Deep Video Frame Interpolation using Cyclic Frame Generation

Yu-Lun Liu[1,2,3], Yi-Tung Liao[1,2], Yen-Yu Lin[1], Yung-Yu Chuang[1,2]

[1]Academia Sinica, [2]National Taiwan University, [3]MediaTek

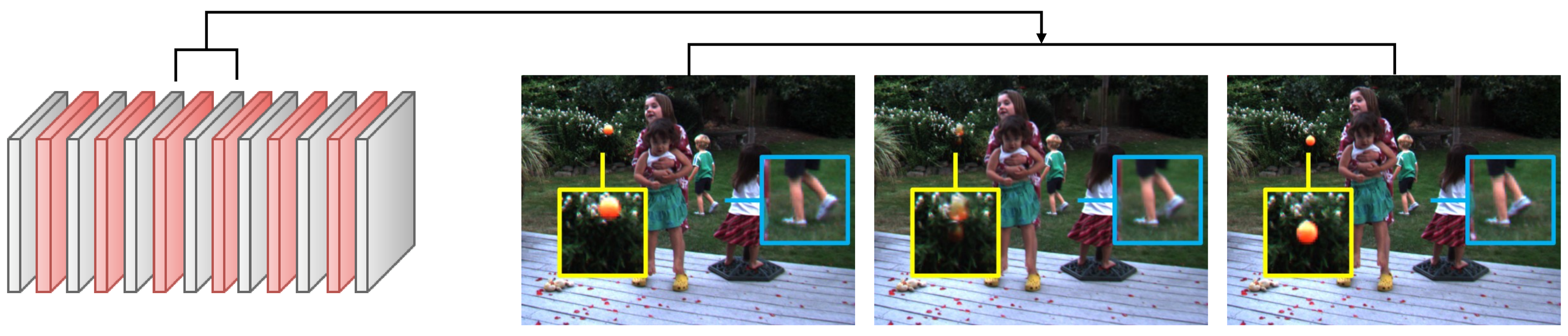
Code Available at: https://github.com/alex04072000/CyclicGen

中央研究院 ACADEMIA SINICA 國立臺灣大學 National Taiwan University MEDIATEK

AAAI-19: Thirty-Third AAAI Conference on Artificial Intelligence

Introduction

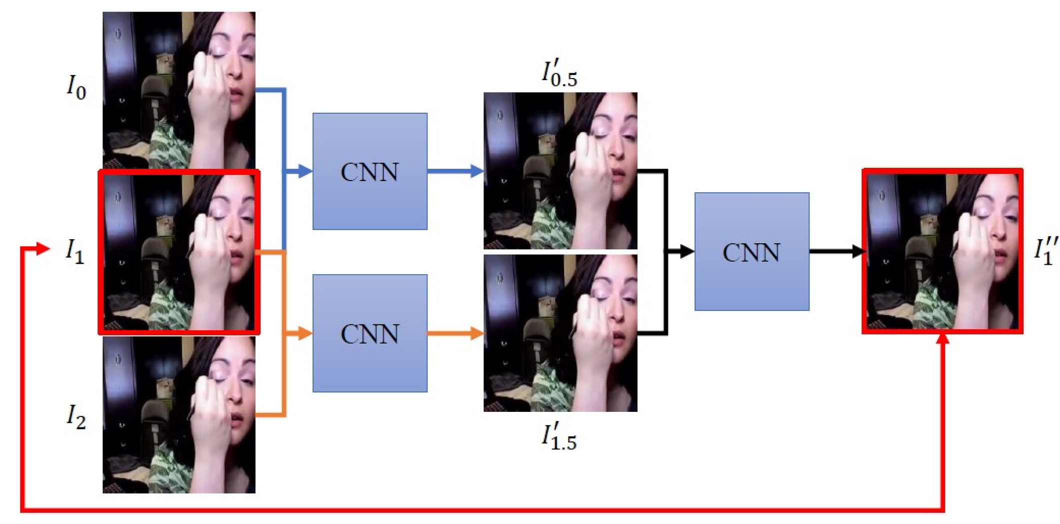
- **Goal:**
  - Predict the intermediate frame between two consecutive frames



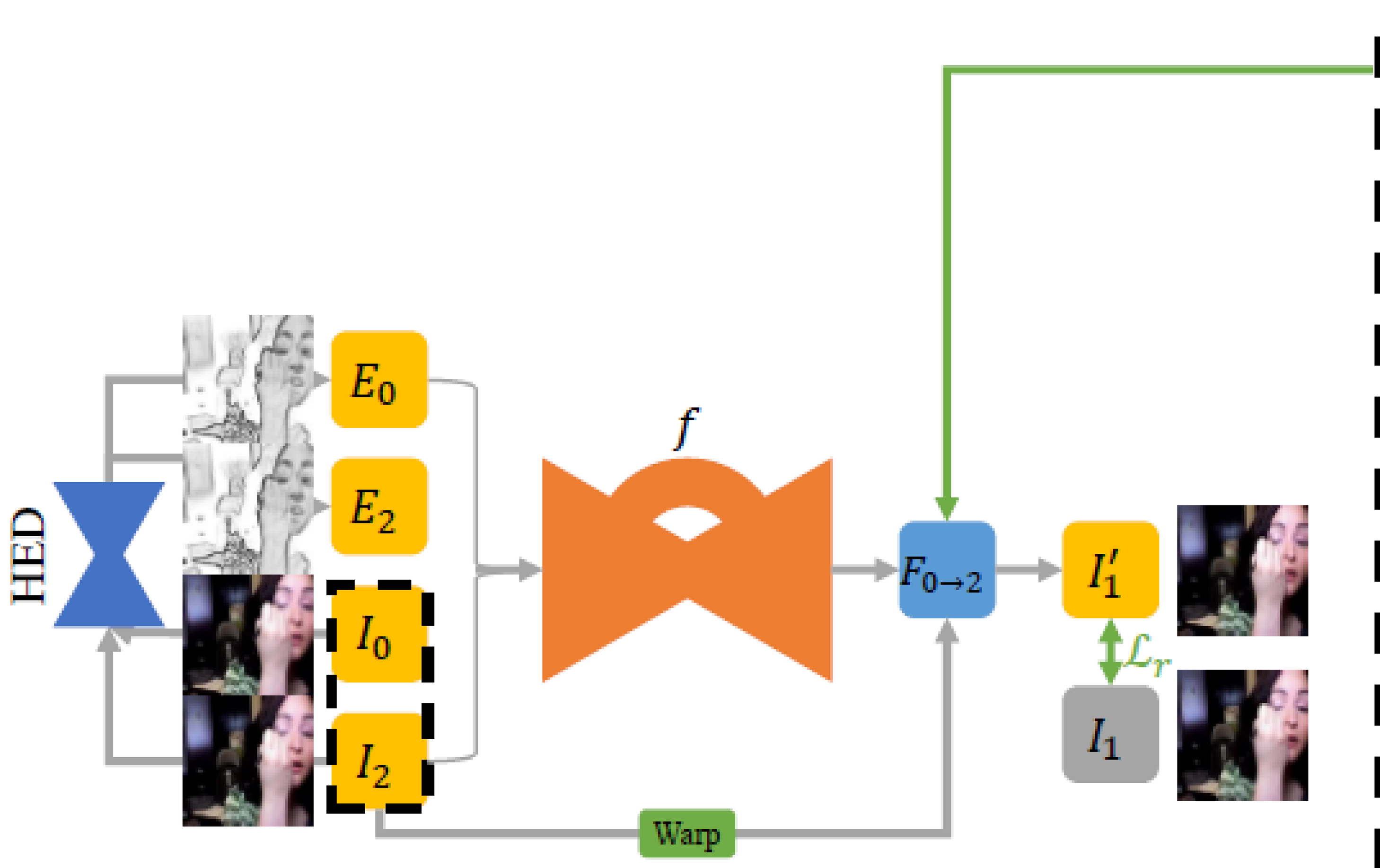
- **Challenges:**
  - Conventional methods → computationally expensive
  - CNN-based methods → artifacts and over-smoothed results

Solution: Enforcing cycle consistency

- **Observation**:
  - Over-smoothed frames or frames with artifacts cannot well reconstruct the original frames

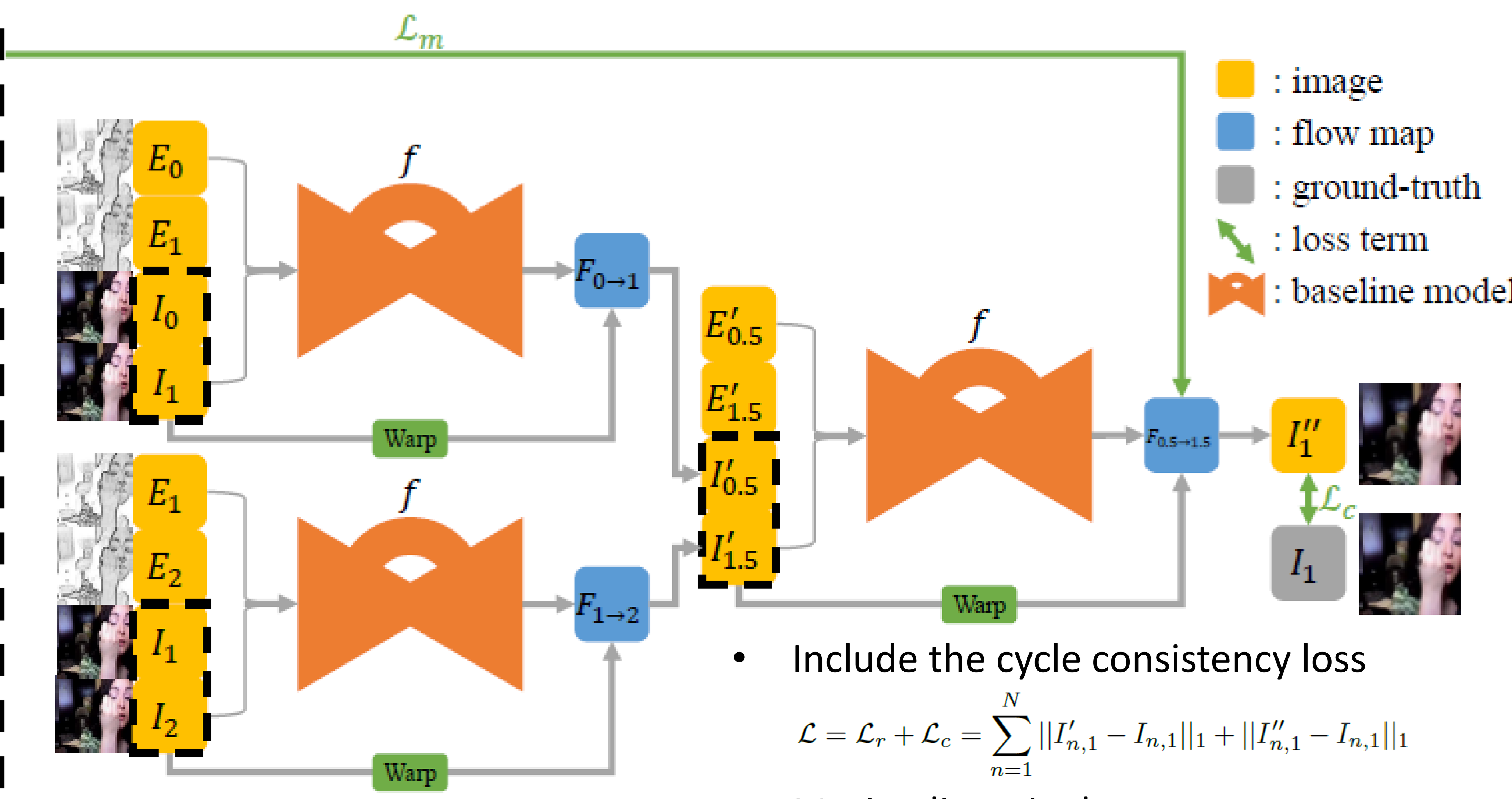


A two-stage training procedure



- **Stage 1**:
  - Pre-train the baseline model
  - Reconstruction loss

$$\mathcal{L}_r = \sum_{n=1}^{N} \|f(I_{n,0}, I_{n,2}) - I_{n,1}\|_1 = \sum_{n=1}^{N} \|I'_{n,1} - I_{n,1}\|_1$$



- **Stage 2**:
  - Include the cycle consistency loss

$$\mathcal{L} = \mathcal{L}_r + \mathcal{L}_c = \sum_{n=1}^{N} \|I'_{n,1} - I_{n,1}\|_1 + \|I''_{n,1} - I_{n,1}\|_1$$

  - Motion linearity loss

$$\mathcal{L}_m = \sum_{n=1}^{N} \|F_{n,0\to2} - 2 \cdot F_{n,0.5\to1.5}\|_2^2$$

  - Edge-guided training

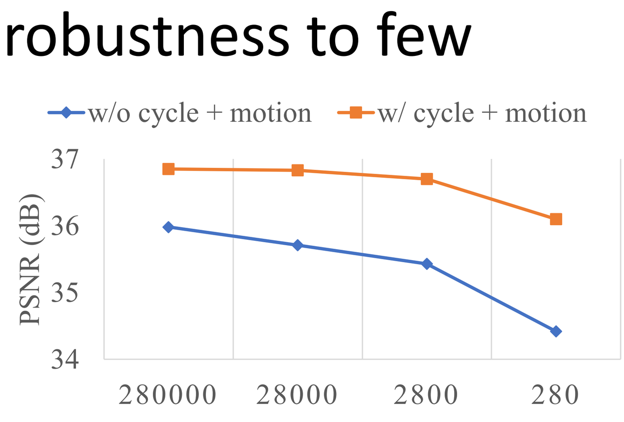
Ablation studies on UCF dataset

- The introduced components help video interpolation

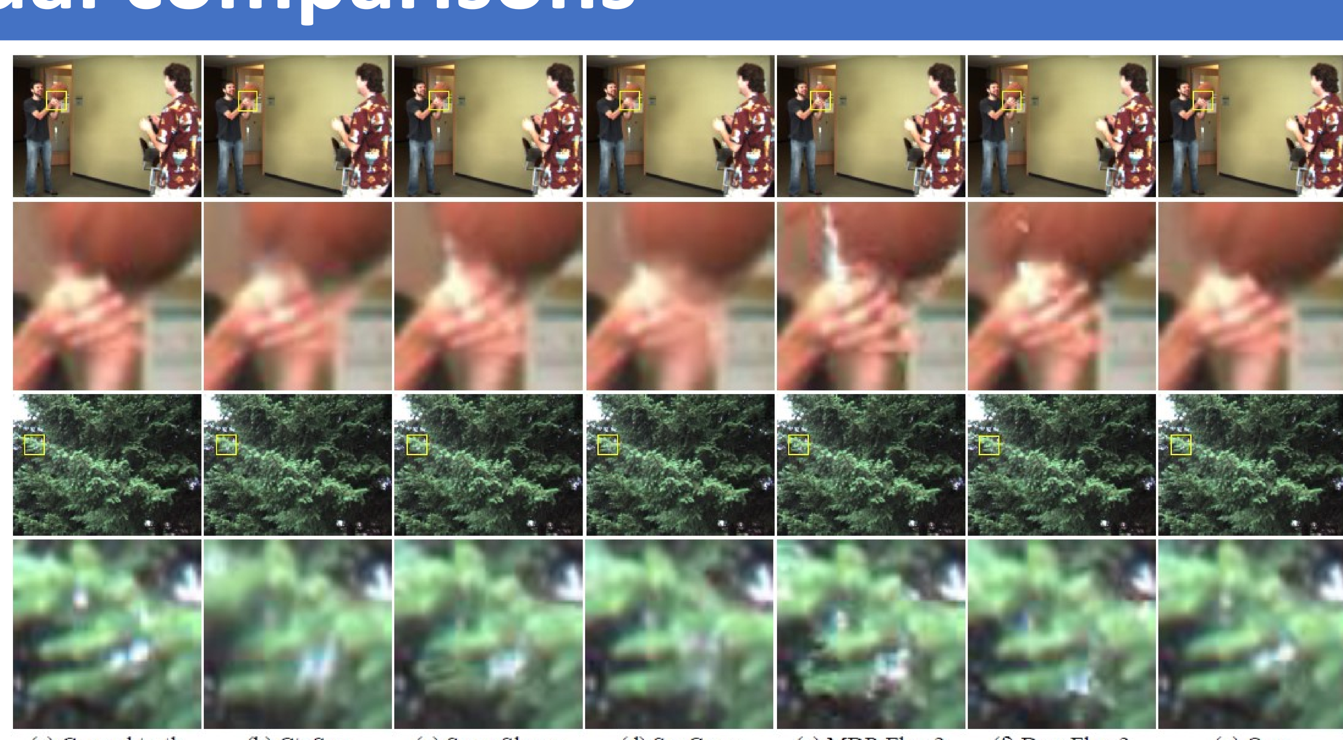
	PSNR	SSIM
Baseline (DVF)	35.89	0.945
+ *Cycle*	36.71 (+0.82)	0.950 (+0.005)
+ *Cycle* + *Motion*	36.85 (+0.96)	0.950 (+0.005)
+ *Cycle* + *Edge*	36.86 (+0.97)	0.952 (+0.007)
full model	**36.96** (+1.07)	**0.953** (+0.008)

- Cycle consistency loss improves the robustness to few training data

Training size	Baseline (DVF)	Ours ($\mathcal{L}_c + \mathcal{L}_m$)
1	35.98	36.85
1/10	35.71 (-0.27)	36.83 (-0.02)
1/100	35.43 (-0.55)	36.70 (-0.15)
1/1000	34.42 (-1.56)	36.10 (-0.75)



Visual comparisons



(a) Ground truth (b) CtxSyn (c) SuperSlomo (d) SepConv (e) MDP-Flow2 (f) DeepFlow (g) Ours

Comparison with SoTAs

- On UCF-101 dataset and a high-quality video *See You Again*

	UCF101		*See You Again*	
	PSNR	SSIM	PSNR	SSIM
DVF	35.89	0.945	40.15	0.958
SepConv	36.49	0.950	41.01	**0.968**
Ours	**36.96**	**0.953**	**41.67**	**0.968**

- On Middlebury dataset

	AVERAGE			Mequon			Schefflera			Urban		
	all	disc.	unt.	all	disc.	unt.	all	disc.	unt.	all	disc.	unt.
Ours	**4.20**	**6.16**	**1.97**	2.26	**3.32**	1.42	3.19	**4.01**	2.21	**2.76**	4.05	1.62
CtxSyn	5.28	8.00	2.19	**2.24**	3.72	**1.04**	**2.96**	4.16	1.35	4.32	**3.42**	3.18
SuperSlomo	5.31	8.39	2.12	2.51	4.32	1.25	3.66	5.06	1.93	2.91	4.00	1.41
SepConv	5.61	8.74	2.33	2.52	4.83	1.11	3.56	5.04	1.90	4.17	4.15	2.86
MDP-Flow2	5.83	9.69	2.15	2.89	5.38	1.19	3.47	5.07	**1.26**	3.66	6.10	2.48
DeepFlow	5.97	9.79	2.05	2.98	5.67	1.22	3.88	5.78	1.52	3.62	5.93	**1.34**

	Teddy			Backyard			Basketball			Dumptruck			Evergreen		
	all	disc.	unt.	all	disc.	unt.	all	disc.	unt.	all	disc.	unt.	all	disc.	unt.
Ours	4.97	5.92	3.79	**8.00**	**9.84**	**3.13**	**3.36**	**5.65**	**2.17**	**4.55**	**9.68**	**1.42**	**4.48**	**6.84**	**1.52**
CtxSyn	**4.21**	**5.46**	**3.00**	9.59	11.9	3.46	5.22	9.76	2.22	7.02	15.4	1.58	6.66	10.2	1.69
SuperSlomo	5.05	6.27	3.66	9.56	11.9	3.30	5.37	10.2	2.24	6.69	15.0	1.53	6.73	10.4	1.66
SepConv	5.41	6.81	3.88	10.2	12.8	3.37	5.47	10.4	2.21	6.88	15.6	1.72	6.63	10.3	1.62
MDP-Flow2	5.20	7.48	3.14	10.2	12.8	3.61	6.13	11.8	2.31	7.36	16.8	1.49	7.75	12.1	1.69
DeepFlow	5.39	7.20	3.17	11.0	13.9	3.63	5.91	11.3	2.29	7.14	16.3	1.49	7.80	12.2	1.70

Conclusion

- We present a novel loss, the cycle consistency loss, which
  - can be integrated with existing video frame interpolation methods and trained end-to-end
  - synthesizes more plausible frames possessing similar characteristics with the original frames
- We propose two extensions, motion linearity loss and edge guided training, that
  - regularize the training procedure
  - further improve model performance
- The proposed approach better utilizes the training data, not only enhancing the interpolation results, but also reaching better performance with less training data.

References:

- Liu, Z.; Yeh, R.; Tang, X.; Liu, Y.; and Agarwala, A. 2017. Video frame synthesis using deep voxel flow. In Proceedings of IEEE ICCV.
- Niklaus, S.; Mai, L.; and Liu, F. 2017b. Video frame interpolation via adaptive separable convolution. In Proceedings of IEEE ICCV.
- Niklaus, S., and Liu, F. 2018. Context-aware synthesis for video frame interpolation. In Proceedings of IEEE CVPR.
- Jiang, H.; Sun, D.; Jampani, V.; Yang, M.-H.; Learned-Miller, E.; and Kautz, J. 2018. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In Proceedings of IEEE CVPR.