Perception-Oriented Video Frame Interpolation via Asymmetric Blending

Guangyang Wu[1], Xin Tao[2], Changlin Li[3], Wenyi Wang[4], Xiaohong Liu[1*], Qingqing Zheng[5*]

[1] Shanghai Jiao Tong University [2] Kuaishou Technology [3] SeeKoo [4] University of Electronic Science and Technology of China

[5] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

CVPR SEATTLE, WA JUNE 17-21, 2024

Paper Link Project Page

***TL;DR:** A new paradigm for VFI: Ensuring stable, high visual quality and efficiency, even with large motion.*

Motivation

Common issues for previous VFI methods:

- Blur and ghosting effects persist.
- Unavoidable motion errors are overlooked.
- Mis-alignment in supervision.

Solution

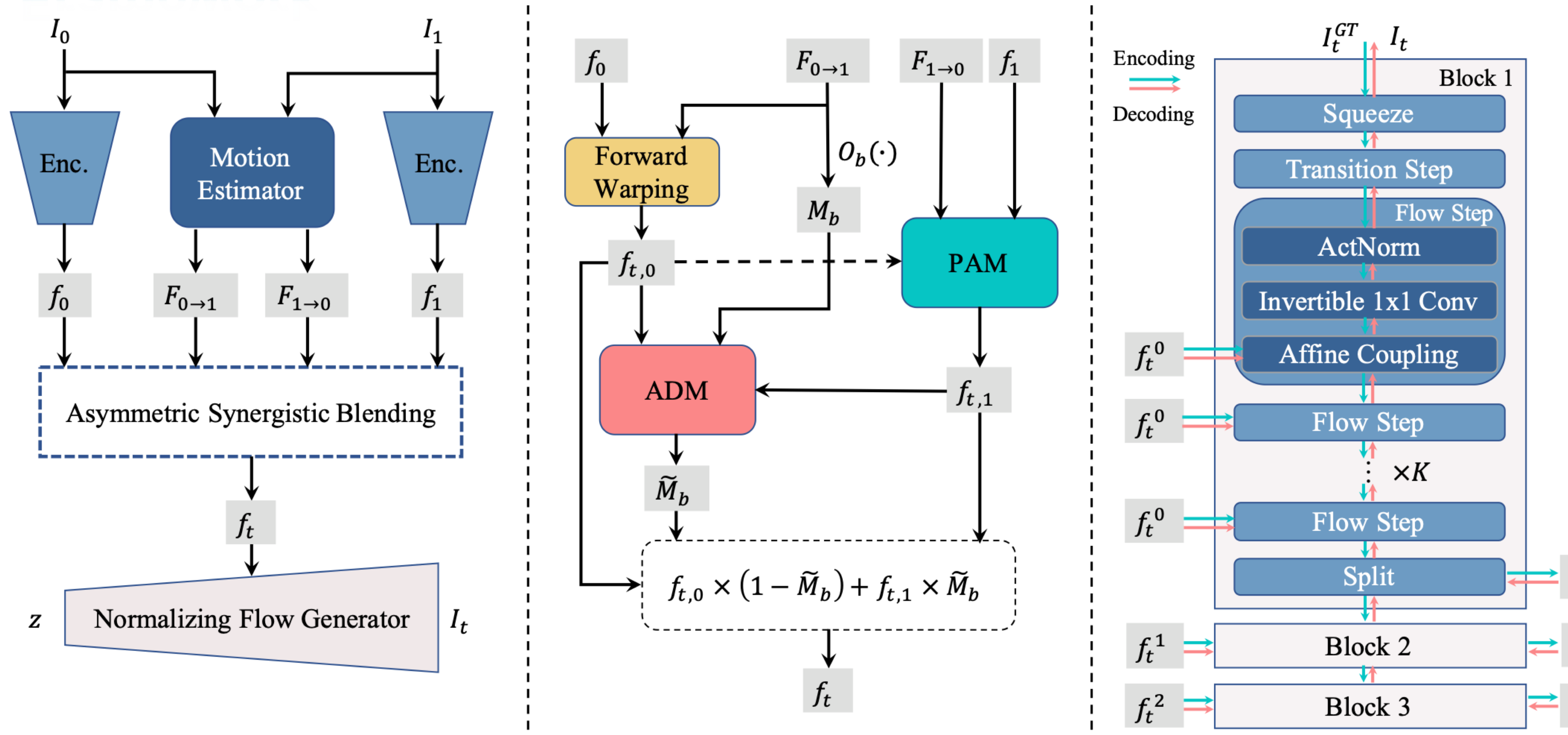
PerVFI address the above issues by:

- Asymmetric synergistical blending scheme.
- Normalizing flow-based network as generator.

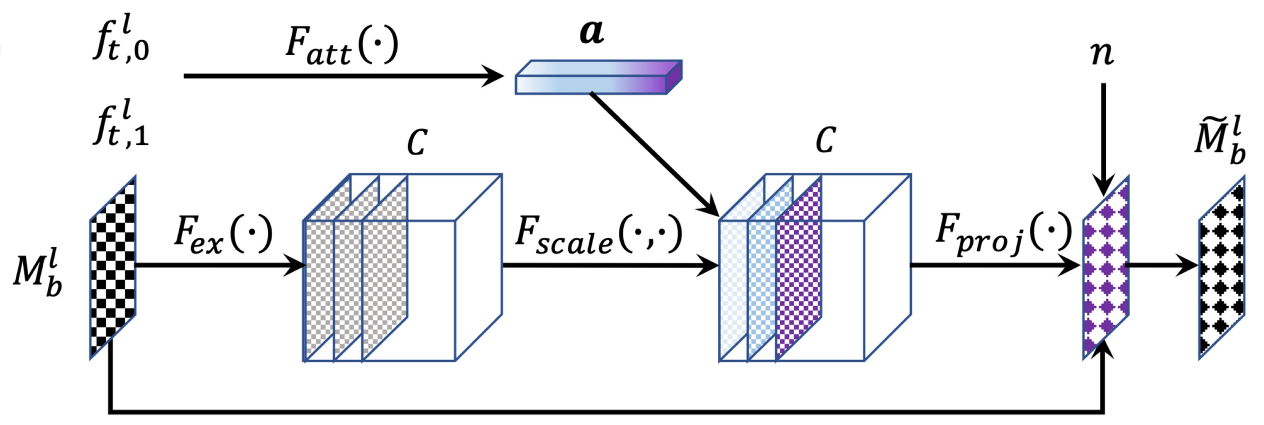
Framework

- Utilizing optical flow as motion input, adaptable to diverse optical flow estimators.
- Leveraging PAM for alignment and ADM for sparsity mask to blend features.
- PAM is used for alignment. ADM is used to apply sparsity during blending.
- Incorporating a normalizing flow-based generator that utilizes multi-scale features as conditions.

Framework



(a) Overall Framework (b) Asymmetric Synergistic Blending (c) Normalizing Flow Generator



(d) ADM Structure. It produces a quasi-binary mask for future blending.

NOTE: This framework is simple yet highly effective. While each component is widely used on its own, their integration leads to exceptional performance.

Experiment Result

Table 1. Performance comparison of VFI algorithms on DAVIS-2017 [41]. The scores for LDMVFI [11] are taken from their paper and indicated with the † symbol. 'OOM' means out of memory. The best values are highlighted in **red** and the second-best values are in blue.

	DAVIS (480P)					DAVIS (1080P)				
	PSNR↑	SSIM ↑	LPIPS↓	FloLPIPS↓	VFIPS↑	PSNR↑	SSIM↑	LPIPS↓	FloLPIPS↓	VFIPS↑
EDSC [5]	26.52	0.784	0.132	0.093	72.62	24.54	0.768	0.205	0.138	51.05
RIFE [17]	26.97	0.807	0.085	0.063	80.19	25.89	0.803	0.134	0.097	62.56
STMFNet [8]	28.55	0.850	0.121	0.086	77.38	27.43	0.844	0.178	0.119	60.25
LDMVFI [11]	25.54 †	-	0.107 †	0.153 †	75.78 †	-	-	-	-	-
VFIFormer [34]	27.33	0.814	0.124	0.090	77.32	OOM	OOM	OOM	OOM	OOM
EMA-VFI [59]	**28.83**	**0.856**	0.127	0.085	78.84	**27.61**	**0.846**	0.203	0.131	60.87
AMT [28]	27.42	0.818	0.101	0.073	80.57	25.72	0.806	0.177	0.122	60.39
PerVFI (ours)	26.83	0.804	**0.077**	**0.058**	**87.51**	26.23	0.808	**0.114**	**0.087**	**72.52**

Table 2. Performance comparison of VFI algorithms on Xiph4K [37] and Vimeo-90K [57]. The best values are highlighted in **red**, while the second-best values are in blue. 'OOM' means out of memory.

	Xiph - 2K			Xiph - "4K"			Vimeo-90K		
	LPIPS↓	FloLPIPS↓	VFIPS↑	LPIPS↓	FloLPIPS↓	VFIPS↑	PSNR↑	SSIM↑	LPIPS↓
EDSC [5]	0.085	0.072	64.73	0.177	0.120	51.24	34.86	0.961	0.027
RIFE [17]	0.041	0.050	65.26	0.099	0.067	54.31	34.16	0.955	0.020
STMFNet [8]	0.110	0.063	65.19	0.245	0.128	53.33	-	-	-
VFIFormer [34]	OOM	OOM	OOM	OOM	OOM	OOM	**36.38**	**0.971**	0.021
EMA-VFI [59]	0.110	0.081	65.12	0.241	0.114	53.57	36.34	0.967	0.026
AMT [28]	0.089	0.055	65.60	0.199	0.114	53.22	35.79	0.968	0.021
PerVFI (ours)	**0.038**	**0.032**	**68.67**	**0.086**	**0.062**	**57.47**	33.89	0.953	**0.018**

Table 3. Comparisons of running time, MACs and number of parameters. The evaluations are conducted on frames of size 512 ⇥ 512 on a NVIDIA RTX 3090 GPU

Methods	Runtime (s) ↓	MACs (G) ↓	#Params (M) ↓
EDSC [3]	0.077	71.826	8.945
XVFI [14]	0.067	70.725	5.577
RIFE [6]	0.037	52.247	2.278
EBME [8]	0.079	43.731	3.908
STMFNet [5]	0.206	876.156	17.960
*PerVFI	0.161	458.340	8.481
GMFlow+PerVFI	0.193	523.758	13.161
RAFT+PerVFI	0.204	836.591	13.739

Table 4. Comparisons of PerVFI with different optical flow estimators. PerVFI can be adaptable to diverse optical flow estimators.

	DAVIS (480P)				
	PSNR↑	SSIM ↑	LPIPS↓	FloLPIPS↓	VFIPS↑
RAFT-small+PerVFI	26.84	0.812	0.080	0.062	81.15
GMFlow+PerVFI	27.13	0.815	0.077	0.058	82.98
GMA+PerVFI	**27.19**	**0.816**	0.0753	0.057	**83.34**
RAFT+PerVFI	27.16	**0.816**	**0.0751**	**0.056**	83.30

Key Insights & Future work

- Using a sparse mask to blend two mis-aligned features linearly is what makes our pipeline "asymmetric." This straightforward approach results in exceptional performance.
- The generator-based method addresses misalignment during supervision, yielding less blurry results.
- Our network structure has not been meticulously optimized, leaving ample room for future improvements in both efficiency and performance.