JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 16, No. 4, December 2022, pp. 185-193

Feeding Longer Frames for Efficient Video Denoising Model

Kavita Arjun Bhosale and Sang-hyo Park*
School of Computer Science and Engineering, Kyungpook National University, Republic of Korea
s.park@knu.ac.kr

Abstract
Recently, deep video denoising networks showed substantially high denoising performance with considerably lower computing times. However, such models may not be able to denoise long-term frames appropriately due to the various characteristics of video motion. In this paper, we propose a method that takes longer input frames and feeds them to the existing architecture. In particular, the proposed method may extract temporal information effectively from neighboring frames to address the long-term frame dependency issue. To demonstrate the performance of the proposed method, we implemented our method on opt of the state-of-the-art video denoising model. Through extensive experiments, the proposed method showed better performance in terms of quality metrics than the existing one, even with higher noise level, showing considerably lower computing times.

Category: Computer Graphics / Image Processing

Keywords: Video denoising; Motion compensation; deep learning; convolutional neural network; noise reduction; signal processing

I. INTRODUCTION

Goal of image and video denoising is to obtain original signal X from noisy observations Y. Mathematically, the problem for additive type noise can be defined as Y= X + N, where X is original signal, N is a noise and Y is available noisy observation. Consequently, many researchers working on video denoising task find it challenging due to its shooting conditions, i.e., low light and small sensors. However, video restoration requires temporal cohesiveness, thereby making video restoration challenging and demanding.

Compared to image denoising, video denoising is discovered by only few researchers. However, image denoising has been studied for a long time with great achievements. Especially, deep learning techniques based on image denoising have drawn significant attention with their excellent performance. In [1], a trainable nonlinear reaction diffusion model has been proposed, which is based on the cascade of shrinkage fields method and offers equally computational efficacy with high restoration quality. Schmidt and Roth proposed an effective image restoration method which combines a random field-based structural design into the image model and the optimization algorithm in a one unit [2]. In addition, well known algorithms are applied for image denoising such as BM3D [3], non-local bayes (NLB) [4], and multi-layer perceptron [5]. Thus, these algorithms should be trained for each noise level.

Another widely held approach involves the use of convolutional neural networks (CNN), such as RBDN [6], DnCNN [7], and FFDNet [8]. Zhang et al. [7] used the denoising CNN (DnCNN) for image denoising, JPEG image blocking and super-resolution. In addition, one of

Open Access http://dx.doi.org/10.5626/JCSE.2022.16.4.185 http://jcse.kiise.org

Received 08 July 2022; **Accepted** 06 December 2022
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

its main features is that it implements with convolutions, batch-normalization, rectified linear unit (ReLU) and residual learning [9]. Since CNN based models are not limited to image denoising, many researchers used CNN based methods as a preprocessing for video compression. Bhosale et al. [10] studied denoising framework DnCNN as preprocessing to improve compression performance of HEVC encoding.

Video denoising has been less widely studied than image denoising, and only a few of them have been demonstrated neural network-based video denoising. In general, most video denoising methods take patch-based approach. For example, widely held V-BM4D [11], and video non-local Bayes (VNLB) [12]. Chen et al. [13] proposed the first approach of recurrent neural networks. Note that these methods are based on grayscale images [14], and it is not sure the denoising models can work effectively with color components. In [15], DVDnet has been proposed to divide the denoising images in two separate steps, estimating motion of neighbor frames. Other recent approach of blind video denoising is proposed by Ehret et al. [16] and ViDeNN [17]. However, contradictory to DVDnet, ViDeNN does not employ motion estimation. Recent state-of-the-art models VNLnet [18] and VNLB [12] shows better performance for small values of noise, whereas DVDnet shows better results for larger noise.

Nowadays, videos are capturing in high-definition camera's which demands high and efficient algorithms for video restoration. Recent approach FastDVDnet [19] built on DVDnet has implemented CNN based algorithm with fast runtimes. In addition, implicitly handle motion with only single trained model on wide range noise level $\sigma \in (5, 50)$. DVDnet is best at larger values of noise, whereas FastDVDnet shows better result with smaller values of noise and small-scale resolution frames. In shortcoming FastDVDnet shows noticeable motion artifacts and noise on long-term frames. Especially, when noise value is higher. By considering these shortcomings we implemented new modified model of FastDVDnet as CNN based denoising techniques has ability to extract more features than patch-based methods.

In this paper, we introduced a modified model of FastDVDnet. Proposed algorithm designed to reduce motion artifacts and noise from long-term frames i.e., first and last frame few frames. However, modified model takes less time to denoise each color frame than original model. i.e., less than 100ms. Our aim is to implement model for high-definition video denoising (HEVC sequences) and improve results with all noise values as its challenging because of high quality frames. Experimental results proved that modified FastDVDnet model has removed noise and achieved better results in terms of quality matrices and in digital format. However, our model not only best for first and last frame but also it works very well with each frame.

The remainder of this paper is structured as follows. In Section 2, we provide summary of original FastDVDnet model. In Section 3, we present the proposed method with network architecture. In Section 4, we offer experimental results with training and testing details of the proposed method. In Section, we present future study, and summary of paper as a conclusion.

II. OVERVIEW OF VIDEO DENOISING MODEL

For video denoising, efficient algorithm necessitates because video contains more information than image. In addition, video processing needs good temporal coherency and low flickering aspects. In order to achieve these, Tassano et al. [19] proposed FastDVDnet algorithm. This model uses temporal information from neighboring frames for determining a given frame of an image sequence. When we need to denoise a give area of pixels (patch), FastDVDnet [19] exploits the characteristics of spatio-temporal neighborhoods. Accordingly, the model can view for related patches from a neighboring frame of the sequence as well. Thus, such temporal neighbors give the model supporting information to denoise the reference frame and helps to reduce flickering.

Baseline video denoising algorithm does not involve of an explicit motion assessment /compensation stage. However, the capacity of object motion handling is fundamentally embedded in the architecture. Indeed, this model composed of a number of modified U-Net blocks [20], which has been shown to have capability to discover misalignment [21], [22]. This cascaded architecture is trained on end to-end, which escapes distortions and artifacts. When denoising a certain frame at a time t, neighboring frames are also given as an input i.e., 2T= 4. Thus, total five sequential frames are used to denoise the central frame t and these frames taken as a triplet of consecutive frames as an input to denoising block 1 of U-net architecture, while all instances of denoising block 1 share similar weight and triplets composed by outputs of these blocks are used as a inputs for denoising block 2 to estimate results of central frame (See Fig. 1A). Parallel to [7,23] noise map is included as input which provide information to the network about noise distribution, and it has been shown improving denoising performance. Contrary to other denoising methods, this algorithm takes no other constraints as inputs apart from the image sequence and input noise map.

Standard Multi-scale and end-to-end training architecture contains two denoising blocks. As shown in Fig. 1A Denoising block 1 and denoising block 2 comprises of U-net architecture. Denoising Block 1 instances share same weight as an input to next. U-Nets are basically a multi-level encoder-decoder architecture including skip connections [24] that forwards output of each encoder layers to the input of the corresponding decoder layers. More detailed

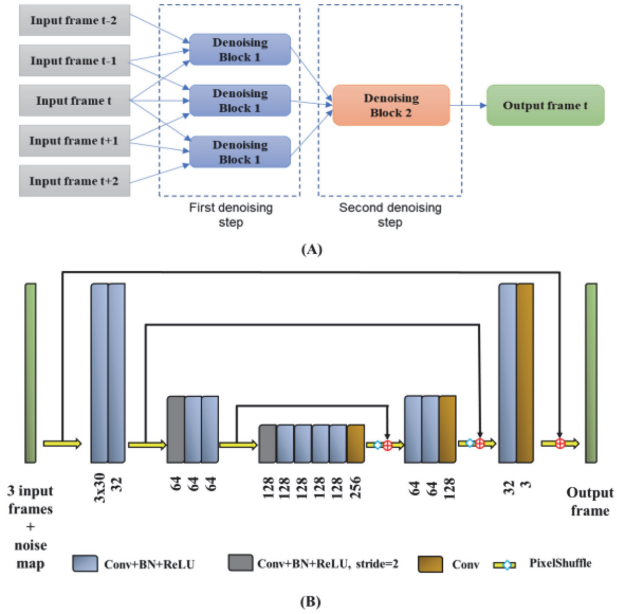

Fig. 1. FastDVDnet model architecture. (A) High-level diagram of the architecture. (B) The denoising blocks of FastDVDnet consist of a modified multi-scale U-Net.

diagram of this architecture is shown in Fig. 2B.

Original state-of-the-art method denoising blocks have some differences from the standard U-Net:

1) This model encoder is designed to take three input frames and noise map as an input.

2) To reduce gridding artifacts, pixel shuffle layer [24] is used in decoder for up sampling.

3) Reduction of memory requirements is obtained by merging the features of the encoder with those of the decoder compressed with a pixel-wise addition operation instead of a channel-wise concatenation.

4) Denoising blocks employ residual learning with residual connection among the middle noisy input frame and the output, which has been monitored to improve the training process [25].

Characteristics of the denoising blocks get a useful compromise among performance and fast running times. There are a total of D=16 convolutional layers that have been used. In most of the layers, the output of convolutions is studied by pointwise ReLU activation function, except for the last layer. Batch normalization (BN) is arranged in the middle of the convolutional and ReLU layers.

III. PROPOSED METHOD

In this section, we proposed a modified model of FastDVDnet based on U-Net architecture. Original model shown worst results with long-term frames mostly,

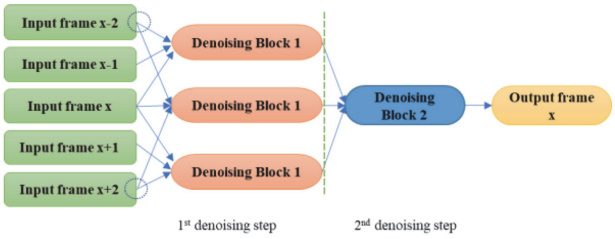

Fig. 2. Modified FastDVDnet (Input5) model architecture.

with higher noise values. As per our assumption it is unable to fetch more information from existent consecutive frames. As shown in Fig. 1A input frame t takes more information from neighboring frames i.e., Input frame t-1 and t+1. In addition, denoising block taken these frames two times. However, input frame t-2 and t+2 frames have unable to fetch more information as these frames taken only one time by denoising blocks.

In baseline model two-step cascaded denoising block assembled with U-net and shares same weight of three input frames which leads in memory reduction of a model and facilitates the training of the network. Comparing to other state-of-art-methods of denoising [11,12], FastDVDnet model is better with low flickering and fast running time

To remove noise and motion artifacts in first and last frame sequence (long-term frames) of original model, we designed model by changing input frame sequence. Fig. 2 shows an architecture of our model. When denoising a certain frame at time x, $\tilde{I}$x, its

2X = 4 adjacent frames are carried as inputs. The aim our model is to minimize artifacts with fast running times. Accordingly, we changed input frame sequence to denoise input frame x. Thus, input of the algorithm shown in (1) below,

$$\mathrm{I} = \left\{\tilde{I}_{x-2}, \tilde{I}_{x-1}, \tilde{I}_{x}, \tilde{I}_{x+1}, \tilde{I}_{x+2}\right\}. \quad (1)$$

Instead of passing three consecutive frames sequentially, we passed two consecutive frame and one random frame among five to each denoising block. By using this idea, we passed input frame x-2 and x+2 frame two times (highlighted in a circle) to denoising block 1 which helps to reduced noise and artifacts of first and last frame. However, input frame x-1 and x+1 we passed one time as both frames are closest to input frame x and it fetch more information than others.

Like FastDVDnet, we used noise map along with input frames as input to denoising block 1. In particular, the noise map is used for distribution of the noise at the input. Further, three frames taken as input to the denoising block 1 which share same weight of all instances. Next, triplet comprised by the outputs of these blocks are

applied as input for denoising block 2. Together, Denoising block 1 and Denoising block 2 share the same architecture. However, the final output is the approximation of the central input frame x.

With our proposed model we seen improved results than FastDVDnet. We analyzed our model with DAVIS, Set 8 and HEVC sequences. Nevertheless, we found that our model is achieved better results not only with FastDVDnet dataset but also HEVC test sequences (ours) as well. See section 4 for more details.

IV. EXPERIMENTAL RESULTS

Similar to FastDVDnet our architecture has been designed in PyTorch [27], a widespread machine learning library. To minimize loss function, ADAM algorithm [28]. is used and all its hyper-parameters set to their standard values. Total 80 epochs and 96 mini batch is set for training.

The mix learning rate is used i.e., it begins at 1e−3 for the initial 50 epochs, then shifts to 1e−4 for the next 10 epochs, and lastly switches to 1e−6 for the remaining of the training. In particular, a learning rate action decay is used in combining with ADAM. Other deep learning methods [29],[30] also used mix learning rate decay and adaptive rate methods with positive results.

Training dataset includes of input- output sets, as shown in (2) below:

$$P_x^j = \{\left(\left(S_x^j,\ M^j\right), I_x^j\right)\}_{j=0}^{m_x}. \quad (2)$$

where, $S_x^j = (I_{x-2}^j,\ I_{x-1}^j,\ I_x^j,\ I_{x+1}^j,\ I_{x+2}^j)$ is the collection of 5 spatial patches cropped of corresponding frames and I_x^j is clean central patch of sequence generated by adding AWGN of $\sigma \in (5, 50)$. And M^j is noise map is constant with all its components equivalent to σ. Spatio-temporal patches are arbitrarily cropped from arbitrarily sampled sequences of the training dataset.

For comparing results, first we trained original

Table 1. PSNR comparison of Set 8, DAVIS, and YUV color dataset. Best results of PSNR of each sequence shown in bold letters while similar or best PSNR of modified model in 'blue'. Best results of long-term frames are highlighted in yellow.

Sequence	Sigma	Avg. PSNR of each sequence		PSNR of first frame		PSNR of last frame	
		Original	Modified (Ours)	Original	Modified (Ours)	Original	Modified (Ours)
Set8 Rafting 960X540	10	36.64	36.67	36.98	37.00	33.75	33.72
	30	30.97	30.97	34.29	34.34	31.12	31.14
	50	28.69	28.69	32.36	32.70	29.37	29.44
Set8 Park joy 960X540	10	32.24	32.32	35.06	35.16	31.88	31.94
	30	28.59	28.25	31.98	32.18	29.63	29.75
	50	26.80	26.36	30.15	30.43	28.16	28.18
DAVIS Helicopter 854x480	10	40.15	40.06	40.56	40.82	40.55	40.64
	30	35.78	35.55	38.02	38.29	38.32	38.50
	50	33.91	33.60	36.54	36.78	36.67	36.83
DAVIS People-sunset 854x480	10	41.02	40.99	44.06	44.19	40.42	40.43
	30	36.66	36.61	40.56	41.78	38.12	38.15
	50	34.67	34.65	36.13	40.13	36.16	36.48
HEVC YUV Basketball-drive 960x540	10	37.37	37.36	41.06	41.08	38.00	38.01
	30	34.03	33.90	38.11	38.29	36.29	36.50
	50	32.14	31.99	34.96	36.44	33.76	35.11
HEVC YUV Tango 960x540	10	39.60	39.41	40.73	40.61	39.88	39.93
	30	35.61	35.38	37.72	37.99	37.81	38.06
	50	33.49	33.30	34.63	34.87	34.68	36.39
HEVC YUV PartyScene 832x480	10	33.74	33.68	35.22	35.20	34.44	34.48
	30	29.72	29.64	31.60	32.19	32.47	32.78
	50	27.60	27.48	29.06	30.11	30.40	31.09

FastDVDnet as per updated libraries i.e., updated version of DALI and Cuda version. Thus, PSNR results may differ than original model test PSNR. To train our modified model of FastDVDnet we used total m_x = 384000 training samples that extracted from DAVIS database. The spatial size of the patches is set to 96x96 with temporal size is 2X+1, i.e., 5. However, Spatial patch size was chosen in such way that subsequent patch size in coarser range of the denoising blocks is 32 × 32. The loss function is shown in (3) below,

$$\mathcal{L}(\theta) = \frac{1}{2m_x}\sum_{j=1}^{m_x} \|\hat{I}_x^j = I_x^j\|^2. \quad (3)$$

$\hat{I}_x^j = f((S_x^j, M^j);\ \theta)$ is the output of model, where θ is the set of all learnable parameters.

To analyze our model, we were used three different testset: DAVIS, Set8 and HEVC. The DAVIS set contains 30 sequences of resolution 854x480, Set8 is composed of 4 sequences of GoPro captured and 4 sequences from Derf 's Test Media collection with downscale resolution 960x540. For HEVC testset we used total 9 sequences i.e., 6 sequences downscale to 960x540 and 3 original resolution 832x480 sequences. In all cases, we adjust the maximum frame limit to 85. We compared our model results with original FastDVDnet model. In addition, we also evaluated PSNR of first and last frame of each sequence using FFmpeg open-source library.

Table 1 shows PSNR comparison results of each dataset. Original model showed better values of average

Table 2. Average PSNR of Set 8, DAVIS, and YUV color dataset. Best results highlighted in 'blue'.

Dataset	Sigma	Avg. PSNR of each dataset		Avg. PSNR of first frame		Avg. PSNR of last frame	
		Original	Modified (Ours)	Original	Modified (Ours)	Original	Modified (Ours)
SET 8	10	36.21	36.20	37.18	37.18	35.69	35.84
	30	31.48	31.32	34.11	34.15	33.48	33.57
	50	29.34	29.12	32.09	32.33	31.56	31.75
DAVIS	10	38.72	38.69	38.97	39.00	38.08	38.06
	30	33.93	33.72	35.83	36.05	35.93	36.01
	50	31.73	31.47	33.65	34.22	33.94	34.26
YUV color dataset	10	36.71	36.64	38.79	38.75	37.67	37.68
	30	32.88	32.72	35.63	35.83	35.69	35.80
	50	30.83	30.66	31.81	33.91	33.37	34.12

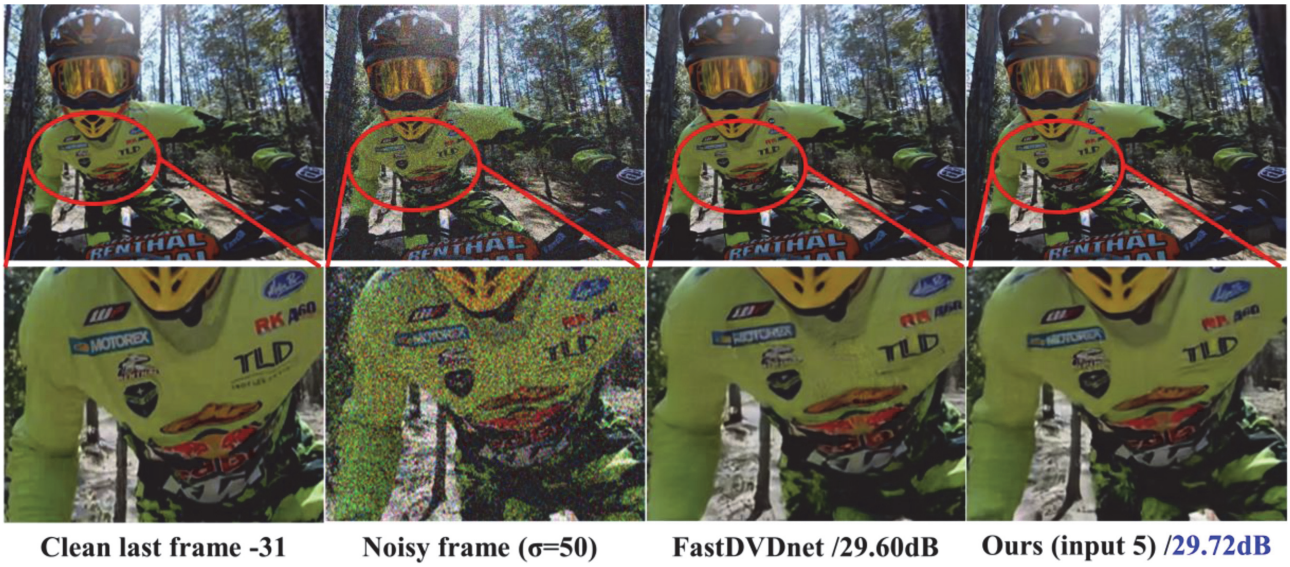

Fig. 3. Comparison results of the SET 8 'Motorbike' sequence (Last frame), σ = 50, (PSNR [dB]) with corresponding zoomed results (best viewed in digital format).

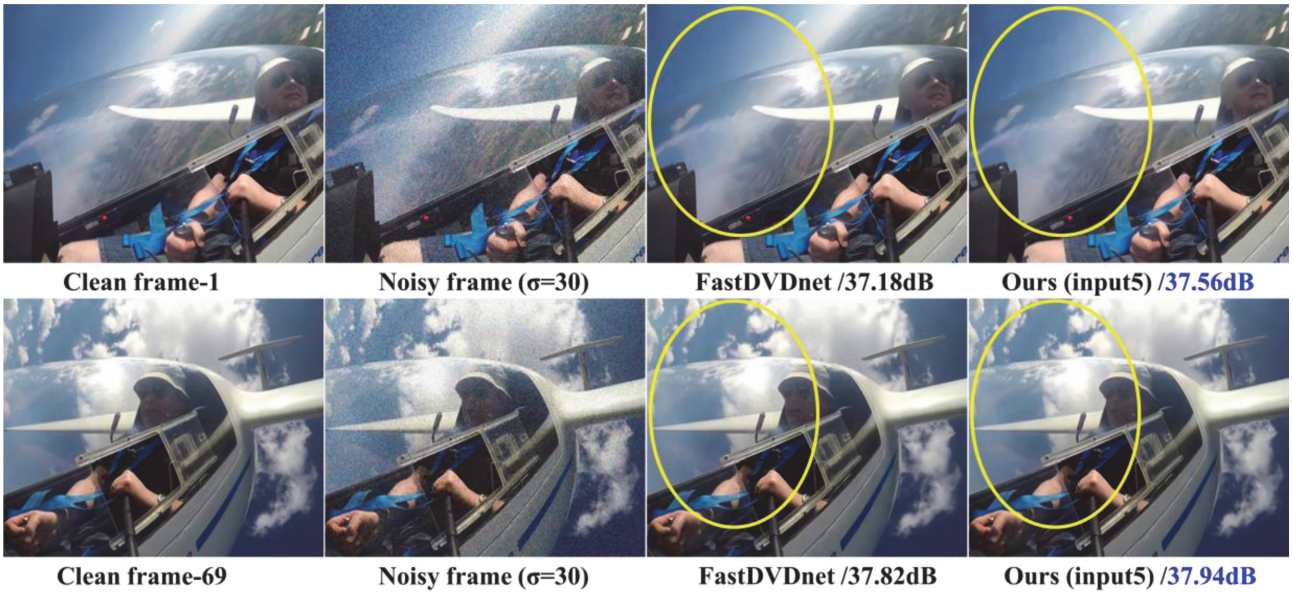

Fig. 4. Comparison of results of the DAVIS 'Aerobatics' sequence (First and last frame), σ = 30, (PSNR [dB]) (best viewed in digital format).

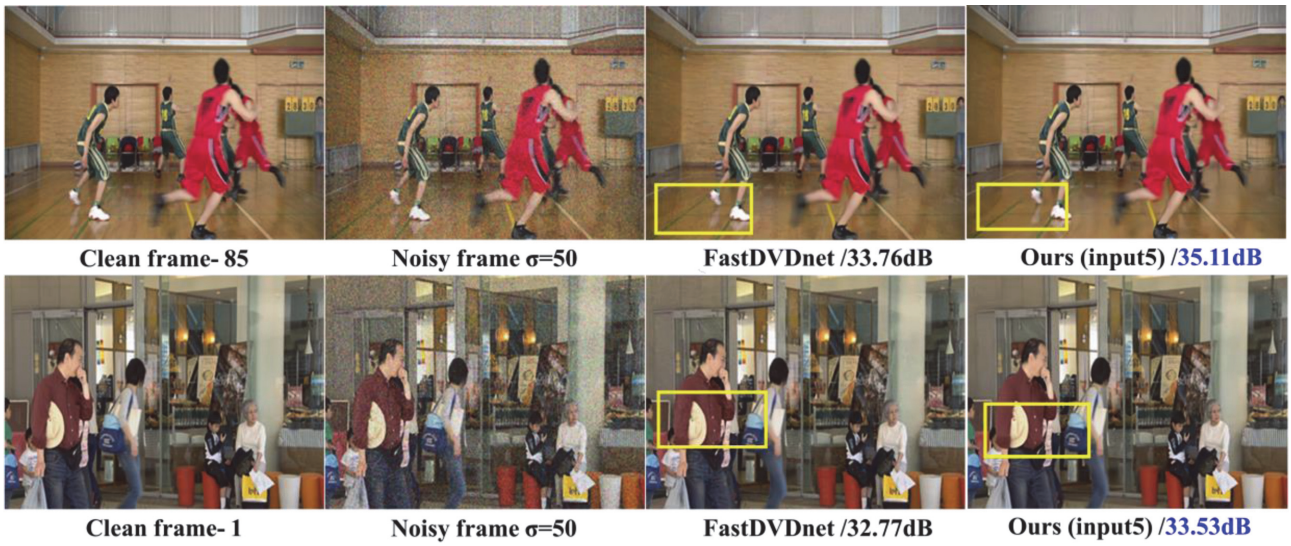

Fig. 5. Comparison results of the HEVC testset 'Basketball Drive- last frame and BQmall - First frame' sequence, σ = 50, (PSNR [dB]) (Best viewed in digital format).

PSNR, but in some cases such as rafting or park joy sequence our model showed better or similar values of PSNR than original. Other than this, HEVC sequences have higher resolution sequences than another dataset still it performs very well in Avg PSNR. For example, Tango sequence, its actual resolution of this is 3840x2160 and we tested them by downscaling it to 960x540 and it performed well as, higher resolution frames are challenging task to denoise, but our model helped to establish better result. Modified model performance is average better in all frames, but we succeeded to achieve better performance in first and last frame in all sequences especially with higher noise values. However, to calculate PSNR of each frame we have converted denoised frames to mp4 video by using FFmpeg library.

Table 2 shows Average PSNR of each dataset. We estimated average PSNR, Avg. PSNR of first frame and last frame of all mentioned dataset on σ 10,30 and 50. Though, Average PSNR of all testset original is better than our model i.e., Avg difference is 0.14dB. But it shows better results with all frames equally. However,

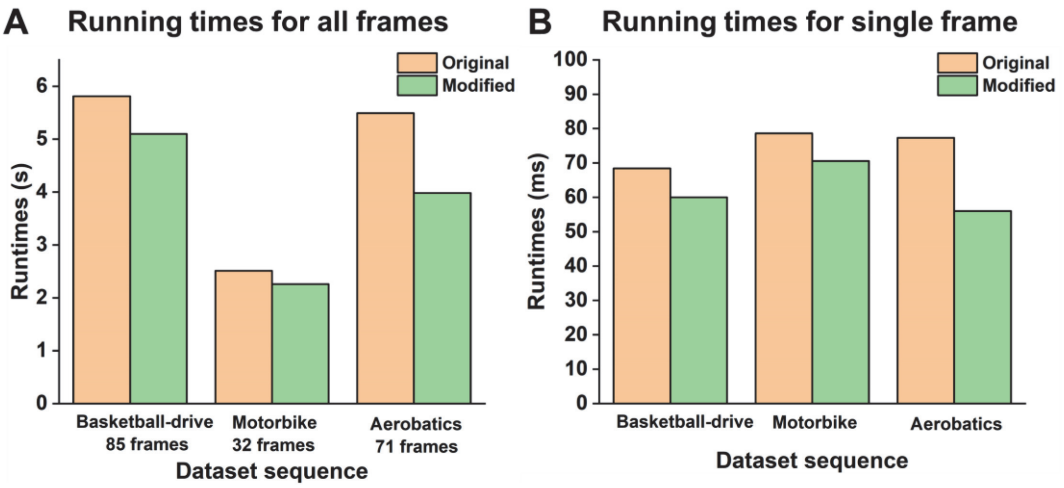

Fig 6. Comparison of running times in original and modified model. **A.** Running times for all frames **B.** for single frame of different dataset sequence.

Avg. PSNR of first and last frame is greater than original model i.e., 0.37dB and 0.19dB respectively. In general, FastDVDnet output sequences shows motion artifacts and noisy results with few frames. For example, Fig. 3 displays last frame of motorbike sequence from Set8 with corresponding zoomed view. However, if we compare subjective result of both models, our network has able to removed noise due to its changed frame sequence

architecture. Similarly, Fig. 4 shows first and last frame of aerobatics sequence. Where, highlighted part of circle results shows noticeably flickering and noisy with FastDVDnet. On the other hand, modified model output very convincing and visually satisfying results. (Best viewed in digital format). In addition to Set 8 and DAVIS we also put subjective results of HEVC sequences. Our aim to test HEVC sequences is to improve the higher resolution challenging sequences with our model. In future, we apply such denoising results for HEVC video compression like [10]. Fig. 5. Shows subjective comparison results on Basketball Drive and BQmall sequences. However, two steps cascaded architecture helped to denoise high resolution sequences. Last frame of BasketballDrive shows flickering results and noise is present in frame. Compared to this, our input 5 model accomplished better results (see highlighted part). In HEVC dataset we also tested three smaller resolution (832x480) sequences which also outperformed with our modified model (see BQmall in Fig. 5). Please note that, we showed higher noise value results because FastDVDnet works well with lower noise value.

Our modified model with simple design architecture and feeding longer frames to each denoising block efficiently achieves fast interpretation time than original model. Though, original model [19] takes 100ms to denoise a 960x540 color frame and our model takes average 60ms to denoise each 960x540 color frame, which is faster than patch based [16,17] and CNN based video denoising method [19]. However, it's not reasonable to compare such running times as testing GPU environments are different. Hence, we have tested FastDVDnet original model and our modified model on DGX server with GPU NVIDIA Tesla. Fig. 6. shows the comparison results of running times with the comparison of FastDVDnet model.

VI. CONCLUSION

In this paper, we proposed a modified model of FastDVDnet to remove noise and motion artifacts from long-term frames. In addition, we also tested our model with higher resolution test set (HEVC) sequences, and we achieved better performance. Our experimental results showed that our modified model is better than the original model with long term frames in terms of quality metrices and in digital format with fast running times. To prove this, we have estimated average PSNR of each dataset and PSNR of first and last frame of each dataset. However, we believe that the proposed model can be helpful in the prepossessing of video compression. In addition, our model could be extended to denoise noisy video sequences without increasing additional memory usage in comparison with the existing FastDVDnet model. Furthermore, it is important to study the effect of various parameters such as addition or deletion of more denoising blocks on such modified models.

AKNOWLEDGMENTS

This research was supported by Kyungpook National University Research Fund, 2020.

REFERENCES

1. Y. Chen and T. Pock, "Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256-1272, 2017.
2. U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 8, pp. 2774-2781, 2014.
3. K. Dabov, A. Foi, and V. Katkovnik, "Image denoising by sparse 3D transformation-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 1-16, 2007.
4. M. Lebrun, A. Buades, and J. M. Morel, "A Nonlocal Bayesian Image Denoising Algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1665-1688, 2013.
5. H.C. Burger, C.J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2392-2399, 2012. https://doi.org/ 10.1109/CVPR.2012.6247952.
6. V. Santhanam, V.I. Morariu, and L.S. Davis, "Generalized deep image to image regression," *in Proc. Computer Vision and Pattern Recognition*, 2016. https://doi.org/10.48550/arXiv.1612.03268.
7. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, 2017.
8. K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608-4622, 2018.
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proc. Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
10. K. A. Bhosale, S. Kuk, and S.-h. Park "Study on deep CNN as preprocessing for video compression", *Proceedings of Society of Photo-Optical Instrumentation Engineers*, Applications of Digital Image Processing XLIV, 118420V, 1 August 2021. https://doi.org/10.1117/12.2596227.
11. M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3952-3966, 2012.
12. P. Arias and J.-M. Morel, "Video denoising via empirical Bayesian estimation of space-time patches," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 1, pp. 70-93, 2018.
13. X. Chen, L. Song, and X. Yang, "Deep rnns for video denoising,", *Proceedings of the SPIE*, Vol. 9971, id. 99711T, pp. 10, 2016. https://doi.org/10.1117/12.2239260.
14. R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *Proceedings of the 30th International Conference on International Conference on Machine Learning*, pp. 1310-1318, 2013. https://doi.org/10.48550/arXiv.1211.5063.
15. M. Tassano, J. Delon, and T. Veit. "DVDnet: A fast network for deep video denoising", *IEEE International Conference on Image Processing*, Sep 2019. https://doi.org/10.48550/arXiv.1906.11890.
16. T. Ehret, A. Davy, J.-M. Morel, G. Facciolo, and P. Arias. "Model-blind video denoising *via* frame-to-frame training", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11369-11378, 2019. https://doi.org/10.48550/arXiv.1811.12766.
17. M. Claus and J. van Gemert. "Videnn: Deep blind video denoising", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1843-1852, 2019. https://doi.org/10.1109/CVPRW.2019.00235.
18. A. Davy, T. Ehret, G. Facciolo, J.-M. Morel, and P. Arias. "Non-local video denoising by CNN", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2409-2413, 2019. https://doi.org/10.1109/ICIP.2019.8803314.
19. M. Tassano, J. Delon and T. Veit. "FastDVDnet: towards real-time deep video denoising without flow estimation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1354-1363, 2020. https://doi.org/10.1109/CVPR42600.2020.00143.
20. O. Ronneberger, P. Fischer, and T. Brox. UNet: "Convolutional Networks for Biomedical Image Segmentation", vol. 9351, *Lecture Notes in Computer Science*, 1362 chapter 28, pp. 234-241. *Springer International Publishing*, 2015.
21. S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang. "Deep high dynamic range imaging with large foreground motions", *In European Conference on Computer Vision*, vol. 11206, pp. 117-132, 2018. https://doi.org/10.1007/978-3-030-01216-8_8.
22. P. Fischer, A. Dosovitskiy, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. "Flownet: Learning optical flow with convolutional networks", *IEEE International Conference on Computer Vision*, pp. 2758-2766, Dec 2015. https://doi.org/10.1109/ICCV.2015.316.
23. M. Gharbi, G. Chaurasia, S. Paris, and F. Durand. "Deep joint demosaicking and denoising", *ACM Transactions on Graphics*, 35(6):1-12, Nov 2016.
24. K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
25. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874-1883, Jun 2016.
26. M. Tassano, J. Delon, and T. Veit. "An analysis and implementation of the FFDnet image denoising method", *Image Processing On Line*, vol, 9, pp.1-25, Jan 2019.
27. A. Paszke, G. Chanan, Z. Lin, S. Gross, E. Yang, L. Antiga, and Z. Devito. "Automatic differentiation in PyTorch", *Advances in Neural Information Processing Systems*, vol 30, pp. 1-4, 2017.
28. D.P. Kingma and J.L. Ba. "ADAM: a Method for Stochastic Optimization", *Proceedings of International Conference on Learning Representations*, pp. 1-15, 2015.
29. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the Inception Architecture for Computer

Vision", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826, Dec 2015.
30. A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht. "The marginal value of adaptive gradient methods in machine learning", *In Advances in Neural Information Processing Systems*, pp. 4148-4158, 2017.

Kavita Arjun Bhosale

Kavita Arjun Bhosale received B.S. and M.S. in computer science from University of Pune, Maharashtra, India in 2012 and 2014, respectively. From 2016 to 2019 she worked as a resource center executive-development in Shipco IT Pvt Ltd-Pune, India and worked on different technologies like Java, Asp.net, Object Oriented PERL Scripting, Web services, JavaScript. Her research interest includes Video processing and deep learning, Ultra-light weight neural network for versatile video coding (VVC) and computer vision.

Sang-hyo Park

Sang-hyo Park received the B.S. and Ph.D. degrees in computer engineering and computer science from Hanyang University, Seoul, South Korea, in 2011 and 2017, respectively. From 2017 to 2018, he held a postdoctoral position with the Intelligent Image Processing Center, Korea Electronics Technology Institute, and a Research Fellow with the Barun ICT Research Center, Yonsei University in 2018. From 2019 to 2020, he held a postdoctoral position with the Department of Electronic and Electrical Engineering, Ewha Womans University. Since 2020, he has been an Assistant Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. His research interests include HEVC, VVC, encoding complexity, immersive video, and deep learning.