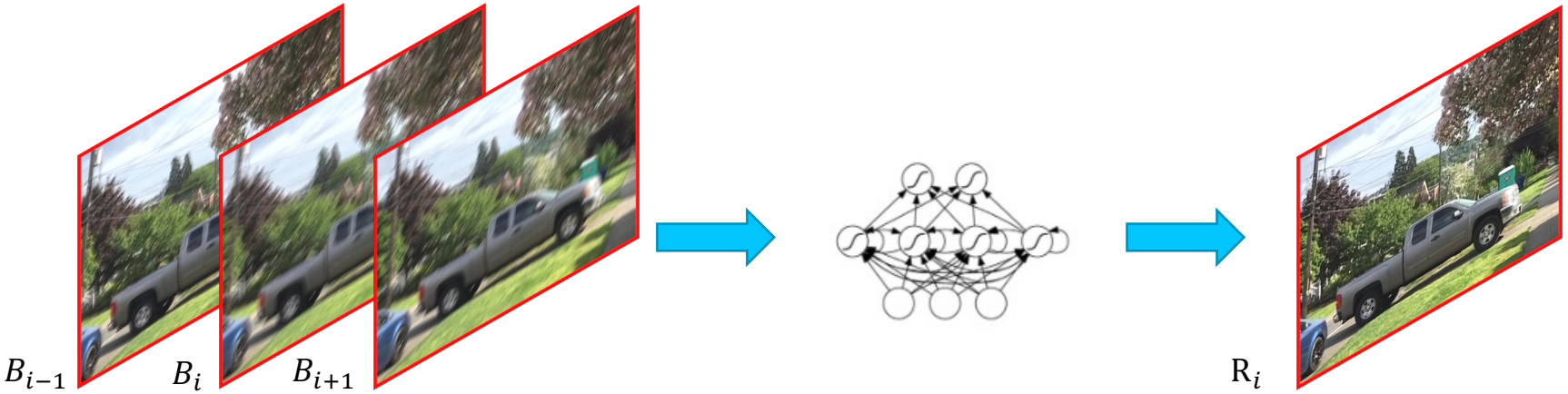
ECCV TEL AVIV 2022

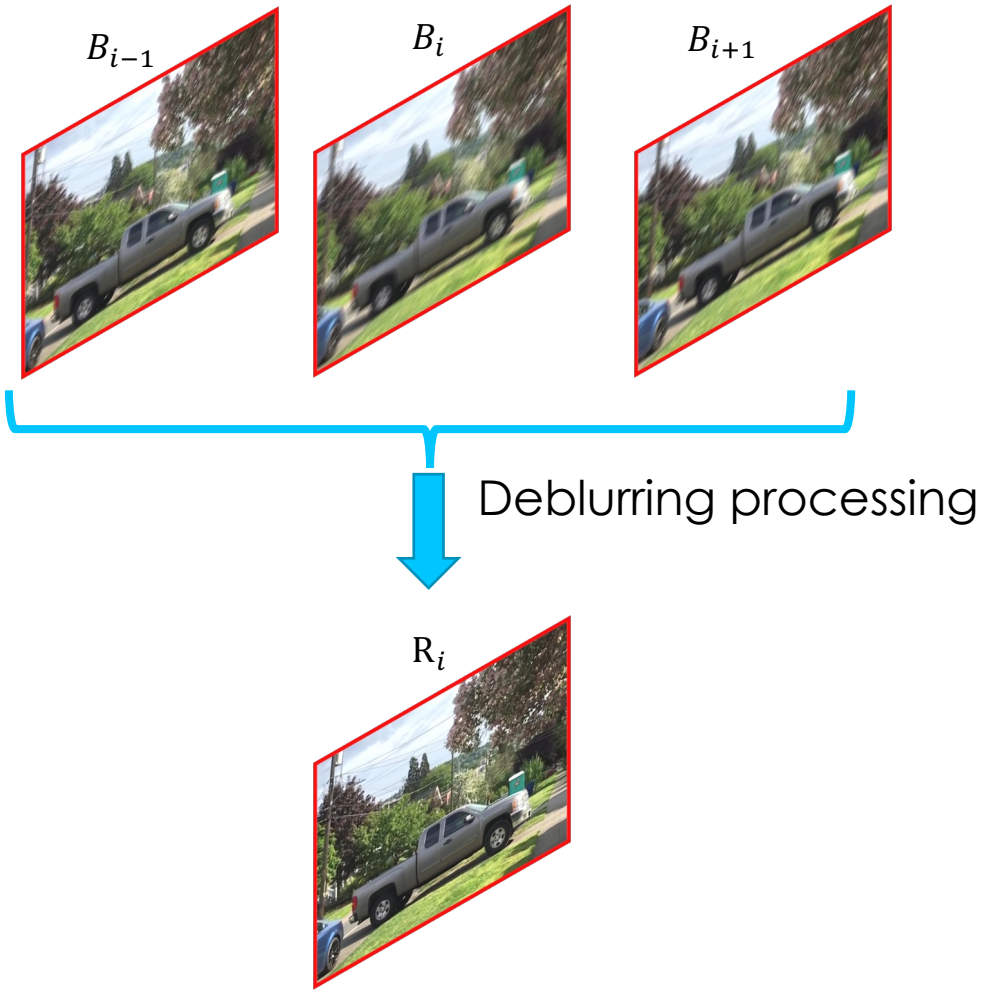
Spatio-Temporal Deformable Attention Network for Video Deblurring

Huicong Zhang, Haozhe Xie, Hongxun Yao

What is Video Deblurring?



What is Video Deblurring?

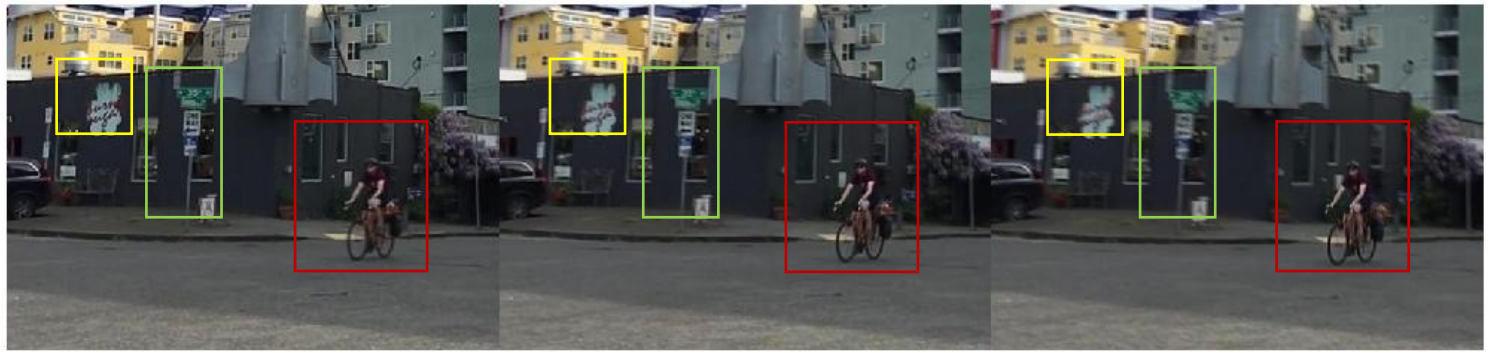


Motivation

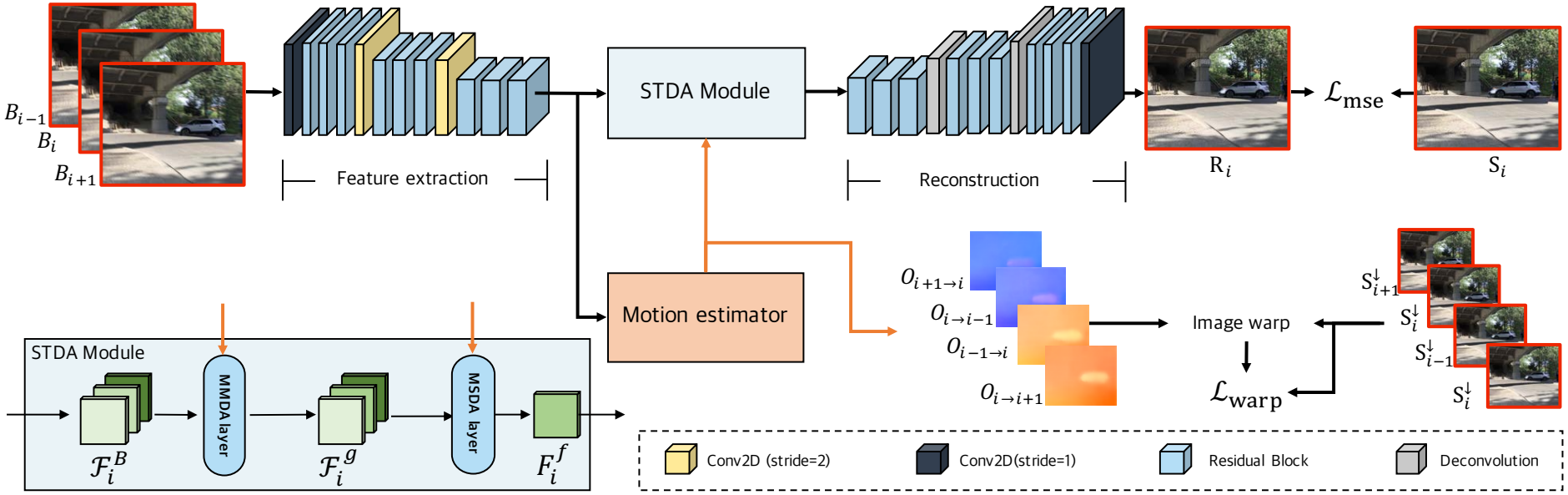
Not all the pixels in video frames are sharp and beneficial for deblurring.

- 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.

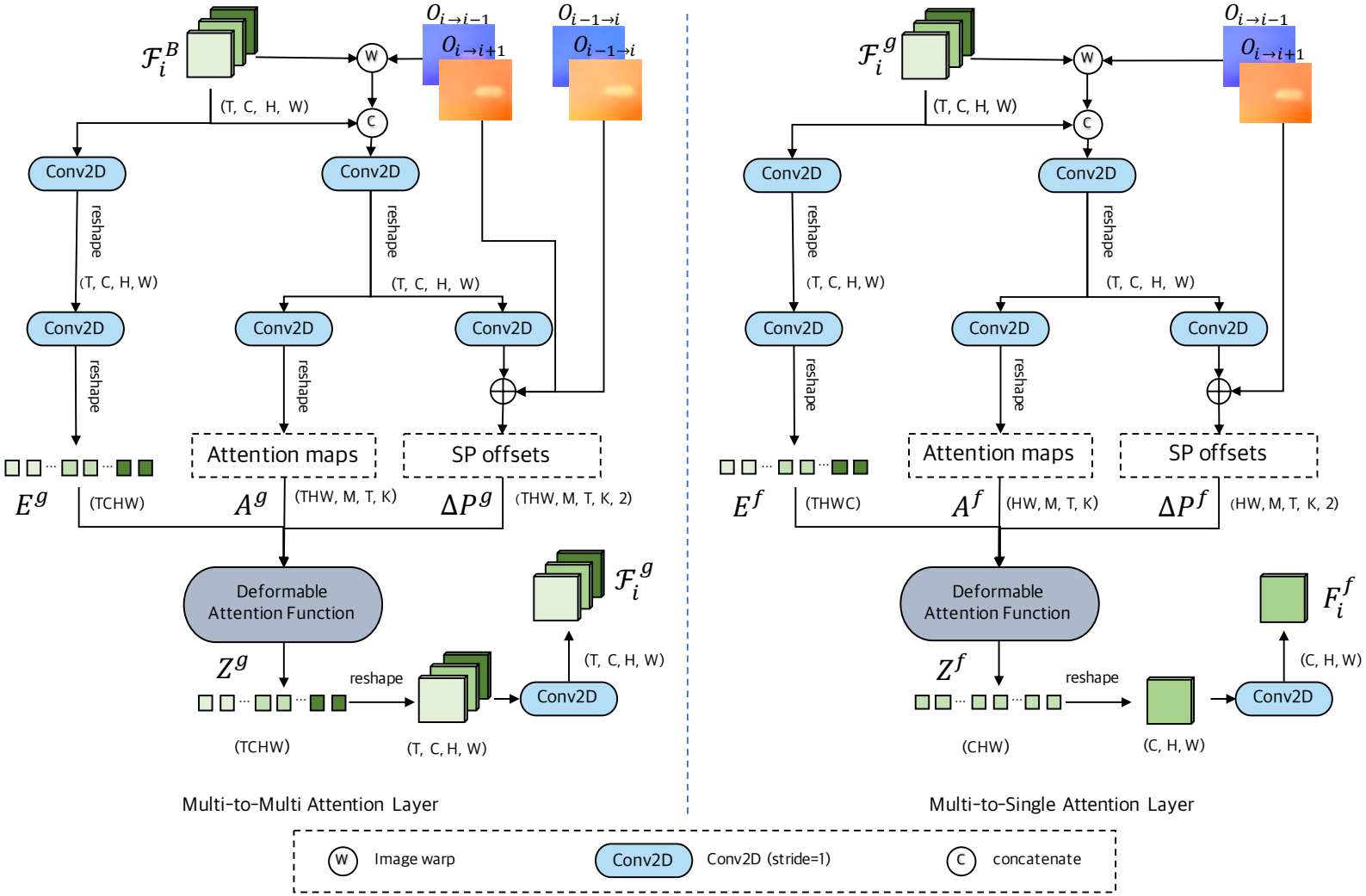
B_{i-1} B_i B_{i+1}



The Proposed Method: STDANet



STDA Module


$\mathcal{F}_i^B$
$O_{i\to i-1}$
$O_{i\to i+1}$
$O_{i-1\to i}$
$O_{i-1\to i}$
W
C
(T, C, H, W)
Conv2D
reshape
(T, C, H, W)
Conv2D
reshape
(T, C, H, W)
Conv2D
Conv2D
Conv2D
reshape
reshape
Attention maps
SP offsets
E^g
(TCHW)
A^g
(THW, M, T, K)
ΔP^g
(THW, M, T, K, 2)
Deformable Attention Function
$\mathcal{F}_i^g$
Z^g
reshape
(T, C, H, W)
Conv2D
(TCHW)
(T, C, H, W)
Multi-to-Multi Attention Layer
$\mathcal{F}_i^g$
$O_{i\to i-1}$
$O_{i\to i+1}$
W
C
(T, C, H, W)
Conv2D
reshape
(T, C, H, W)
Conv2D
reshape
(T, C, H, W)
Conv2D
Conv2D
Conv2D
reshape
reshape
Attention maps
SP offsets
E^f
(THWC)
A^f
(HW, M, T, K)
ΔP^f
(HW, M, T, K, 2)
Deformable Attention Function
F_i^f
Z^f
reshape
(C, H, W)
Conv2D
(CHW)
(C, H, W)
Multi-to-Single Attention Layer
W
Image warp
Conv2D
Conv2D (stride=1)
C
concatenate

Network Loss

MSE loss

$$\mathcal{L}_{mse} = \| \boldsymbol{R} - \boldsymbol{S} \|^2$$

Warp loss

$$\mathcal{L}_{warp} = \| \boldsymbol{S}_i^{\downarrow} - \boldsymbol{Warp}\left(\boldsymbol{S}_i^{\downarrow}, \boldsymbol{O}_{i\to j}\right) \|^2$$

Total Loss

$$\mathcal{L}_{total} = \mathcal{L}_{mse} + \boldsymbol{\gamma}\mathcal{L}_{warp}$$

Qualitative Results

Quantitative Results on DVD

Method	PSNR	SSIM
SRN	30.53	0.8940
IFI-RNN-L	31.67	0.9160
STFAN	31.24	0.9340
EDVR	31.82	0.9160
TSP	32.13	0.9270
PVDNet	32.31	0.9260
ARVo	32.80	0.9352
Ours	32.63	0.9300
Ours*	**33.05**	**0.9374**

Quantitative Results on GoPro

Method	PSNR	SSIM
SRN	30.61	0.9080
IFI-RNN-L	31.05	0.9110
STFAN	28.59	0.8608
EDVR	31.54	0.9260
TSP	31.67	0.9279
PVDNet	31.52	0.9210
PVDNet-L	31.98	0.9280
Ours	**32.29**	**0.9313**
Ours*	**32.62**	**0.9375**

Quantitative Results on BSD

Method	1ms–8ms		2ms–16ms		3ms–24ms	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
IFIRNN	33.00	0.9330	31.53	0.9190	30.89	0.9170
ESTRNN	33.36	0.9370	31.95	0.9250	31.39	0.9260
EDVR	32.79	0.9264	31.99	0.9129	31.53	0.9192
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	**34.21**	**0.9446**	**33.13**	**0.9388**	**32.65**	**0.9409**
Ours*	**34.32**	**0.9456**	**33.27**	**0.9420**	**32.83**	**0.9443**

Demo

THANK YOU

Project Page:

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