A Novel Encoder-Decoder Network with Multi-domain Information Fusion for Video Deblurring

Peiqi Xie, Jinhong He, Chengyun Song and Minglong Xue⋆

Chongqing University of Technology, Chongqing, 400054, China

paige@stu.cqut.edu.cn, hejh@stu.cqut.edu.cn, scyer123@163.com, xueml@cqut.edu.cn

Introduction

background

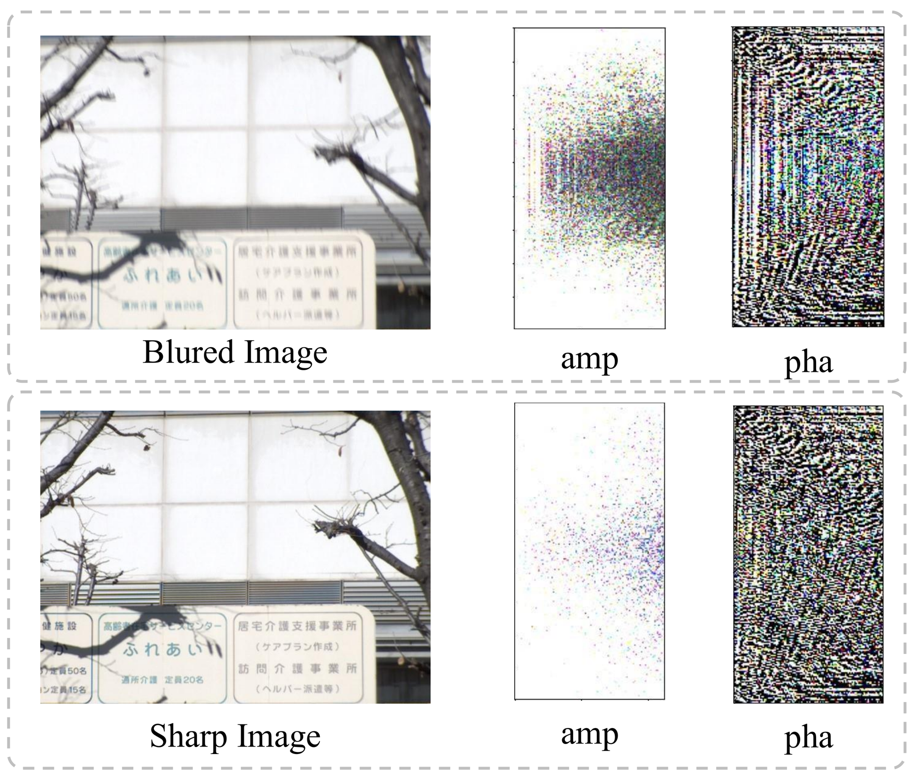
- Existing video deblurring algorithms often overlook the utilization of frequency domain information.
- We believe that our baseline does not adequately capture or utilize the global information across frames.

abstract

- we propose the Spatial-Temporal Frequency domain Fusion network (STFFNet) and improve the network from three key aspects.
- Experimental results demonstrate that the proposed method achieves state-of-the-art deblurring performance on benchmark datasets.
- The code is available at: https://github.com/Paige-Norton/STFFNet.

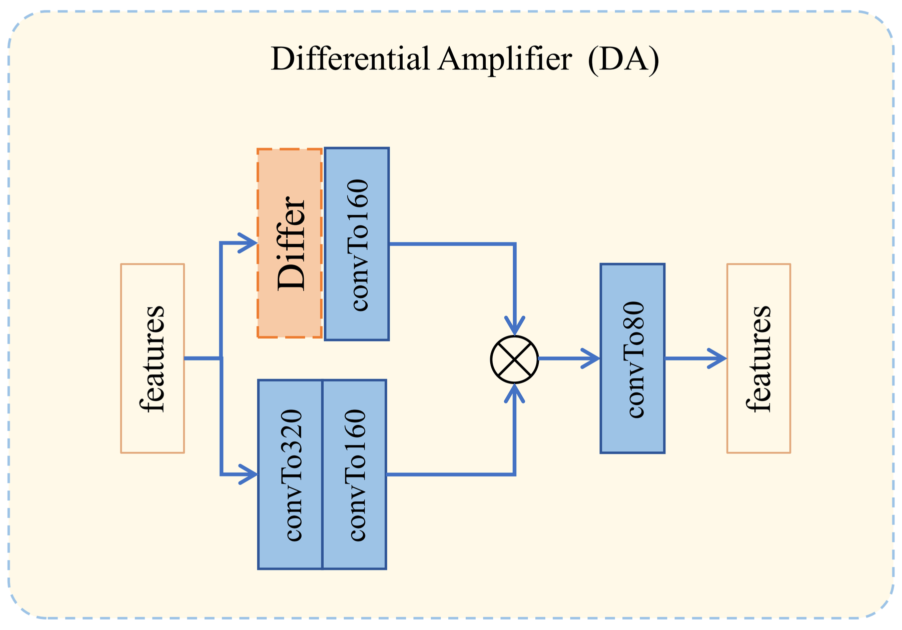
Methods & results

There is a large gap between blurred and clear pictures in the frequency domain



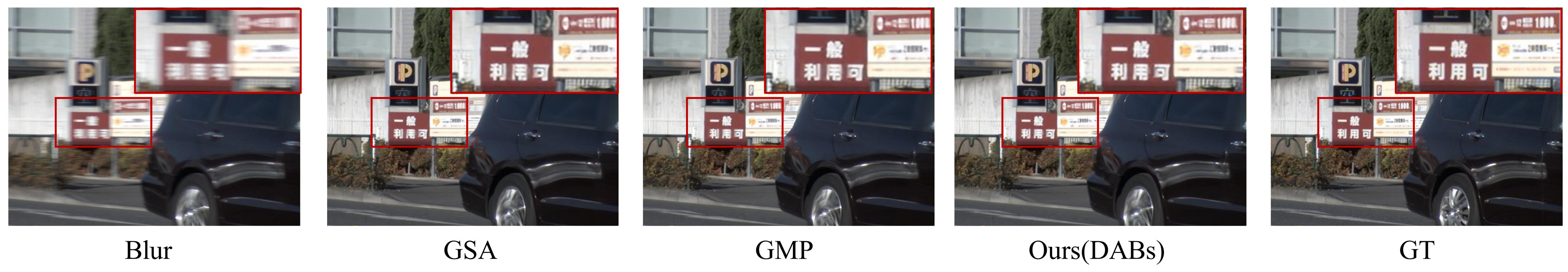
We transform the blurred image and the ground truth image to the frequency domain and observe their differences in amplitude (amp) and phase (pha) information.

DA focuses more on frame-to-frame differences

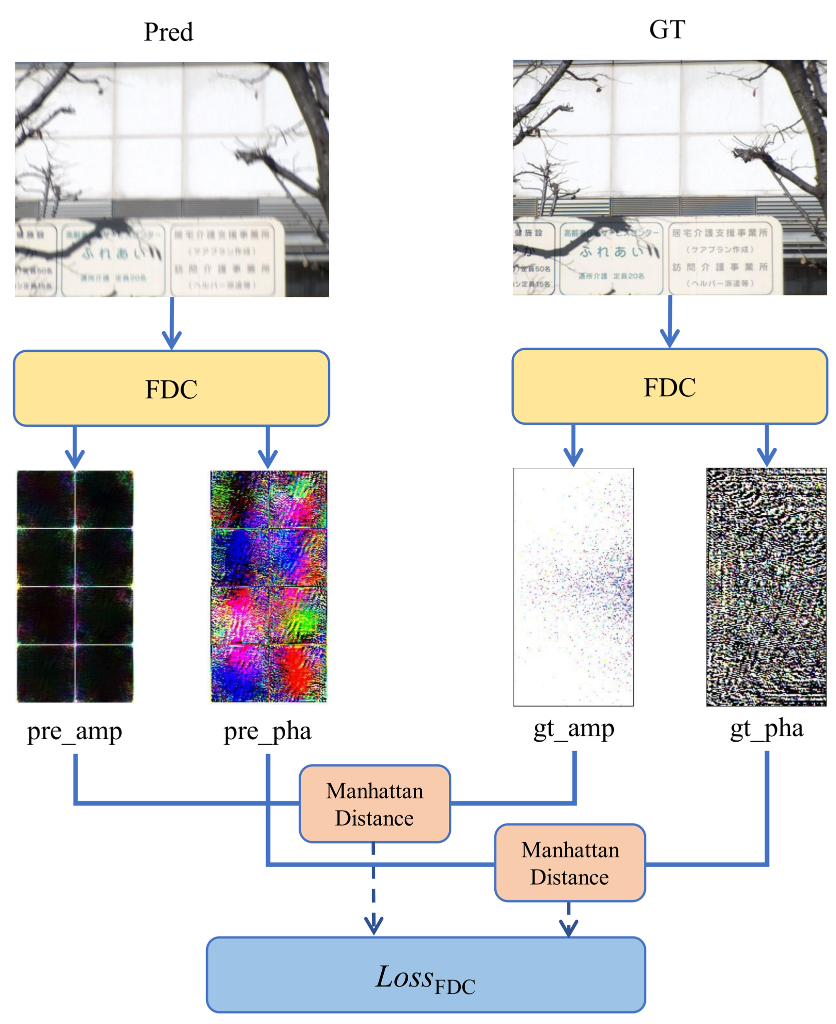


We have a branch outside of feature fusion (Diifer) to focus on feature differences.

Feature fusion module (DABs) excels in comparative methods

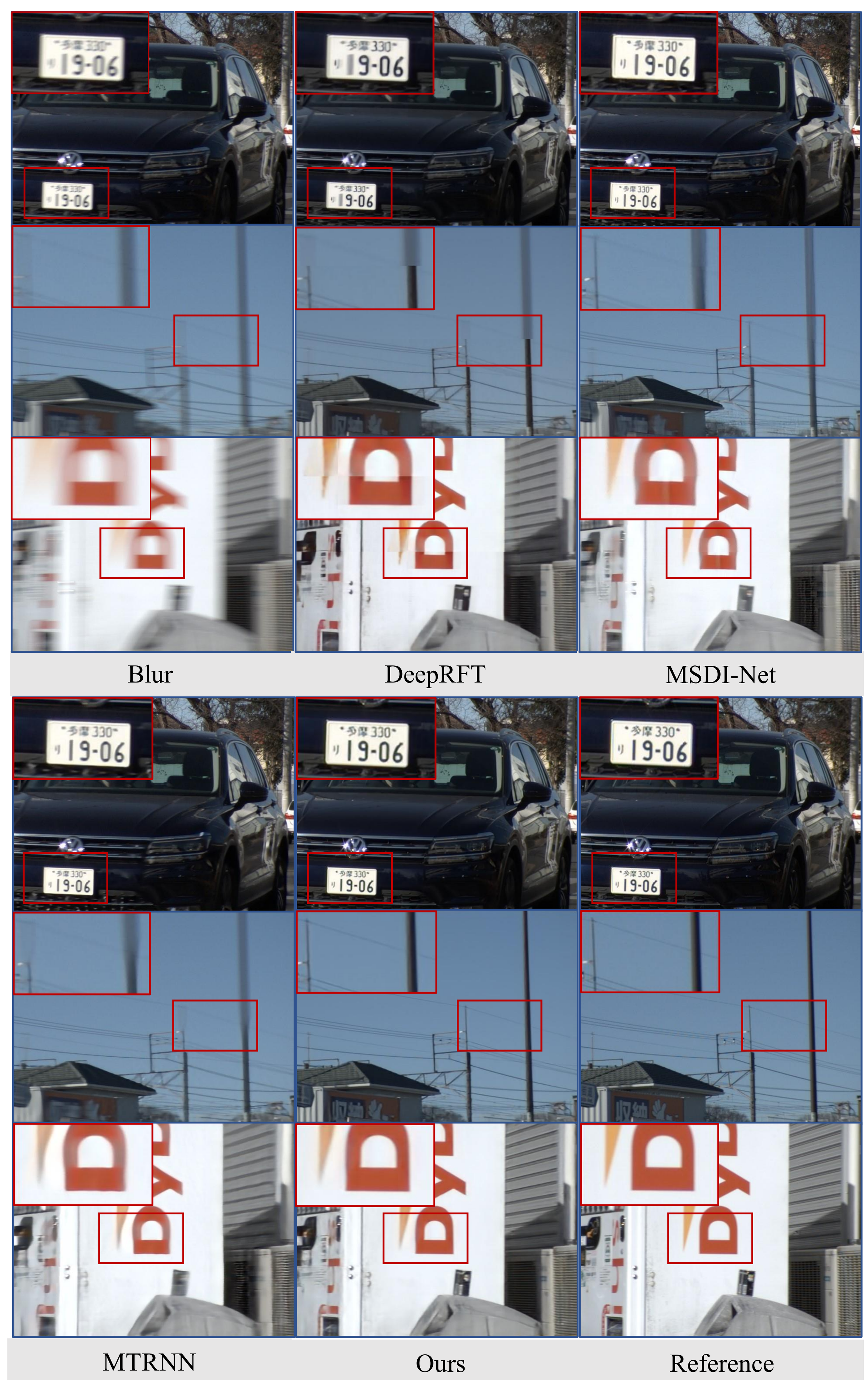


Better recovery of blurred images using frequency domain information



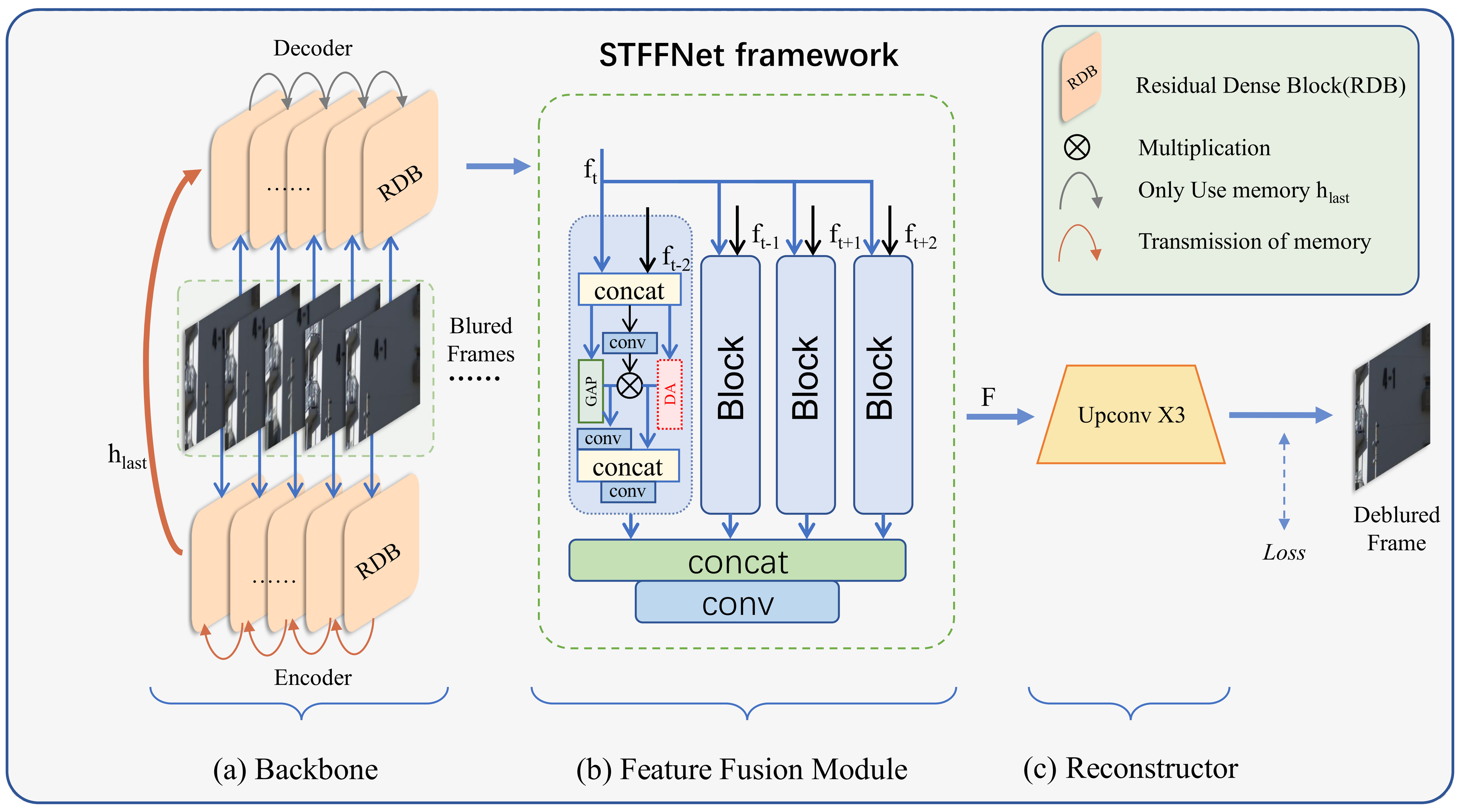
We use Frequency Domain Converters (FDC) for extracting the amplitude and phase of the image. In addition, they are used to compute the Manhattan distance between blurred and clear images in terms of amplitude and phase to guide image recovery.

Our approach excels in comparative methods



Methods	Reference	1ms8ms		2ms16ms		3ms24ms	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
STRCNN	(ICCV 2017)[11]	32.20	0.924	30.33	0.902	29.42	0.893
DBN	(CVPR 2017)[7]	33.22	0.935	31.75	0.922	31.21	0.922
IFI-RNN	(CVPR 2017)[9]	33.00	0.933	31.53	0.919	30.89	0.917
SRN	(ICCV 2018)[27]	31.84	0.917	29.95	0.891	28.92	0.882
STFAN	(ICCV 2019)[10]	32.78	0.922	32.19	0.919	29.47	0.872
MTRNN	(ECCV 2020)[15]	28.06	0.868	26.85	0.841	27.17	0.866
CDVD-TSP	(CVPR 2020)[28]	33.54	**0.942**	32.16	0.926	31.58	0.926
MSDI-Net	(ECCV 2022)[29]	28.40	0.885	27.87	0.865	28.03	0.875
DeepRFT	(AAAI 2023)[30])	29.81	0.902	29.76	0.910	28.14	0.890
ESTRNN	(IJCV 2023)[24]	33.36	0.937	31.95	0.925	31.39	0.926
Ours		**33.68**	0.938	**33.03**	**0.938**	**31.66**	**0.929**

STFFNet framework



Conclusion

This paper proposes an efficient video deblurring method (STFFNet) for perceptually oriented and metrically favourable enhancement. Specifically, we first explore using an Encoder-Decoder to construct a novel backbone. It combines global features while generating current frame features to extract more profound and broader information for better recovery. In addition, we develop a new feature fusion module to speed up the fitting and improve the modelling results. Finally, we added frequency domain information to the network to make the network more focused on high-frequency information, resulting in more explicit images. Experimental results show that STFFNet performs up to the current state-of-the-art methods in the benchmark datasets.