IFRNet: Intermediate Feature Refine Network for Efficient Frame Interpolation Supplementary Material

Lingtong Kong[1*], Boyuan Jiang[2*], Donghao Luo[2], Wenqing Chu[2], Xiaoming Huang[2],
Ying Tai[2], Chengjie Wang[2], Jie Yang[1†]
[1]Shanghai Jiao Tong University, China, [2]Youtu Lab, Tencent
{ltkong, jieyang}@sjtu.edu.cn
{byronjiang, michaelluo, wenqingchu, skyhuang, yingtai, jasoncjwang}@tencent.com

Figure 10. **Qualitative results of IFRNet for 8× interpolation on GoPro [9] and Adobe240 [13] test datasets.** Please watch the video with Adobe Reader. Each video has 9 frames where the first and the last frames are input, and the middle 7 frames are predicted by IFRNet.

In the supplementary, we first present multi-frame interpolation experiments of IFRNet. Second, qualitative video comparisions with other advanced VFI approaches are displayed. Third, we depict structure details of IFRNet and its variants. Fourth, we provide more visual examples and analysis of middle components for better understanding the workflow of IFRNet. Finally, we show the screenshot of VFI results on the Middlebury benchmark. Please note that the numbering within this supplementary has manually been adjusted to continue the ones in our main paper.

6. Multi-Frame Interpolation

Different from other multi-frame interpolation methods which scales optical flow [1, 5] or interpolates middle frames recursively [2, 7], IFRNet can predict multiple intermediate frames by proposed one-channel temporal encoding mask T, which is one of the input of the coarsest decoder $\mathcal{D}^4$. The temporal encoding is a conditional input signal whose values are all the same and set to t, where $t \in \{1/8, 2/8, \ldots, 7/8\}$ in 8× interpolation setting. Also, proposed task-oriented flow distillation loss and feature space geometry consistency loss still work for any intermediate time instance t. To evaluate IFRNet for 8× interpolation, we use the train/test split of FLAVR [6], where we train IFRNet on GoPro [9] training set with the same learning schedule and loss functions as our main paper. Then we test the pre-trained model on GoPro testing and Adobe240 [13] datasets whose results are listed in Table 5.

Method	GoPro [9]		Adobe240 [13]		Time
	PSNR	SSIM	PSNR	SSIM	(s)
DVF [8]	21.94	0.776	28.23	0.896	0.87
SuperSloMo [5]	28.52	0.891	30.66	0.931	0.44
DAIN [1]	29.00	0.910	29.50	0.910	4.10
IFRNet (Ours)	**29.84**	**0.920**	**31.93**	**0.943**	**0.16**

Table 5. **Quantitative comparison for 8× interpolation.**

IFRNet outperforms all of the other SOTA methods

∗ Equal contribution. This work was done when Lingtong Kong was an intern at Tencent Youtu Lab. Code is available at https://github.com/ltkong218/IFRNet.

† Corresponding author: Jie Yang (jieyang@sjtu.edu.cn). This research is partly supported by NSFC, China (No: 61876107, U1803261).

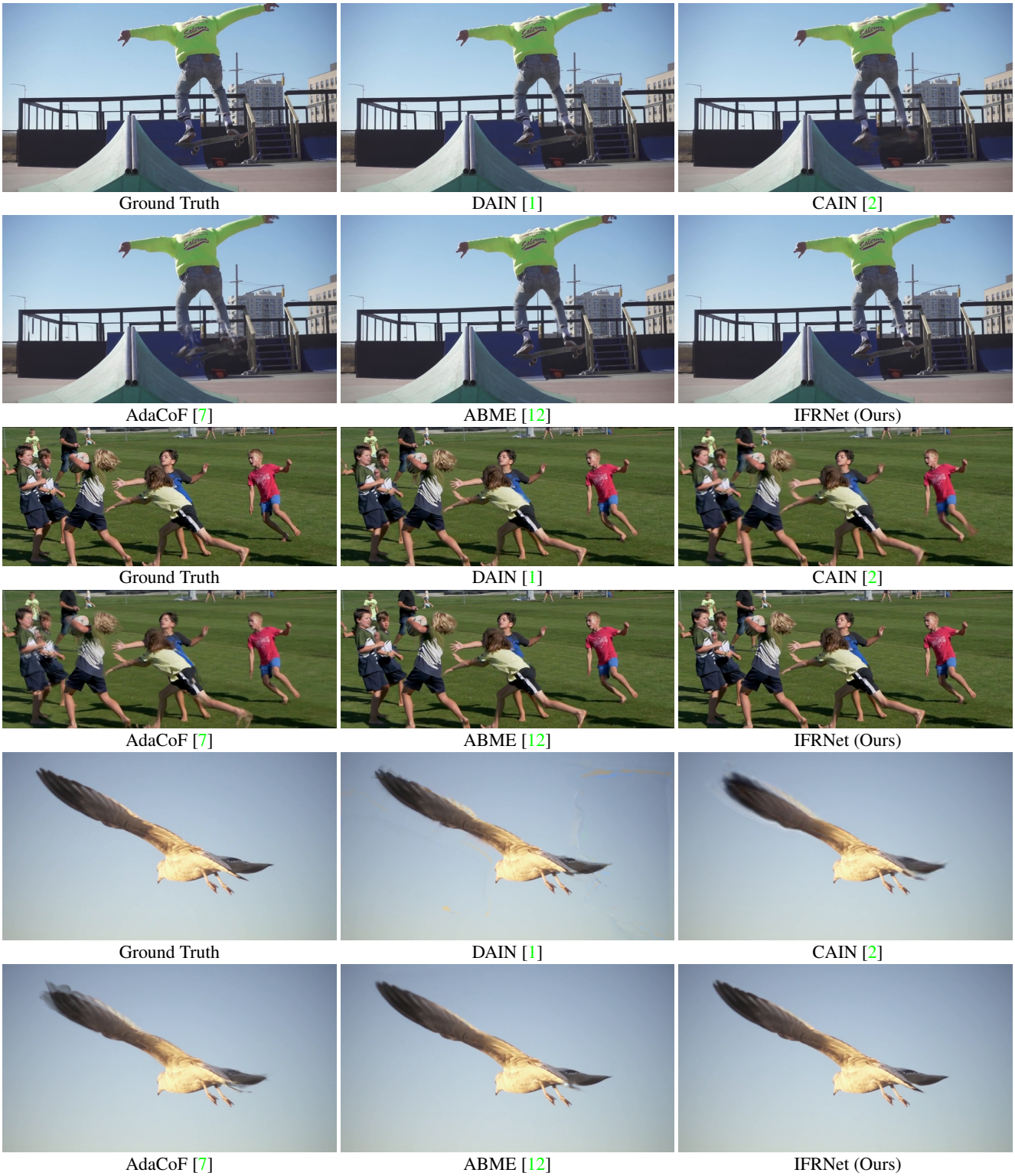


Figure 11. **Video comparison on SNU-FILM [2] dataset**. Please watch the video with Adobe Reader and zoom in for best view.

with 2 input frames on both GoPro and Adobe240 datasets in both PSNR and SSIM metrics. For example, IFRNet achieves **0.84** dB better results than DAIN [1] on GoPro and exceeds SuperSloMo [5] by **1.27** dB on Adobe240. Thanks to the modularity character of IFRNet, the encoder only needs a single forward pass, while the decoders infer 7

times with different temporal embedding to convert videos from 30 fps into 240 fps. Therefore, the speed advantage of IFRNet is still or even more obvious than other approaches. Figure 10 gives some qualitative results of IFRNet for 8× interpolation, demonstrating its superior ability for frame rate up-conversion and slow motion generation.

7. Video Comparison

In this part, we qualitatively compare interpolated videos by proposed IFRNet against other open source VFI methods on SNU-FILM [2] dataset, whose results are shown in Figure 11. As can be seen, our approach can generate motion boundary and texture details faithfully thanks to the powerfulness of gradually refined intermediate feature.

8. Network Architecture

In this section, we present the structure details of five sub-networks of IFRNet, *i.e.*, pyramid encoder $\mathcal{E}$ and coarse-to-fine decoders $\mathcal{D}^4, \mathcal{D}^3, \mathcal{D}^2, \mathcal{D}^1$. In each following figure, arguments of 'Conv' and 'Deconv' from left to right are input channels, output channels, kernel size, stride and padding, respectively. Dimensions of input and output tensors from left to right stand for feature channels, height and width, separately. A PReLU [4] follows each 'Conv' layer, while there is no activation after each 'Deconv' layer. In practice, the intermediate flow fields are estimated in a residual manner, which is not reflected in the figures to emphasize the primary network structure. We take input frames with spatial size of 640×480 as example.

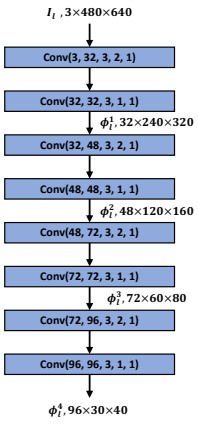


Figure 12. **Details of the pyramid encoder $\mathcal{E}$.** The two input frames $I_l, l \in \{0, 1\}$ are encoded by the same Siamese network.

As for IFRNet large and IFRNet small, feature channels from the first to the fourth pyramid levels are set to 64, 96, 144, 192 and 24, 36, 54, 72, respectively. Correspondingly, channel numbers in multiple decoders are adjusted. Also,

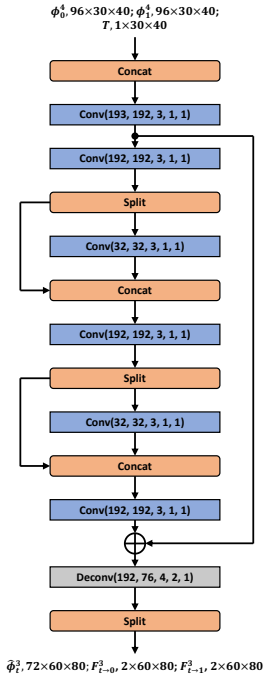


Figure 13. **Details of the bottom decoder $\mathcal{D}^4$.**

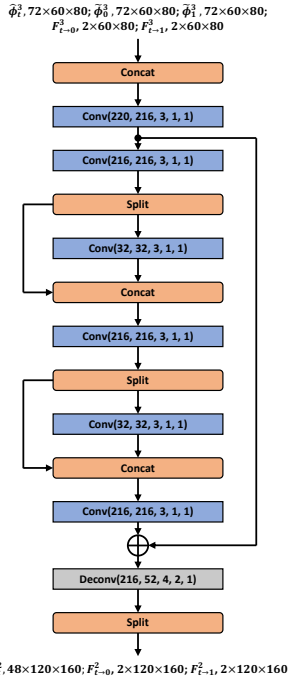


Figure 14. **Details of the middle decoder $\mathcal{D}^3$.**

feature channels of the third and the fifth convolution layers in coarse-to-fine decoders of IFRNet large and IFRNet small are set to 64 and 24, separately.

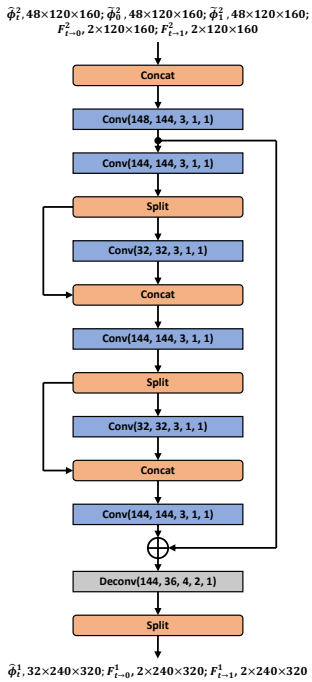


Figure 15. **Details of the middle decoder $\mathcal{D}^2$.**

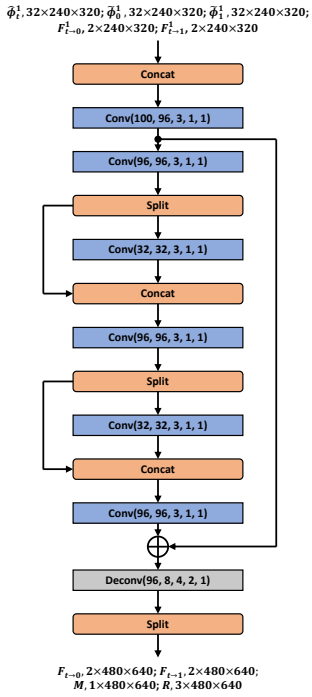


Figure 16. **Details of the top decoder $\mathcal{D}^1$.**

9. Visualization and Discussion

Figure 17 presents some visual examples to show the robustness masks in proposed task-oriented flow distillation loss, which can decrease the adverse impacts while focusing on the useful knowledge for better frame interpolation. It seems that intermediate flow prediction of IFRNet behaves smoother and contains less artifacts than flow prediction of pseudo label, that helps to achieve better VFI accuracy.

Figure 17. **Illustration of task-oriented flow distillation.** From top to bottom rows are ground truth frame I_t^{gt}, pseudo label of intermediate flow fields $F_{t\to0}^p$, $F_{t\to1}^p$, predicted intermediate flow fields $F_{t\to0}$, $F_{t\to1}$, task-oriented robustness masks P_0, P_1. Darker color in P_0, P_1 approaches to 1, while brighter color tends to 0. Each column represents a separate example on Vimeo90K [15] dataset. Zoom in for best view.

Figure 18. **Illustration of mean feature map of intermediate feature $\hat{\phi}_t^1$ w/o and w/ $\mathcal{L}_g$.** From top to bottom rows are ground truth frame I_t^{gt}, mean feature map of $\hat{\phi}_t^1$ w/o $\mathcal{L}_g$, mean feature map of $\hat{\phi}_t^1$ w/ $\mathcal{L}_g$. Each column represents a separate example on Vimeo90K [15] dataset. Zoom in for best view.

Figure 18 depicts more visual results of mean feature maps of intermediate feature w/o and w/ proposed geometry consistency loss, demonstrating its effect on regularizing refined intermediate feature to keep better structure layout.

Figure 19 gives visual understanding of frame interpola-

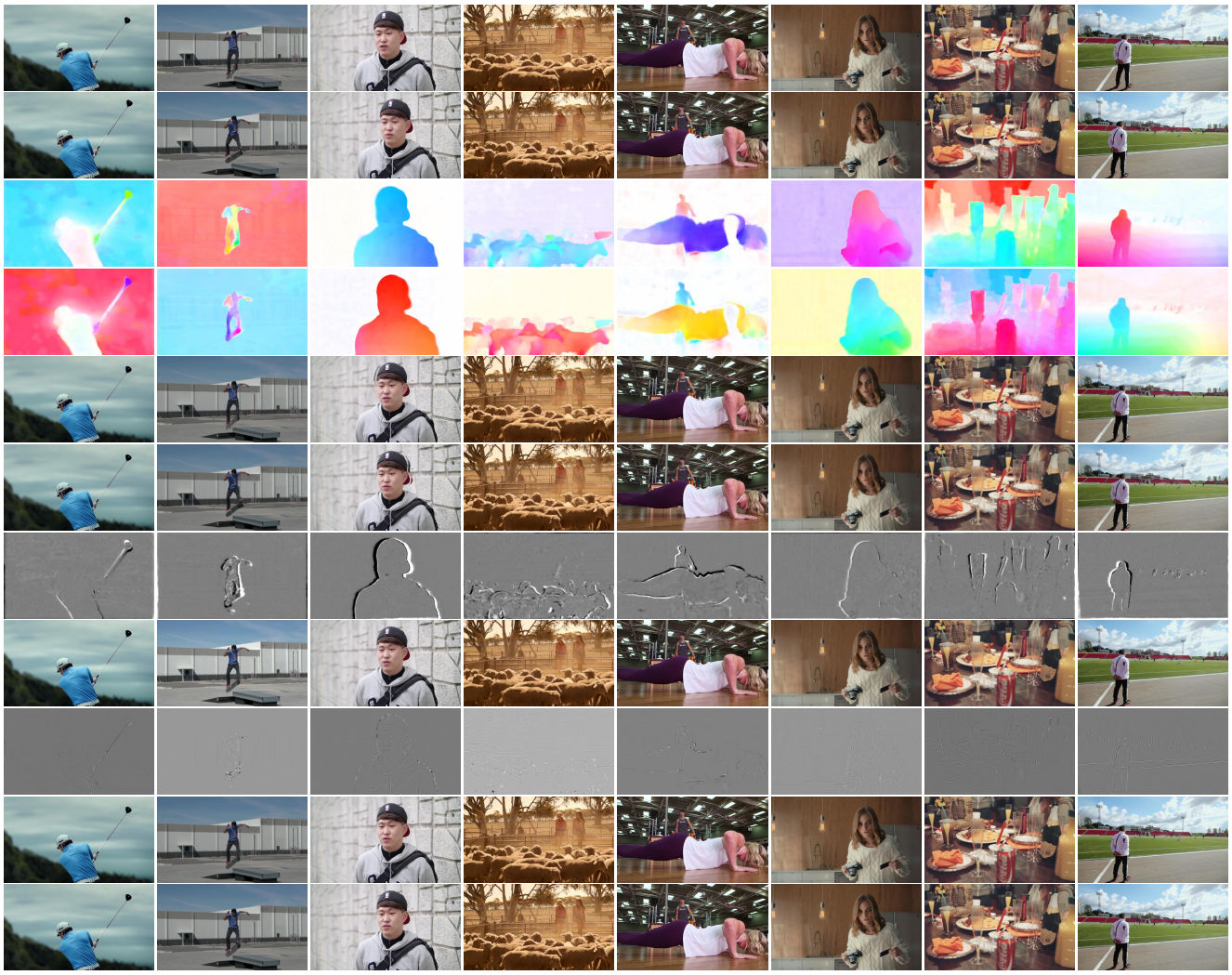
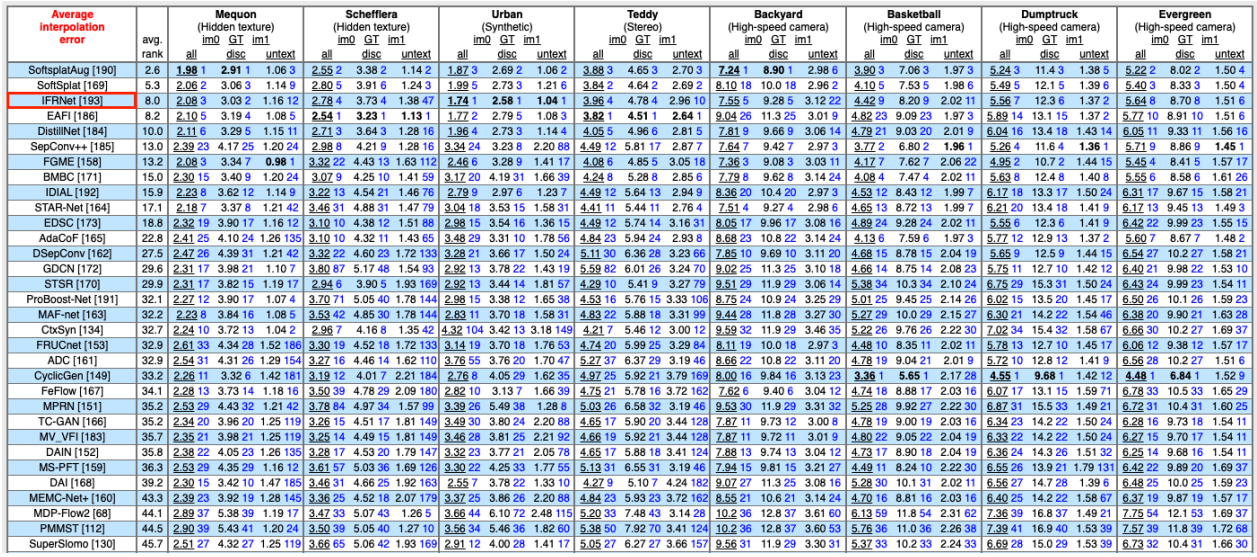


Figure 19. **Illustration of intermediate components of IFRNet.** From top to bottom rows are input frames I_0, I_1, predicted intermediate flow fields $F_{t\to0}, F_{t\to1}$, warped input frames $\tilde{I}_0, \tilde{I}_1$, merge mask M, merged frame $\hat{I}_t'$, residual R, final prediction $\hat{I}_t$ and ground truth I_t^{gt}, where merged frame is calculated by $\hat{I}_t' = M \odot \tilde{I}_0 + (1 - M) \odot \tilde{I}_1$. For better visualization of residual R, we multiply it by 10 and add a bias of 0.5. Each column represents a separate example on Vimeo90K [15] dataset. Zoom in for best view.

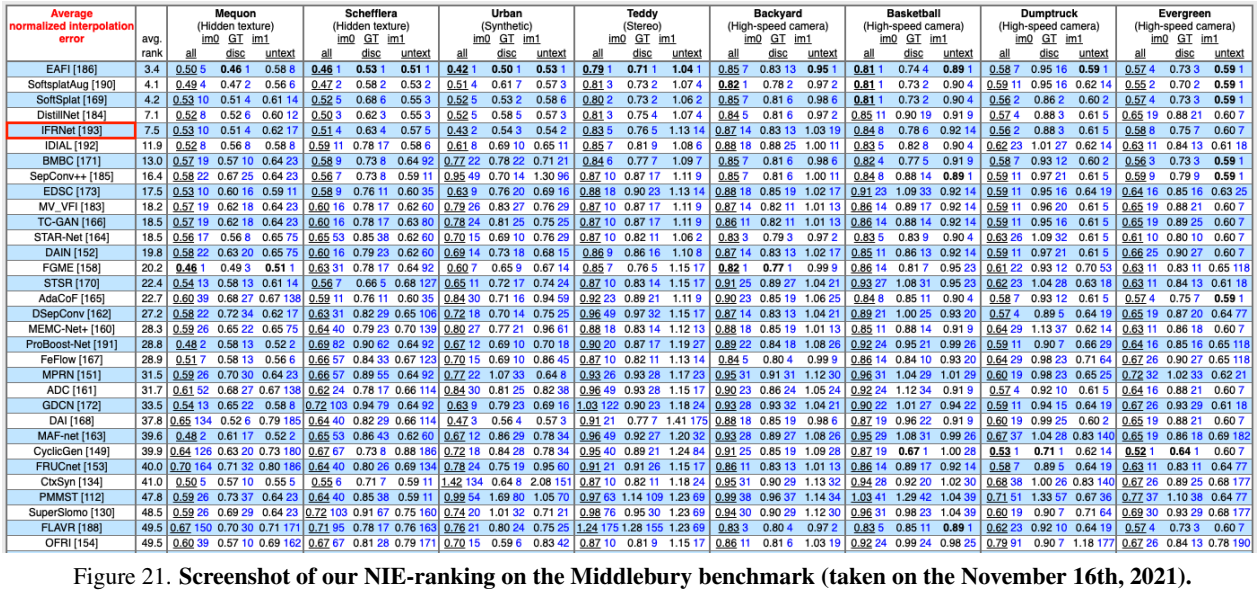
tion process of IFRNet. Thanks to the reference anchor information offered by intermediate feature together with effective supervision provided by geometry consistency loss and task-oriented flow distillation loss, IFRNet can estimate relatively good intermediate flow with clear motion boundary. Further, we can see that merge mask M can identify occluded regions of warped frames by adjusting the mixing weight, where it tends to average the candidate regions when both views are visible. Finally, residual R can compensate for some contextual details, which usually response at motion boundary and image edges. Different from other flow-based VFI methods that take cascaded structure design, merge mask M and residual R in IFRNet share the same encoder-decoder with intermediate optical flow, making proposed architecture achieve better VFI accuracy while being more lightweight and fast.

Readers may think our IFRNet is similar with PWC-Net [14] which is designed for optical flow. However, It is non-trivial to adapt PWC-Net for frame interpolation, since previous related works employ it as one of many components. We summarize their difference in several aspects: **1)** Anchor feature in PWC-Net is extracted by the encoder, while in IFRNet, it is reconstructed by the decoder. **2)** Besides motion information in intermediate feature, there are occlusion, texture and temporal information in it. **3)** PWC-Net designed for motion estimation, is optimized only by flow regression loss with strong augmentation. However, IFRNet designed for frame synthesizing, is optimized in a multi-target manner with weak data augmentation.



Average interpolation error	avg. rank	Mequon (Hidden texture) im0 GT im1 all	Mequon disc	Mequon untext	Schefflera (Hidden texture) all	Schefflera disc	Schefflera untext	Urban (Synthetic) all	Urban disc	Urban untext	Teddy (Stereo) all	Teddy disc	Teddy untext	Backyard (High-speed camera) all	Backyard disc	Backyard untext	Basketball (High-speed camera) all	Basketball disc	Basketball untext	Dumptruck (High-speed camera) all	Dumptruck disc	Dumptruck untext	Evergreen (High-speed camera) all	Evergreen disc	Evergreen untext
SoftsplatAug [190]	2.6	**1.98** 1	**2.91** 1	1.06 3	2.55 2	3.38 2	1.14 2	1.87 3	2.69 2	1.06 2	3.88 3	4.65 3	2.70 3	**7.24** 1	**8.90** 1	2.98 6	3.90 3	7.06 3	1.97 3	5.24 3	11.4 3	1.38 5	5.22 2	8.02 2	1.50 4
SoftSplat [169]	5.3	2.06 2	3.06 3	1.14 9	2.80 5	3.91 6	1.24 3	1.99 5	2.73 3	1.21 6	3.84 2	4.64 2	2.69 2	8.10 18	10.0 18	2.96 2	4.10 5	7.53 5	1.98 6	5.49 5	12.1 5	1.39 6	5.40 3	8.33 3	1.50 4
IFRNet [193]	8.0	2.08 3	3.03 2	1.16 12	2.78 4	3.73 4	1.38 47	**1.74** 1	**2.58** 1	**1.04** 1	3.96 4	4.78 4	2.96 10	7.55 5	9.28 5	3.12 22	4.42 9	8.20 9	2.02 11	5.56 7	12.3 6	1.37 2	5.64 8	8.70 8	1.51 6
EAFI [186]	8.2	2.10 5	3.19 4	1.08 5	**2.54** 1	**3.23** 1	**1.13** 1	1.77 2	2.79 5	1.08 3	**3.82** 1	**4.51** 1	**2.64** 1	9.04 26	11.3 25	3.01 9	4.82 23	9.09 23	1.97 3	5.89 14	13.1 15	1.37 2	5.77 10	8.91 10	1.51 6
DistillNet [184]	10.0	2.11 6	3.29 5	1.15 11	2.71 3	3.64 3	1.28 16	1.96 4	2.73 3	1.14 4	4.05 5	4.96 6	2.81 5	7.81 9	9.66 9	3.06 14	4.79 21	9.03 20	2.01 9	6.04 16	13.4 18	1.43 14	6.05 11	9.33 11	1.56 16
SepConv++ [185]	13.0	2.39 23	4.17 25	1.20 24	2.98 8	4.21 9	1.28 16	3.34 24	3.23 8	2.20 88	4.49 12	5.81 17	2.87 7	7.64 7	9.42 7	2.97 3	3.77 2	6.80 2	**1.96** 1	5.26 4	11.6 4	**1.36** 1	5.71 9	8.86 9	**1.45** 1
FGME [158]	13.2	2.08 3	3.34 7	**0.98** 1	3.32 22	4.43 13	1.63 112	2.46 6	3.28 9	1.41 17	4.08 6	4.85 5	3.05 18	7.36 3	9.08 3	3.03 11	4.17 7	7.62 7	2.06 22	4.95 2	10.7 2	1.44 15	5.45 4	8.41 5	1.57 17
BMBC [171]	15.0	2.30 15	3.40 9	1.20 24	3.07 9	4.25 10	1.41 59	3.17 20	4.19 31	1.66 39	4.24 8	5.28 8	2.85 6	7.79 8	9.62 8	3.14 24	4.08 4	7.47 4	2.02 11	5.63 8	12.4 8	1.40 8	5.55 6	8.58 6	1.61 26
IDIAL [192]	15.9	2.23 8	3.62 12	1.14 9	3.22 13	4.54 21	1.46 76	2.79 9	2.97 6	1.23 7	4.49 12	5.64 13	2.94 9	8.36 20	10.4 20	2.97 3	4.53 12	8.43 12	1.99 7	6.17 18	13.3 17	1.50 24	6.31 17	9.67 15	1.58 21
STAR-Net [164]	17.1	2.18 7	3.37 8	1.21 42	3.46 31	4.88 31	1.47 79	3.04 18	3.53 15	1.58 31	4.41 11	5.44 11	2.76 4	7.51 4	9.27 4	2.98 6	4.65 13	8.72 13	1.99 7	6.21 20	13.4 18	1.41 9	6.17 13	9.45 13	1.49 3
EDSC [173]	18.8	2.32 19	3.90 17	1.16 12	3.10 10	4.38 12	1.51 88	2.98 15	3.54 16	1.36 15	4.49 12	5.74 14	3.16 31	8.05 17	9.96 17	3.08 16	4.89 24	9.28 24	2.02 11	5.55 6	12.3 6	1.41 9	6.42 22	9.99 23	1.55 15
AdaCoF [165]	22.8	2.41 25	4.10 24	1.26 135	3.10 10	4.32 11	1.43 65	3.48 29	3.31 10	1.78 56	4.84 23	5.94 24	2.93 8	8.68 23	10.8 22	3.14 24	4.13 6	7.59 6	1.97 3	5.77 12	12.9 13	1.37 2	5.60 7	8.67 7	1.48 2
DSepConv [162]	27.5	2.47 26	4.39 31	1.21 42	3.32 22	4.60 23	1.72 133	3.28 21	3.66 17	1.50 24	5.11 30	6.36 28	3.23 66	7.85 10	9.69 10	3.11 20	4.68 15	8.78 15	2.04 19	5.65 9	12.5 9	1.44 15	6.54 27	10.2 27	1.58 21
GDCN [172]	29.6	2.31 17	3.98 21	1.10 7	3.80 87	5.17 48	1.54 93	2.92 13	3.78 22	1.43 19	5.59 82	6.01 26	3.24 70	9.02 25	11.3 25	3.10 18	4.66 14	8.75 14	2.08 23	5.75 11	12.7 10	1.42 12	6.40 21	9.98 22	1.53 10
STSR [170]	29.9	2.31 17	3.82 15	1.19 17	2.94 6	3.90 5	1.93 169	2.92 13	3.44 14	1.81 57	4.29 10	5.41 9	3.27 79	9.51 29	11.9 29	3.06 14	5.38 34	10.3 34	2.10 24	6.75 29	15.3 31	1.50 24	6.43 24	9.99 23	1.54 11
ProBoost-Net [191]	32.1	2.27 12	3.90 17	1.07 4	3.70 71	5.05 40	1.78 144	2.98 15	3.38 12	1.65 38	4.53 16	5.76 15	3.33 106	8.75 24	10.9 24	3.25 29	5.01 25	9.45 25	2.14 26	6.02 15	13.5 20	1.45 17	6.50 26	10.1 26	1.59 23
MAF-net [163]	32.2	2.23 8	3.84 16	1.08 5	3.53 42	4.85 30	1.78 144	2.83 11	3.70 18	1.58 31	4.83 22	5.88 18	3.31 99	9.44 28	11.8 28	3.27 30	5.27 29	10.0 29	2.15 27	6.30 21	14.2 22	1.54 46	6.38 20	9.90 21	1.63 28
CtxSyn [134]	32.7	2.24 10	3.72 13	1.04 2	2.96 7	4.16 8	1.35 42	4.32 104	3.42 13	3.18 149	4.21 7	5.46 12	3.00 12	9.59 32	11.9 29	3.46 35	5.22 26	9.76 26	2.22 30	7.02 34	15.4 32	1.58 67	6.66 30	10.2 27	1.69 37
FRUCnet [153]	32.9	2.61 33	4.34 28	1.52 186	3.30 19	4.52 18	1.72 133	3.14 19	3.70 18	1.76 53	4.74 20	5.99 25	3.29 84	8.11 19	10.0 18	2.97 3	4.48 10	8.35 11	2.02 11	5.78 13	12.7 10	1.45 17	6.06 12	9.38 12	1.57 17
ADC [161]	32.9	2.54 31	4.31 26	1.29 154	3.27 16	4.46 14	1.62 110	3.76 55	3.76 20	1.70 47	5.27 37	6.37 29	3.19 46	8.66 22	10.8 22	3.11 20	4.78 19	9.04 21	2.01 9	5.72 10	12.8 12	1.41 9	6.56 28	10.2 27	1.51 6
CyclicGen [149]	33.2	2.26 11	3.32 6	1.42 181	3.19 12	4.01 7	2.21 184	2.76 8	4.05 29	1.62 35	4.97 25	5.92 21	3.79 169	8.00 16	9.84 16	3.13 23	**3.36** 1	**5.65** 1	2.17 28	**4.55** 1	**9.68** 1	1.42 12	**4.48** 1	**6.84** 1	1.52 9
FeFlow [167]	34.1	2.28 13	3.73 14	1.18 16	3.50 39	4.78 29	2.09 180	2.82 10	3.13 7	1.66 39	4.75 21	5.78 16	3.72 162	7.62 6	9.40 6	3.04 12	4.74 18	8.88 17	2.03 16	6.07 17	13.1 15	1.59 71	6.78 33	10.5 33	1.65 29
MPRN [151]	35.2	2.53 29	4.43 32	1.21 42	3.78 84	4.97 34	1.57 99	3.39 26	5.49 38	1.28 8	5.03 26	6.58 32	3.19 46	9.53 30	11.9 29	3.31 32	5.25 28	9.92 27	2.22 30	6.87 31	15.5 33	1.49 21	6.72 31	10.4 31	1.60 25
TC-GAN [166]	35.2	2.34 20	3.96 20	1.25 119	3.26 15	4.51 17	1.81 149	3.49 30	3.80 24	2.20 88	4.65 17	5.90 20	3.44 128	7.87 11	9.73 12	3.00 8	4.78 19	9.00 19	2.03 16	6.34 23	14.2 22	1.50 24	6.28 16	9.73 18	1.54 11
MV_VFI [183]	35.7	2.35 21	3.98 21	1.25 119	3.25 14	4.49 15	1.81 149	3.46 28	3.81 25	2.21 92	4.66 19	5.92 21	3.44 128	7.87 11	9.72 11	3.01 9	4.80 22	9.05 22	2.04 19	6.33 22	14.2 22	1.50 24	6.27 15	9.70 17	1.54 11
DAIN [152]	35.8	2.38 22	4.05 23	1.26 135	3.28 17	4.53 20	1.79 147	3.32 23	3.77 21	2.05 78	4.65 17	5.88 18	3.41 124	7.88 13	9.74 13	3.04 12	4.73 17	8.90 18	2.04 19	6.36 24	14.3 26	1.51 32	6.25 14	9.68 16	1.54 11
MS-PFT [159]	36.3	2.53 29	4.35 29	1.16 12	3.61 57	5.03 36	1.69 126	3.30 22	4.25 33	1.77 55	5.13 31	6.55 31	3.19 46	7.94 15	9.81 15	3.21 27	4.49 11	8.24 10	2.22 30	6.55 26	13.9 21	1.79 131	6.42 22	9.89 20	1.69 37
DAI [168]	39.2	2.30 15	3.42 10	1.47 185	3.46 31	4.66 25	1.92 163	2.55 7	3.78 22	1.33 10	4.27 9	5.10 7	4.24 182	9.07 27	11.3 25	3.08 16	5.28 30	10.1 31	2.02 11	6.56 27	14.7 28	1.39 6	6.48 25	10.0 25	1.59 23
MEMC-Net+ [160]	43.3	2.39 23	3.92 19	1.28 145	3.36 25	4.52 18	2.07 179	3.37 25	3.86 26	2.20 88	4.84 23	5.93 23	3.72 162	8.55 21	10.6 21	3.14 24	4.70 16	8.81 16	2.03 16	6.40 25	14.2 22	1.58 67	6.37 19	9.87 19	1.57 17
MDP-Flow2 [68]	44.1	2.89 37	5.38 39	1.19 17	3.47 33	5.07 43	1.26 5	3.66 44	6.10 72	2.48 115	5.20 33	7.48 43	3.14 28	10.2 36	12.8 37	3.61 60	6.13 59	11.8 54	2.31 62	7.36 39	16.8 37	1.49 21	7.75 54	12.1 53	1.69 37
PMMST [112]	44.5	2.90 39	5.43 41	1.20 24	3.50 39	5.05 40	1.27 10	3.56 34	5.46 36	1.82 60	5.38 50	7.92 70	3.41 124	10.2 36	12.8 37	3.60 53	5.76 36	11.0 36	2.26 38	7.39 41	16.9 40	1.53 39	7.57 39	11.8 39	1.72 68
SuperSlomo [130]	45.7	2.51 27	4.32 27	1.25 119	3.66 65	5.06 42	1.93 169	2.91 12	4.00 28	1.41 17	5.05 27	6.27 27	3.66 157	9.56 31	11.9 29	3.30 31	5.37 33	10.2 33	2.24 33	6.69 28	15.0 29	1.53 39	6.73 32	10.4 31	1.66 30

Figure 20. **Screenshot of our IE-ranking on the Middlebury benchmark (taken on the November 16th, 2021).**



Average normalized interpolation error	avg. rank	Mequon (Hidden texture) im0 GT im1 all	Mequon disc	Mequon untext	Schefflera (Hidden texture) all	Schefflera disc	Schefflera untext	Urban (Synthetic) all	Urban disc	Urban untext	Teddy (Stereo) all	Teddy disc	Teddy untext	Backyard (High-speed camera) all	Backyard disc	Backyard untext	Basketball (High-speed camera) all	Basketball disc	Basketball untext	Dumptruck (High-speed camera) all	Dumptruck disc	Dumptruck untext	Evergreen (High-speed camera) all	Evergreen disc	Evergreen untext
EAFI [186]	3.4	0.50 5	**0.46** 1	0.58 8	**0.46** 1	**0.53** 1	**0.51** 1	**0.42** 1	**0.50** 1	**0.53** 1	**0.79** 1	**0.71** 1	**1.04** 1	0.85 7	0.83 13	**0.95** 1	**0.81** 1	0.74 4	**0.89** 1	0.58 7	0.95 16	**0.59** 1	0.57 4	0.73 3	**0.59** 1
SoftsplatAug [190]	4.1	0.49 4	0.47 2	0.56 6	0.47 2	0.58 2	0.53 2	0.51 4	0.61 7	0.57 3	0.81 3	0.73 2	1.07 4	**0.82** 1	0.78 2	0.97 2	**0.81** 1	0.73 2	0.90 4	0.59 11	0.95 16	0.62 14	0.55 2	0.70 2	**0.59** 1
SoftSplat [169]	4.2	0.53 10	0.51 4	0.61 14	0.52 5	0.68 6	0.55 3	0.52 5	0.53 2	0.58 6	0.80 2	0.73 2	1.06 2	0.85 7	0.81 6	0.98 6	**0.81** 1	0.73 2	0.90 4	0.56 2	0.86 2	0.60 2	0.57 4	0.73 3	**0.59** 1
DistillNet [184]	7.1	0.52 8	0.52 6	0.60 12	0.50 3	0.62 3	0.55 3	0.52 5	0.58 5	0.57 3	0.81 3	0.75 4	1.07 4	0.84 5	0.81 6	0.97 2	0.85 11	0.90 19	0.91 9	0.57 4	0.88 3	0.61 5	0.65 19	0.88 21	0.60 7
IFRNet [193]	7.5	0.53 10	0.51 4	0.62 17	0.51 4	0.63 4	0.57 5	0.43 2	0.54 3	0.54 2	0.83 5	0.76 5	1.13 14	0.87 14	0.83 13	1.03 19	0.84 8	0.78 6	0.92 14	0.56 2	0.88 3	0.61 5	0.58 8	0.75 7	0.60 7
IDIAL [192]	11.9	0.52 8	0.56 8	0.58 8	0.59 11	0.78 17	0.58 6	0.61 8	0.69 10	0.65 11	0.85 7	0.81 9	1.08 6	0.88 18	0.88 25	1.00 11	0.83 5	0.82 8	0.90 4	0.62 23	1.01 27	0.62 14	0.63 11	0.84 13	0.61 18
BMBC [171]	13.0	0.57 19	0.57 10	0.64 23	0.58 9	0.73 8	0.64 92	0.77 22	0.78 22	0.71 21	0.84 6	0.77 7	1.09 7	0.85 7	0.81 6	0.98 6	0.82 4	0.77 5	0.91 9	0.58 7	0.93 12	0.60 2	0.56 3	0.73 3	**0.59** 1
SepConv++ [185]	16.4	0.58 22	0.67 25	0.64 23	0.56 7	0.73 8	0.59 11	0.95 49	0.70 14	1.30 96	0.87 10	0.87 17	1.11 9	0.85 7	0.81 6	1.00 11	0.84 8	0.88 14	**0.89** 1	0.59 11	0.97 21	0.61 5	0.59 9	0.79 9	**0.59** 1
EDSC [173]	17.5	0.53 10	0.60 16	0.59 11	0.58 9	0.76 11	0.60 35	0.63 9	0.76 20	0.69 16	0.88 18	0.90 23	1.13 14	0.88 18	0.85 19	1.02 17	0.91 23	1.09 33	0.92 14	0.59 11	0.95 16	0.64 19	0.64 16	0.85 16	0.63 25
MV_VFI [183]	18.2	0.57 19	0.62 18	0.64 23	0.60 16	0.78 17	0.62 60	0.79 26	0.83 27	0.76 29	0.87 10	0.87 17	1.11 9	0.87 14	0.82 11	1.01 13	0.86 14	0.89 17	0.92 14	0.59 11	0.96 20	0.61 5	0.65 19	0.88 21	0.60 7
TC-GAN [166]	18.5	0.57 19	0.62 18	0.64 23	0.60 16	0.78 17	0.63 80	0.78 24	0.81 25	0.75 25	0.87 10	0.87 17	1.11 9	0.86 11	0.82 11	1.01 13	0.86 14	0.88 14	0.92 14	0.59 11	0.95 16	0.61 5	0.65 19	0.89 25	0.60 7
STAR-Net [164]	18.5	0.56 17	0.56 8	0.65 75	0.65 53	0.85 38	0.62 60	0.70 15	0.69 10	0.76 29	0.87 10	0.82 11	1.06 2	0.83 3	0.79 3	0.97 2	0.83 5	0.83 9	0.90 4	0.63 26	1.09 32	0.61 5	0.61 10	0.80 10	0.60 7
DAIN [152]	19.8	0.58 22	0.63 20	0.65 75	0.60 16	0.79 23	0.62 60	0.69 14	0.73 18	0.68 15	0.86 9	0.86 16	1.10 8	0.87 14	0.83 13	1.02 17	0.85 11	0.86 13	0.92 14	0.59 11	0.97 21	0.61 5	0.66 25	0.90 27	0.60 7
FGME [158]	20.2	**0.46** 1	0.49 3	**0.51** 1	0.63 31	0.78 17	0.64 92	0.60 7	0.65 9	0.67 14	0.85 7	0.76 5	1.15 17	**0.82** 1	**0.77** 1	0.99 9	0.86 14	0.81 7	0.95 23	0.61 22	0.93 12	0.70 53	0.63 11	0.83 11	0.65 118
STSR [170]	22.4	0.54 13	0.58 13	0.61 14	0.56 7	0.66 5	0.68 127	0.65 11	0.72 17	0.74 24	0.87 10	0.83 14	1.15 17	0.91 25	0.89 27	1.04 21	0.93 27	1.08 31	0.95 23	0.62 23	1.04 28	0.63 18	0.63 11	0.84 13	0.61 18
AdaCoF [165]	22.7	0.60 39	0.68 27	0.67 138	0.59 11	0.76 11	0.60 35	0.84 30	0.71 16	0.94 59	0.92 23	0.89 21	1.11 9	0.90 23	0.85 19	1.06 25	0.84 8	0.85 11	0.90 4	0.58 7	0.93 12	0.61 5	0.57 4	0.75 7	**0.59** 1
DSepConv [162]	27.2	0.58 22	0.72 34	0.62 17	0.63 31	0.82 29	0.65 106	0.72 18	0.70 14	0.75 25	0.96 49	0.97 32	1.15 17	0.87 14	0.83 13	1.04 21	0.89 21	1.00 25	0.93 20	0.57 4	0.89 5	0.64 19	0.65 19	0.87 20	0.64 77
MEMC-Net+ [160]	28.3	0.59 26	0.65 22	0.65 75	0.64 40	0.79 23	0.70 139	0.80 27	0.77 21	0.96 61	0.88 18	0.83 14	1.12 13	0.88 18	0.85 19	1.01 13	0.85 11	0.88 14	0.91 9	0.64 29	1.13 37	0.62 14	0.63 11	0.86 18	0.60 7
ProBoost-Net [191]	28.8	0.48 2	0.58 13	0.52 2	0.69 82	0.90 62	0.64 92	0.67 12	0.69 10	0.70 18	0.90 20	0.87 17	1.19 27	0.89 22	0.84 18	1.08 26	0.92 24	0.95 21	0.99 26	0.59 11	0.90 7	0.66 29	0.64 16	0.85 16	0.65 118
FeFlow [167]	28.9	0.51 7	0.58 13	0.56 6	0.66 57	0.84 33	0.67 123	0.70 15	0.69 10	0.86 45	0.87 10	0.82 11	1.13 14	0.84 5	0.80 4	0.99 9	0.86 14	0.84 10	0.93 20	0.64 29	0.98 23	0.71 64	0.67 26	0.90 27	0.65 118
MPRN [151]	31.5	0.59 26	0.70 30	0.64 23	0.66 57	0.89 55	0.64 92	0.77 22	1.07 33	0.64 8	0.93 26	0.93 28	1.17 23	0.95 31	0.91 31	1.12 30	0.96 31	1.04 29	1.01 29	0.60 19	0.98 23	0.65 25	0.72 32	1.02 33	0.62 21
ADC [161]	31.7	0.61 52	0.68 27	0.67 138	0.62 24	0.78 17	0.66 114	0.84 30	0.81 25	0.82 38	0.96 49	0.93 28	1.15 17	0.90 23	0.86 24	1.05 24	0.92 24	1.12 34	0.91 9	0.57 4	0.92 10	0.61 5	0.64 16	0.88 21	0.60 7
GDCN [172]	33.5	0.54 13	0.65 22	0.58 8	0.72 103	0.94 79	0.64 92	0.63 9	0.79 23	0.69 16	1.03 122	0.90 23	1.18 24	0.93 28	0.93 32	1.04 21	0.90 22	1.01 27	0.94 22	0.59 11	0.94 15	0.64 19	0.67 26	0.93 29	0.61 18
DAI [168]	37.8	0.65 134	0.52 6	0.79 185	0.64 40	0.82 29	0.66 114	0.47 3	0.56 4	0.57 3	0.91 21	0.77 7	1.41 175	0.88 18	0.85 19	0.98 6	0.87 19	0.96 22	0.91 9	0.60 19	0.99 25	0.60 2	0.65 19	0.88 21	0.60 7
MAF-net [163]	39.6	0.48 2	0.61 17	0.52 2	0.65 53	0.86 43	0.62 60	0.67 12	0.86 29	0.78 34	0.96 49	0.92 27	1.20 32	0.93 28	0.89 27	1.08 26	0.95 29	1.08 31	0.99 26	0.67 37	1.04 28	0.83 140	0.65 19	0.86 18	0.69 182
CyclicGen [149]	39.9	0.64 126	0.63 20	0.73 180	0.67 67	0.73 8	0.88 186	0.72 18	0.84 28	0.78 34	0.95 40	0.89 21	1.24 84	0.91 25	0.85 19	1.09 28	0.87 19	**0.67** 1	1.00 28	**0.53** 1	**0.71** 1	0.62 14	**0.52** 1	**0.64** 1	0.60 7
FRUCnet [153]	40.0	0.70 164	0.71 32	0.80 186	0.64 40	0.80 26	0.69 134	0.78 24	0.75 19	0.95 60	0.91 21	0.91 26	1.15 17	0.86 11	0.83 13	1.01 13	0.86 14	0.89 17	0.92 14	0.58 7	0.89 5	0.64 19	0.63 11	0.83 11	0.64 77
CtxSyn [134]	41.0	0.50 5	0.57 10	0.55 5	0.55 6	0.71 7	0.59 11	1.42 134	0.64 8	2.08 151	0.87 10	0.82 11	1.18 24	0.95 31	0.90 29	1.13 32	0.94 28	0.92 20	1.02 30	0.68 38	1.00 26	0.83 140	0.67 26	0.89 25	0.68 177
PMMST [112]	47.8	0.59 26	0.73 37	0.64 23	0.64 40	0.85 38	0.59 11	0.99 54	1.69 80	1.05 70	0.97 63	1.14 109	1.23 69	0.99 38	0.96 37	1.14 34	1.03 41	1.29 42	1.04 39	0.71 51	1.33 57	0.67 36	0.77 37	1.10 38	0.64 77
SuperSlomo [130]	48.5	0.59 26	0.69 29	0.64 23	0.72 103	0.91 67	0.75 160	0.74 20	1.01 32	0.71 21	0.98 76	0.95 30	1.23 69	0.94 30	0.90 29	1.12 30	0.96 31	0.98 23	1.04 39	0.60 19	0.90 7	0.71 64	0.69 30	0.93 29	0.68 177
FLAVR [188]	49.5	0.67 150	0.70 30	0.71 171	0.71 95	0.78 17	0.76 163	0.76 21	0.80 24	0.75 25	1.24 175	1.28 155	1.23 69	0.83 3	0.80 4	0.97 2	0.83 5	0.85 11	**0.89** 1	0.62 23	0.92 10	0.64 19	0.57 4	0.73 3	0.60 7
OFRI [154]	49.5	0.60 39	0.57 10	0.69 162	0.67 67	0.81 28	0.79 171	0.70 15	0.59 6	0.83 42	0.87 10	0.81 9	1.15 17	0.86 11	0.81 6	1.03 19	0.92 24	0.99 24	0.98 25	0.79 91	0.90 7	1.18 177	0.67 26	0.84 13	0.78 190

Figure 21. **Screenshot of our NIE-ranking on the Middlebury benchmark (taken on the November 16th, 2021).**

10. Screenshots of the Middlebury Benchmark

We take screenshots of the online Middlebury benchmark for VFI on the November 16th, 2021, whose results are shown in Figure 20 and Figure 21. Since the average rank is a relative indicator, previous methods [1, 3, 10, 11] usually report average IE (interpolation error) and average NIE (normalized interpolation error) for comparison. As summarized in Table 2 in our main paper, proposed IFRNet large model achieves best results on both IE and NIE metrics among all published VFI methods that are trained on Vimeo90K [15] dataset. Moreover, IFRNet large runs several times faster than previous state-of-the-art algorithms [10, 12], demonstrating the superior VFI accuracy and fast inference speed of proposed approaches.

References

[1] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-aware video frame interpolation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 6

[2] Myungsub Choi, Heewon Kim, Bohyung Han, Ning Xu, and Kyoung Mu Lee. Channel attention is all you need for video frame interpolation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. 1, 2, 3

[3] Shurui Gui, Chaoyue Wang, Qihua Chen, and Dacheng Tao. Featureflow: Robust video interpolation via structure-to-texture generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.

Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. 3

[5] Huaizu Jiang, Deqing Sun, Varan Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2

[6] Tarun Kalluri, Deepak Pathak, Manmohan Chandraker, and Du Tran. Flavr: Flow-agnostic video representations for fast frame interpolation. In *Arxiv*, 2021. 1

[7] Hyeongmin Lee, Taeoh Kim, Tae-young Chung, Daehyun Pak, Yuseok Ban, and Sangyoun Lee. Adacof: Adaptive collaboration of flows for video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2

[8] Ziwei Liu, Raymond A. Yeh, Xiaoou Tang, Yiming Liu, and Aseem Agarwala. Video frame synthesis using deep voxel flow. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 1

[9] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 1

[10] Simon Niklaus and Feng Liu. Softmax splatting for video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6

[11] Junheum Park, Keunsoo Ko, Chul Lee, and Chang-Su Kim. Bmbc: Bilateral motion estimation with bilateral cost volume for video interpolation. In *European Conference on Computer Vision*, 2020. 6

[12] Junheum Park, Chul Lee, and Chang-Su Kim. Asymmetric bilateral motion estimation for video frame interpolation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2, 6

[13] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[14] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 5

[15] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision (IJCV)*, 2019. 4, 5, 6