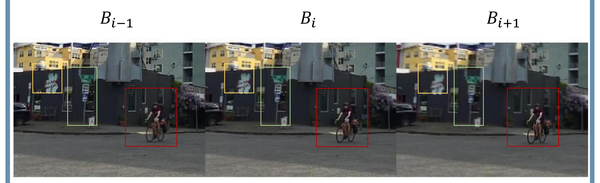
Spatio-Temporal Deformable Attention Network for Video Deblurring

Huicong Zhang, Haozhe Xie, Hongxun Yao

ECCV TEL AVIV 2022

Project Page
https://vilab.hit.edu.cn/projects/stdan

Motivation



Not all the pixels in video frames are sharp and beneficial for deblurring.

- In the i frame, the man riding a bicycle *(highlighted with a red bounding box)* is sharp. However, the same regions are blurry in the $i-1$ and $i+1$ frames.
- In contrast, the green and yellow bounding box regions are sharp in the $i-1$ frame and blurry in the other frames.

Observation

- Only the sharp pixel information of each frame is collected for restoration.

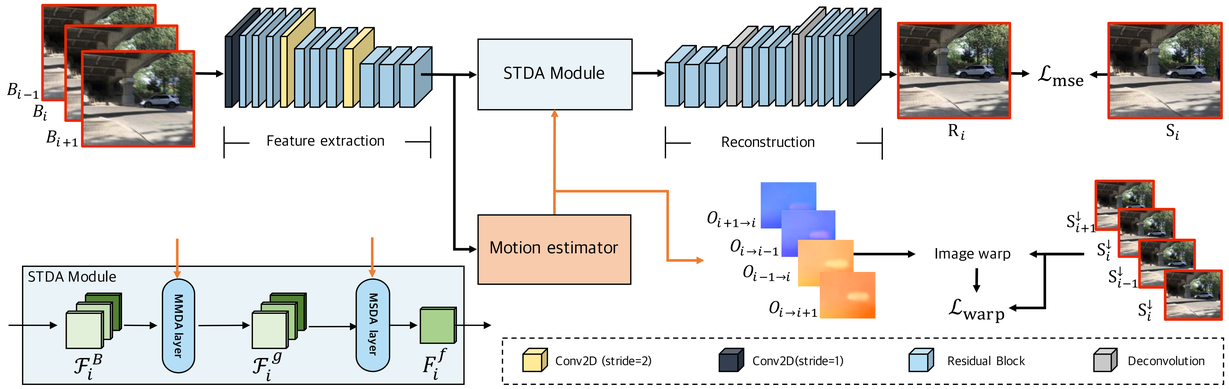
Solution

- The information of sharp pixels is aggregated by considering the pixel-wise blur levels of the video frames.

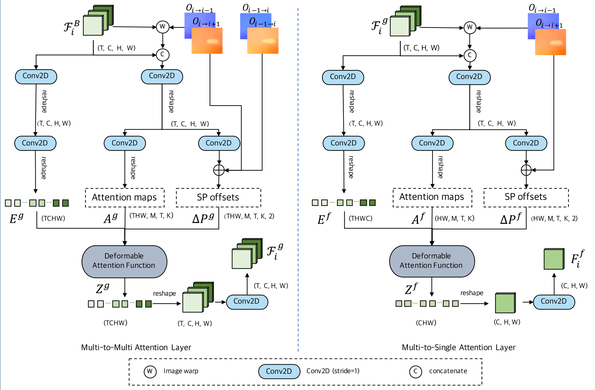
Contribution

- We propose a spatio-temporal deformable attention (STDA) module which aggregates the information of sharp pixels in the input consecutive video frames and eliminates the effects of blurry pixels introduced from input consecutive video frames.
- We present a spatio-temporal deformable attention network (STDANet) equipped with motion estimator and the proposed STDA module, where motion estimator predicts coarse optical flows and provides base offsets to find sharp pixels in adjacent frames.
- We quantitatively and qualitatively evaluate STDANet on the DVD, GoPro, and BSD datasets. The experimental results indicate that STDANet performs favorably against state-of-the-art methods with comparable computational complexity.

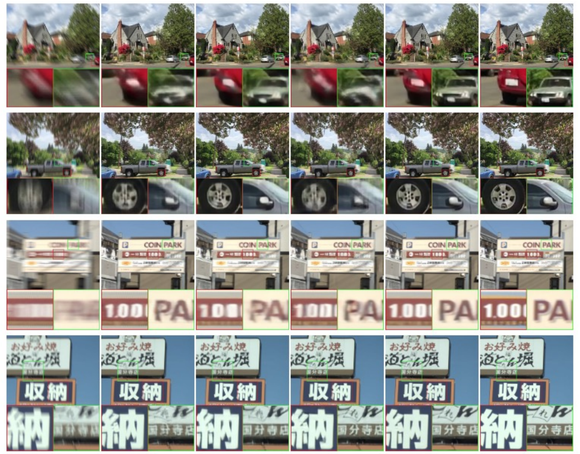
The Proposed Method: STDANet



MMA and MSA layers



Qualitative Evaluation



(a) Input (b) EDVR (c) TSP (d) PVDNet-L (e) Ours (f) GT

Quantitative Evaluation

DVD dataset

Method	PSNR	SSIM
EDVR	31.82	0.9160
TSP	32.13	0.9270
PVDNet	32.31	0.9260
ARVo	32.80	0.9352
Ours*	**33.05**	**0.9374**

GoPro dataset

Method	PSNR	SSIM
TSP	31.67	0.9279
PVDNet-L	31.98	0.9280
Ours*	**32.62**	**0.9375**

BSD dataset

Method	1ms–8ms		2ms–16ms		3ms–24ms	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ESTRNN	33.36	0.9370	31.95	0.9250	31.39	0.9260
TSP	33.62	0.9419	32.19	0.9285	31.68	0.9266
PVDNet-L	33.93	0.9392	32.46	0.9290	31.87	0.9293
Ours*	**34.32**	**0.9456**	**33.27**	**0.9420**	**32.83**	**0.9443**

The visualization of attention maps