Boost Video Frame Interpolation via Motion Adaptation

Haoning Wu[1], Xiaoyun Zhang[1], Weidi Xie[1,2], Ya Zhang[1,2], Yanfeng Wang[1,2]

[1]CMIC, Shanghai Jiao Tong University [2]Shanghai AI Laboratory

Shanghai Jiao Tong University

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

BMVC 2023

Contributions

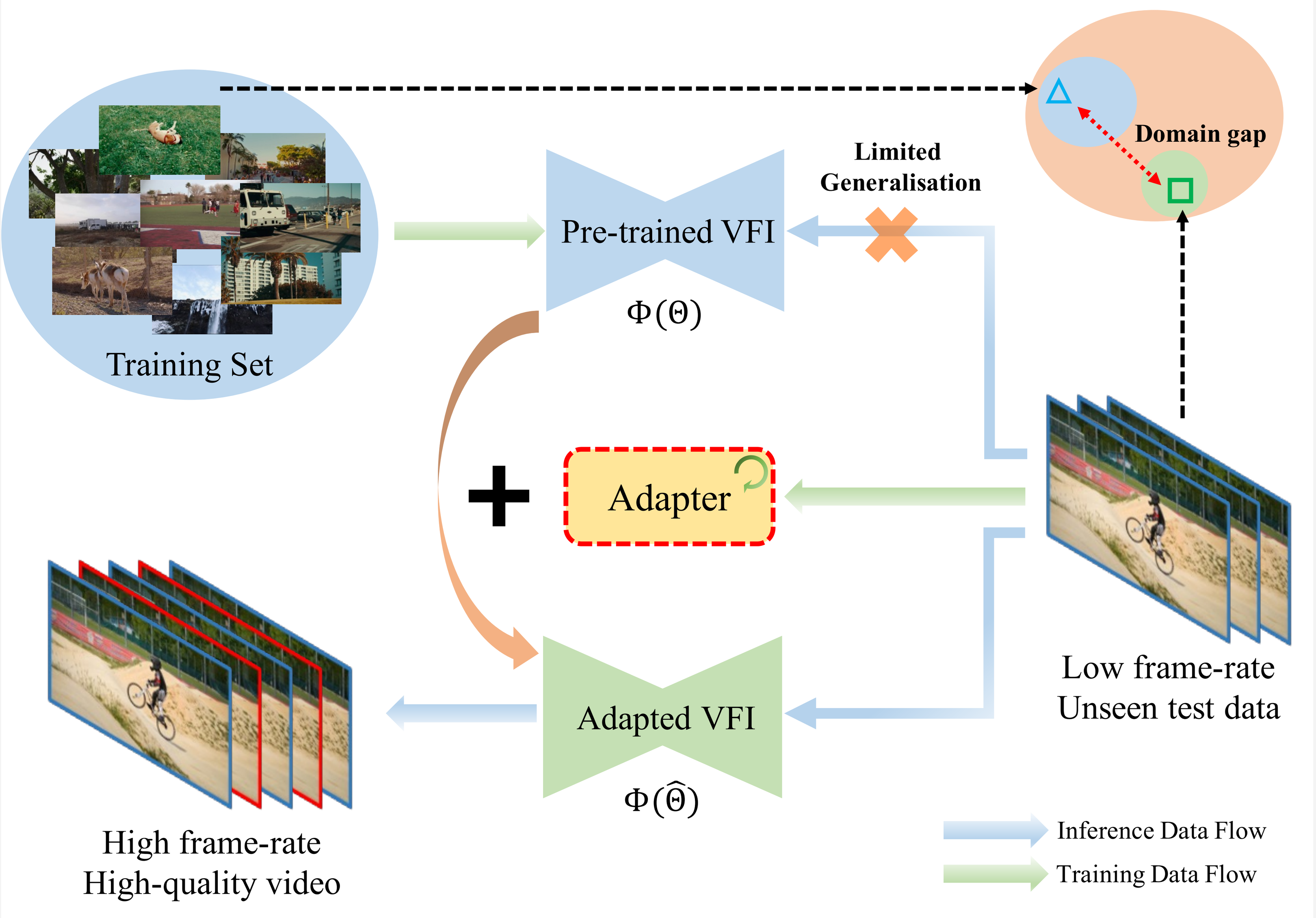
- **Optimization-based VFI**: improves the generalization ability of existing VFI models towards various unseen video scenarios.
- **Steady Adaptation**: cycle-consistency adaptation fully utilizes the inter-frame consistency to learn motion characteristics within video sequences.
- **Efficient Adaptation**: VFIAdapter significantly improves the efficiency of motion adaptation.
- **Significant Gain**: our boosted VFI models achieve SOTA performance, and even outperform approaches with extra inputs.

Qualitative Results

Inputs (Overlay) | FLAVR | VFIT-B | UPRNet | UPRNet-ours-plugin | UPRNet-ours-e2e++ | Ground Truth

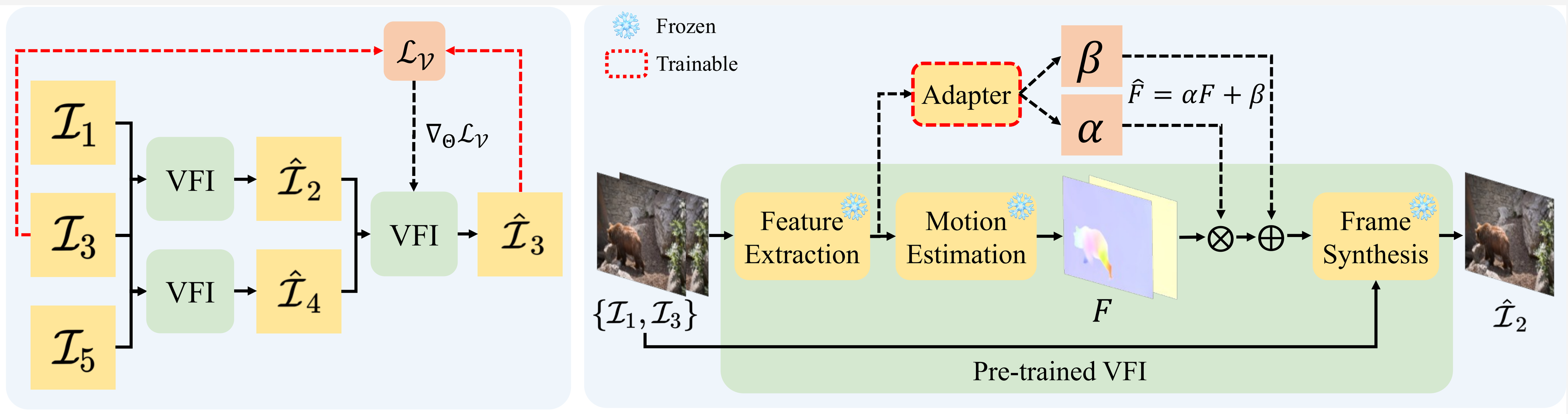
Project Page | *Paper* | *Code*

Optimization-based Video Frame Interpolation



- Our motivation is to adapt models to **unseen motion characteristics** via **optimisation-based VFI**.
- Based on model parameters learnt on a training set, optimisation-based VFI further optimizes the parameters with **each given low frame-rate test video** to **boost the interpolation performance**.

Video Frame Interpolation Adapter (VFIAdapter)



(a) Cycle-consistency adaptation strategy (b) Plug-in adapter module

- (a) **Cycle-consistency adaptation** can learn specific motion characteristics by fully utilizing **inter-frame consistency within video sequences**.
- (b) Lightweight **plugin-in *adapter*** is proposed to effectively refine the estimated motion flow with **minimal tuning cost**, leading to **significant efficiency improvement**.

Experiment Results

- Quantitative (PSNR/SSIM) comparison with representative methods.

Methods	Adaptation e2e	Adaptation plugin	Vimeo90K [50]	DAVIS [36]	SNU-FILM [6] Easy	SNU-FILM [6] Medium	SNU-FILM [6] Hard	SNU-FILM [6] Extreme
SepConv [32]	✗	✗	33.72 / 0.9639	26.65 / 0.8611	40.21 / 0.9909	35.45 / 0.9785	29.62 / 0.9302	24.16 / 0.8457
SepConv-ours-e2e	✓	✗	33.96 / 0.9650	26.83 / 0.8639	40.41 / 0.9911	35.71 / 0.9794	29.80 / 0.9313	24.26 / 0.8479
EDSC [3]	✗	✗	34.55 / 0.9677	26.83 / 0.8578	40.66 / 0.9915	35.77 / 0.9795	29.75 / 0.9301	24.12 / 0.8420
EDSC-ours-e2e	✓	✗	34.73 / 0.9685	26.96 / 0.8600	40.88 / 0.9917	35.98 / 0.9803	29.85 / 0.9313	24.19 / 0.8436
RIFE [11]	✗	✗	35.28 / 0.9704	27.61 / 0.8760	40.74 / 0.9916	36.18 / 0.9808	30.30 / 0.9368	24.62 / 0.8531
RIFE-ours-e2e	✓	✗	35.57 / 0.9717	27.81 / 0.8798	40.95 / 0.9918	36.58 / 0.9816	30.49 / 0.9386	24.71 / 0.8549
RIFE-ours-e2e++	✓	✗	35.93 / 0.9733	28.10 / 0.8850	41.20 / 0.9924	36.94 / 0.9835	30.83 / 0.9430	24.87 / 0.8589
RIFE-ours-plugin	✗	✓	35.56 / 0.9714	27.76 / 0.8771	40.99 / 0.9918	36.55 / 0.9825	30.48 / 0.9387	24.64 / 0.8533
IFRNet [19]	✗	✗	35.86 / 0.9729	28.03 / 0.8851	40.91 / 0.9918	36.58 / 0.9816	30.75 / 0.9403	24.85 / 0.8590
IFRNet-ours-e2e	✓	✗	36.38 / 0.9753	28.45 / 0.8936	41.21 / 0.9921	37.03 / 0.9832	31.10 / 0.9440	25.03 / 0.8634
IFRNet-ours-e2e++	✓	✗	36.68 / 0.9760	28.78 / 0.8995	41.48 / 0.9923	37.57 / 0.9850	31.45 / 0.9482	25.22 / 0.8694
IFRNet-ours-plugin	✗	✓	36.01 / 0.9734	28.16 / 0.8825	41.06 / 0.9920	36.92 / 0.9834	30.88 / 0.9404	24.93 / 0.8599
UPRNet [15]	✗	✗	36.07 / 0.9735	28.38 / 0.8914	41.01 / 0.9919	36.80 / 0.9819	31.22 / 0.9422	25.39 / 0.8648
UPRNet-ours-e2e	✓	✗	36.68 / 0.9758	28.84 / 0.8997	41.31 / 0.9923	37.24 / 0.9836	31.66 / 0.9464	25.64 / 0.8699
UPRNet-ours-e2e++	✓	✗	**36.90** / 0.9768	**29.15 / 0.9062**	**41.48 / 0.9925**	**37.66 / 0.9855**	**32.00 / 0.9519**	**25.99 / 0.8798**
UPRNet-ours-plugin	✗	✓	36.44 / 0.9751	28.69 / 0.8945	41.32 / 0.9923	37.38 / 0.9843	31.64 / 0.9448	25.69 / 0.8705
VFIformer [27]	✗	✗	36.14 / 0.9738	28.33 / 0.8898	40.93 / 0.9918	36.53 / 0.9815	30.52 / 0.9392	24.92 / 0.8580
EMA-VFI [51]	✗	✗	36.23 / 0.9740	28.07 / 0.8826	41.04 / 0.9921	36.73 / 0.9821	30.88 / 0.9400	24.92 / 0.8580
FLAVR [18]	✗	✗	36.22 / 0.9746	27.97 / 0.8806	41.09 / 0.9918	36.85 / 0.9830	31.10 / 0.9456	25.23 / 0.8676
VFIT-S [40]	✗	✗	36.42 / 0.9760	28.46 / 0.8926	41.15 / 0.9920	37.07 / 0.9845	31.39 / 0.9501	25.52 / 0.8717
VFIT-B [40]	✗	✗	36.89 / **0.9775**	28.60 / 0.8945	41.24 / 0.9921	37.06 / 0.9839	31.39 / 0.9501	25.61 / 0.8731

- Quantitative (PSNR/SSIM) comparison of **adaptation strategies**.

Strategies	#Adaptations	SepConv [32]	EDSC [3]	RIFE [11]	IFRNet [19]	UPRNet [15]
Original	0	33.72 / 0.9639	34.55 / 0.9677	35.28 / 0.9704	35.86 / 0.9729	36.07 / 0.9735
Naïve	5	33.77 / 0.9641	34.62 / 0.9679	35.36 / 0.9708	35.95 / 0.9734	36.23 / 0.9744
	10	33.83 / 0.9644	34.69 / 0.9683	35.45 / 0.9713	35.81 / 0.9731	36.16 / 0.9747
	20	33.91 / 0.9647	34.80 / 0.9687	35.45 / 0.9715	35.03 / 0.9685	35.79 / 0.9737
	30	33.95 / 0.9648	34.85 / 0.9688	35.33 / 0.9710	34.09 / 0.9615	35.51 / 0.9721
Cycle	5	33.83 / 0.9644	34.63 / 0.9680	35.41 / 0.9710	36.14 / 0.9741	36.49 / 0.9750
	10	33.96 / 0.9650	34.73 / 0.9685	35.57 / 0.9717	36.38 / 0.9753	36.68 / 0.9758
	20	34.17 / 0.9659	34.94 / 0.9693	35.80 / 0.9728	36.60 / 0.9759	36.84 / 0.9766
	30	**34.29 / 0.9662**	**35.06 / 0.9699**	**35.93 / 0.9733**	**36.68 / 0.9760**	**36.90 / 0.9768**

- Ablation study on **end-to-end** and **plug-in adapter adaptation**.

Methods	#Finetuning Parameters	Adaptation Time (ms) Vimeo90K	Adaptation Time (ms) DAVIS	Adaptation Time (ms) SNU-FILM	Inference Time (ms) Vimeo90K	Inference Time (ms) DAVIS	Inference Time (ms) SNU-FILM
RIFE-ours-e2e	10.21M	145.6	162.7	260.8	10.94	12.74	23.61
RIFE-ours-plugin	0.087M	83.13	86.84	125.4	11.79	14.67	24.79
IFRNet-ours-e2e	18.79M	107.7	196.2	403.3	18.61	25.94	55.54
IFRNet-ours-plugin	0.676M	39.08	73.79	158.1	19.11	29.32	61.58
UPRNet-ours-e2e	6.260M	285.5	507.0	1487.8	28.33	49.90	90.85
UPRNet-ours-plugin	0.009M	162.0	237.6	872.7	29.20	50.72	92.60

- **Motion Field Visualization**.

UPRNet (Pre-trained) | UPRNet (Adapted) | RGB Ground Truth | Motion before adaptation | Motion after adaptation | Motion Ground Truth