Softmax Splatting for Video Frame Interpolation
— Supplementary Material —

Simon Niklaus
Portland State University
sniklaus@pdx.edu

Feng Liu
Portland State University
fliu@cs.pdx.edu

Please note that the section numbering within this supplementary document has manually been adjusted to match the relevant sections in the main paper.

3.1. Forward Warping via Softmax Splatting

Our proposed softmax splatting shares resemblance to the softmax function and is hence invariant to translations β with respect to Z which can easily be shown as follows.

$$\text{let } \boldsymbol{u} = \boldsymbol{p} - \big(\boldsymbol{q} + F_{0 \to t}[\boldsymbol{q}]\big) \tag{1}$$

$$b(\boldsymbol{u}) = \max(0, 1 - |\boldsymbol{u}_x|) \cdot \max(0, 1 - |\boldsymbol{u}_y|) \tag{2}$$

$$I_t^\sigma[\boldsymbol{p}] = \frac{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}] + \beta) \cdot b(\boldsymbol{u}) \cdot I_0[\boldsymbol{q}]}{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}] + \beta) \cdot b(\boldsymbol{u})} \tag{3}$$

$$= \frac{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}]) \cdot \exp(\beta) \cdot b(\boldsymbol{u}) \cdot I_0[\boldsymbol{q}]}{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}]) \cdot \exp(\beta) \cdot b(\boldsymbol{u})} \tag{4}$$

$$= \frac{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}]) \cdot b(\boldsymbol{u}) \cdot I_0[\boldsymbol{q}]}{\sum_{\forall \boldsymbol{q} \in I_0} \exp(Z[\boldsymbol{q}]) \cdot b(\boldsymbol{u})} \tag{5}$$

This property is particularly important when mapping multiple pixels to the same location. For example, if Z represents depth then moving then changing $Z' = Z + \beta$ should not affect the way the mapping ambiguity is resolved.

4.3. Quantitative Evaluation

As shown in Table 1 as well as Table 2 and summarized in Table 3, our proposed approach ranks first in the relevant Middlebury benchmark for optical flow [1]. Specifically, our approach achieves an average rank of 2.5 in terms of the interpolation error, compared to 5.4 achieved by the second highest ranked method. Similarly, in terms of the normalized interpolation error, the average rank of our approach and the second highest ranked method are 2.0 and 10.8, respectively. Note that the second-best method differs, the second best method in terms of the interpolation error has not yet been published while DAIN [2] is the second best method in terms of the normalized interpolation error. These results demonstrate that our images synthesis approach, empowered by softmax platting, is able to consistently produce high-quality interpolation results.

Since common datasets to evaluate frame interpolation are subject to a relatively low resolution, we additionally incorporated 4K video clips from Xiph. These video clips are commonly used to assess video compression and contain challenging scenarios, such as significant bokeh that mixes foreground and background regions or fine-details that are reflective and subject to sudden brightness changes. For the purpose of evaluating frame interpolation techniques, we selected the eight 4K video clips with the most amount of inter-frame motion. Please see our main paper for more details where we state the average metrics. For completeness, we show the per-clip metrics in Tables 4–6.

4.4. Qualitative Evaluation

Since videos are at the heart of our approach, we provide a qualitative comparison in the supplementary video. We additionally provide still results on challenging examples from DAVIS [7] in Figures 1–3. These support the findings of the quantitative evaluation and show difficult scenarios where our proposed approach produces high-quality results whereas competing techniques are subject to visual artifacts. Please note that for methods with models trained using different losses, we only show the results for the models that focus on perceptual quality.

Average interpolation error	avg. rank	Mequon (Hidden texture) im0 GT im1			Schefflera (Hidden texture) im0 GT im1			Urban (Synthetic) im0 GT im1			Teddy (Stereo) im0 GT im1			Backyard (High-speed camera) im0 GT im1			Basketball (High-speed camera) im0 GT im1			Dumptruck (High-speed camera) im0 GT im1			Evergreen (High-speed camera) im0 GT im1		
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
SoftSplat [170]	2.5	2.06 2	3.06 2	1.14 6	**2.80** 1	**3.91** 1	**1.24** 1	1.99 2	2.73 2	1.21 3	**3.84** 1	**4.64** 1	**2.69** 1	8.10 9	10.0 9	**2.96** 1	4.10 2	7.53 2	**1.98** 1	5.49 3	12.1 3	**1.39** 1	5.40 2	8.33 2	1.50 2
PyrWarp [161]	5.4	**1.99** 1	**3.00** 1	0.99 2	3.02 3	4.11 3	1.39 41	**1.81** 1	**2.54** 1	**1.13** 1	3.95 2	4.73 2	2.77 3	7.51 2	9.25 2	3.06 6	4.46 4	8.27 5	2.05 8	5.65 4	12.5 4	1.51 16	5.84 4	9.02 4	1.60 10
FGME [163]	8.2	2.08 3	3.34 4	**0.98** 1	3.32 11	4.43 5	1.63 96	2.46 3	3.28 3	1.41 11	4.08 3	4.85 3	3.05 11	**7.36** 1	**9.08** 1	3.03 4	4.17 3	7.62 3	2.06 9	4.95 2	10.7 2	1.44 5	5.45 3	8.41 3	1.57 6
STAR-Net [169]	10.2	2.18 4	3.37 5	1.21 28	3.46 20	4.88 17	1.47 66	3.04 8	3.53 6	1.58 24	4.41 5	5.44 5	2.76 2	7.51 2	9.27 3	2.98 3	4.65 7	8.72 7	1.99 2	6.21 8	13.4 8	1.41 2	6.17 6	9.45 6	**1.49** 1
DSepConv [167]	16.3	2.47 11	4.39 16	1.21 28	3.32 11	4.60 11	1.72 115	3.28 10	3.66 7	1.50 17	5.11 16	6.36 13	3.23 52	7.85 4	9.69 4	3.11 8	4.68 8	8.78 8	2.04 6	5.65 4	12.5 4	1.44 5	6.54 12	10.2 12	1.58 9
MAF-net [168]	20.3	2.23 5	3.84 8	1.08 4	3.53 27	4.85 16	1.78 127	2.83 5	3.70 8	1.58 24	4.83 9	5.88 7	3.31 83	9.44 13	11.8 13	3.27 14	5.27 15	10.0 15	2.15 11	6.30 9	14.2 10	1.54 31	6.38 10	9.90 11	1.63 13
MPRN [155]	20.6	2.53 14	4.43 17	1.21 28	3.78 64	4.97 20	1.57 83	3.39 14	5.49 22	1.28 4	5.03 12	6.58 17	3.19 32	9.53 14	11.9 14	3.31 16	5.25 14	9.92 13	2.22 14	6.87 15	15.5 17	1.49 9	6.72 16	10.4 16	1.60 10
CtxSyn [136]	20.9	2.24 6	3.72 7	1.04 3	2.96 2	4.16 4	1.35 33	4.32 83	3.42 5	3.18 125	4.21 4	5.46 6	3.00 5	9.59 16	11.9 14	3.46 19	5.22 12	9.76 12	2.22 14	7.02 18	15.4 16	1.58 50	6.66 15	10.2 12	1.69 21
ADC [166]	21.5	2.54 16	4.31 11	1.29 133	3.27 5	4.46 6	1.62 94	3.76 38	3.76 10	1.70 36	5.27 22	6.37 14	3.19 32	8.66 12	10.8 12	3.11 8	4.78 11	9.04 11	2.01 3	5.72 6	12.8 7	1.41 2	6.56 13	10.2 12	1.51 3
FRUCnet [157]	22.6	2.61 18	4.34 13	1.52 163	3.30 8	4.52 7	1.72 115	3.14 9	3.70 8	1.76 42	4.74 8	5.99 11	3.29 68	8.11 10	10.0 9	2.97 2	4.48 5	8.35 6	2.02 4	5.78 7	12.7 6	1.45 7	6.06 5	9.38 5	1.57 6
MS-PFT [164]	23.6	2.53 14	4.35 14	1.16 7	3.61 40	5.03 22	1.69 110	3.30 11	4.25 17	1.77 44	5.13 17	6.55 16	3.19 32	7.94 7	9.81 7	3.21 12	4.49 6	8.24 4	2.22 14	6.55 12	13.9 9	1.79 110	6.42 11	9.89 10	1.69 21
DAIN [156]	24.4	2.38 9	4.05 10	1.26 117	3.28 6	4.53 9	1.79 129	3.32 12	3.77 11	2.05 64	4.65 7	5.88 7	3.41 106	7.88 5	9.74 5	3.04 5	4.73 10	8.90 10	2.04 6	6.36 10	14.3 12	1.51 16	6.25 7	9.68 8	1.54 5
CyclicGen [153]	25.0	2.26 7	3.32 3	1.42 159	3.19 4	4.01 2	2.21 161	2.76 4	4.05 15	1.62 28	4.97 11	5.92 9	3.79 146	8.00 8	9.84 8	3.13 10	**3.36** 1	**5.65** 1	2.17 12	**4.55** 1	**9.68** 1	1.42 4	**4.48** 1	**6.84** 1	1.52 4
MDP-Flow2 [68]	29.4	2.89 22	5.38 23	1.19 9	3.47 21	5.07 28	1.26 2	3.66 27	6.10 57	2.48 94	5.20 18	7.48 28	3.14 19	10.2 20	12.8 22	3.61 41	6.13 42	11.8 38	2.31 44	7.36 23	16.8 21	1.49 9	7.75 39	12.1 38	1.69 21
PMMST [114]	29.7	2.90 23	5.43 25	1.20 13	3.50 26	5.05 26	1.27 7	3.56 18	5.46 20	1.82 47	5.38 35	7.92 55	3.41 106	10.2 20	12.8 22	3.60 35	5.76 20	11.0 20	2.26 22	7.39 24	16.9 24	1.53 24	7.57 24	11.8 24	1.72 53
TOF-M [154]	30.5	2.54 16	4.35 14	1.16 7	3.70 54	5.19 32	1.88 136	3.43 15	3.89 13	1.93 54	5.05 13	6.43 15	3.39 100	9.84 17	12.3 17	3.42 18	5.34 17	10.0 15	2.28 30	6.88 16	15.2 15	1.61 59	7.14 19	11.0 19	1.69 21
MEMC-Net+ [165]	30.6	2.39 10	3.92 9	1.28 126	3.36 14	4.52 7	2.07 157	3.37 13	3.86 12	2.20 75	4.84 10	5.93 10	3.72 140	8.55 11	10.6 11	3.14 11	4.70 9	8.81 9	2.03 5	6.40 11	14.2 10	1.58 50	6.37 9	9.87 9	1.57 6
SuperSlomo [132]	30.7	2.51 12	4.32 12	1.25 104	3.66 48	5.06 27	1.93 148	2.91 6	4.00 14	1.41 11	5.05 13	6.27 12	3.66 135	9.56 15	11.9 14	3.30 15	5.37 18	10.2 18	2.24 17	6.69 13	15.0 14	1.53 24	6.73 17	10.4 16	1.66 14
OFRI [158]	36.0	2.28 8	3.45 6	1.35 152	3.44 17	4.57 10	2.13 159	3.02 7	3.34 4	1.73 39	4.51 6	5.42 4	3.88 147	7.89 6	9.76 6	3.10 7	5.24 13	9.92 13	2.10 10	6.78 14	14.3 12	1.82 114	6.32 8	9.62 7	1.75 95
SepConv-v1 [127]	38.4	2.52 13	4.83 18	1.11 5	3.56 36	5.04 24	1.90 139	4.17 69	4.15 16	2.86 112	5.41 44	6.81 18	3.88 147	10.2 20	12.8 22	3.37 17	5.47 19	10.4 19	2.21 13	6.88 16	15.6 18	1.72 96	6.63 14	10.3 15	1.62 12
DeepFlow [86]	39.2	2.98 35	5.67 40	1.22 59	3.88 77	5.78 76	1.52 74	3.62 20	5.93 49	1.34 6	5.39 40	7.20 23	3.17 23	11.0 53	13.9 60	3.63 52	5.91 26	11.3 25	2.29 35	7.14 19	16.3 19	1.49 9	7.80 47	12.2 45	1.70 30
CBF [12]	44.3	2.83 19	5.20 19	1.23 79	3.97 90	5.79 79	1.56 79	3.62 20	5.47 21	1.60 27	5.21 19	7.12 20	3.29 68	10.1 18	12.6 18	3.62 46	5.97 29	11.5 29	2.31 44	7.76 58	17.8 59	1.61 59	7.60 27	11.9 27	1.76 109
DeepFlow2 [108]	44.8	2.99 38	5.65 37	1.22 59	3.88 77	5.79 79	1.48 68	3.62 20	6.03 51	1.34 6	5.38 35	7.44 27	3.22 45	11.0 53	13.8 53	3.67 61	5.83 21	11.2 21	2.25 21	7.60 39	17.4 40	1.50 12	7.82 48	12.2 45	1.77 118
NN-field [71]	46.8	2.98 35	5.70 41	1.20 13	3.31 10	4.73 14	1.26 2	4.69 107	5.91 47	2.03 63	5.99 107	9.13 124	3.57 127	10.3 26	12.8 22	3.60 35	6.24 53	12.0 52	2.31 44	7.39 24	16.9 24	1.54 31	7.69 35	12.0 33	1.72 53

Table 1: Screenshot of our IE-ranking in the Middlebury benchmark (taken on the 7th of November, currently private).

Average norm. interpolation error	avg. rank	Mequon (Hidden texture) im0 GT im1			Schefflera (Hidden texture) im0 GT im1			Urban (Synthetic) im0 GT im1			Teddy (Stereo) im0 GT im1			Backyard (High-speed camera) im0 GT im1			Basketball (High-speed camera) im0 GT im1			Dumptruck (High-speed camera) im0 GT im1			Evergreen (High-speed camera) im0 GT im1		
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
SoftSplat [170]	2.0	0.53 5	0.51 3	0.61 7	**0.52** 1	**0.68** 1	**0.55** 1	0.52 2	**0.53** 1	**0.58** 1	**0.80** 1	**0.73** 1	**1.06** 1	0.85 4	0.81 4	0.98 2	**0.81** 1	0.73 2	**0.90** 1	0.56 2	0.86 2	**0.60** 1	0.57 2	0.73 2	**0.59** 1
DAIN [156]	10.8	0.58 10	0.63 8	0.65 58	0.60 7	0.79 11	0.62 48	0.69 5	0.73 8	0.68 10	0.86 4	0.86 8	1.10 3	0.87 7	0.83 6	1.02 7	0.85 3	0.86 6	0.92 5	0.59 6	0.97 10	0.61 2	0.66 11	0.90 13	0.60 2
STAR-Net [169]	11.1	0.56 8	0.56 4	0.65 58	0.65 37	0.85 22	0.62 48	0.70 6	0.69 6	0.76 18	0.87 5	0.82 5	**1.06** 1	0.83 2	0.79 2	**0.97** 1	0.83 2	0.83 5	**0.90** 1	0.63 10	1.09 16	0.61 2	0.61 3	0.80 3	0.60 2
FGME [163]	13.4	**0.46** 1	**0.49** 1	**0.51** 1	0.63 16	0.78 9	0.64 77	0.60 3	0.65 5	0.67 9	0.85 3	0.76 3	1.15 6	**0.82** 1	**0.77** 1	0.99 3	0.86 6	0.81 4	0.95 9	0.61 9	0.93 8	0.70 35	0.63 5	0.83 5	0.65 101
PyrWarp [161]	14.7	0.47 2	**0.49** 1	0.52 2	0.57 3	0.72 3	0.60 24	**0.51** 1	0.58 2	0.62 2	0.82 2	0.75 2	1.10 3	0.84 3	0.79 2	1.00 4	0.85 3	0.75 3	0.94 8	0.64 13	0.93 8	0.81 113	0.62 4	0.80 3	0.67 142
DSepConv [167]	16.0	0.58 10	0.72 18	0.62 9	0.63 16	0.82 17	0.65 87	0.72 8	0.70 7	0.75 16	0.96 31	0.97 17	1.15 6	0.87 7	0.83 6	1.04 9	0.89 9	1.00 12	0.93 7	0.57 3	0.89 3	0.64 7	0.65 9	0.87 10	0.64 61
MEMC-Net+ [165]	17.2	0.59 13	0.65 10	0.65 58	0.64 25	0.79 11	0.70 116	0.80 13	0.77 10	0.96 47	0.88 8	0.83 7	1.12 5	0.88 9	0.85 9	1.01 5	0.85 3	0.88 7	0.91 3	0.64 13	1.13 21	0.62 5	0.63 5	0.86 8	0.60 2
MPRN [155]	17.5	0.59 13	0.70 15	0.64 12	0.66 41	0.89 37	0.64 77	0.77 11	1.07 17	0.64 4	0.93 12	0.93 13	1.17 11	0.95 15	0.91 16	1.12 14	0.96 15	1.04 15	1.01 13	0.60 7	0.98 11	0.65 10	0.72 16	1.02 17	0.62 7
ADC [166]	20.1	0.61 36	0.68 13	0.67 120	0.62 12	0.78 9	0.66 94	0.84 16	0.81 11	0.82 26	0.96 31	0.93 13	1.15 6	0.90 10	0.86 11	1.05 10	0.92 10	1.12 18	0.91 3	0.57 3	0.92 7	0.61 2	0.64 8	0.88 11	0.60 2
MAF-net [168]	26.2	0.48 3	0.61 7	0.52 2	0.65 37	0.86 25	0.62 48	0.67 4	0.86 13	0.78 22	0.96 31	0.92 12	1.20 18	0.93 13	0.89 13	1.08 11	0.95 13	1.08 17	0.99 11	0.67 21	1.04 13	0.83 119	0.65 9	0.86 8	0.69 160
FRUCnet [157]	28.5	0.70 143	0.71 16	0.80 163	0.64 25	0.80 14	0.69 110	0.78 12	0.75 9	0.95 46	0.91 9	0.91 11	1.15 6	0.86 5	0.83 6	1.01 5	0.86 6	0.89 8	0.92 5	0.58 5	0.89 3	0.64 7	0.63 5	0.83 5	0.64 61
CyclicGen [153]	28.8	0.64 109	0.63 8	0.73 158	0.67 49	0.73 4	0.88 163	0.72 8	0.84 12	0.78 22	0.95 24	0.89 9	1.24 62	0.91 11	0.85 9	1.09 12	0.87 8	**0.67** 1	1.00 12	**0.53** 1	**0.71** 1	0.62 5	**0.52** 1	**0.64** 1	0.60 2
CtxSyn [136]	29.2	0.50 4	0.57 5	0.55 5	0.55 2	0.71 2	0.59 5	1.42 112	0.64 4	2.08 127	0.87 5	0.82 5	1.18 12	0.95 15	0.90 14	1.13 16	0.94 12	0.92 9	1.02 14	0.68 22	1.00 12	0.83 119	0.67 12	0.89 12	0.68 156
PMMST [114]	32.5	0.59 13	0.73 21	0.64 12	0.64 25	0.85 22	0.59 5	0.99 39	1.69 65	1.05 56	0.97 44	1.14 90	1.23 52	0.99 22	0.96 21	1.14 18	1.03 25	1.29 26	1.04 23	0.71 36	1.33 42	0.67 20	0.77 21	1.10 22	0.64 61
SuperSlomo [132]	33.5	0.59 13	0.69 14	0.64 12	0.72 82	0.91 48	0.75 138	0.74 10	1.01 16	0.71 13	0.98 57	0.95 15	1.23 52	0.94 14	0.90 14	1.12 14	0.96 15	0.98 10	1.04 23	0.60 7	0.90 5	0.71 47	0.69 14	0.93 14	0.68 156
MS-PFT [164]	34.7	0.56 8	0.67 12	0.61 7	0.69 63	0.90 44	0.70 116	0.82 14	0.92 14	0.89 36	0.95 24	1.00 21	1.18 12	0.91 11	0.88 12	1.10 13	0.95 13	1.00 12	1.03 16	0.72 41	1.05 14	0.92 136	0.72 16	0.98 16	0.70 162
NN-field [71]	36.7	0.59 13	0.77 32	0.64 12	0.59 4	0.77 7	0.58 2	1.09 59	1.77 75	1.16 70	1.00 80	1.18 105	1.26 86	0.98 18	0.95 18	1.14 18	1.08 44	1.46 51	1.05 37	0.68 22	1.26 31	0.70 35	0.78 24	1.12 25	0.63 12
OFRI [158]	37.7	0.60 25	0.57 5	0.69 141	0.67 49	0.81 16	0.79 148	0.70 6	0.59 3	0.83 30	0.87 5	0.81 4	1.15 6	0.86 5	0.81 4	1.03 8	0.92 10	0.99 11	0.98 10	0.79 73	0.90 5	1.18 155	0.67 12	0.84 7	0.78 167
MDP-Flow2 [68]	37.8	0.59 13	0.72 18	0.63 10	0.62 12	0.85 22	0.58 2	1.24 91	2.52 125	1.61 103	0.94 15	1.05 42	1.24 62	0.98 18	0.96 21	1.16 49	1.09 52	1.49 57	1.05 37	0.70 33	1.32 40	0.68 25	0.78 24	1.12 25	0.63 12
SepConv-v1 [127]	41.9	0.54 6	0.81 61	0.54 4	0.67 49	0.91 48	0.63 66	1.07 56	1.18 20	1.07 60	1.03 102	1.04 37	1.28 111	0.99 22	0.96 21	1.14 18	0.96 15	1.01 14	1.02 14	0.68 22	1.15 23	0.76 84	0.70 15	0.96 15	0.66 122
TOF-M [154]	45.9	0.55 7	0.66 11	0.60 6	0.72 82	0.95 61	0.71 126	0.92 29	0.92 14	1.08 61	0.96 31	0.90 10	1.24 62	0.97 17	0.94 17	1.14 18	0.98 18	1.06 16	1.05 37	0.72 41	1.18 26	0.83 119	0.90 66	1.30 64	0.70 162
CombBMOF [113]	46.1	0.62 59	0.80 52	0.65 58	0.63 16	0.86 25	0.59 5	1.11 62	2.06 99	1.19 75	1.00 80	1.14 90	1.28 111	1.02 34	1.02 38	1.15 29	1.08 44	1.43 44	1.04 23	0.71 36	1.33 42	0.70 35	0.75 18	1.06 19	0.63 12
NNF-Local [87]	47.1	0.58 10	0.71 16	0.63 10	0.59 4	0.76 5	0.58 2	1.21 80	2.31 113	1.51 99	0.98 57	1.13 81	1.25 72	0.98 18	0.95 18	1.14 18	1.13 77	1.61 86	1.07 57	0.75 58	1.45 65	0.88 129	0.77 21	1.10 22	0.63 12
ALD-Flow [66]	47.3	0.62 59	0.81 61	0.66 97	0.70 68	0.99 73	0.62 48	0.87 19	1.28 26	0.65 7	0.94 15	1.01 24	1.21 24	1.09 74	1.12 74	1.54 101	1.03 25	1.24 23	1.07 57	0.64 13	1.12 18	0.65 10	0.97 103	1.44 104	0.63 12

Table 2: Screenshot of our NE-ranking in the Middlebury benchmark (taken on the 7th of November, currently private).

	Avg. Rank		Mequon		Schefflera		Urban		Teddy		Backyard		Basketball		Dumptruck		Evergreen	
	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓	IE ↓	NE ↓
SepConv - $\mathcal{L}_1$ [6]	38.4	41.9	2.52	0.54	3.56	0.67	4.17	1.07	5.41	1.03	10.20	0.99	5.47	0.96	6.88	0.68	6.63	0.70
ToFlow [8]	30.5	45.9	2.54	0.55	3.70	0.72	3.43	0.92	5.05	0.96	9.84	0.97	5.34	0.98	6.88	0.72	7.14	0.90
Super SloMo [3]	30.7	33.5	2.51	0.59	3.66	0.72	2.91	0.74	5.05	0.98	9.56	0.94	5.37	0.96	6.69	0.60	6.73	0.69
CyclicGen [4]	25.0	28.8	2.26	0.64	3.19	0.67	2.76	0.72	4.97	0.95	8.00	0.91	3.36	0.87	4.55	0.53	4.48	0.52
CtxSyn - $\mathcal{L}_{Lap}$ [5]	20.9	29.2	2.24	0.50	2.96	0.55	4.32	1.42	4.21	0.87	9.59	0.95	5.22	0.94	7.02	0.68	6.66	0.67
DAIN [2]	24.4	10.8	2.38	0.58	3.28	0.60	3.32	0.69	4.65	0.86	7.88	0.87	4.73	0.85	6.36	0.59	6.25	0.66
Ours - $\mathcal{L}_{Lap}$	2.5	2.0	2.06	0.53	2.80	0.52	1.99	0.52	3.84	0.80	8.10	0.85	4.10	0.81	5.49	0.56	5.40	0.57

Table 3: Quantitative results in the interpolation category of the Middlebury benchmark for optical flow.

	Boxing 2K PSNR ↑	Boxing “4K” PSNR ↑	Crosswalk 2K PSNR ↑	Crosswalk “4K” PSNR ↑	Driving 2K PSNR ↑	Driving “4K” PSNR ↑	Market-1 2K PSNR ↑	Market-1 “4K” PSNR ↑	Market-2 2K PSNR ↑	Market-2 “4K” PSNR ↑	RitualDance 2K PSNR ↑	RitualDance “4K” PSNR ↑	Square 2K PSNR ↑	Square “4K” PSNR ↑	Tango 2K PSNR ↑	Tango “4K” PSNR ↑
SepConv - $\mathcal{L}_1$ [6]	36.75	33.35	36.34	32.79	34.95	33.32	32.03	31.21	36.49	34.62	28.15	23.12	38.63	36.16	34.85	31.90
SepConv - $\mathcal{L}_F$ [6]	36.54	33.15	35.69	32.10	34.84	33.24	31.69	30.68	36.26	34.27	27.95	23.06	38.41	35.84	34.43	31.08
ToFlow [8]	36.75	33.52	33.54	31.42	34.84	33.38	30.87	29.41	34.24	30.30	28.11	22.61	38.87	36.59	34.24	28.67
CyclicGen [4]	36.51	32.95	33.73	31.37	34.74	33.46	30.02	28.69	29.89	27.91	28.21	22.95	37.44	35.13	33.45	29.60
CtxSyn - $\mathcal{L}_{Lap}$ [5]	37.41	33.55	38.14	34.03	34.92	32.71	32.93	31.97	38.33	37.18	28.47	23.08	39.35	37.09	36.16	34.21
CtxSyn - $\mathcal{L}_F$ [5]	36.68	32.88	37.40	33.01	34.56	32.45	32.20	31.10	37.94	36.62	28.24	23.10	38.87	36.61	35.36	33.10
DAIN [2]	37.74	34.75	38.81	35.90	35.14	33.60	33.06	31.99	38.03	36.49	29.16	23.91	39.50	37.00	36.14	34.28
Ours - $\mathcal{L}_{Lap}$	38.48	35.40	38.71	33.69	35.68	33.82	33.34	32.16	39.61	37.86	29.20	23.67	41.23	38.00	36.73	34.16
Ours - $\mathcal{L}_F$	37.45	34.08	37.15	32.36	35.15	33.15	32.16	30.95	39.06	36.64	28.72	23.29	40.65	37.04	35.57	32.51

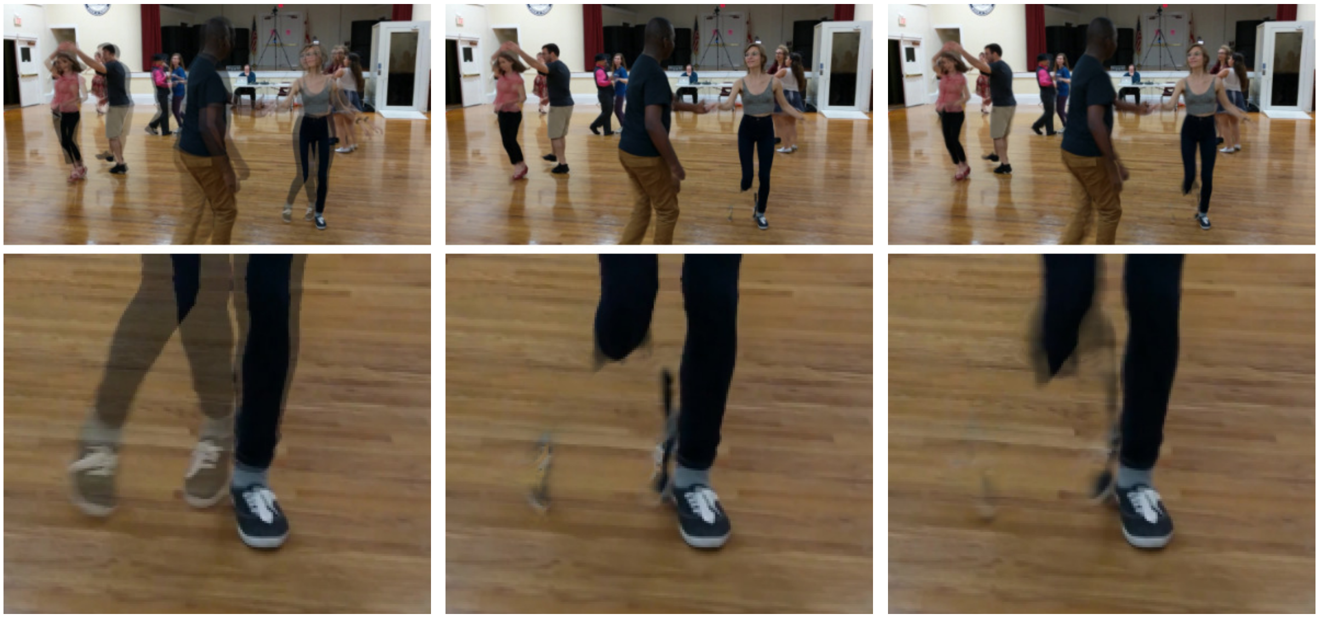
Table 4: Quantitative comparison in terms of PSNR on the eight 4K clips from Xiph with the most inter-frame motion.

	Boxing 2K SSIM ↑	Boxing “4K” SSIM ↑	Crosswalk 2K SSIM ↑	Crosswalk “4K” SSIM ↑	Driving 2K SSIM ↑	Driving “4K” SSIM ↑	Market-1 2K SSIM ↑	Market-1 “4K” SSIM ↑	Market-2 2K SSIM ↑	Market-2 “4K” SSIM ↑	RitualDance 2K SSIM ↑	RitualDance “4K” SSIM ↑	Square 2K SSIM ↑	Square “4K” SSIM ↑	Tango 2K SSIM ↑	Tango “4K” SSIM ↑
SepConv - $\mathcal{L}_1$ [6]	0.946	0.903	0.937	0.890	0.939	0.916	0.823	0.752	0.963	0.925	0.926	0.858	0.978	0.952	0.919	0.844
SepConv - $\mathcal{L}_F$ [6]	0.940	0.886	0.926	0.871	0.936	0.912	0.805	0.718	0.960	0.914	0.917	0.846	0.976	0.947	0.906	0.812
ToFlow [8]	0.947	0.900	0.925	0.874	0.935	0.917	0.805	0.703	0.947	0.834	0.928	0.852	0.978	0.953	0.915	0.816
CyclicGen [4]	0.946	0.894	0.924	0.867	0.930	0.916	0.764	0.658	0.847	0.749	0.923	0.850	0.972	0.938	0.904	0.815
CtxSyn - $\mathcal{L}_{Lap}$ [5]	0.952	0.908	0.944	0.898	0.940	0.917	0.841	0.771	0.972	0.944	0.932	0.866	0.980	0.956	0.928	0.863
CtxSyn - $\mathcal{L}_F$ [5]	0.934	0.862	0.926	0.858	0.934	0.905	0.804	0.715	0.968	0.931	0.919	0.842	0.977	0.947	0.904	0.796
DAIN [2]	0.956	0.915	0.948	0.905	0.942	0.921	0.846	0.770	0.972	0.945	0.939	0.879	0.981	0.958	0.932	0.866
Ours - $\mathcal{L}_{Lap}$	0.958	0.922	0.951	0.905	0.947	0.924	0.853	0.782	0.975	0.949	0.944	0.888	0.984	0.962	0.936	0.872
Ours - $\mathcal{L}_F$	0.936	0.874	0.921	0.850	0.936	0.907	0.797	0.706	0.970	0.925	0.923	0.848	0.980	0.946	0.902	0.793

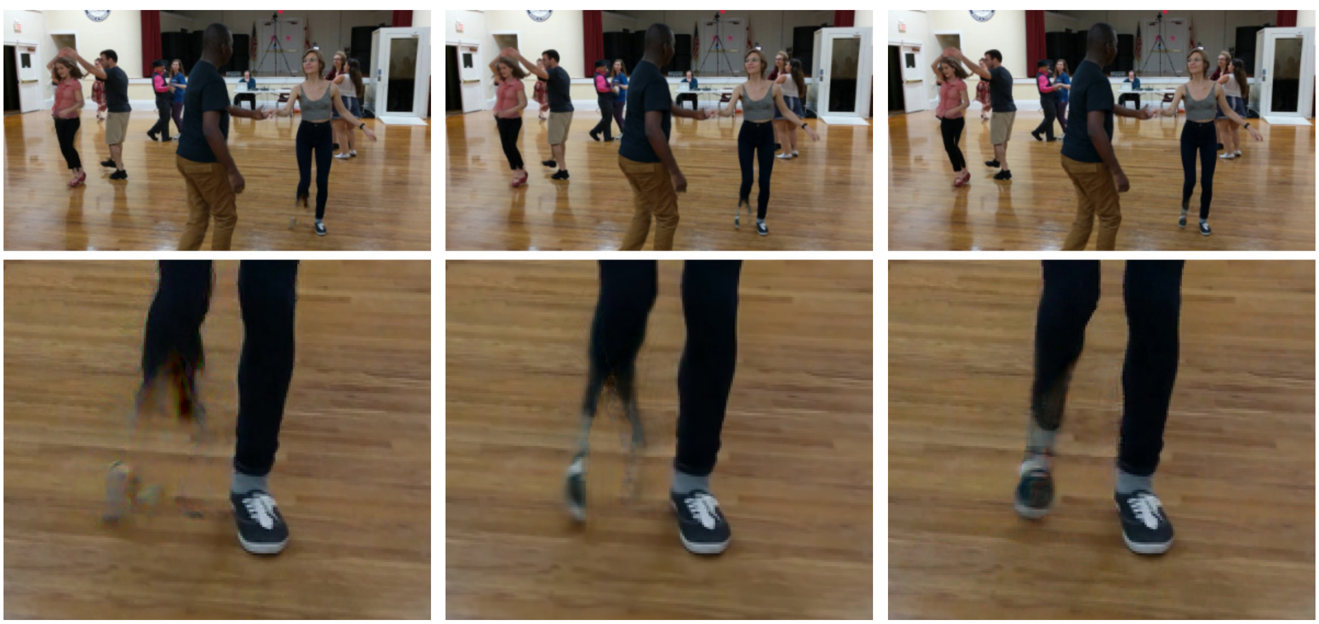
Table 5: Quantitative comparison in terms of SSIM on the eight 4K clips from Xiph with the most inter-frame motion.

	Boxing 2K LPIPS ↓	Boxing “4K” LPIPS ↓	Crosswalk 2K LPIPS ↓	Crosswalk “4K” LPIPS ↓	Driving 2K LPIPS ↓	Driving “4K” LPIPS ↓	Market-1 2K LPIPS ↓	Market-1 “4K” LPIPS ↓	Market-2 2K LPIPS ↓	Market-2 “4K” LPIPS ↓	RitualDance 2K LPIPS ↓	RitualDance “4K” LPIPS ↓	Square 2K LPIPS ↓	Square “4K” LPIPS ↓	Tango 2K LPIPS ↓	Tango “4K” LPIPS ↓
SepConv - $\mathcal{L}_1$ [6]	0.037	0.133	0.094	0.208	0.038	0.061	0.156	0.331	0.026	0.106	0.100	0.239	0.012	0.040	0.071	0.234
SepConv - $\mathcal{L}_F$ [6]	0.021	0.057	0.046	0.111	0.037	0.055	0.096	0.187	0.017	0.062	0.063	0.171	0.010	0.027	0.037	0.109
ToFlow [8]	0.035	0.088	0.088	0.152	0.047	0.062	0.139	0.238	0.027	0.118	0.075	0.189	0.013	0.038	0.063	0.171
CyclicGen [4]	0.039	0.090	0.085	0.142	0.068	0.067	0.185	0.271	0.108	0.169	0.077	0.177	0.029	0.060	0.073	0.159
CtxSyn - $\mathcal{L}_{Lap}$ [5]	0.053	0.172	0.103	0.215	0.040	0.074	0.185	0.354	0.017	0.073	0.086	0.223	0.012	0.045	0.085	0.244
CtxSyn - $\mathcal{L}_F$ [5]	0.021	0.056	0.031	0.082	0.038	0.058	0.078	0.154	0.013	0.036	0.061	0.174	0.009	0.022	0.031	0.066
DAIN [2]	0.063	0.161	0.139	0.214	0.045	0.073	0.200	0.329	0.020	0.088	0.091	0.200	0.013	0.051	0.105	0.243
Ours - $\mathcal{L}_{Lap}$	0.078	0.226	0.177	0.322	0.053	0.095	0.273	0.455	0.022	0.118	0.105	0.238	0.013	0.067	0.131	0.349
Ours - $\mathcal{L}_F$	0.016	0.043	0.026	0.079	0.033	0.053	0.074	0.148	0.010	0.034	0.045	0.136	0.006	0.018	0.024	0.059

Table 6: Quantitative comparison in terms of LPIPS on the eight 4K clips from Xiph with the most inter-frame motion.



Overlayed Input | ToFlow [8] | CyclicGen [4]



CtxSyn - $\mathcal{L}_F$ [5] | DAIN [2] | Ours - $\mathcal{L}_F$

Figure 1: Qualitative comparison, please also consider our supplementary video demo to see this example in motion.

Overlayed Input | ToFlow [8] | CyclicGen [4]

CtxSyn - $\mathcal{L}_F$ [5] | DAIN [2] | Ours - $\mathcal{L}_F$

Figure 2: Qualitative comparison, please also consider our supplementary video demo to see this example in motion.

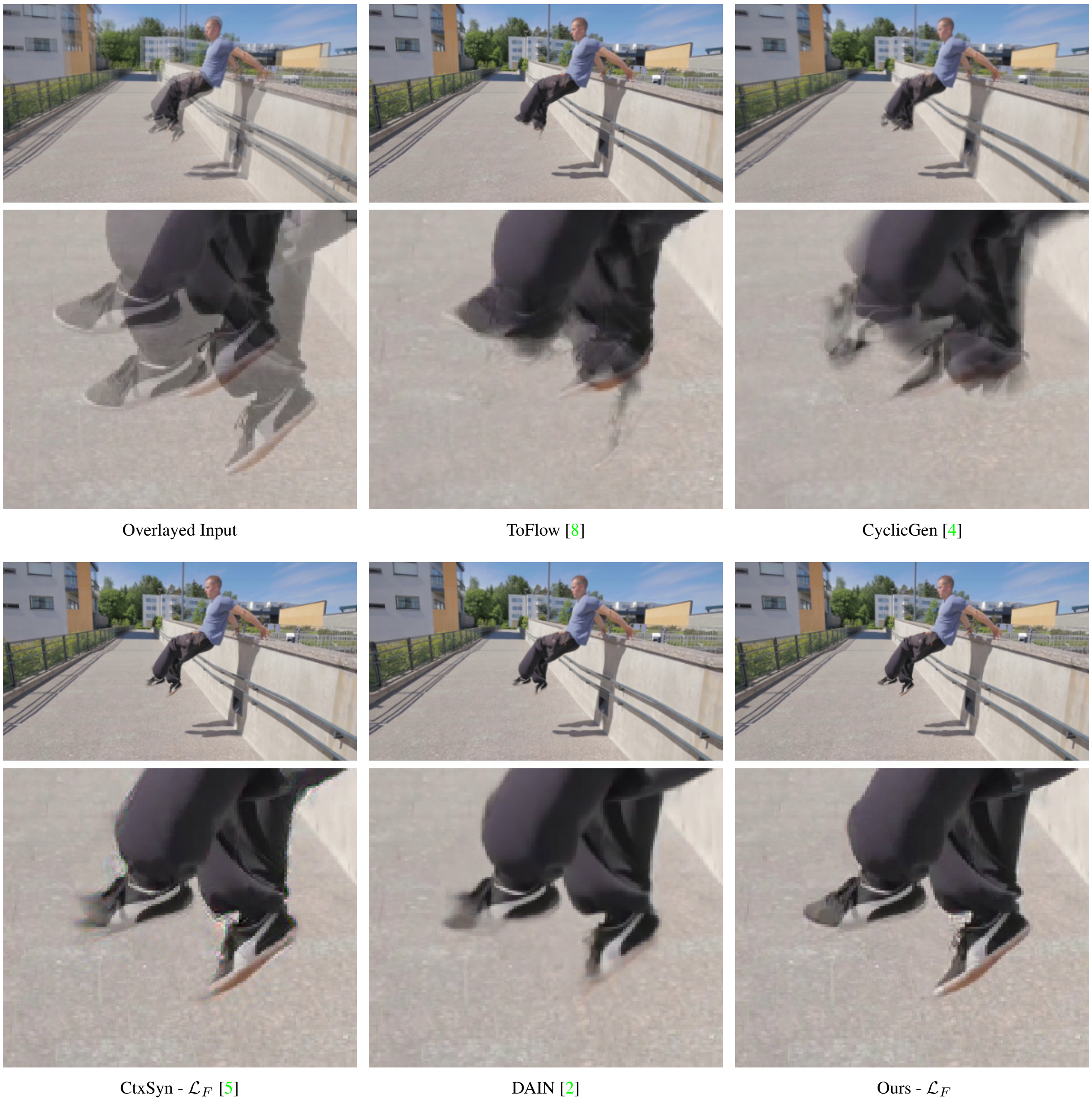


Figure 3: Qualitative comparison, please also consider our supplementary video demo to see this example in motion.

References

[1] Simon Baker, Daniel Scharstein, J. P. Lewis, Stefan Roth, Michael J. Black, and Richard Szeliski. A Database and Evaluation Methodology for Optical Flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. 1

[2] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-Aware Video Frame Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2, 3, 4, 5, 6

[3] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik G. Learned-Miller, and Jan Kautz. Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2

[4] Yu-Lun Liu, Yi-Tung Liao, Yen-Yu Lin, and Yung-Yu Chuang. Deep Video Frame Interpolation Using Cyclic Frame Generation. In *AAAI Conference on Artificial Intelligence*, 2019. 2, 3, 4, 5, 6

[5] Simon Niklaus and Feng Liu. Context-Aware Synthesis for Video Frame Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2, 3, 4, 5, 6

[6] Simon Niklaus, Long Mai, and Feng Liu. Video Frame Interpolation via Adaptive Separable Convolution. In *IEEE International Conference on Computer Vision*, 2017. 2, 3

[7] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus H. Gross, and Alexander Sorkine-Hornung. A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 1

[8] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T. Freeman. Video Enhancement With Task-Oriented Flow. *International Journal of Computer Vision*, 127(8):1106–1125, 2019. 2, 3, 4, 5, 6