Paper Reading Report-01

Shreya Chawla
u7195872

Abstract

*This is my reading report for the paper titled: "***Deep Video Deblurring for Hand-held Cameras***", authored by Shuochen Su (University of British Columbia) et al, and published in IEEE CVPR 2017.*

All ENGN8501 submissions will be subject to ANU's TurnitIn plagiarism check, against both the original paper, internet resources, as well as all other students' submissions. So please make sure that you must write your own reports, and declare the following originality statement:

I, Shreya Chawla, hereby confirm that I am the sole author of this report and that I have compiled it in my own words.

1. Problem Statement

This research paper [1] attempts at deblurring of motion blur from camera shake problem in videos using paired blurry and sharp images. This a common phenomenon where even after video stabilization, the motion blur remains resulting in "jumping" artifacts. Deep learning was applied to dynamic scene deblurring by constructing datasets with high-speed cameras.

Dongwei Ren et. al. [3] proposed an all inclusive method of dealing with all types of blur. Jochen Gast et. al. [2] improved the results of [1] by 2.11dB, thus reaching and even surpassing the quality of complex state-of-the-art networks on standard datasets.

2. Summary of the paper's main contributions

Three versions of dataset with varying degrees of alignment are created, and are used to train the model. Upon comparison, the proposed method works well even without accurate frame-to-frame alignment making it significantly faster and robust. The main contribution is their method to solve the problem.

3. Method and Experiment

The paper introduces DeblurNet (DBN) - a CNN based encoder-decoder network trained end-to-end for multi-image video deblurring. Symmetric skip connections between corresponding layers in encoder and decoder counterparts of the network significantly are added. The dataset comprises of real-world sharp video, collected by high frame rate cameras, accumulated to synthetically approximate a longer exposure. The key idea is to increase the receptive field while being easy to train.

They experimented on differently learned configurations based on various alignment types: no-alignment, frame-wise homography alignment, and optical flow alignment.

PSNR/MSSIM measurements for quantitative comparisons and visual qualitative comparisons are made with state-of-art approaches. Their results are comparable without tuning and without the explicit need for challenging image alignment unlike their predecessors. Additionally, their methodology generalizes well to a wide range of scenarios. Also, the approach is computationally very efficient as image is deblured in a single forward pass.

4. Critical Analysis

4.1. Significance of the paper's contributions

Their work is the first end-to-end data-driven approach to video deblurring, given a short stack of neighboring video frames. Prior to this work, the main challenge was alignment of differently blurred frames. The warping based techniques were not robust in addition to high computational cost. Unlike prior techniques, which explicitly specify how to aggregate multiple frames, their approach learns to fuse. They extended their work to motion blur caused by object motion as well.

4.2. Validity of the authors' main claims

They claim high efficiency of their method as it is able to process one frame within a second using NVidia Titan X GPU contrary to previous approaches on CPUs. But their claim is weak as they did not make the comparison on same hardware device that is they made compassion of time with existing methods on a CPU while theirs' uses a GPU.

Limitations of the previous works are overcome. The qualitative and quantitative comparisons justify their claims to be at par or at times better than existing model designs.

The quantitative analysis of no-alignment, homography alignment, and optical flow alignment shows the necessity of applying different alignment varying input blurriness. Analysis of the three representative learned filters at F0 shows that DeblurNet learns to preserve color tone, extract edges, and detect warping artifacts.

4.3. Limitation and weaknesses

The dataset is collected with a high frame rate camera of 240 fps. The synthetic blur effect is the integration of multiple sharp frame instances. But in real life, most video content plays at 24, 30 or 60 frames per second. Hence the real blurred frames are different from the examples in the training set resulting in the DBN+FLOW trained model to fail when the input frames are too blurry. [1]

This can be addressed by using low and high fps camera simultaneously to capture the blurry image and ground truth respectively as in RealBlur dataset. [4]

The frames are typically degraded from multiple visual phenomenons for instance motion blur, low resolution, fast object motion, and small aperture with a wide depth of field, compression artifacts, noise, etc. which are difficult to precisely model. [5] Hence, the testing and training dataset should be extended to cover more cases for a better comparison.

4.4. Extension and future work

This paper can be extended to incorporate other types of blurring in videos. Varying training patch sizes and sequence lengths of frames can also be compared. [2] Shallower CNNs could be tested against the proposed model.

Another variation to be tested are the different fusion strategies. While late fusion occasionally helps with challenging cases where DBN+NOALIGN fails, this improvement is not consistent, which might be related to the failure of artifact or misalignment detection without the presence of reference frame. [1] The effect of late vs early fusion could also be studied.

This work can be applied for deblurring CCTV, drone, satellite footage where the objects captured as well as sensor are moving.

4.5. Is the paper stimulating or inspiring ?

The autoencoder network with skip connections accelerates the convergence and helps generate much sharper video frames, which is very interesting. The analysis presented is also very simulating. The paper is divided into several section with appropriate headings.

4.6. Conclusion and personal reflection

Despite several limitations of dataset like lighting, the results are quite comparable to current works. The model is able to generalize well to majority of experimented varying inputs.

An alternate solution could be to use GANs [5] or to test against other datasets like REDS dataset [6]. More cutting edge techniques could be combined and experimented on.

To conclude, this paper presents CNN based efficient autoencoder technique for deblurring.

References

[1] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. *Deep Video Deblurring for Hand-held Cameras*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pages 1279-1288, IEEE, 2017. 1, 2

[2] Jochen Gast, and Stefan Roth. *Deep Video Deblurring: The Devil is in the Details*. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019. 1, 2

[3] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo *Neural Blind Deconvolution Using Deep Priors*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020. 1

[4] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. *Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms*. In Proc. IEEE European Conf. on Comput. Vis. (ECCV), 2020. 2

[5] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. *Deblurring by Realistic Blurring*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020. 2

[6] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. *NTIRE 2019 challenges on video deblurring and superresolution: Dataset and study*. In CVPR Workshops, 2019. 2