves finer details without introducing blur and noise (red boxes in (a)-(d))

frames [47]. It requires video frames to be either static or captured at discrete time instances, as shown in Fig. 2(b), and cannot provide training data with complex natural motion. The second group, e.g., DAVIS [29], creates noisy-clean pairs with added synthetic noise. It first collects clean videos and introduces noise (e.g., Gaussian noise) in the sRGB domain. The added noise may not be the same or even similar to that produced by cameras, considering complex distributions of real noise [27,51].

如图2(b)所示，它要求视频帧是静态的或在离散的时间实例中捕获的，并且不能提供具有复杂自然运动的训练数据。第二组，例如DA VIS[29]，用添加的合成噪声创建噪声-清洁对。它首先收集干净的视频并在sRGB域中引入噪声（例如高斯噪声）。考虑到真实噪声的复杂分布，添加的噪声可能与摄像机产生的噪声不一样，甚至不相似[27,51]。

In addition, as explained in [51,47,2], shot noise in RAW format can be modeled by Poisson distributions and read noise follows Gaussian distributions. Recent work [50] further shows that the denoising networks trained with the Poisson-Gaussian noise distribution assumption achieve comparable results with networks trained on data degraded by real noise in the RAW domain. However, no existing datasets take this important fact into consideration yet when collecting data.

此外，正如[51,47,2]中所解释的，RA W格式的射击噪声可以用泊松分布来建模，而读取噪声则遵循高斯分布。

最近的工作[50]进一步表明，用泊松-高斯噪声分布假设训练的去噪网络与在RA W域中被真实噪声降解的数据上训练的网络取得了相当的结果。然而，现有的数据集在收集数据时还没有考虑到这个重要的事实。

In this paper, we construct a new powerful dataset, "Practical Video Denoising Dataset" (PVDD), by collecting clean videos with complicated natural motion, followed by the corruption of realistic noise in RAW format to form RAW video pairs. The final sRGB video pairs are produced with a calibrated ISP processing pipeline. It contains 200 noisy-clean RAW/sRGB video pairs with rich dynamics, as shown in Figs. 2(a) and 3, which greatly diversify temporal

通过收集具有复杂自然运动的干净视频，然后用RA W格式的现实噪声进行破坏，形成RA W视频对。最终的sRGB视频对是通过一个经过校准的ISP处理管道产生的。如图2（a）和图3所示，它包含了200个具有丰富动态的噪声清洁RA W/sRGB视频对，这些视频对极大地丰富了时间上的变化。

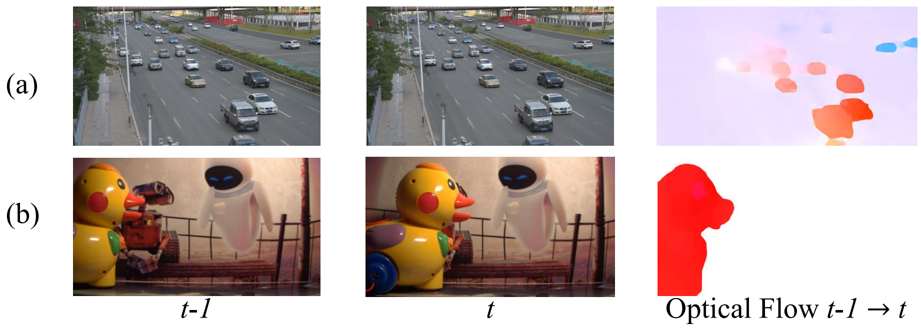


Fig. 2: Example frames and corresponding optical flow maps from PVDD and CRVD. (a) Frames from PVDD are captured in video mode, containing complex natural motion. (b) Frames from CRVD are captured at discrete time instances and only contain simple and overly large frame-to-frame foreground motion

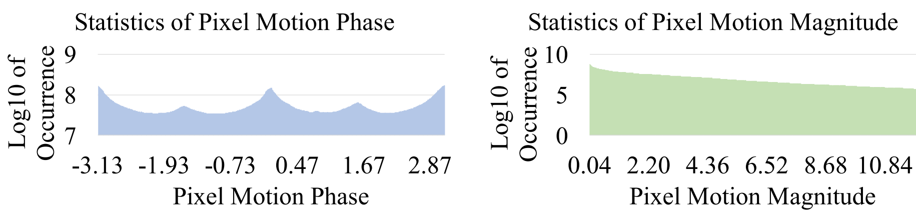


Fig. 3: Motion statistics in PVDD in terms of motion phase and magnitude, exhibiting the richness of natural motion. The per-pixel flow in every frame of PVDD is computed via a representative OpenCV algorithm [15]

local motion when used in system training. The clips also cover a variety of scenes, as illustrated in Fig. 4, for better generalization for all networks. Table 1 provides a simple high-level comparison of PVDD with existing datasets – the advantage is obvious.

In addition to the ISP modules utilized in previous work [2,19], we incorporate a color temperature module [32] and a filmic tone-mapper module [31] to better approximate the camera ISP pipeline (Fig. 6(a)). Details will be presented in Sec. 3.2. The resulting synthetic noises are realistic since ISP modules are calibrated, and noises added in the RAW domain are in accord with a realistic noise model.

除了以前工作中使用的ISP模块[2,19]，我们还加入了一个色温模块[32]和一个电影色调映射模块[31]，以更好地接近相机ISP管道（图6(a)）。细节将在第3.2节中介绍。由于ISP模块是经过校准的，所产生的合成噪声是真实的，而在RA W域中添加的噪声是符合真实的噪声模型的

PVDD can help the denoising for noisy RAW videos or sRGB videos as shown in Fig. 1, while its performance towards real-world sRGB videos with agnostic post-processing can be further improved by analyzing and simulating the impact of various degradations in the practical post-processing. To this, we propose a shuffle-based practical degradation model (SPDM) for simulation. Extensive ablation studies show such a model's effectiveness on training networks for real-noise-corrupted videos with the complicated and agnostic post-processing. *Our dataset and code will be made publicly available.*

如图1所示，PVDD可以帮助对嘈杂的RA W视频或sRGB视频进行去噪，而通过分析和模拟实际后处理中各种退化的影响，可以进一步提高其对现实世界sRGB视频的不可知后处理的性能。为此，我们提出了一个基于洗牌的实用退化模型（SPDM）进行模拟。广泛的消融研究表明，这样的模型对训练网络的有效性，适用于真实噪声干扰的视频，并具有复杂和不可知的后期处理。我们的数据集和代码将公开提供

In conclusion, our contribution is fourfold.

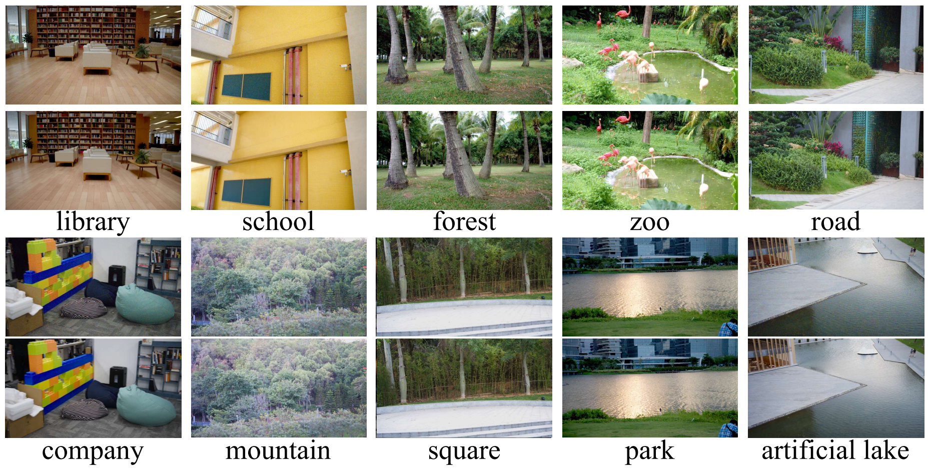


Fig. 4: Visual examples to demonstrate the diversity of our dataset, e.g., regarding different indoor/outdoor scenes

Table 1: Comparison between PVDD and other representative datasets

Dataset	Realistic Noise	Real Motions	Number of Scene	Resolution	sRGB	RAW
CRVD	√	×	11	1080P	×	√
DAVIS	×	√	90	480P	√	×
Ours	√	√	200	1080P	√	√

- We propose an information-rich dataset, named PVDD, for video denoising in RAW and sRGB domains. It incorporates complex natural motion and realistic noise.
- Our PVDD serves as a practical platform to evaluate video denoising networks and has the potential to further promote research along this line.
- To further effectively deal with complex noise patterns in the sRGB domain with post-processing, we designed SPDM, a shuffle-based degradation model.
- We conduct extensive experiments with representative deep learning architectures including both quantitative and qualitative evaluation, and user study covering 30 individuals. They manifest the superiority of the proposed dataset and the new degradation model.

2 Related Work

Video denoising datasets. Unlike image denoising datasets [27,48,5,2,3,30,9,8], video denoising datasets are far less, but still can be categorized into two kinds. The first kind collects video pairs by utilizing long exposure for clean videos and short exposure for noisy videos, e.g., SMID [8], or creating clean video frames by averaging static neighboring noisy frames, e.g., CRVD [47]. These datasets contain only static or sparse frames captured in the temporal dimension. The second kind creates noisy-clean pairs by noise synthesis, e.g., DAVIS [29]. It adds noise (e.g., Gaussian noise) to clean videos.

与图像去噪数据集[27,48,5,2,3,30,9,8]不同，视频去噪数据集要少得多，但仍可分为两类。

第一种是通过对干净的视频进行长曝光，对嘈杂的视频进行短曝光来收集视频对，例如SMID[8]，或者通过对静态相邻的嘈杂帧进行平均来创建干净的视频帧，例如CR VD[47]。这些数据集只包含在时间维度上捕获的静态或稀疏的帧。第二种是通过噪声合成创建噪声-清洁对，例如DA VIS [29]。它将噪声（例如高斯噪声）添加到干净的视频中。

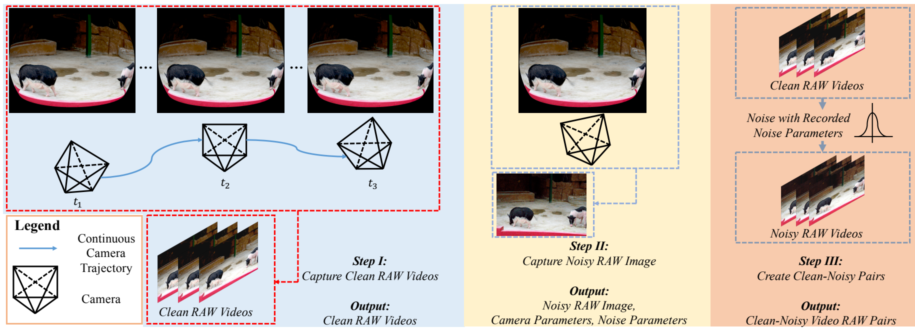


Fig. 5: Overview of the proposed pipeline to create dynamic noisy-clean video pairs in RAW format

Video denoising methods. Most video denoising techniques estimate pixel motions between adjacent frames via, for example, non-local matching [23,4,11,38], optical flow [46,6], kernel-prediction networks [26,44,45] and deformable convolutions [40]. Since accurate per-pixel motion estimation is challenging, methods with implicit motion modeling are also proposed [10,35,36,24]. These approaches either take simultaneously multiple frames as input, which are jointly processed by the model [10,35,36,47], or use information from previous frames as the additional input [17,13,24]. Results of supervised methods on real noisy videos largely depends on the training data. Self-supervised video denoising [14,12,33,21] was explored. The result still falls behind that of supervised ones.

大多数视频去噪技术通过非局部匹配[23,4,11,38]、光流[46,6]、核预测网络[26,44,45]和可变形卷积[40]来估计相邻帧间的像素运动。由于精确的每像素运动估计具有挑战性，因此也提出了隐含运动建模的方法[10,35,36,24]。这些方法或者同时采用多个帧作为输入，由模型共同处理[10,35,36,47]，或者使用以前的帧的信息作为额外的输入[17,13,24]。监督方法在真实噪声视频上的结果很大程度上取决于训练数据。自监督视频去噪[14,12,33,21]得到了探索。其结果仍然落后于有监督的方法。

3 Proposed PVDD Dataset

Our PVDD, which is the main contribution of this work, contains 200 videos in both sRGB and RAW formats (captured with a NIKON Z7 II mirrorless camera), and each video clip spans 8s to 15s.

3.1 Data Collection Setup

For the RAW video collection, we use a NIKON Z7 II mirrorless camera and an Atomos Ninja V recording monitor. The monitor is attached to the camera by an HDMI interface and is utilized to transport, encode and store the continuous video stream into an Apple ProRes RAW video (*Apple ProRes RAW is a well-recognized video codec standard directly applied to the camera sensor's data*). All videos are captured in resolution 1,920×1,080@25p and 1,920×1,080@60p format. The captured RAW videos can be further converted into sRGB videos by a calibrated camera ISP.

使用NIKON Z7 II无反光镜相机和Atomos Ninja V录音监视器。监视器通过HDMI接口连接到相机上，用于传输、编码并将连续视频流存储为Apple ProRes RA W视频（Apple ProRes RA W是一种公认的视频编解码标准，直接应用于相机传感器的数据）

We capture the video by hand-held shooting to introduce additional camera motion with more complexity than conventional rig capture. We control the camera motion slowly to avoid visible motion jittering or motion blur. Camera setting, such as aperture, ISO and focal length, is adjusted to minimize the noise level of the captured videos. For example, the aperture size is usually set between 4 and 6 to increase the amount of incident light received by the sensor, and the

手持拍摄来捕捉视频，以引入额外的摄像机运动，比传统的钻机捕捉更复杂

我们缓慢地控制摄像机的运动，以避免可见的运动抖动或运动模糊。摄像机的设置，如光圈、ISO和焦距，被调整到最小化所拍摄视频的噪音水平。例如，光圈大小通常设置在4至6之间，以增加传感器接收的入射光量，ISO调节至400，模拟噪声

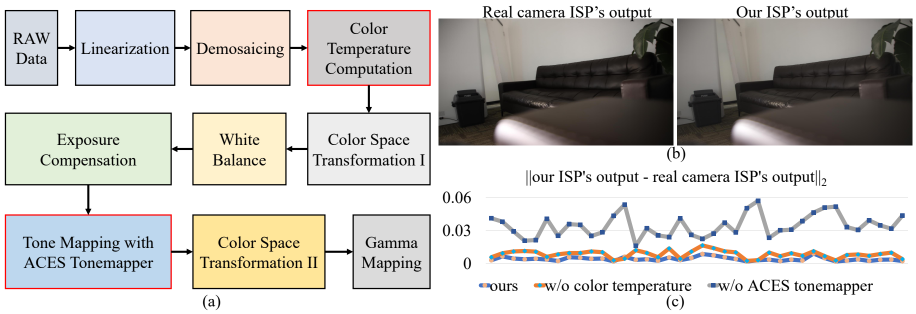


Fig. 6: (a) Modified ISP pipeline to produce sRGB videos. (b) Comparison between sRGB images produced by our modified synthetic ISP and the real camera ISP. (c) Average normalized per-pixel L_2 distance between our ISP output and that of real camera ISP across 38 different scenes. Removing either the color temperature computation module or ACES tonemapper increases this distance

ISO is set below 400 to eliminate random noise. Focal length is adjusted by the camera's built-in auto-focus module to avoid out-of-focus blur. Furthermore, we also compute the SNR for the videos in our PVDD dataset, obtaining the signal by denoising them by SOTA video denoising network [38]. With SNR value as 48.46dB, it demonstrates that our captured videos can be employed as the ground truth.

焦距由相机的内置自动对焦模块调整，以避免焦外模糊。此外，我们还计算了PVDD数据集中的视频的信噪比，通过SOTA视频去噪网络[38]对其进行去噪处理获得信号。信噪比值为48.46dB，这表明我们拍摄的视频可以被用作地面实况。

3.2 Realistic Noise Synthesis

RAW noise synthesis. During collection, we capture one clean video for every scene, as shown in Fig. 5. We obtain a set of readout noise parameters σ_r and shot noise parameters σ_s at different ISO settings, which are pre-calibrated and recorded in the camera by the manufacturer.

我们获得一组不同ISO设置下的读出噪声参数σr和拍摄噪声参数σs，这些参数由制造商预先校准并记录在相机中。

To produce noisy RAW videos, we first randomly select (σ_r^K, σ_s^K) corresponding to a particular ISO K and then add Gauss-Poisson noise accordingly to the clean RAW videos. Suppose the clean RAW video is denoted as Y. The process to obtain noisy observation X can be expressed by

$$X_c \sim \mathcal{P}((\sigma_{s,c}^K)^2 Y_c) + \mathcal{N}(0, (\sigma_{r,c}^K)^2), c \in \{\text{R, G, B}\}, \tag{1}$$

where $\mathcal{P}$ represents the Poisson distribution with mean Y_c and variance $(\sigma_{s,c}^K)^2 Y_c$, $\mathcal{N}$ denotes the Gaussian distribution with zero mean and variance $(\sigma_{r,c}^K)^2$.

The noise pattern in X is found close to real noises since 1) the noise distribution in the RAW domain can be well approximated by the Gauss-Poisson distribution [51,47,2]; 2) noise distribution model parameters are obtained from the actual calibration of the camera.

sRGB video pairs. To obtain noisy-clean video pairs in the sRGB domain, we adopt an integrated camera ISP pipeline as shown in Fig. 6(a). Camera parameters in the ISP for each clean RAW video are provided by one additional

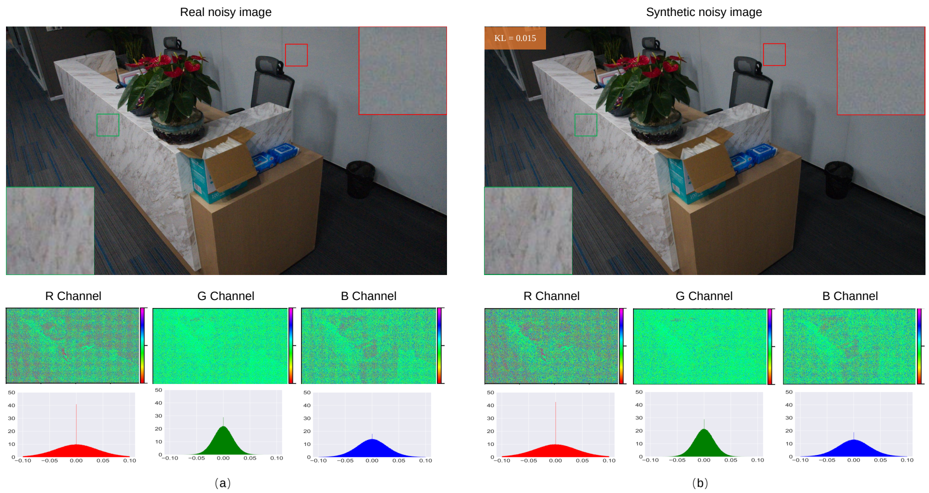


Fig. 7: Illustration of similarity between the real and synthetic noise distributions in sRGB domain.

noisy RAW image as shown in Fig. 5. Our ISP is built upon the ISP utilized in [2] and our improvement is the following.

In order to better approximate a real camera ISP, we 1) compute the color temperature and serve it as a hyper-parameter to assist the color space transformation module with an accurate color transformation matrix; 2) choose an ACES tonemapper [31] as our tone mapping function and apply it to ProPhoto RGB color space, which has a wide color gamut [34]. The example in Fig. 6(b)&(c) demonstrates the effectiveness of our improvement.

为了更好地接近真实的相机ISP，我们1）计算色温，并将其作为一个超参数，以准确的色彩转换矩阵来协助色彩空间转换模块；2）选择ACES色调映射器[31]作为我们的色调映射函数，并将其应用于ProPhoto RGB色彩空间，它具有广泛的色域[34]。图6(b)和(c)中的例子显示了我们的改进效果

3.3 Evaluation Dataset

For the quantitative evaluation of our noise synthesis and the superiority over existing video denoising datasets, we also collect an evaluation dataset, a.k.a. *Static15*, consisting of 15 different static scenes. This dataset is produced by capturing static noisy videos by the tripod-mounted NIKON Z7 II camera, whereas the corresponding clean video is derived by averaging neighboring noisy video frames [27,48,5,2]. Especially, we average 50 consecutive noisy frames to obtain a clean frame whose SNR value is 50.87dB proving clean frames' quality.

为了定量评估我们的噪声合成和对现有视频去噪数据集的优越性，我们还收集了一个评估数据集，又称Static15，由15个不同的静态场景组成。

这个数据集是由安装在三脚架上的尼康Z7 II相机捕捉静态噪声视频产生的，而相应的干净视频是由相邻的噪声视频帧的平均值得到的[27,48,5,2]。特别是，我们对50个连续的噪声帧进行平均，得到一个干净的帧，其SNR值为50.87dB，证明了干净帧的质量。

Note that we already generate noise on clean frames, and thus can compare it with the captured noise. As shown in Fig. 7, the noise distribution of the synthetic noisy frame is surprisingly close to actual captured ones in the sRGB domain. We compute the KL divergence between synthesized and captured real noise in the sRGB domain across all video clips on Static15 (noise value range [0,1]). The average normalized per-pixel KL divergence is 0.034 over sRGB and 0.041 over RAW, indicating that the synthesized noise is realistic under the same camera setting(a KL divergence value below 0.05 is acceptable [1]).

请注意，我们已经在干净的帧上产生了噪声，因此可以将其与捕获的噪声进行比较。如图7所示，合成噪声帧的噪声分布与sRGB域中实际捕获的噪声出奇地接近。我们计算了Static15上所有视频片段在sRGB域中合成的噪声和捕获的真实噪声之间的KL分歧（噪声值范围[0,1]）。平均每像素的归一化KL分歧在sRGB上为0.034，在RA W上为0.041，表明在相同的相机设置下，合成的噪声是真实的（KL分歧值低于0.05是可以接受的[1]）

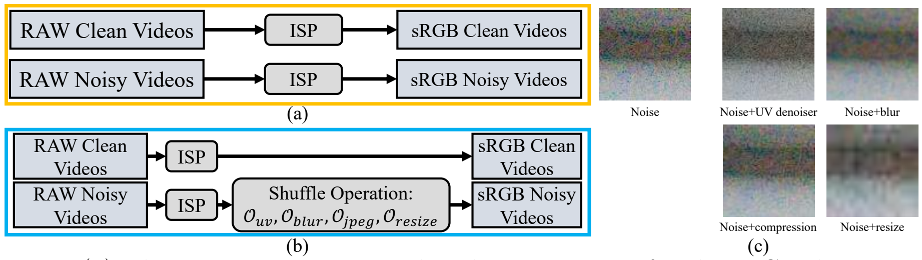


Fig. 8: (a) The common training and evaluation setting for the sRGB domain in current video denoising works that obtain sRGB videos from ISP without post-processing. (b) We, in this paper, consider dealing with noisy videos undergoing real and complicated post-processing pipelines, and thus propose the SPDM to further enhance the corresponding performance for PVDD. (c) There are various degradations for real-world videos with post-processing. They should be considered in the training of denoising networks to handle real-world degradations in the post-processing

3.4 Editable Property of PVDD

Our empirical results in Sec. 6 will show that the proposed PVDD under the Gauss-Poisson configuration outperforms other video denoising datasets regarding the training performance of denoise. Existing noise models [50,7,52,39,27,43] can also be utilized to create more sophisticated noise distributions in RAW domain. Parameters of the utilized ISP modules can be adjusted to make sRGB training data suitable for different target camera devices.

在PVDD数据集上使用后处理手段重新获得新数据

Further, extra degradation models in the practical post-processing can be applied on PVDD to obtain new sRGB data, and training with them can further improve the performance on real-world videos undergoing complex and unknown post-processing. We will present the extension in Sec. 4.

4 Shuffle-Based Practical Degradation Model

As shown in Sec. 6.1, PVDD can better help the training of denoising networks for noisy sRGB videos than existing video denoising datasets. In addition, we note that a noisy sRGB video from the ISP is very likely to be further processed with a post-processing pipeline in practice, i.e., noise in the sRGB domain is further distorted by various types of degradation, e.g., UV denoiser, blur, compression, and resizing, as shown in Fig. 8(c). In order to further improve the denoising performance of our PVDD towards such cases, we consider the simulation of all these degradations, which will be detailed in this section.

如图8(c)所示，sRGB域的噪声被各种类型的退化进一步扭曲，例如UV去噪器、模糊、压缩和调整大小。为了进一步提高我们的PVDD对这种情况的去噪性能，我们考虑对所有这些退化进行模拟，这将在本节中详细说明。

UV denoiser is achieved by a bilateral filter [37] in the YUV color space for reducing the chroma noise. In a modern camera/smartphone, it is likely to turn a video from the ISP into YUV space and then apply a chroma denoiser module to it as the post-processing [18]. Blur is simulated either by an isotropic Gaussian kernel or a generalized Gaussian kernel[22]. Moreover, compression and scaling are general post-processing operations, benefiting video transmission and storage, e.g., video sharing on the Internet. For compression, we adopt an image codec (e.g., JPEG compression) with various compression factors to degrade image quality. For resizing, we adopt commonly-used downscaling operators such as bilinear and bicubic interpolations. Inspired by [49,41], we incorporate a shuffle-based degradation strategy to generate the training data, as shown in Fig. 8(b).

Let $\mathcal{O}_{uv}$, $\mathcal{O}_{blur}$, $\mathcal{O}_{jpeg}$, and $\mathcal{O}_{resize}$ denote the above mentioned degradation operations respectively. For each degradation, we use a probability parameter p, as well as a magnitude parameter λ to determine how this degradation should be applied to a given video clip. We denote a degradation operation as

$$\mathcal{O}(x; p, \lambda) = \begin{cases} \mathcal{O}(x; \lambda) & : \text{with probability } p \\ x & : \text{with probability } 1-p \end{cases}. \tag{2}$$

Given a set of sRGB video clips with only synthetic sensor noise, i.e., $x_1, ..., x_b$, we apply different degradation processes. For a particular video clip $x_i, i \in [1, b]$, we first randomly shuffle the order of the degradation operators $\mathcal{O}_{uv}$, $\mathcal{O}_{blur}$, $\mathcal{O}_{jpeg}$, and $\mathcal{O}_{resize}$. We then apply them one by one according to the shuffled order on x_i. We follow the common practice considering the fact that, in general, compression is applied as the final post-processing to introduce an extra degradation.

5 Datasets and Networks for Evaluation

5.1 Datasets

Training data. We train representative state-of-the-art video denoising networks on both PVDD and popular benchmark datasets and compare the performance under the same network structure. Previous datasets include CRVD [47] (including synthesized data part [25]) and DAVIS [29]. Note that SMID [8] dataset is not considered since its noisy and clean frames have different illumination values, making training difficult.意SMID[8]数据集没有被考虑，因为它的噪声帧和清洁帧有不同的照度值，使训练变得困难。

We divide all the videos in PVDD into the training collection (90%) and the testing collection (10%). As it is not necessary to employ all the frames (71,714 in total) for training, for each video in the collection, we select 2 or 3 clips (each contains 25 continuous frames) randomly to form the actual training set.

As for CRVD, RAW data is transferred to the sRGB domain using the default ISP implemented by SID [9]. As for DAVIS, we employ two settings. Firstly, we directly add Gaussian noise on DAVIS sRGB videos, which is the conventional setting [36,20], to produce "DAVIS I". In the second setting [26,16,47,28], we first use the unprocessing technique [5] to turn sRGB videos into the RAW format, and add Gauss-Poisson noise in the RAW domain to form pairs. The noisy RAW videos are finally converted back to sRGB. We name this dataset "DAVIS II".

至于CR VD，RA W数据是使用SID[9]实现的默认ISP转移到sRGB域的。

至于DAVIS，我们采用了两种设置。首先，我们直接在DAVIS sRGB视频上添加高斯噪声，这是传统的设置[36,20]，以产生 "DAVIS I"。

在第二种设置中[26,16,47,28]，我们首先使用未处理技术[5]将sRGB视频变成RA W格式，并在RA W域中添加高斯-泊松噪声以形成对。最后将有噪声的RA W视频转换回sRGB。我们将这个数据集命名为 "DAVIS II"

Evaluation data. Different from existing works, we consider two settings for evaluation. The first is the denoising task for noisy sRGB videos obtained from

Table 2: Quantitative comparison in the sRGB domain, (b) means the blind models. We report PSNR/SSIM

	Dynamic20			Static15		
	PVDD	DAVIS II	CRVD	PVDD	DAVIS II	CRVD
VNLnet	**33.02/0.907**	30.59/0.819	29.31/0.757	**35.35/0.915**	34.98/0.909	32.34/0.878
FastDVDnet	**33.20/0.893**	30.83/0.801	29.29/0.763	**35.98/0.931**	35.19/0.924	34.81/0.920
RViDeNet	**32.14/0.888**	30.89/0.847	28.23/0.738	**35.71/0.941**	35.02/0.934	32.93/0.892
EDVR	**30.89/0.794**	28.38/0.789	28.49/0.758	**34.82/0.890**	33.69/0.859	34.22/0.871
VNLnet (b)	**32.55/0.899**	30.55/0.826	29.25/0.823	**34.97/0.926**	34.94/0.921	32.23/0.864
FastDVDnet (b)	**33.10/0.889**	30.66/0.808	28.69/0.742	**35.32/0.916**	34.42/0.910	34.04/0.907
RViDeNet (b)	**30.97/0.849**	30.04/0.860	27.41/0.726	**36.39/0.960**	35.80/0.951	34.43/0.918
EDVR (b)	**30.17/0.770**	28.31/0.751	28.30/0.750	**34.28/0.865**	33.66/0.858	34.19/0.853

ISP without post-processing as shown in Fig. 8(a), which is the common setting for current works [47,36]. Moreover, we also consider the evaluation setting with post-processing as stated in Sec. 4.

与现有的工作不同，我们考虑在两种情况下进行评估。首先，如图8(a)所示，对从ISP获得的嘈杂sRGB视频进行去噪处理，这也是目前作品的常见设置[47,36]。此外，我们还考虑了第4节中所述的带有后处理的评估设置

We use several types of data for evaluation. The first is selected from the testing set of PVDD to contain only dynamic scenes. 20 dynamic clean video clips are picked from the testing set of PVDD to form *Dynamic20*. For each clip, three noisy versions are generated with noise levels of “Heavy”, “Medium”, and “Light”, corresponding to ISO of 20,000, 8,000, and 2,500.

The second type is *Static15* (as explained in Sec. 3), which consists of 15 static videos with real camera noise. These noisy videos are captured with 1,920×1,080 resolution at 60 FPS under various ISO settings ranging from 16,000 to 25,600. The corresponding clean version is directly obtained by averaging consecutive noisy frames.

Note that both Dynamic20 and Static15 are constructed without post-processing pipelines. And as stated in Sec. 4, go beyond existing works, we aim to evaluate the denoising performance on videos with post-processing in the sRGB domain, and the corresponding post-processing operations should be agnostic and not appear during the training to simulate the practical situation. To this, we process the videos in Dynamic20 and Static15 with the real-world application, i.e., YouTube. Especially, we upload the sRGB videos in Dynamic20 and Static15 to YouTube, and APIs in YouTube will process these videos with several operations (e.g., compression), and we then download them with the option of “1080P” to obtain the videos undergoing complex and agnostic post-processing since the processing pipeline in the YouTube is black-box. Significantly, such a procedure to obtain videos with post-processing satisfies the practical requirement since it simulates the propagation of videos on the internet where the video post-processing would always occur. The obtained videos undergoing post-processing are called *Dynamic20-P* and *Static15-P*.

为了超越现有的工作，我们的目标是评估在sRGB域进行后处理的视频的去噪性能，相应的后处理操作应该是不可知的，在训练过程中不会出现，以模拟实际情况。

我们将Dynamic20和Static15中的sRGB视频上传到YouTube，YouTube的API将对这些视频进行若干操作（如压缩），然后我们用"1080P "的选项下载这些视频，以获得经过复杂和不可知的后期处理的视频，因为YouTube的处理管道是黑箱的。重要的是，这样一个获得视频后处理的程序满足了实际需求，因为它模拟了视频在互联网上的传播，而视频后处理总是会发生的。经过后期处理的视频被称为动态20-P和静态15-P。

To further evaluate the generalization of models trained on competing datasets for noisy videos undergoing complicated post-processing, we select 15 noisy videos that undergo various real-world post-processing operations in different applications to form *General15*. Each video in General15 suffers from noise contamination from the blind post-processing pipeline.

为了进一步评估在竞争性数据集上训练的模型对经历复杂后期处理的噪声视频的通用性，我们选择了15个在不同应用中经历各种真实世界后期处理操作的噪声视频，组成General15。General15中的每个视频都受到了盲目的后期处理管道的噪声污染。

Table 3: Quantitative comparison on DAVIS I and DAVIS II

	Dynamic20		Static15	
	DAVIS I	DAVIS II	DAVIS I	DAVIS II
VNLnet	29.33/0.762	**30.59**/**0.819**	**35.09**/0.902	34.98/**0.909**
FastDVDnet	29.85/0.793	**30.83**/**0.801**	**35.66**/0.920	35.19/**0.924**
RViDeNet	30.69/**0.849**	**30.89**/0.847	34.48/0.931	**35.02**/**0.934**
EDVR	**29.31**/0.755	28.38/**0.789**	33.41/0.852	**33.69**/**0.859**
VNLnet (b)	29.11/0.753	**30.55**/**0.826**	34.10/0.887	**34.94**/**0.921**
FastDVDnet (b)	30.05/0.803	**30.66**/**0.808**	**34.62**/0.904	34.42/**0.910**
RViDeNet (b)	**30.35**/0.837	30.04/**0.860**	**35.95**/0.950	35.80/**0.951**
EDVR (b)	**29.28**/0.750	28.31/ **0.751**	33.36/0.848	**33.66**/**0.858**

5.2 Metrics

For the evaluation on datasets with ground truth (Dynamic20 and Static15), we compute the PSNR and SSIM [42] between model output and ground truth. For the evaluation on General15 with no ground truth clean frames, we adopt user study instead. The user study is conducted on General15 with 30 participants. Each participant simultaneously sees videos produced by different denoising models in an arbitrary order, and ranks them according to the subjective visual quality. More details will be presented in Sec 6.1.

在没有地面真相的General15上的评估，我们采用用户研究来代替。用户研究是在General15上进行的，有30名参与者。每个参与者同时看到由不同的去噪模型按任意顺序制作的视频，并根据主观的视觉质量对它们进行排名

5.3 Video Denoising Networks

We evaluate our dataset and others using state-of-the-art denoising frameworks with publicly available training and evaluation codes, including VNLnet [11], FastDVDnet [36], EDVR [40], and RViDeNet [47].

We follow the original setup proposed by the authors to train these networks, i.e., 15 consecutive frames for VNLnet, 5 for FastDVDnet and EDVR, and 3 for RViDeNet, respectively. We use the official implementation for training and evaluation, and all training configurations remain as the default ones. We train both blind (without noise level prior) and non-blind (with noise level prior) versions for each dataset-network setup.

我们为每个数据集-网络设置训练盲法（无噪声水平先验）和非盲法（有噪声水平先验）版本。

6 Experiments

In this section, we conduct extensive experiments to demonstrate the superiority of our dataset and the effectiveness of our designed SPDM.

6.1 Comparison on sRGB Videos

sRGB videos are commonly used for evaluation in the video denoising task.

Quantitative analysis. We evaluate models trained with PVDD, CRVD, and DAVIS II (blind and non-blind versions) on both Static15 and Dynamic20 with only sensor noise degradation and report results in Table 2. All models trained with our dataset significantly outperform themselves trained with other datasets.

在静态15和动态20上的表现，只有传感器噪声退化，并在表2中报告结果。

Table 4: Quantitative comparison on the sRGB domain for Dynamic20-P, Static15-P, and General15. We report values of PSNR, SSIM, and user study's results 在sRGB域对Dynamic20-P、Static15-P和General15进行定量比较。我们报告了PSNR、SSIM的数值，以及用户研究的结果

	Dynamic20-P				Static15-P				General15
	VNLNet	FastDVDnet	RViDeNet	EDVR	VNLNet	FastDVDnet	RViDeNet	EDVR	Avg. ranking
PVDD	**30.00/0.843**	**30.27/0.844**	**30.31/0.861**	**30.00/0.803**	**34.37/0.910**	**35.50/0.927**	**34.98/0.938**	**33.91/0.893**	**2.14**
DAVIS I	28.54/0.805	27.65/0.796	28.50/0.816	27.43/0.753	31.58/0.864	33.25/0.914	33.88/0.905	32.93/0.898	2.23
DAVIS II	27.84/0.807	27.86/0.803	28.73/0.848	27.60/0.790	31.81/0.872	33.64/0.916	33.28/0.915	33.06/0.903	2.51
CRVD	26.59/0.788	27.34/0.755	27.46/0.766	27.44/0.750	30.97/0.863	32.25/0.906	32.45/0.871	33.19/0.873	2.74
	Dynamic20-P				Static15-P				General15
	VNLNet(b)	FastDVDnet(b)	RViDeNet(b)	EDVR(b)	VNLNet(b)	FastDVDnet(b)	RViDeNet(b)	EDVR(b)	Avg. ranking
PVDD	**29.72/0.840**	**29.94/0.834**	**30.59/0.864**	**29.77/0.804**	**34.04/0.925**	**35.14/0.939**	**34.42/0.953**	**34.67/0.903**	**2.10**
DAVIS I	27.67/0.797	27.67/0.801	28.58/0.831	27.46/0.756	32.72/0.880	32.71/0.904	32.17/0.928	33.01/0.852	2.34
DAVIS II	27.75/0.816	27.79/0.802	28.65/0.846	27.61/0.721	32.74/0.905	33.05/0.916	34.02/0.933	33.12/0.855	2.42
CRVD	26.63/0.769	27.17/0.759	27.07/0.714	27.13/0.754	29.78/0.828	33.59/0.898	33.43/0.903	33.17/0.853	2.90

For example, compared with blind FastDVDnet trained on the datasets of CRVD and DAVIS II, the same model trained on our dataset achieves more than *2 dB* gain on Dynamic20 and *1 dB* gain on Static15 (in terms of PSNR), which is significant. The advantage stems from the large scale of PVDD that contains natural motion and realistic noise.

Moreover, the performance drop is observed from Static15 to Dynamic20 for all models (Table 2). Models trained on PVDD is 3.348/0.057 (PSNR/SSIM) lower on average, while the models trained on CRVD and DAVIS II reduce 5.028/0.131 and 4.682/0.096, respectively. It is obvious that the decrease on PVDD is smaller than that on CRVD and DAVIS II, verifying the usefulness of diverse motion in our dataset. The advantages of DAVIS II over DAVIS I (as displayed in Table 3) show the inaccuracy when applying Gaussian noise directly to the sRGB domain – considering initial RAW images and using the ISP to obtain the sRGB data work much better.

DA VIS II相对于DA VIS I的优势（如表3所示）显示了将高斯噪声直接应用于sRGB域时的不准确性--考虑初始RA W图像并使用ISP获得sRGB数据的效果要好得多。

Additionally, the improvement is also achieved on the Dynamic20-P and Static15-P datasets, which can measure the denoising performance on real-world videos with complex and blind post-processing pipelines, as shown in Table 4. The model trained on our dataset yields a higher PSNR/SSIM since the videos in our dataset cover many types of camera/scene dynamics and realistic noise distributions. Improvement on Dynamic20-P and Static15-P shows the ability of the model trained on PVDD to deal with real-world videos undergoing complex and unknown degeneration in the post-processing. Although the superior performance of PVDD has been proved, in Sec. 6.2, we would demonstrate that the performance on noisy videos with post-processing can be further improved with our designed SPDM in Sec. 6.2.

在PVDD上训练的模型有能力处理在后处理中经历复杂和未知退化的真实世界视频

尽管PVDD的优越性能已经得到证明，但在第6.2节中，我们将证明用我们设计的SPDM可以进一步提高后处理的噪声视频的性能。

Qualitative analysis. We provide visual comparisons on the datasets with real noise. In Fig. 9, we show visual comparison on test videos in the General15 dataset. The model trained with PVDD yields greater visual quality. For example, FastDVDnet in Fig. 9 trained with PVDD generates the most pleasing results in both smooth areas (most noise eliminated) and regions with rich textures (details preserved). Results from training with the CRVD dataset show

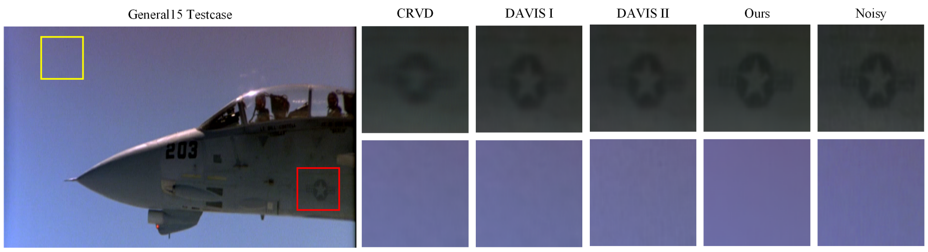


Fig. 9: Comparison of results of FastDVDnet trained on different datasets in "General15". "Noisy" is the noisy input. The model trained with PVDD yields better visual results on the smooth area (yellow rectangle) and the region with complex texture (red rectangle)

FastDVDnet在 "General15 "中不同数据集上训练的结果比较。"Noisy "是噪声输入。用PVDD训练的模型在光滑区域（黄色矩形）和具有复杂纹理的区域（红色矩形）产生了更好的视觉效果

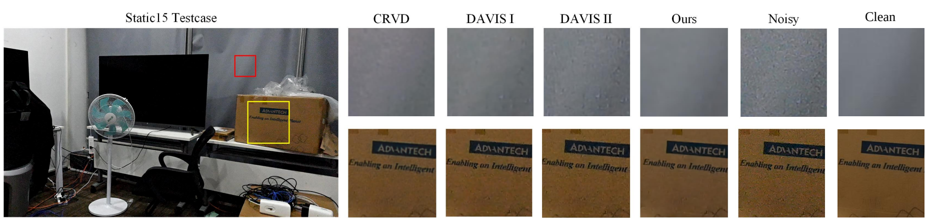


Fig. 10: Comparison of results of VNLnet trained on different datasets. We use the "Static15" evaluation set

blurry artifacts, while those with the DAVIS dataset remain much noise. Visual comparison on Static15 is shown in Fig. 10.

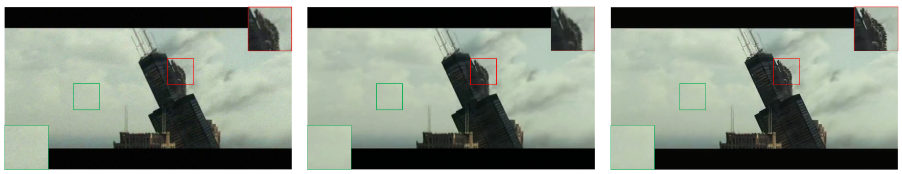
User study. For General15, we also conduct a user study. As earlier introduced, we present simultaneous results (in arbitrary order) produced by models trained with PVDD, CRVD, DAVIS I, and DAVIS II using the same network architecture and training configuration, and let the audience rank the video quality from 1 to 4 in the descending order. Each participant performs 8 (for networks)$\times$4 (for videos)$=$ 32 evaluations. We report the average ranking (Avg. ranking) of each dataset in Table 4. The results support that the networks trained on our dataset achieve better quality in terms of human perception. Moreover, we conduct a t-test for user studies, employing the T-TEST function in MS Excel. With a significant level of 0.001, all t-test results are statistically significant, since all p-values are smaller than 0.001.

6.2 Effects of SPDM

In this section, we provide additional experiments to demonstrate the effectiveness of the proposed SPDM for improving the generalization ability of trained networks on real-world noisy videos with post-processing. Since SPDM is utilized to handle the video denoising with real, complex, and agnostic post-processing, the experiments are conducted on Dynamic20-P, Static15-P, and General15.

Table 5: Comparison of model performance with/without SPDM on Dynamic20-P, Static15-P, and General15 in the sRGB domain. We report the value of PSNR/SSIM and the user study's results

	Dynamic20-P				Static15-P				General15
	VNLNet	FastDVDnet	RViDeNet	EDVR	VNLNet	FastDVDnet	RViDeNet	EDVR	Avg. ranking
PVDD	30.00/0.843	30.27/0.844	30.31/0.861	30.00/0.803	34.37/0.910	35.50/0.927	34.98/0.938	33.91/0.893	1.67
PVDD+SPDM	**31.04/0.856**	**32.16/0.862**	**31.74/0.878**	**30.82/0.814**	**35.02/0.925**	**35.97/0.934**	**35.52/0.947**	**35.02/0.916**	**1.33**
	Dynamic20-P				Static15-P				General15
	VNLNet(b)	FastDVDnet(b)	RViDeNet(b)	EDVR(b)	VNLNet(b)	FastDVDnet(b)	RViDeNet(b)	EDVR(b)	Avg. ranking
PVDD	29.72/0.840	29.94/0.834	30.59/0.864	29.77/0.804	34.04/0.925	35.14/0.939	34.42/0.953	34.67/0.903	1.75
PVDD+SPDM	**30.83/0.851**	**31.05/0.847**	**32.16/0.870**	**30.58/0.812**	**34.91/0.933**	**35.72/0.943**	**35.26/0.958**	**35.31/0.916**	**1.25**



(a) input (b) EDVR+PVDD (c) EDVR+PVDD+SPDM

Fig. 11: Qualitative comparison regarding the model trained with PVDD (b) and the same model trained with PVDD and SPDM (c). Both (b) and (c) eliminate noise, while SPDM leads to better denoising and details preserving effect (see red rectangles)

We experiment with two versions of our sRGB dataset: the first is the default one created from the RAW domain with the ISP pipeline shown in Fig. 8(a); the second is with additional SPDM as shown in Fig. 8(b). The quantitative results on the Dynamic20-P and Static15-P datasets are presented in Table 5. The improvement in terms of PSNR and SSIM shows that networks trained with SPDM have better generalization ability on real-world noisy videos with the post-processing.

One result on General15 produced by EDVR with noise level prior is shown in Fig. 11. The visual quality improvement brought by PVDD and SPDM is obvious – fine details are better preserved with reduced noise in rich texture.

User study. We simultaneously present results on General15 (in arbitrary order) produced by models trained on PVDD with and without SPDM using the same network architecture and training configuration, and let the participants vote for the video with better visual quality. The voting results are reported in Table 5. The models trained with SPDM perform consistently better on real-world noisy videos with the complex and agnostic post-processing.

7 Conclusion

In this work, we have introduced to the community a large-scale dataset for video denoising in both RAW and sRGB domains, named PVDD. The significant contribution of this dataset is that it provides training videos with rich dynamic motion as well as realistic noise. To handle complex degeneration in the practical video post-processing, we propose an extensible degradation pipeline SPDM. Extensive experiments demonstrate the superiority of the proposed dataset to train video denoising networks handling noisy sRGB videos. Additional experiments are conducted to show the effectiveness of the proposed SPDM. Significantly, PVDD provides a practical benchmark for different video denoising networks in both sRGB and RAW domains.

References

1. Abdelhamed, A., Brubaker, M.A., Brown, M.S.: Noise flow: Noise modeling with conditional normalizing flows. In: Int. Conf. Comput. Vis. pp. 3165–3173 (2019)
2. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: IEEE Conf. Comput. Vis. Pattern Recog. (2018)
3. Anaya, J., Barbu, A.: RENOIR–a dataset for real low-light image noise reduction. Journal of Visual Communication and Image Representation (2018)
4. Arias, P., Morel, J.M.: Kalman filtering of patches for frame-recursive video denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. Worksh. (2019)
5. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. (2019)
6. Buades, A., Lisani, J.L.: Enhancement of noisy and compressed videos by optical flow and non-local denoising. IEEE Trans. Circuit Syst. Video Technol. (2019)
7. Chang, K.C., Wang, R., Lin, H.J., Liu, Y.L., Chen, C.P., Chang, Y.L., Chen, H.T.: Learning camera-aware noise models. In: Eur. Conf. Comput. Vis. (2020)
8. Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: Int. Conf. Comput. Vis. (2019)
9. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: IEEE Conf. Comput. Vis. Pattern Recog. (2018)
10. Claus, M., Van Gemert, J.: ViDeNN: Deep blind video denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. Worksh. (2019)
11. Davy, A., Ehret, T., Morel, J.M., Arias, P., Facciolo, G.: A non-local CNN for video denoising. In: IEEE Int. Conf. Image Process. (2019)
12. Dewil, V., Anger, J., Davy, A., Ehret, T., Facciolo, G., Arias, P.: Self-supervised training for blind multi-frame video denoising. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2021)
13. Ehmann, J., Chu, L.C., Tsai, S.F., Liang, C.K.: Real-time video denoising on mobile phones. In: IEEE Int. Conf. Image Process. (2018)
14. Ehret, T., Davy, A., Morel, J.M., Facciolo, G., Arias, P.: Model-blind video denoising via frame-to-frame training. In: IEEE Conf. Comput. Vis. Pattern Recog. (2019)
15. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Scandinavian conference on Image analysis (2003)
16. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. IEEE Trans. Image Process. (2008)
17. Godard, C., Matzen, K., Uyttendaele, M.: Deep burst denoising. In: Eur. Conf. Comput. Vis. (2018)
18. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Trans. Graph. (2016)

19. Karaimer, H.C., Brown, M.S.: A software platform for manipulating the camera imaging pipeline. In: Eur. Conf. Comput. Vis. (2016)
20. Lee, S., Cho, D., Kim, J., Kim, T.H.: Restore from restored: Video restoration with pseudo clean video. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021)
21. Li, Y., Guo, B., Wen, J., Xia, Z., Liu, S., Han, Y.: Learning model-blind temporal denoisers without ground truths. In: ICASSP (2021)
22. Liu, Y.Q., Du, X., Shen, H.L., Chen, S.J.: Estimating generalized gaussian blur kernels for out-of-focus image deblurring. IEEE Trans. Circuit Syst. Video Technol. (2020)
23. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. IEEE Trans. Image Process. (2012)
24. Maggioni, M., Huang, Y., Li, C., Xiao, S., Fu, Z., Song, F.: Efficient multi-stage video denoising with recurrent spatio-temporal fusion. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021)
25. Milan, A., Leal-Taixé, L., Reid, I., Roth, S., Schindler, K.: MOT16: A benchmark for multi-object tracking. arXiv:1603.00831 (2016)
26. Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R., Carroll, R.: Burst denoising with kernel prediction networks. In: IEEE Conf. Comput. Vis. Pattern Recog. (2018)
27. Nam, S., Hwang, Y., Matsushita, Y., Kim, S.J.: A holistic approach to cross-channel image noise modeling and its application to image denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)
28. Paliwal, A., Zeng, L., Kalantari, N.K.: Multi-stage raw video denoising with adversarial loss and gradient mask. In: IEEE International Conference on Computational Photography (2021)
29. Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., Sorkine-Hornung, A.: A benchmark dataset and evaluation methodology for video object segmentation. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)
30. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: IEEE Conf. Comput. Vis. Pattern Recog. (2017)
31. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques (2002)
32. Robertson, A.R.: Computation of correlated color temperature and distribution temperature. JOSA (1968)
33. Sheth, D.Y., Mohan, S., Vincent, J.L., Manzorro, R., Crozier, P.A., Khapra, M.M., Simoncelli, E.P., Fernandez-Granda, C.: Unsupervised deep video denoising. In: Int. Conf. Comput. Vis. (2021)
34. Spaulding, K.E., Woolfe, G.J., Giorgianni, E.J.: 'reference input/output medium metric rgb color encodings. Proc. IS&T PICS (2000)
35. Tassano, M., Delon, J., Veit, T.: DVDnet: A fast network for deep video denoising. In: IEEE Int. Conf. Image Process. (2019)
36. Tassano, M., Delon, J., Veit, T.: FastDVDnet: Towards real-time deep video denoising without flow estimation. In: IEEE Conf. Comput. Vis. Pattern Recog. (2020)
37. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Int. Conf. Comput. Vis. (1998)
38. Vaksman, G., Elad, M., Milanfar, P.: Patch craft: Video denoising by deep modeling and patch matching. In: Int. Conf. Comput. Vis. (2021)

39. Wang, W., Chen, X., Yang, C., Li, X., Hu, X., Yue, T.: Enhancing low light videos by exploring high sensitivity camera noise. In: Int. Conf. Comput. Vis. (2019)
40. Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: EDVR: Video restoration with enhanced deformable convolutional networks. In: IEEE Conf. Comput. Vis. Pattern Recog. Worksh. (2019)
41. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In: Int. Conf. Comput. Vis. Worksh. (2021)
42. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. (2004)
43. Wei, K., Fu, Y., Yang, J., Huang, H.: A physics-based noise formation model for extreme low-light Raw denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. (2020)
44. Xu, X., Li, M., Sun, W.: Learning deformable kernels for image and video denoising. arXiv:1904.06903 (2019)
45. Xu, X., Li, M., Sun, W., Yang, M.H.: Learning spatial and spatio-temporal pixel aggregations for image and video denoising. IEEE Trans. Image Process. (2020)
46. Yu, S., Park, B., Park, J., Jeong, J.: Joint learning of blind video denoising and optical flow estimation. In: IEEE Conf. Comput. Vis. Pattern Recog. Worksh. (2020)
47. Yue, H., Cao, C., Liao, L., Chu, R., Yang, J.: Supervised Raw video denoising with a benchmark dataset on dynamic scenes. In: IEEE Conf. Comput. Vis. Pattern Recog. (2020)
48. Yue, H., Liu, J., Yang, J., Nguyen, T.Q., Wu, F.: High ISO JPEG image denoising by deep fusion of collaborative and convolutional filtering. IEEE Trans. Image Process. (2019)
49. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: Int. Conf. Comput. Vis. (2021)
50. Zhang, Y., Qin, H., Wang, X., Li, H.: Rethinking noise synthesis and modeling in Raw denoising. In: Int. Conf. Comput. Vis. (2021)
51. Zhou, Y., Jiao, J., Huang, H., Wang, Y., Wang, J., Shi, H., Huang, T.: When AWGN-based denoiser meets real noises. In: AAAI (2020)
52. Zhu, F., Chen, G., Heng, P.A.: From noise modeling to blind image denoising. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)