DeFMO: Deblurring and Shape Recovery of Fast Moving Objects

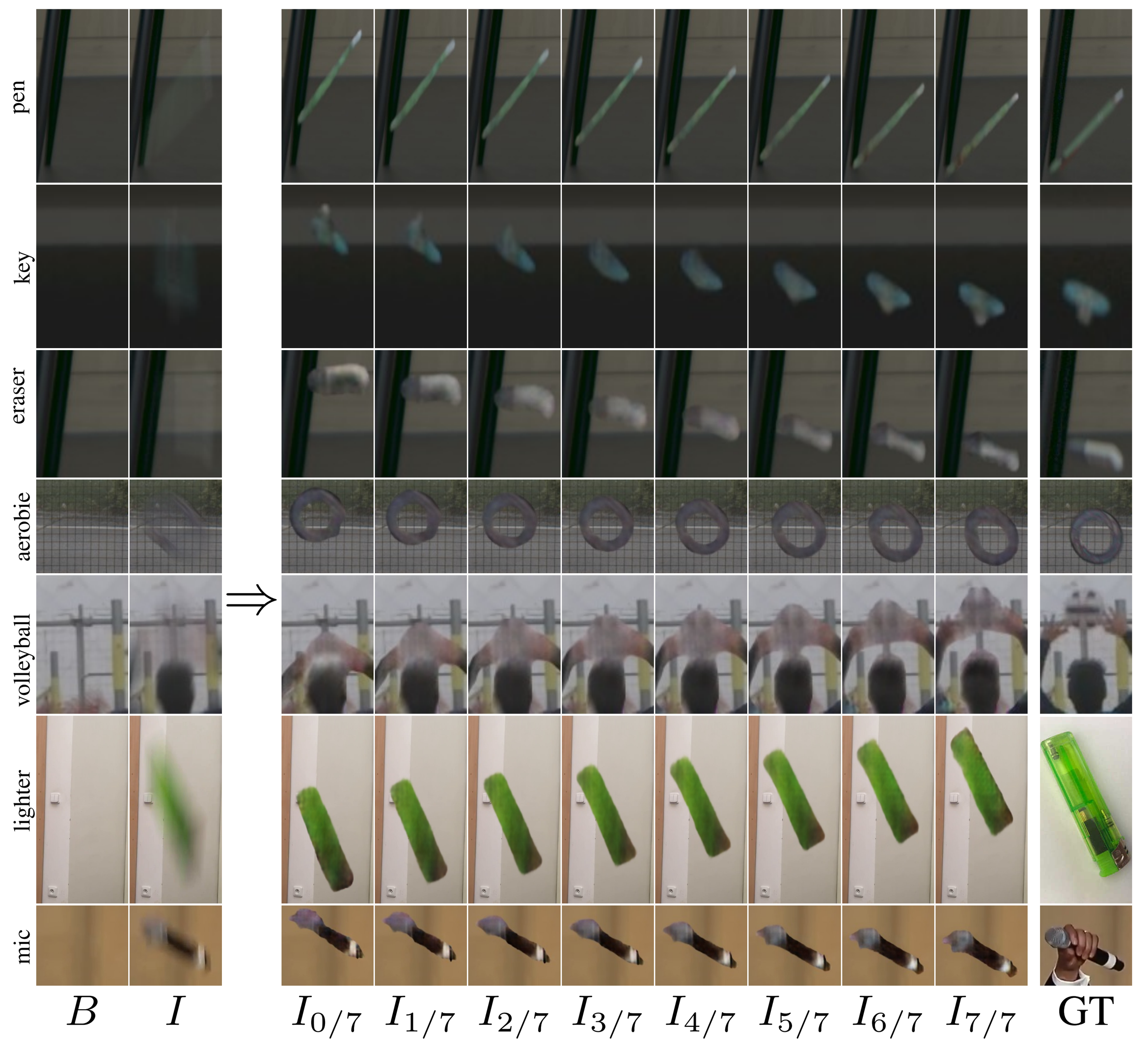
Denys Rozumnyi[1,4], Martin R. Oswald[1], Vittorio Ferrari[2], Jiří Matas[4], Marc Pollefeys[1,3]

[1]Department of Computer Science, ETH Zurich [3]Microsoft Mixed Reality and AI Zurich Lab
[2]Google Research [4]Visual Recognition Group, Czech Technical University in Prague

CVPR VIRTUAL JUNE 19-25

Introduction:

- **Setting**: Motion blur is exceptionally ambiguous for fast moving objects (FMOs) – objects that move over a distance larger than their size within the camera exposure time.
- **Input**: image I with an object moving fast and thus appearing blurred, background B without the object.
- **Task**: reconstruct sub-frames as if this was a short video captured by a high-speed camera (temporal super-resolution).



- Image formation model for an FMO with constant appearance F and shape M moving along trajectory H over background B:

$$I = H * F + (1 - H * M) \cdot B$$

- For time-varying F_i and M_i, TbD-3D defined a sub-frame formation model; cannot deal well will complex objects or complex motions:

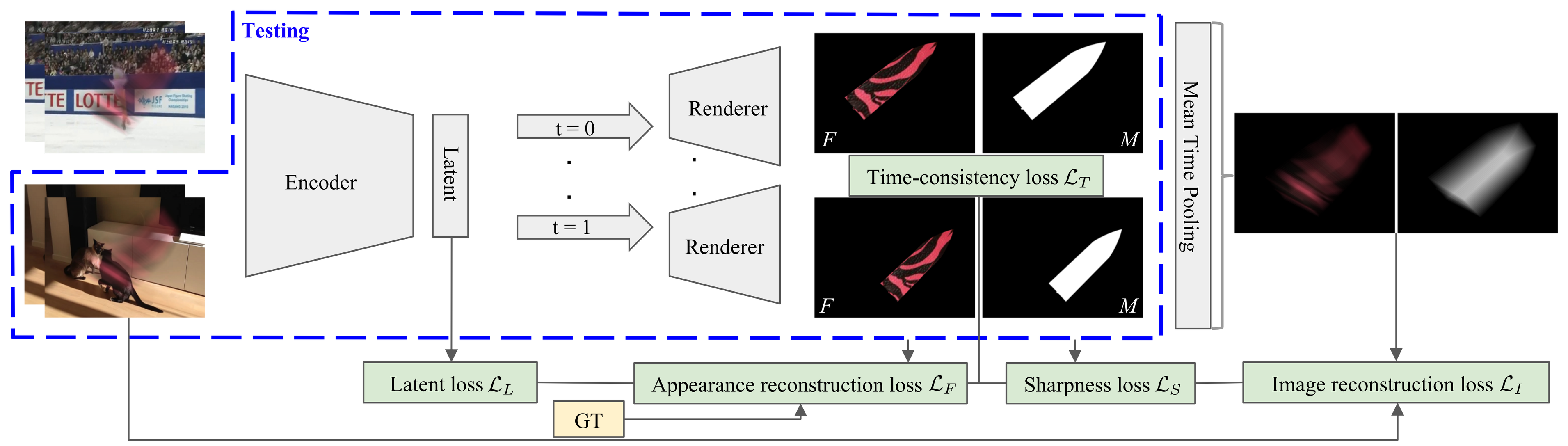
$$I = \sum_i H_i * F_i + \left(1 - \sum_i H_i * M_i\right) \cdot B$$

Method:

- We propose a new generalized image formation model with FMOs, which generalizes all previous ones:

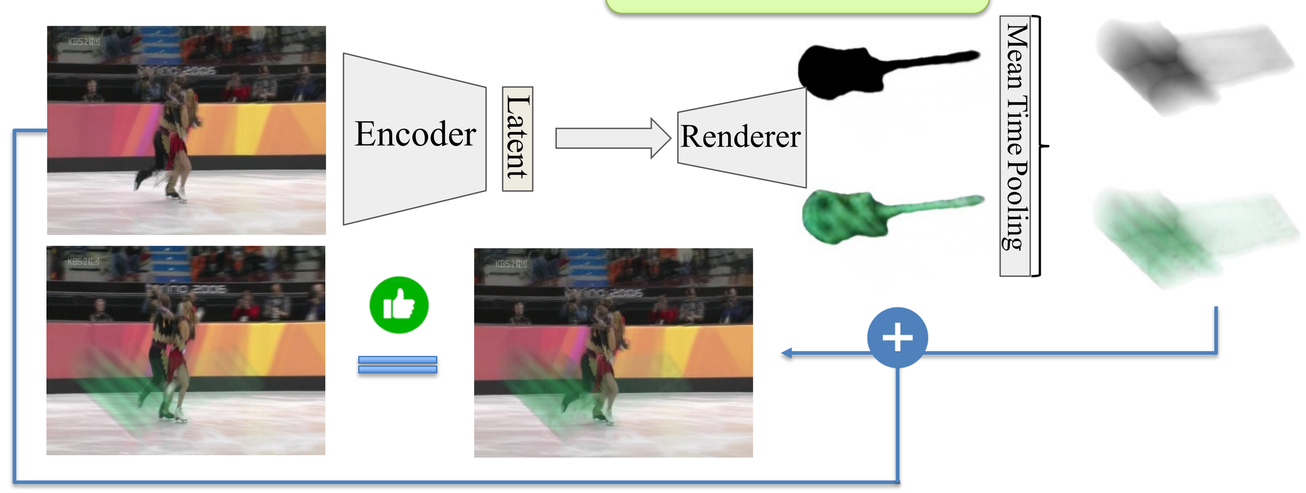
$$I_{t_0:t_1} = \int_{t_0}^{t_1} F_t M_t \,\mathrm{d}t + \left(1 - \int_{t_0}^{t_1} M_t \,\mathrm{d}t\right) B$$

- We treat F_t and M_t as a rendering network.

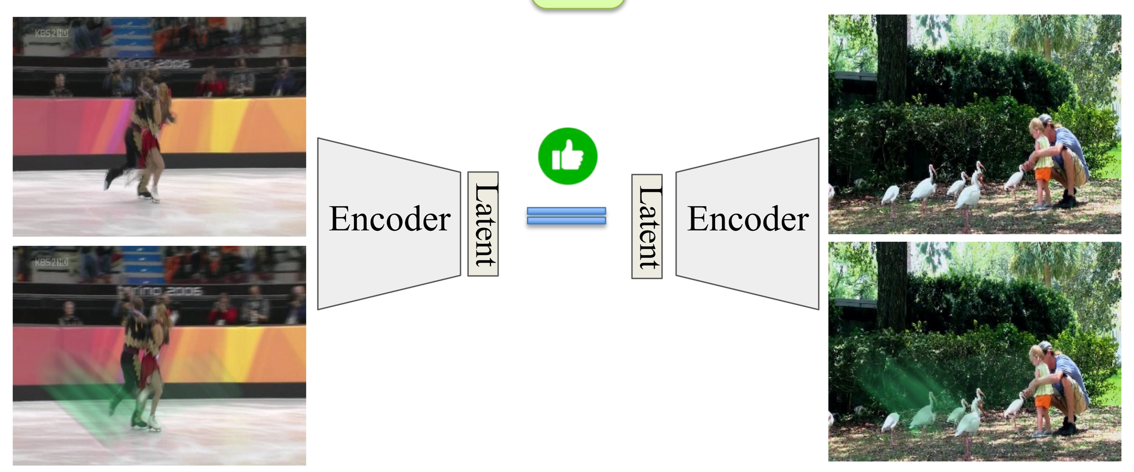


- We have 5 loss terms: a supervised loss and 4 self-supervised ones.
- Sub-frame appearance reconstruction $\mathcal{L}_F$: $(F_t, M_t) \approx (\tilde{F}_t, \tilde{M}_t)$

- Image reconstruction $\mathcal{L}_I$: $I \approx I_{0:1}$



- Latent learning loss $\mathcal{L}_L$



- Time-consistency
  $\mathcal{L}_T$: $(F_t, M_t) \approx (F_{t+\mathrm{d}t}, M_{t+\mathrm{d}t})$
- Sharpness
  $\mathcal{L}_S = \text{entropy}(M_t)$

Experiments:

- We evaluate on TbD, TbD-3D, and Falling Objects datasets.

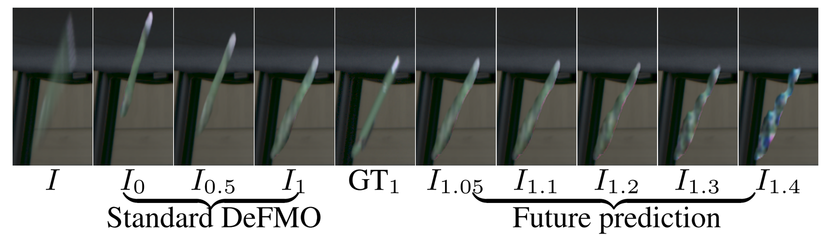
Dataset	Score	Inputs B	Inputs I	Jin *et al.*	DeblurGAN-v2	TbD	TbD-3D	Proposed DeFMO	Traj. Oracle TbD-3D-Or.
Falling	TIoU↑	N/A	N/A	N/A	N/A	0.539	0.539	**0.684**	1.000
	PSNR↑	19.71	23.76	23.54	23.36	20.53	23.42	**26.83**	23.38
	SSIM↑	0.456	0.594	0.575	0.588	0.591	0.671	**0.753**	0.692
TbD-3D	TIoU↑	N/A	N/A	N/A	N/A	0.598	0.598	**0.879**	1.000
	PSNR↑	19.81	24.80	24.52	23.58	18.84	23.13	**26.23**	24.84
	SSIM↑	0.426	0.640	0.590	0.603	0.504	0.651	**0.699**	0.705
TbD	TIoU↑	N/A	N/A	N/A	N/A	0.542	0.542	**0.550**	1.000
	PSNR↑	21.48	25.06	24.90	24.27	23.22	25.21	**25.57**	26.36
	SSIM↑	0.466	0.568	0.530	0.537	0.605	**0.674**	0.602	0.712
Runtime (on 240 × 320)		N/A	N/A	2 fps	10 fps	0.01 fps	0.001 fps	**20 fps**	0.001 fps

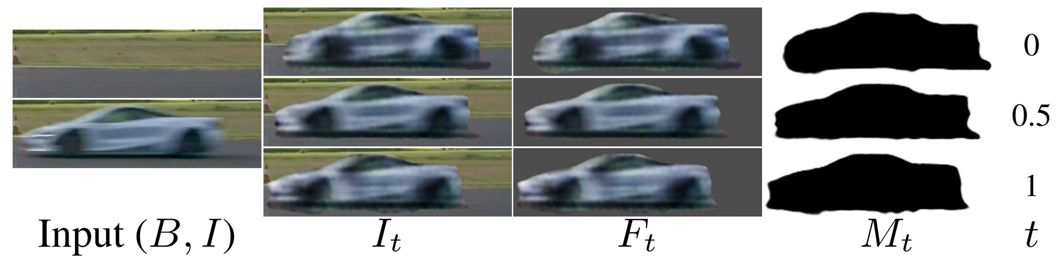
We beat everything.

- Ablation study shows that each loss brings improvement.

Appear. rec. $\mathcal{L}_F$	Image rec. $\mathcal{L}_I$	Time-cons. $\mathcal{L}_T$	Latent learn. $\mathcal{L}_L$	Sharpness $\mathcal{L}_S$	Train PSNR↑	Val PSNR↑	Test – Falling Objects PSNR↑	SSIM↑	TIoU↑
TbD-3D					6.58	6.59	23.0	.695	.545
✓	✗	✗	✗	✗	**23.0**	**22.6**	24.8	.691	.684
✓	✓	✗	✗	✗	22.5	22.2	25.5	.705	.653
✗	✓	✓	✓	✓	11.6	10.9	19.7	.459	.347
✓	✗	✓	✓	✓	12.3	12.2	17.5	.362	.489
✓	✓	✗	✓	✓	21.6	21.5	25.8	.743	.673
✓	✓	✓	✗	✓	22.2	22.0	26.1	.739	.678
✓	✓	✓	✓	✗	22.4	22.2	**26.4**	.750	.676
✓	✓	✓	✓	✓	22.5	22.4	**26.4**	**.753**	**.703**

- The proposed model handles FMOs with complex shapes and significant appearance changes within one video frame.





Conclusion:

- We proposed a novel generative model for disentangling and deblurring of fast moving objects.
- Code is published on GitHub: https://github.com/rozumden/DeFMO
- Benchmark: https://github.com/rozumden/fmo-deblurring-benchmark

References:
[TbD] Kotera et al. Intra-frame Object Tracking by Deblatting, ICCV VOT 2019
[TbD-3D] Rozumnyi et al. Sub-frame Appearance and 6D Pose Estimation of Fast Moving Objects, CVPR 2020
[Falling Objects] Kotera et al. Restoration of Fast Moving Objects, TIP 2020