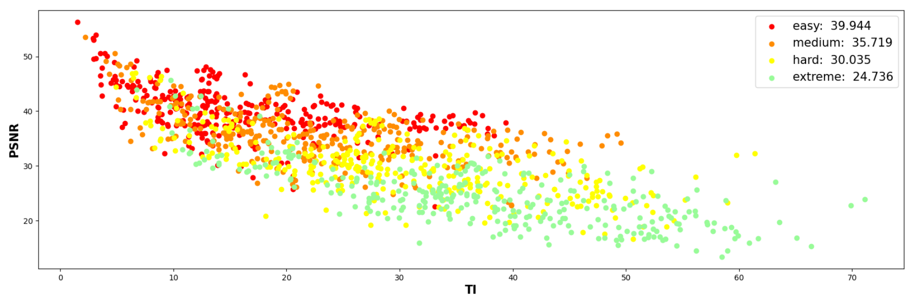
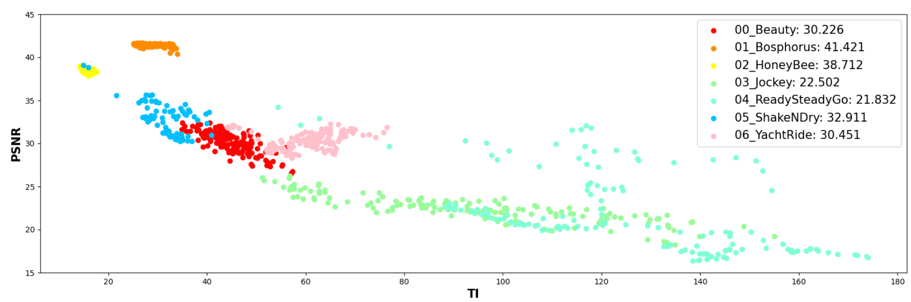
7 Appendix

7.1 Relation between VFI results and TI for various benchmark

We further show the relation between TI and the interpolation results for various benchmarks to demonstrate the argument in Sec. 3.1 that motion complexity needs to be considered in VFI.



(a) Relation between TI and PSNR in SNU-FILM [4]



(b) Relation between TI and PSNR in UVG [25]

Fig. 8: Relation between VFI results and Temporal Information for various benchmarks

Fig. 8 shows the relation between TI and VFI results for the SNU-FILM and UVG benchmarks. For both benchmarks, we can see that interpolation performance degrades for frames with high TI value that contain a lot of motion between two consecutive frames. In particular, for SNU-FILM, which is divided into easy, medium, hard, and extreme subsets based on the temporal gap between two frames, the magnitude of motion increases as the gap increases. As a result, the TI value increases and the PSNR degrades, as shown in Fig. 8a.

7.2 Additional Subjective Results

Additional subjective results on SNU-FILM [4] and UVG [25] benchmarks with various motions are shown in Fig. 9 to demonstrate that our proposed methods effectively interpolate all regions when applied to the existing network. The first column compares subjective results for UVG data, and the other columns compare results for extreme, a subset of SNU-FILM data. As shown in Fig. 9, interpolation performance is improved compared to the existing RIFE in all motion complexity regions. Furthermore, we can see that the subjective quality is improved in regions containing text and building exteriors, and even regions with a lot of detail are interpolated successfully. These results show that our proposed methods interpolate all regions of the video effectively.

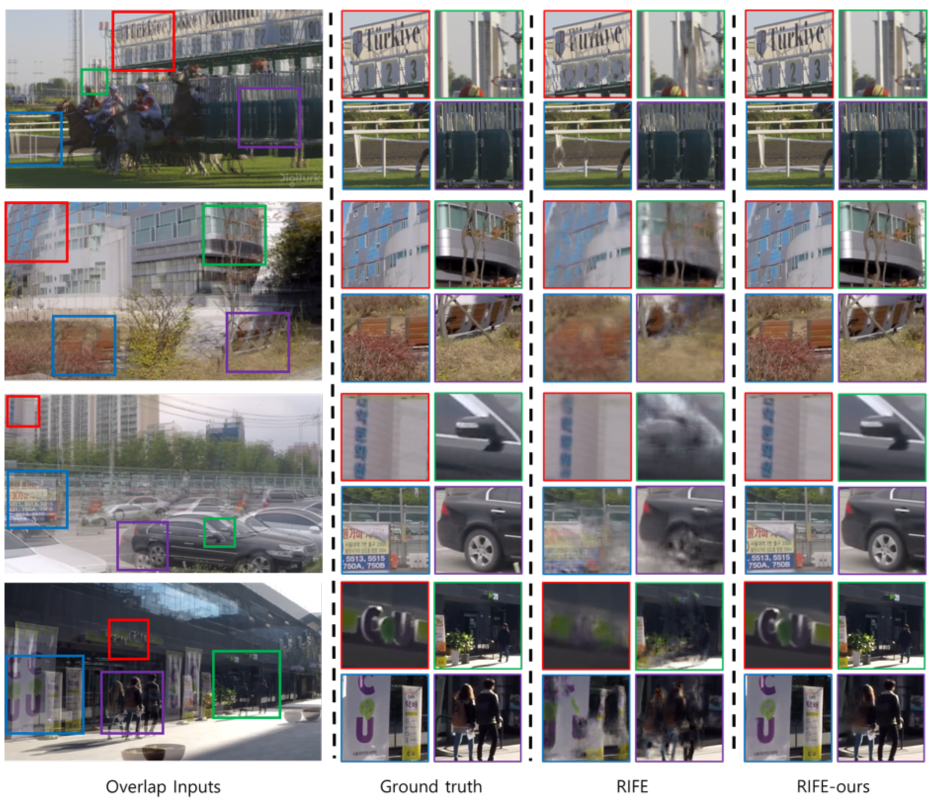


Fig. 9: Additional subjective results of proposed IAM-VFI

Table 5: Quantitative results comparison on benchmark datasets. "†" indicates the results we obtained by retraining to compare results.

Algorithm	HD	SNU-FILM				UVG	Xiph	
		Easy	Medium	Hard	Extreme		2K	4K
TDPNet† [39]	32.25/0.944	39.99/0.990	35.85/0.979	30.16/0.933	24.95/0.855	31.18/0.902	36.41/0.902	33.43/0.9431
TDPNet$_{ours}$	**32.52/0.947**	**40.19/0.991**	**36.11/0.980**	**30.62/0.937**	**25.37/0.862**	**32.15/0.915**	**36.72/0.967**	**34.18/0.946**
EBME† [17]	32.13/0.94	39.97/0.991	35.73/0.99	30.29/0.979	25.16/0.861	32.77/0.915	36.13/0.965	33.62/0.943
EBME-H† [17]	32.44/0.944	40.17/0.991	35.94/0.979	**30.49/0.936**	25.28/0.862	32.68/0.916	36.61/0.969	33.91/0.946
EBME$_{ours}$	**32.50/0.945**	**40.17/0.991**	**35.95/0.979**	30.48/0.936	**25.42/0.863**	**33.30/0.924**	**36.61/0.969**	**34.01/0.946**

Table 6: Computation and inference time comparison of our proposed method and existing VFI algorithms

Algorithm	Parameters (M)	FLOPs (G)	Runtime (ms)
TDPNet† [39]	10.4	166.7	88.5
TDPNet$_{ours}$	20.4	400.8	123
EBME† [17]	3.9	164.98	62
EBME-H† [17]	4.0	481.12	111
EBME$_{ours}$	5.8	368.48	95

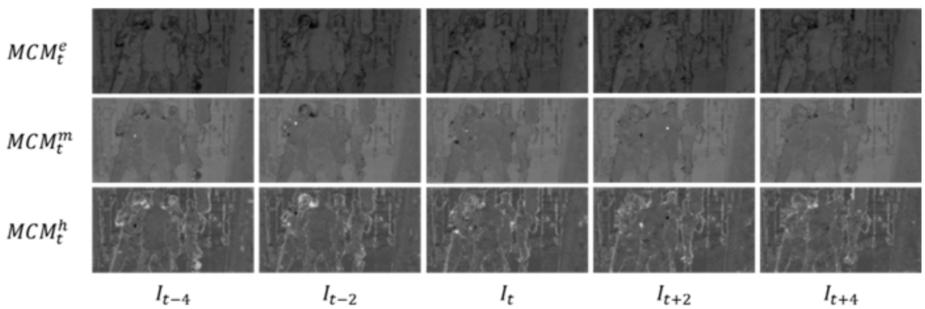
7.3 Additional Experiments and Efficiency Comparison

To demonstrate that our proposed methods are universally applicable and effective for most flow-based VFI algorithms in addition to RIFE and EMA-VFI algorithms, we also applied it to TDPNet [39] and EBME [17]. The results of additional experiments are summarised in Tab. 5, which shows that our proposed approach improves the performance of both TDPNet and EBME. In particular, the performance improves significantly on the UVG and Xiph4K benchmarks, which are high resolution and contain a variety of motions, and on extreme, a subclip of SNU-FILM with a large motion distribution. And compared to EBME-H, which is proposed to handle large resolutions, we can see that applying our proposed method to EBME outperforms on all benchmarks except hard, which is a subset of SNU-FILM.

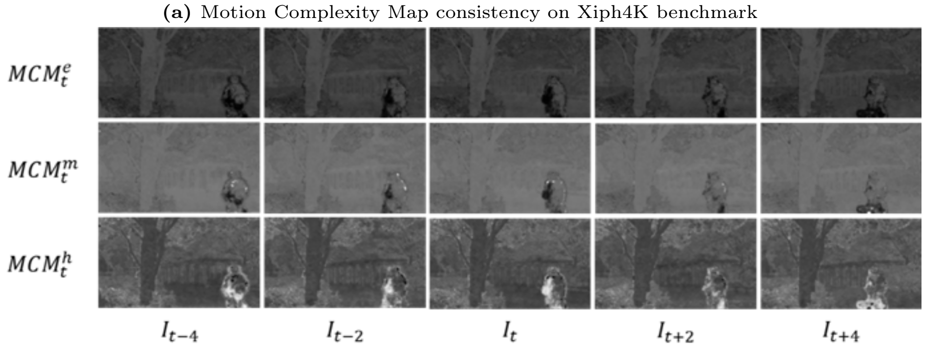
We performed a comparison about the additional complexity cost (FLOPs), parameters, and runtime compared to the existing VFI algorithms to show the efficiency of our proposed methods. We summarized the results of the comparison in Tab. 6. The table shows that the proposed method increases FLOPs, parameters, and runtime compared to the existing network since it needs to have an optimized flow estimation network for each motion complexity. However, when comparing the results of applying our proposed method to EBME with EBME-H, which is proposed to handle large resolutions, it can be seen that the computation is less than that of EBME-H, while the performance is higher on average. Moreover, we can see that the inference time for the VFI network to interpolate intermediate frames is also faster than EBME-H. As a result, our proposed methods can efficiently improve performance at the cost of a slight increase in computational complexity or inference time over the existing VFI algorithms.

7.4 Consistency of Motion Complexity Map

We visualize the MCM to show that it performed well in all continuous sequences of the video, not only in certain specific frames. Fig. 10 shows each channel of the MCM estimating a probability value for whether each region in the frame to be interpolated is easy, medium, or hard. It can be seen that the MCM performs well in estimating the motion complexity of each region in consecutive frames without any interruption.



(a) Motion Complexity Map consistency on Xiph4K benchmark



(b) Motion Complexity Map consistency on HD benchmark

Fig. 10: Motion Complexity Map consistency