A Study on Motion-robust Video Deblurring

Jianfeng Xu[1,a)] Kazuyuki Tasaka[1,b)]

Abstract: Most existing video deblurring works focus on the use of temporal redundancy and lack utilization of the prior information about data itself, resulting in strong dataset dependency and limited generalization ability in handling challenging scenarios, like blur in low contrast or severe motion areas, and non-uniform blur. Therefore, we propose a PRiOr-enlightened MOTION-robust video deblurring model (PROMOTION) suitable for both global and local blurry scenarios. On the one hand, we use 3D group convolution to efficiently encode heterogeneous prior information (including illumination, structure, and motion priors), which enhances the model's blur perception while mitigating the output's artifacts. On the other hand, we design the priors representing blur distribution, which enables our model to better handle non-uniform blur in spatio-temporal domain. In addition to the classical camera shake caused blurry scenes, we also prove the generalization of the model for local blur in real scenario, resulting in better accuracy of hand pose estimation.

Keywords: Video deblurring, heterogeneous prior information, motion-robust

1. Introduction

In recent years, more and more deblurring works have emerged. In terms of image deblurring, a major trend is to use multi-scale structure, such as [1], [12], [21], or use pyramid structure, such as [11], for the purpose of enabling the network to have varying receptive fields, so as to effectively deal with different degrees of blur. For more complex video deblurring, in addition to conventionally aligning multiple frames and deblurring the center frame like EDVR [22] and DeBlurNet [20], it is often focused on how to make full use of inter-frame redundancy, such as using recurrent structure [7], [13], 3D convolution [4], [24], or encoding each frame separately and then aggregating them to decode [14]. Therefore, research on the utilization of multi-scale receptive fields and inter-frame information can be considered relatively mature. Different from the research perspective of these works, we study the importance of prior information to video deblurring.

Deblurring challenging blurs, including blur in low-illumination and severe motion (e.g. close to cameras) areas, non-uniform blur (e.g. local blur), and other special cases, has been not well solved by conventional video deblurring yet is a very important problem. This inspires us to dig out the information of scenes themselves when designing priors.

As the typical representative of conventional methods, EDVR mainly consists of predeblur, alignment, fusion, and reconstruction modules, as the golden background area shown in Figure 4. However, by analyzing the design and deblurring results, we have some interesting observations as below: 1) Fail to effectively and deeply utilize temporal correlation to help deblur. The aligned feature is obtained by just concating. This lacks the mining of motion information. On the other hand, 2D convolution also has

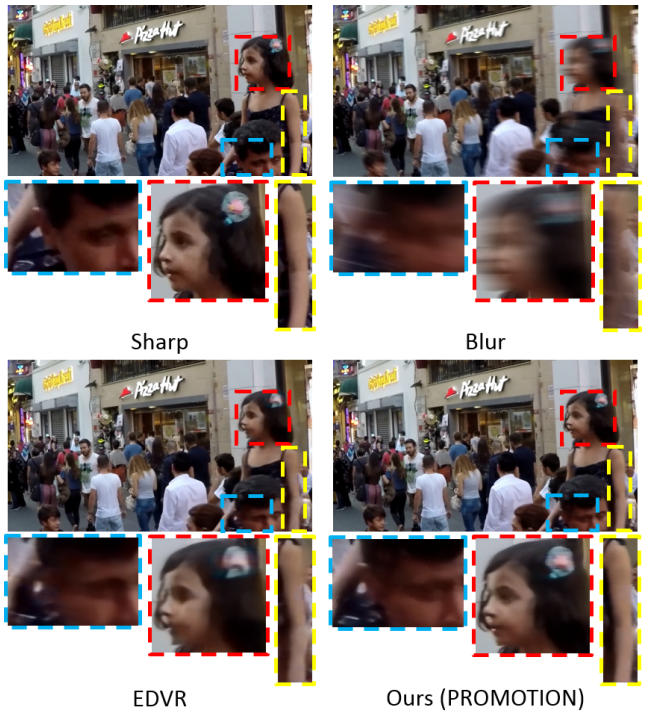

Fig. 1 Performance illustration of our method. Compared with EDVR [22] which has limited generalization ability of handling challenging blurs and tends to introduce structural distortion, due to the utilization of priors, our method can deliever more visual pleasing results, especially with more detail recovery and better fidelity in structure.

[1] KDDI Research, Inc., Ohara 2–1–15, Fujimino, Saitma 356–8502, Japan
[a)] ji-xu@kddi-research.jp
[b)] ka-tasaka@kddi-research.jp

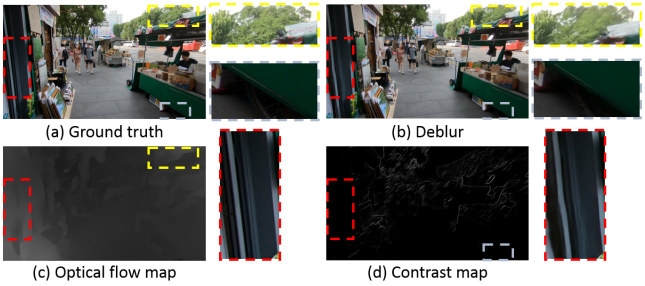

Fig. 2 Illustration of the weaknesses of the typical conventional method EDVR [22], including structure distortion, weak deblurring performance on large optical flow or low contrast areas. Since it lacks the effective constraint of heterogeneous prior information.

Fig. 3 Illustration that EDVR seriouly worsens the sharp input since it cannot discriminate the sharp information.

limited capability in modelling long-term temporal dependency. 2) The deblurred frames are prone to be infected with structure distortion, such as straight lines becoming twisty, as the red dashed boxes shown in Figure 2 (a) and (b). This is because the model does not have the constraint of structure priors. 3) The larger the optical flow, the more difficult it is to eliminate blur. We observe the red and yellow blocks presented in Figure 2 (a)(b)(c). It is not difficult to find these areas with large motion have a poor recovery of detail, such as the leaves. This is because the model lacks the prior information of motion, which makes it unable to adjust the intensity of deblurring adaptively. 4) The lower the contrast, the harder it is to remove blur. Similarly as the red and blue boxes shown in Figure 2 (a)(b)(d), the low contrast areas on the door edge and the tire are not well restored, since the model lacks the ability to perceive the illumination distribution of scenes. 5) The original sharp information cannot be preserved. For example in Figure 3, if we input the sharp frame mixed with blur frames, the output is seriously worse than the input. We inference the model actually lacks of the ability to distinguish sharp and blurred inputs (non-uniform blur). 6) The loss design, that is, Charbonnier loss [2] which calculates the error of each pixel indiscriminately, cannot also effectively avoid artifacts and reflect the influence from non-uniform blur.

Based on the above analysis, we enhance the model's perception about scenes from the use of prior information, the way of feature extraction, and the design of constraint. And further make a series of optimizations accordingly. Our contributions are as follows:

1) We utilize 3D group convolution to encode heterogeneous prior information to explicitly supplement the model's multidimensional perception about scenes. The heterogeneous priors include structure prior, motion prior, and illumination prior, which effectively constrain the artifacts of the model and enhance the detail recovery of challenging blurs. This point corresponds to observation 1) to 4).

2) In the spatio-temporal dimension, we increase the model's ability to distinguish sharp and blurred inputs. In temporal dimension, the embedding prior vector is embedded in aligned feature to represent the blur degree of each frame. In spatio dimension, optical flow based attention information is used to indicate the blur degree of different regions within a single frame. This point corresponds to observation 5).

3) We design the dual loss function, which considers the effects from pixel level and perceptual level simultaneously, to effectively constrain the blur removal and guarantee subjective quality. This point corresponds to observation 6).

4) A small-scale video deblurring dataset for hand pose estimation is constructed. And we further verified that the estimation accuracy is improved after deblurring. This indirectly reflects the effectiveness of our method for local blurry scenes.

The use of priori information brings a gift to video deblurring task, especially in terms of subjective effect, as shown in Figure 1 where our results have more detail restoration and structure fidelity.

2. Related work

Since the first end-to-end data-driven video deblurring method DeBlurNet was propopsed [20], in the past two years, the success of deep learning has brought significant promotion to video deblurring [3], [4], [7], [13], [19], [20], [22], [23], [24], [27]. Different from image deblurring which only focuses on the mining of spatial information, such as using multi-scale receptive field [1], [12], [21], video deblurring also focuses on how to effectively use redundant information in the temporal domain to assist deblurring.

The conventional approach is to stack multiple frames together as a single input and then uses 2D convolutions to extract features [4], [20], [22]. Compared to the classic DeBlurNet [20], [4] has more consideration about the complexity and params of the model, which also wins 2nd place in the NTIRE19 challenge. Before going through 2D convolutions, EDVR aligns adjacent frames to better aggregate information [22].

Some works use recurrent neural networks (RNNs) for sequential data processing [7], [13], [23]. [7] designs a spatio-temporal recurrent network which extends the receptive field while keeping the network small, and uses dynamic temporal blending to enforce temporal consistency. [13] presents a RNN-based video deblurring method that exploits both the intra-frame and inter-frame recurrent schemes and updates the hidden state multiple times internally during a single time-step. Another spatially variant RNN for dynamic scene deblurring is proposed in [23],

where the weights of the RNN are learned by a deep CNN.

Another part studies how to extract pixel-wise information to handle the spatially variant blur [3], [19], [27]. Similar to CycleGAN's circular idea [28], [3] first deblurs each frame separately, then estimates optical flow and pixel-wise blur kernels to reblur the estimated sharp images, which makes the network fine-tuned via self-supervised learning. Similarly, [27] proposes a spatial-temporal network for video deblurring based on filter adaptive convolutional layers, and the network is able to dynamically generate element-wise alignment and deblurring filters in order. [19] presents a motion deblurring kernel learning network that predicts the per-pixel deblur kernel and a residual image with two novel base blocks named residual down-up and residual up-down blocks.

In addition, 3D convolution has been also used for video deblurring recently to extract spatio-temporal information simultaneously [24]. [24] applies 3D convolutions to capture jointly spatial and temporal information, and uses a discriminator for adversarial training.

3. Proposed method

The overall diagram of our proposed PRiOr-enlightened MOTION-robust video deblurring (PROMOTION) method is presented in Figure 4. Given 5 consecutive input frames $I_{[t-2:t+2]}$ where t is the center frame's number in the sequence, we denote the center frame as I_t and the other frames as neighboring frames. The aim of video deblurring is to restore a sharp center frame $\hat{O}_t$ which is close to its corresponding ground truth O_t. On the one hand, we explicitly calculate the heterogeneous prior information of the input frame stack and encode it with 3D convolution to obtain a prior feature map, which is used to supplement the features output from the temporal and spatial attention (TSA) fusion module. This is described in Sec. 3.1. On the other hand, we design the embedding prior vector to rectify the aligned feature and introduce the optical flow based attention information in the loss function, to increase the model's perception ability of uneven blur distribution in spatio-temporal dimension. These are described in Sec. 3.2 and Sec. 3.4 separately. Finally, in order to make the model more flexible to learn different patterns, channel attention technology is introduced into the basic residual blocks, which is described in Sec. 3.3.

3.1 Heterogeneous prior information

As discussed earlier in the introduction, we use the illumination, structure and motion priors of scenes to improve the model's structural fidelity and detail recovery abilities for low contrast and large optical flow regions.

Contrast group. For each frame I_i in the input stack, we calculate its contrast map G_i^c as follows:

$$G_i^{c'}(p,q) = \frac{1}{4} \sum_{(\hat{p},\hat{q}) \in N_4(p,q)} (G_i(p,q) - G_i(\hat{p},\hat{q}))^2 \quad (1)$$

Table 1 Parameters of the heterogeneous prior encoding network. Note stride is the same in row, column and depth directions.

Layer	In_channel	Out_channel	Kernel Size	Stride	Group
3D Conv1	3	9	3*5*5	1	3
MaxPool	9	9	2*2	2	-
3D Conv2	9	27	3*5*5	1	9
MaxPool	27	27	2*2	2	-
2D Conv	27	128	1*1	1	-

$$G_i^c = \frac{G_i^{c'}}{max(G_i^{c'})} \quad (2)$$

where G_i is the gray map of I_i, and $N_4(p,q)$ denotes the 4-neighborhood of pixel (p,q). From the contrast map in Figure 5, it can be seen that there are larger activation values for the high illumination regions and smaller activation values for the low ones. This indicates that contrast maps are able to effectively reflect the illumination distribution of scenes.

Gradient group. Similarly, we calculate the gradient map G_i^g for each input frame as the equation (3) given, and use it to increase the sensitivity of the model to structure information.

$$G_i^g(p,q) = [G_i(p,q) - G_i(p+1,q)] + [G_i(p,q) - G_i(p,q+1)] \quad (3)$$

As the pink box shown in the upper right corner of Figure 4, gradient map can highlight structural information in a scene, such as regular lines on the wall. This allows the model to have better structural fidelity.

Optical flow group. Instead of using all the optical flow maps of input stack, we only estimate the optical flow map of center frame as baseline by using [8], while for the neighbor frames, we use the difference maps between center frame and each neighbor frame to represent the relative motion information. This is to save the calculation time of optical flow.

Then 3D group convolutions are used to encode the spatio-temporal information more efficiently. The detail parameters of the heterogeneous prior encoding network are presented in Table 1.

3.2 Embedding prior vector

In order to enable the model to place corresponding emphasis on frames with different degrees of blur in temporal dimension, embedding prior vector is designed to explicitly tell the model which frames it should pay more attention to.

Specifically, we first filter each frame with the Laplacian operator of size $3 * 3$, as shown in Figure 6. It is not difficult to find the blurry image on the left has a smooth filtered map, while the clear one on the right has a sharper filtered map. Then, we calculate the variance of these filtered maps $var(Lap(G_i))$ to reflect the blur degree of each frame. The more blurry the frame, its variance is smaller. Therefore, the embedding prior vector can be represented by:

$$V_{emb} = \frac{5 \cdot [var(Lap(G_i))^{-1}]}{\sum_{i=t-2}^{t+2} var(Lap(G_i))^{-1}}, \quad i \in [t-2, t+2] \quad (4)$$

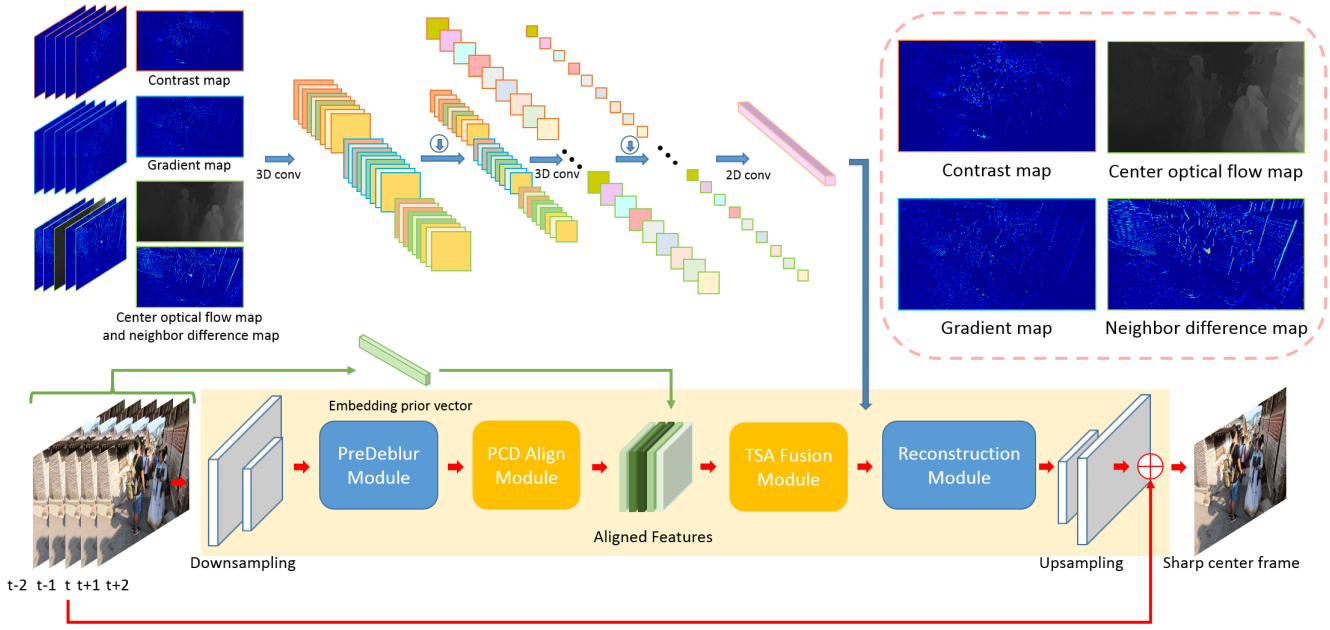

Fig. 4 The diagram of our proposed PRiOr-enlightened MOTION-robust video deblurring (PROMOTION) method. The modules in the golden background area are consistent with those in EDVR [22]. Recommend reading in color version.

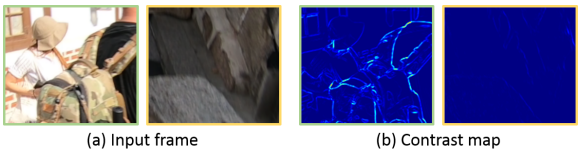

Fig. 5 An example that contrast map effectively reflects the illumination distribution. The green and yellow boxes represent high and low illumination areas, respectively.

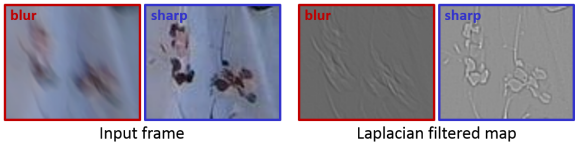

Fig. 6 The Laplacian filtered maps of blur and sharp example frames.

Finally this vector is multiplied by the aligned feature in depth and we obtain the refined aligned feature.

3.3 Channel attention enhanced residual block

Inspired by the success of Residual Channel Attention Networks (RCAN) in super resolution [26], we introduce the channel attention mechanism in the residual basic block, which only consists of two convolution layers and one ReLU layer originally. As shown in Figure 7, on the one hand, the dimensions of channel are first compressed and then restored, therefore the effective information can be amplified. On the other hand, by adaptively learning the importance of each channel, the model can express various patterns more flexibly.

3.4 Dual loss function

In order to increase the naturalness of deblurred videos, we constrain the training from two complementary levels, namely pixel level and perceptual level.

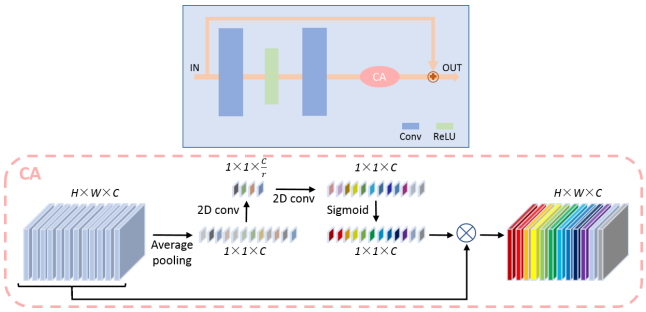

Fig. 7 Illustration of channel attention used in the residual basic block. "CA" denotes channel attention.

At the pixel level, we believe that the blurred areas are often the areas suffering from motion or change, while these are exactly the focus of people when watch videos. For example, in the hand pose estimation scenario, compared with the still areas in the background, people usually pay more attention to the movement of hands. As a result, the optical flow information of the current frame is used as a representation of attention to increase the model's sense ability for non-uniform blur in spatio domain, which complements the embedding prior vector handling in temporal domain. Therefore, the pixel-wise loss can be represented as below:

$$L_{cb} = \frac{1}{HW} \sum \sqrt{((\hat{O}_t - O_t) \otimes (1 + w_{att}))^2 + \epsilon} \quad (5)$$

where $\otimes$ denotes element-wise multiplication and ϵ is 10^{-6}. w_{att} is the normalized optical flow map. H and W are the height and width of one frame.

At the perceptual level, in order to describe the subjective differences between deblurred frames and ground truth, we use neural network-based perceptual similarity to represent such high-level distance [25] as below:

$$L_{ps} = f_{ps}(\hat{O}_t, O_t) \quad (6)$$

Table 2 Performance comparison under the second dataset division of REDS dataset.

Metric	DeblurGAN	Nah's	SRN	DeBlurNet	EDVR	Ours
PSNR	24.09	26.16	26.98	26.55	34.80	35.10
SSIM	0.7482	0.8249	0.8141	0.8066	0.9487	0.9565

where f_{ps} is the perceptual similarity network, whose output is a score ranging from 0 to 1.

Therefore, the overall loss function is:

$$L = L_{cb} + \lambda L_{ps} \tag{7}$$

Here λ is a balance factor that adjusts the relative importance of pixel-level and perceptual-level losses. We empirically set it as 0.1. In this way, the model can not only handle non-uniform blur, but also guarantee the subjective quality of deblurred videos.

4. Experiments

In this section, we first verify the motion robustness of our model in the global and local blurry scenarios. For global blur, such as camera shake, two well-known video deblurred datasets for natural scenes are used for evaluation in Sec. 4.1. For local blur, in Sec. 4.2, we consider the specific application scenario, hand pose estimation, and construct a small-scale video deblurring dataset for hand pose estimation. In this case, the estimation accuracy is also used as one of the measurement of deblurring effect.

4.1 Global blur

REDS dataset. This dataset used in the NTIRE19 challenge includes 270 videos available online right now. Following the same setting as EDVR, which wins the championship in the NTIRE19 challenge, 266 training videos and 4 specific videos for testing [22]. Each video has 100 frames. And their resolution is 720*1280 and frame rate is 24 fps.

As presented in Table 2, compared with the original EDVR method and other four kinds of deblurring methods, namely DeblurGAN [10], Nah's [12], SRN [21], and DeBlurNet [20], our model continues getting the best performance both in terms of PSNR and SSIM.

GoPro dataset. To further illustrate the robustness of our model, we also test it on another video deblurring dataset for natural scenes named GoPro [12]. This dataset has blur sources similar to the REDS dataset, namely camera shake and object motion. 22 training sequences and 11 testing sequences are included in it. Each sequence has unequal lengths, but the resolution is the same one 720*1280. It should be noted that the dataset provides blurry and sharp image pairs, and blurry images include both gamma corrected and linear CRF versions.

For the evaluation of EDVR and our model on GoPro dataset, we finetune the models pretrained on REDS dataset. As Table 4 shown, where all the models are tested on the linear CRF version, we compare with both video deblurring [19], [22], [23] and image debluring methods [11], [12], [21]. Our model outperforms the state-of-the-art methods by a large margin, whatever in terms of video or image deblurring. Compared with on REDS dataset, we attribute this to both the effectiveness of prior information and the lower image quality of GoPro dataset. First, using the prior information can alleviate the model's dependence on datasets and assist the model in obtaining information about the current data distribution more directly and comprehensively. Second, according to the shielding effect, we can inference that on the basis of low quality, the improvement is obvious, while on the basis of high quality, the difference is more unconspicuous.

Table 3 Performance comparison on GoPro dataset.

Metric	Nah's	DeblurGAN-v2	SRN	Zhang's	Sim's	EDVR	Ours
PSNR	28.62	29.55	30.26	29.19	31.34	30.20	33.25
SSIM	0.9094	0.9340	0.9342	0.9306	0.9474	0.9109	0.9481
Note	Image deblurring			Video deblurring			

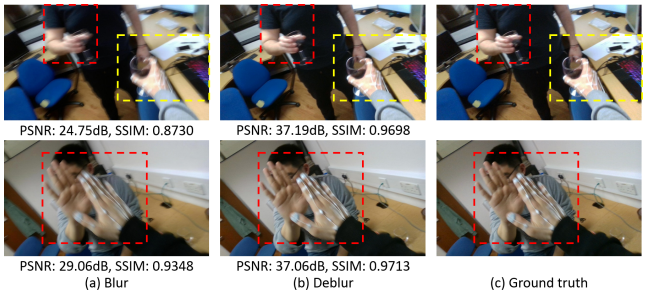

Fig. 8 Deblurring performance on the synthesized video deblurring dataset for hand pose estimation. Zoom in for best view.

Table 4 Performance comparison on GoPro dataset.

Metric	Nah's	DeblurGAN-v2	SRN	Zhang's	Sim's	EDVR	Ours
PSNR	28.62	29.55	30.26	29.19	31.34	30.20	33.25
SSIM	0.9094	0.9340	0.9342	0.9306	0.9474	0.9109	0.9481
Note	Image deblurring			Video deblurring			

Table 5 The performance improvement for hand pose estimation after deblurring.

	PSNR	SSIM	RMSE	ABSE
gain	6.0171	0.0287	0.3110	0.2653

4.2 Local blur

In order to prove the generalization of the model, we consider the specific local blurry scene, hand pose estimation. First, a video deblurring dataset for hand pose estimation is synthesized. Then we test the improvement of estimation accuracy after deblurring to indirectly reflect the deblurring effect.

Based on the existing hand pose estimation dataset named First-Person Hand Action Benchmark [5], we synthesize blur for these sequences. This dataset provides RGB-D frames and their corresponding hand pose labels. All the sequences have a frame rate of 30 fps and resolution of 1080*1920. Then we follow the synthesis process of the REDS dataset [14]. The synthesized blur effect is shown in Figure 8 (a) and (c).

We finetune the model pretrained on REDS dataset, and use the state-of-the-art hand pose estimation method [6] to measure the accuracy. Here we calculate the RMSE and

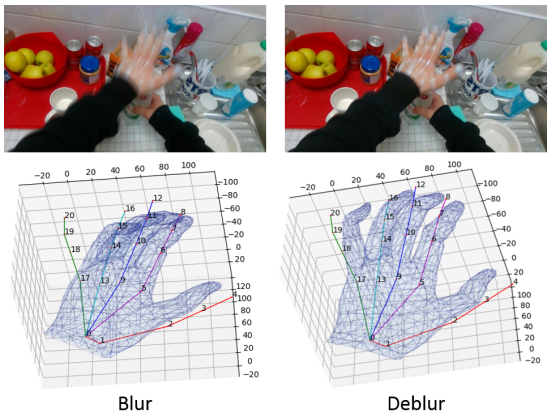

Fig. 9 Performance of hand pose estimation before and after deblurring. The improvement of RMSE and ABSE of joint location estimatioin after deblurring are 11.9015 and 9.4662, respectively.

absolute error (ABSE) between the locations of the estimated joints and ground truth to represent the accuracy. As expected, in the Table 5, all the indicators improve after deblurring. Visualized results are given in Figure 8 (b) and Figure 9. In Figure 9, skeleton represents ground truth, and mesh represents the estimated result. It is not difficult to see that for the deblurred frame, skeleton and mesh have a better overlap.

5. Conclusion

To better handle challenging blurs, we first introduce prior information in video deblurring, and propose a PRiOr-enlightened MOTION-robust video deblurring (PROMOTION) model. Specifically, 3D group convolutions are used to better encode heterogeneous priors, including illuminance, structure, and motion priors, which are proven to be related with deblurring. Then, embeding prior vector and optical-flow based attention prior are used to increase the model's ability to recognize spatio-temporal non-uniform blur. Experimental results on two globally blurred datasets show our method can achieve the state-of-the-art performance. In addition, for specific applications suffering local blur, such as hand pose estimation, we also demonstrate our method can bring performance gains to the task by preprocessing the source data.

References

[1] Y. Bai, H. Jia, M. Jiang, X. Liu, X. Xie, and W. Gao. Single image blind deblurring using multi-scale latent structure prior. IEEE Transactions on Circuits and Systems for Video Technology, pages 1–1, 2019.
[2] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In ICIP, volume 2, pages 168–172 vol.2, Nov 1994.
[3] H. Chen, J. Gu, O. Gallo, M. Liu, A. Veeraraghavan, and J. Kautz. Reblur2deblur: Deblurring videos via self-supervised learning. In ICCP, pages 1–9, May 2018.
[4] Yuchen Fan, Jiahui Yu, Ding Liu, and Thomas S. Huang. An empirical investigation of efficient spatio-temporal modeling in video restoration. In CVPR Workshops, June 2019.
[5] Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In CVPR, June 2018.
[6] Yana Hasson, Gul Varol, Dimitrios Tzionas, Igor Kalevatykh, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning joint reconstruction of hands and manipulated objects. In CVPR, June 2019.
[7] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In ICCV, Oct 2017.
[8] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In CVPR, July 2017.
[9] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1):221–231, Jan 2013.
[10] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Ji Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In CVPR, June 2018.
[11] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better, 2019.
[12] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In CVPR, July 2017.
[13] Seungjun Nah, Sanghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video deblurring. In CVPR, June 2019.
[14] Seungjun Nah, Radu Timofte, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring: Methods and results. In CVPR Workshops, June 2019.
[15] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In ICCV, Oct 2017.
[16] Markus Oberweger and Vincent Lepetit. Deepprior++: Improving fast and accurate 3d hand pose estimation. In ICCV Workshops, Oct 2017.
[17] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring text images via l0-regularized intensity and gradient prior. In CVPR, June 2014.
[18] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In CVPR, June 2016.
[19] Hyeonjun Sim and Munchurl Kim. A deep motion deblurring network based on per-pixel adaptive kernels with residual down-up and up-down modules. In CVPR Workshops, June 2019.
[20] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In CVPR, July 2017.
[21] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In CVPR, June 2018.
[22] Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In CVPR Workshops, June 2019.
[23] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson W.H. Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In CVPR, June 2018.
[24] K. Zhang, W. Luo, Y. Zhong, L. Ma, W. Liu, and H. Li. Adversarial spatio-temporal learning for video deblurring. IEEE Transactions on Image Processing, 28(1):291–301, Jan 2019.
[25] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In CVPR, June 2018.
[26] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In ECCV, September 2018.
[27] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. CoRR, abs/1904.12257, 2019.
[28] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In ICCV, Oct 2017.