Optical Flow Estimation from a Single Motion-blurred Image

Dawit Mureja Argaw, Junsik Kim, Francois Rameau, Jae Won Cho, In So Kweon

Robotics and Computer Vision Laboratory, KAIST

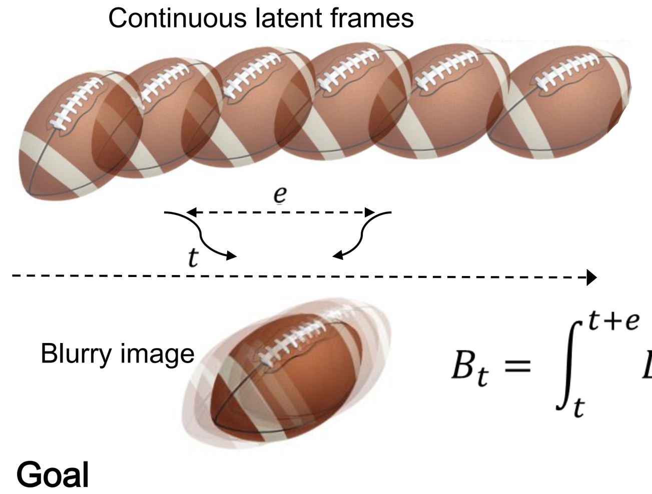
AAAI-21
Feb. 2-9

This work was supported by NAVER Labs

Introduction

Motion blur

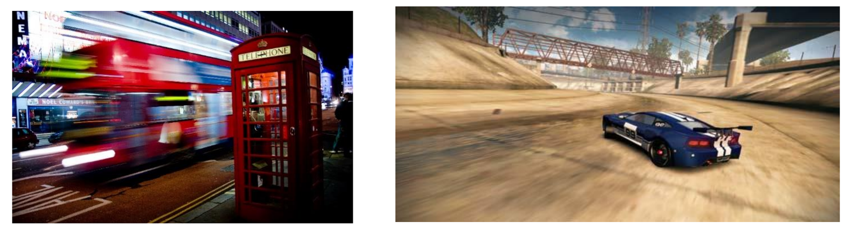
- Temporal aggregate of continuous latent frames due to sudden camera shake or dynamic motion of objects in the scene during exposure time of a camera



$$B_t = \int_t^{t+e} L(\tau)\, d\tau$$

Applications of motion blur

- In photography
  - Artistic effect
- In Games
  - Adds realism and cinematic look

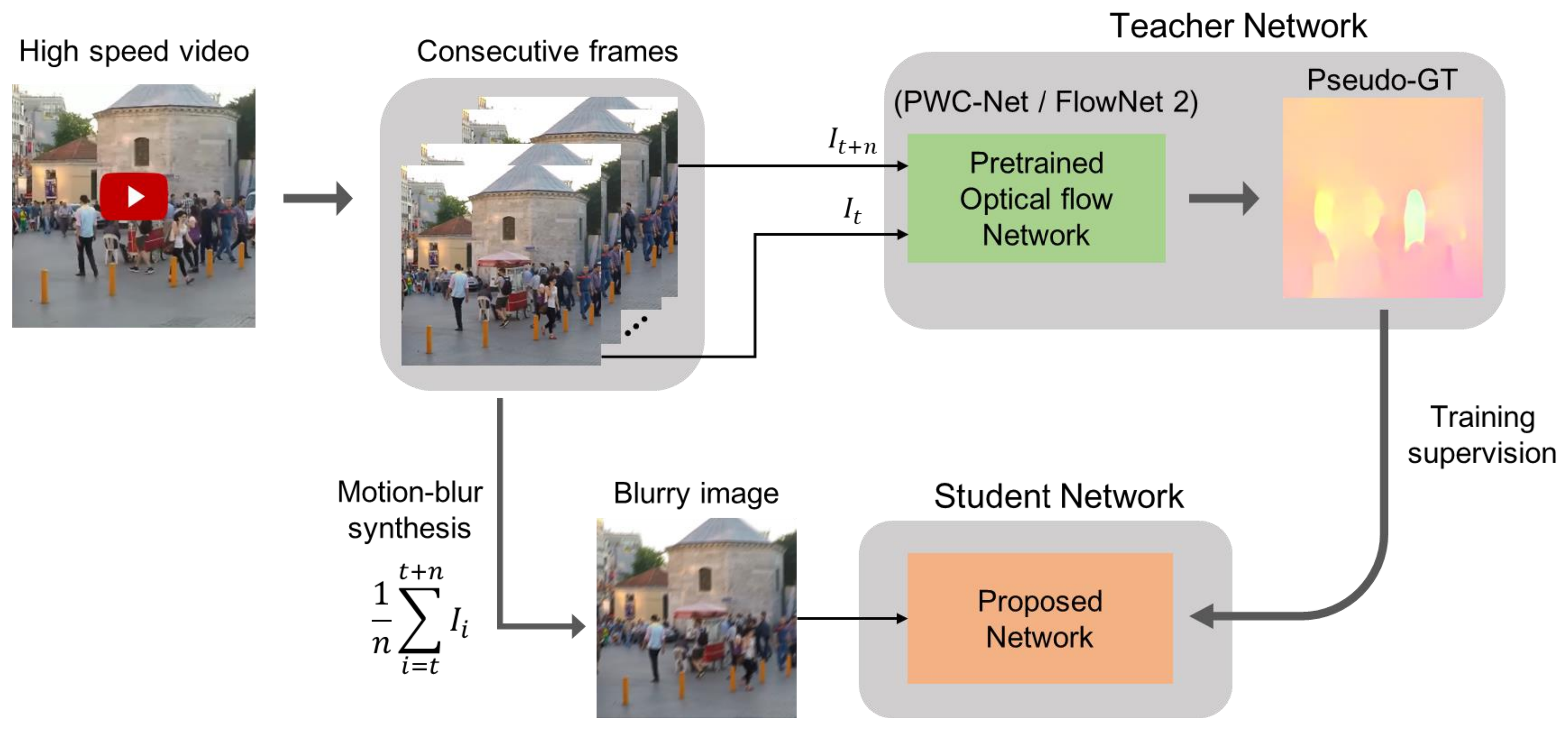


- In most computer vision research
  - Unwanted artifact
  - Image and Video deblurring

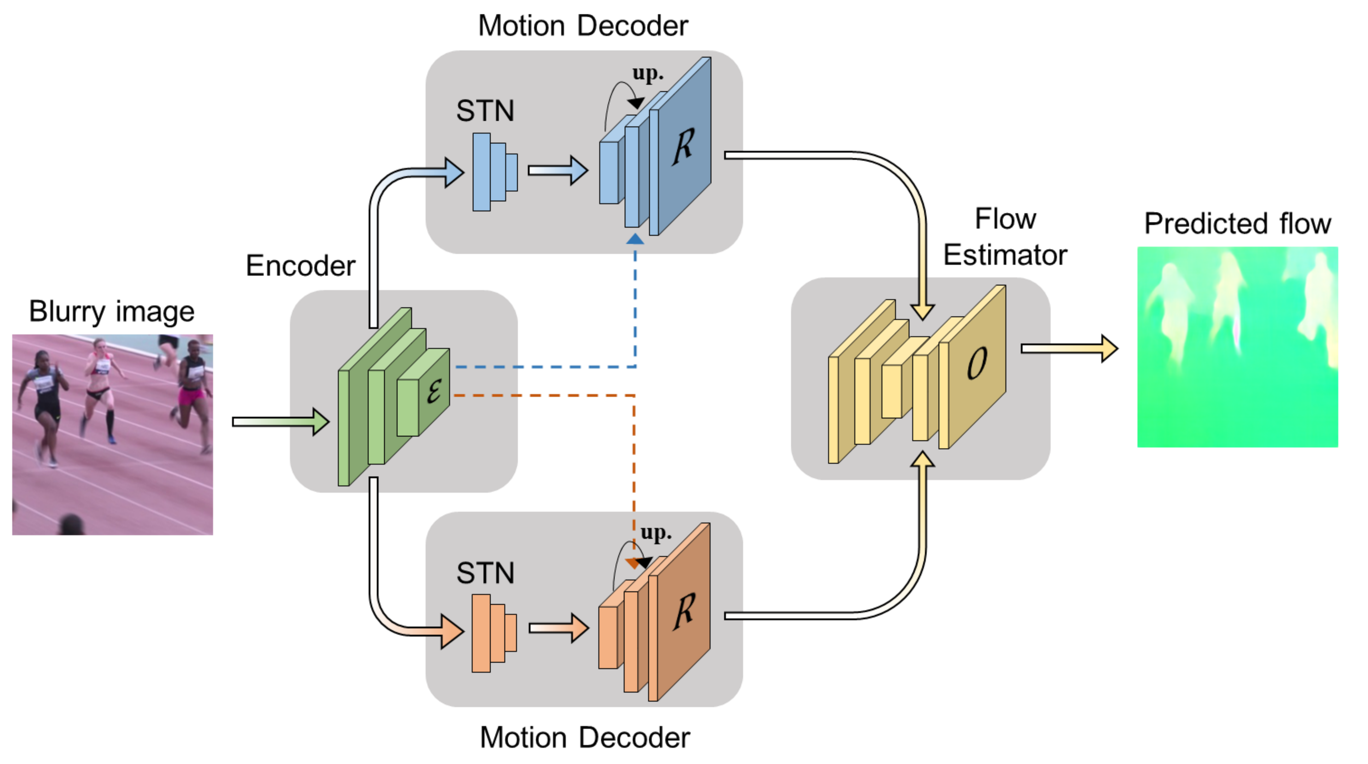
Goal

- Our work aims to explore the potential of motion blur
  - Can a deep neural network learn motion from blur? What are its challenges?
  - Can the learned motion be useful in computer vision (CV) applications?

Problem Formulation



Proposed Architecture



Methodology

Feature encoder

- Top-down feature extraction
- Feed-forward CNN network with 6 convolutional blocks

$$\{U_e^1, \ldots, U_e^k\} = \mathcal{E}(I)$$

Feature decoder

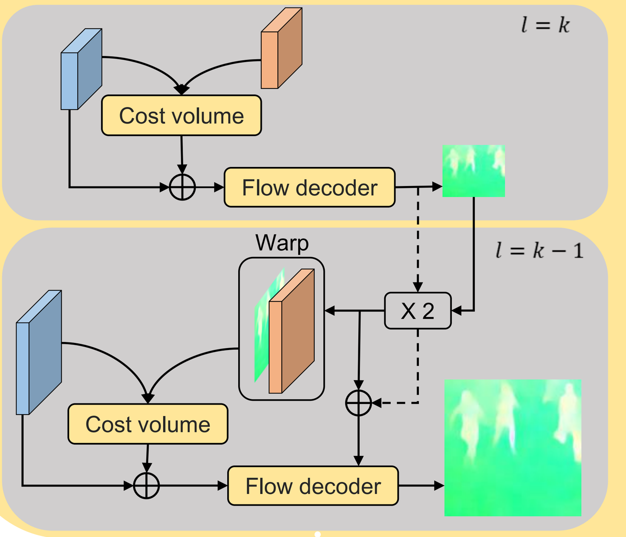
- Bottom-up feature decoding
- Decode features by learning motion from blur
- Spatial Transformer Network (STN) [11]
  - 2D Euclidean transformation

$$U_\theta^l = \mathrm{STN}_{[R|T]}^l\left(U_e^l\right)$$

  - Non-local motion
- Feature refining block
  - Compensate for locally-varying motion
  - 5 densely-connected convolutional layers

$$V^l = R^l\left(U_\theta^l \,||\, U_e^l \,||\, \mathbf{up}.\left(V^{l+1}\right)\right)$$

Optical flow estimator



- Coarse-to-fine flow prediction
- Backward warping

$$\hat{V}_1^l = \mathrm{warp}\left(V_2^l, \mathbf{up}.\left(f^{l+1}\right)\right)$$

- Cost volume

$$C = \mathrm{Correlation}\left(V_1^l, \hat{V}_1^l\right)$$

- Flow decoder
  - 5 densely-connected convolutional layers

$$V_f^l = D^l\left(C \,||\, V_1^l \,||\, \mathbf{up}.\left(V_f^{l+1}\right) || \, \mathbf{up}.\left(f^{l+1}\right)\right)$$

- Flow prediction and refinement

$$f^l = P^l\left(V_f^l\right)$$

  - Refine flow using a context network [1]

Network training

- Loss function
  - Multi-scale endpoint error

$$L = \sum_{l=1}^{l=k} w^l . \left|f^l - \hat{f}^l\right|_2$$

- Temporal ambiguity
  - Averaging does not preserve temporal order
  - Estimating the correct flow direction is an intractable problem
  - Measuring the quality of the predicted flows

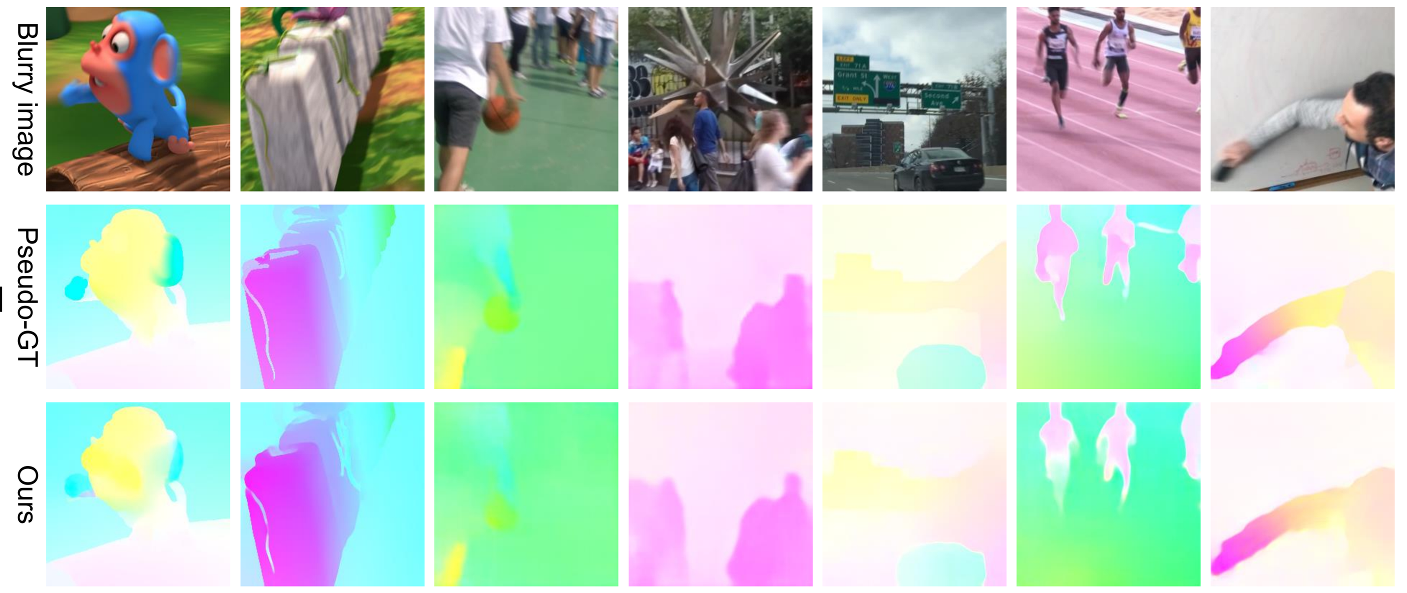
$$\text{min/max} \quad \{\mathrm{EPE}(f, \hat{f}_{t\to t+n}), \mathrm{EPE}(f, \hat{f}_{t+n\to t})\}$$

Experiments and Results

Dataset

- Synthetic scene blur dataset
  - Monkaa [6]
  - Blur synthesis via interpolation + averaging
- Real scene blur dataset
  - GOPRO [7] and Need for Speed (NfS) [8]
  - Blur synthesis by averaging 7 consecutive frames
- Other blur datasets
  - Blur Detection [9] Dataset and SONY RX V [10]

Qualitative results on motion-blurred images from different datasets



Quantitative results

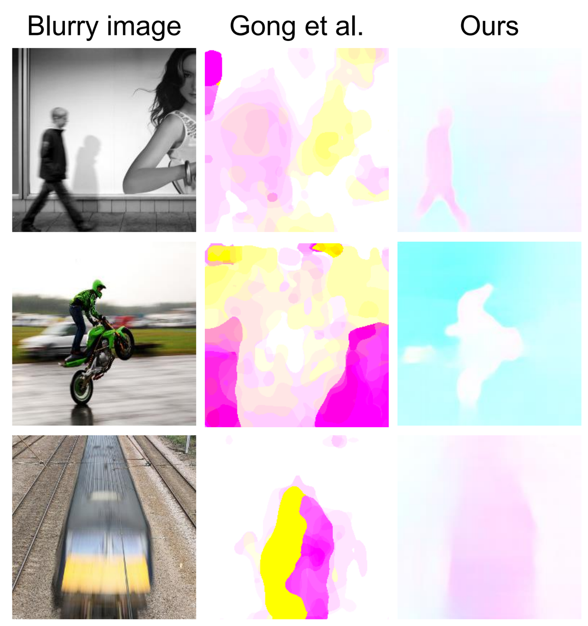
Table 1: Quantitative evaluation. For simplicity, we refer to our Blur to Flow network as B2F-net.

Method	Monkaa	GoPro	NfS
B2F-net	1.158	2.077	-
B2F-net + *fine tuning*	-	2.038	1.958

Analysis

Comparison with motion flow estimation approaches

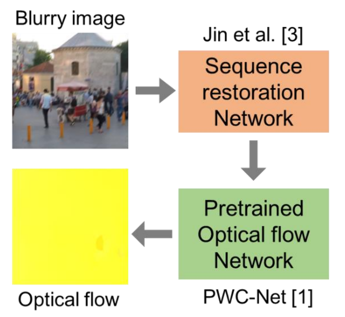
- Motion blur → Motion flow
  - For heterogeneous blur removal
- Gong et al. [2]
  - Classification task
  - Constraints on direction & magnitude
  - Simulated synthetic blurs
- Ours
  - Regression task
  - No motion constraints
  - Real high-speed video blurs



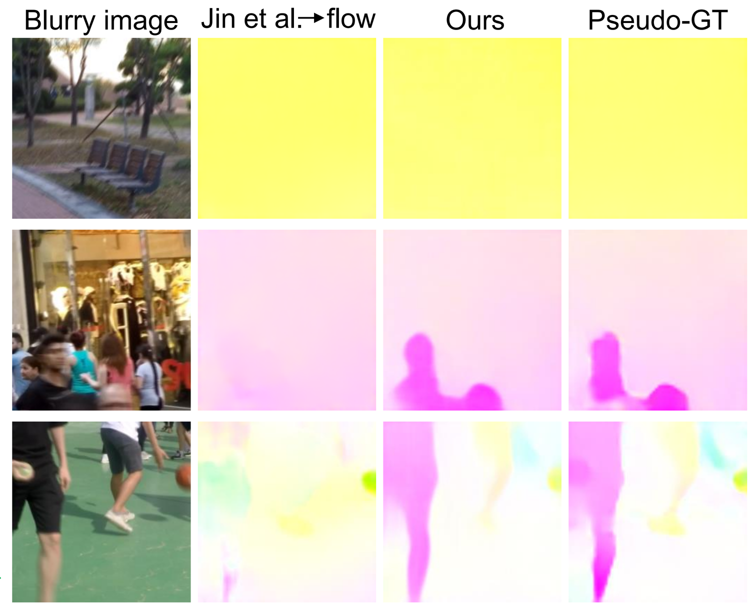
- Our approach generalizes better

Comparison with optical flow from restored frames

- Problem formulation



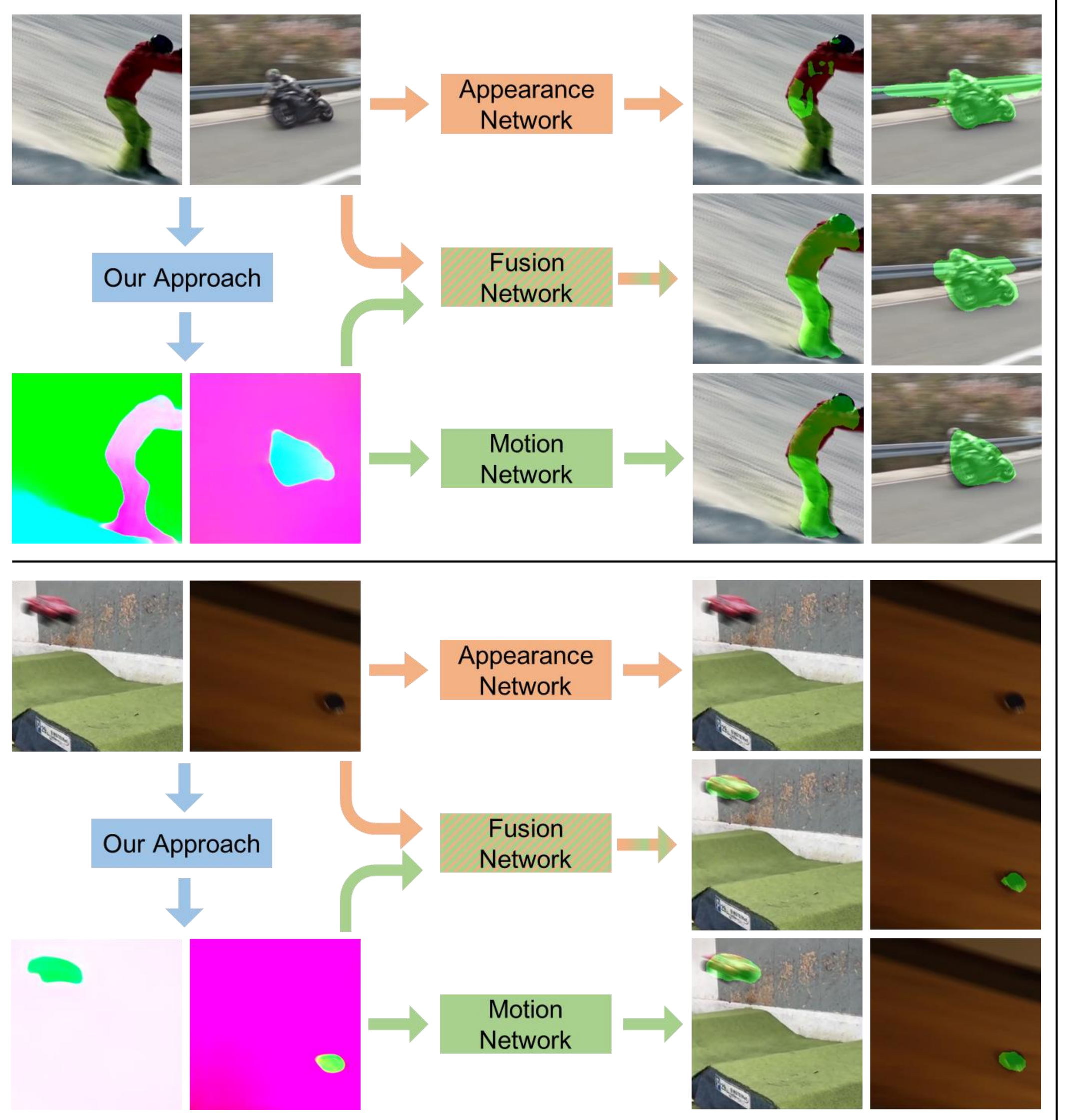
- Jin et al.[3] → flow [1]
  - Blur artifacts
  - Static scene: ✓
  - Dynamic scene: ✗
- Ours
  - Static scene: ✓
  - Dynamic scene: ✓



Downstream tasks

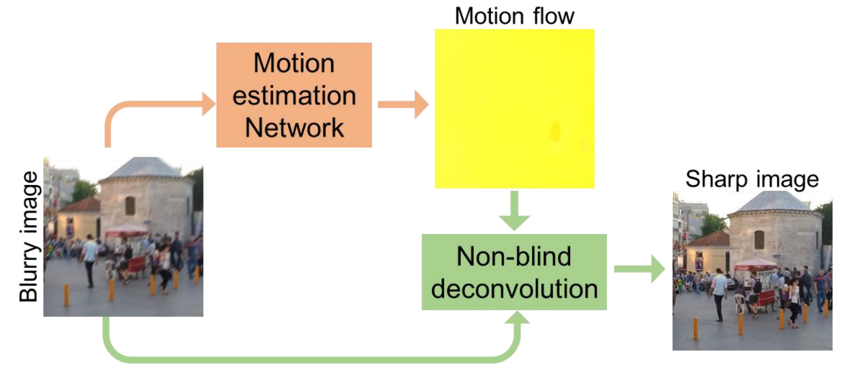
Moving object segmentation in a blurry scenario

- FusionSeg model [4] - Combine appearance and motion information
  - 3 streams (appearance, motion and fusion of the two)
- Object boundaries and appearance cues are corrupted by blur
  - Inaccurate segmentation masks
  - Failure to segment small objects



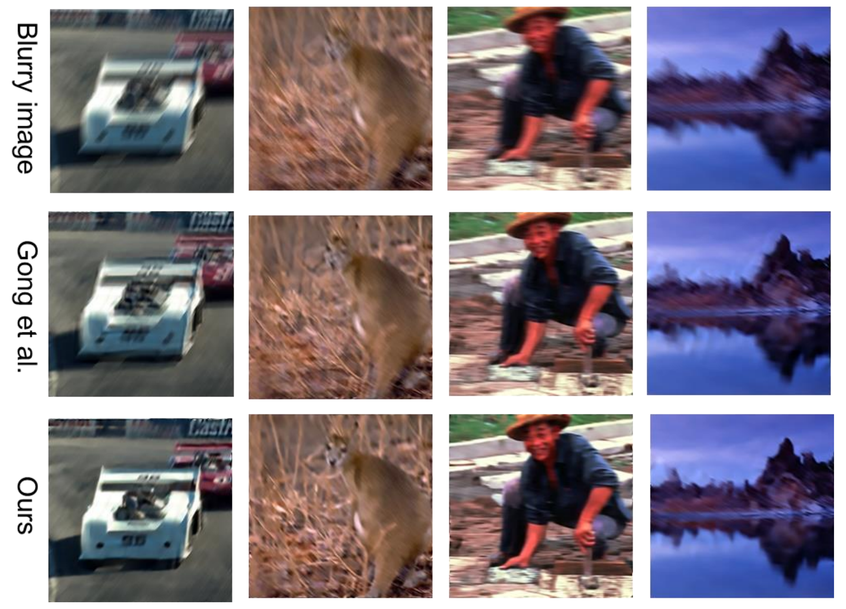
- Optical flow predicted by our network comes to rescue

Motion-blur removal via non-blind deconvolution



- Comparison with previous works [2,5]

	BSD-M		BSD-S	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Sun *et al.*	22.97	0.674	20.53	0.530
Gong *et al.*	23.88	0.718	21.85	0.625
Ours	**25.23**	**0.786**	**23.41**	**0.714**



- Motion estimated by our approach results in sharper images

Ablation studies

Network components

- Feature decoding
  - Direct flow estimation from encoded features
  - U-net network [1]
  - 32% EPE increase
- Motion decoders
  - Both STN and feature refining block (RB) boost network performance

Table: Ablation on different network components

STN	RB	EPE (↓)
✓	✓	**2.077**
✗	✓	2.263
✓	✗	2.383
✗	✗	2.748

Parameter size

- Symmetric motion assumption
  - Single STN and inverse transformation
  - Reduces the # of STNs by half
- Weight shared RB
  - # of parameters reduced by ≈45%
- Performance decreases compared to the baseline model

Table: Ablation on symmetric motion and weight sharing

STN	RB	EPE (↓)
✓(symmetric)	✓	2.216
✓	✓(shared)	2.309
✓(symmetric)	✓(shared)	2.351

References

[1]. Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. PWCNet: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume, CVPR 2018
[2]. Gong, D.; Yang, J.; Liu, L.; Zhang, Y.; Reid, I.; Shen, C.;van den Hengel, A.; and Shi, Q. From Motion Blur to Motion Flow: A Deep Learning Solution for Removing Heterogeneous Motion Blur, CVPR 2017
[3]. Jin, M., Meishvili, G., Favaro, P.: Learning to extract a video sequence from a single motion-blurred image, CVPR 2018
[4]. Jain, Suyog Dutt and Xiong, Bo and Grauman, Kristen, FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos, CVPR 2017
[5]. Sun, J.; Cao, W.; Xu, Z.; and Ponce, J. Learning a convolutional neural network for non-uniform motion blur removal, CVPR 2015
[6]. N.Mayer; E.Ilg; P.H¨ausser; P.Fischer; D.Cremers; A.Dosovitskiy; and T.Brox. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation, CVPR 2016
[7]. Nah,S.;Kim,T.H.;andLee,K.M. Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring, CVPR 2017
[8]. Galoogahi, H. K.; Fagg, A.; Huang, C.; Ramanan, D.; and Lucey, S. Need for Speed: A Benchmark for Higher Frame Rate Object Tracking, ICCV 2017
[9]. Jianping, S.; Li, X.; and Jiaya, J. Discriminative Blur Detection Features, CVPR 2014
[10]. Jin, M., Hu, Z., Favaro, P.: Learning to extract flawless slow motion from blurry videos, CVPR 2019
[11]. Jaderberg, M.; Simonyan, K.; Zisserman, A.; et al. Spatial transformer networks, NIPS 2015