ST-MFNet: A Spatio-Temporal Multi-Flow Network for Frame Interpolation

Duolikun Danier Fan Zhang David Bull

University of Bristol

University of BRISTOL

BRISTOL VISION INSTITUTE

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Introduction

Background: Existing video frame interpolation methods face two major challenges: large motion and dynamic textures.

- **Large motion**: large pixel displacement; difficult to capture with limited receptive field.
- **Dynamic texture**: complex motion; non-smooth optical field.

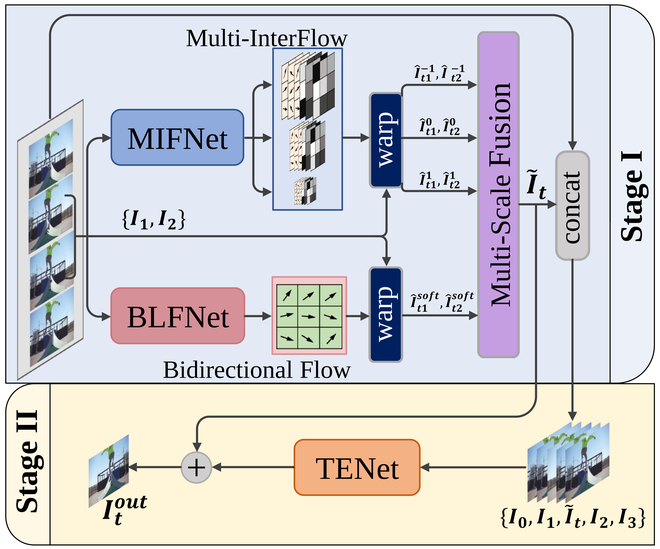
Motivation:

- **Flow-based VFI**: use optical flow to warp frames.
  - *Pros*: large motion capturing due to coarse-to-fine architectures.
  - *Cons*: one-to-one mapping and smoothness assumption not suited for complex motion.
- **Kernel-based VFI**: predict local adaptive convolution kernels to synthesize pixels.
  - *Pros*: many-to-one mapping enable flexible pixel transformation to model complex motion.
  - *Cons*: fixed kernel size limits large motion capturing.

Design Idea:

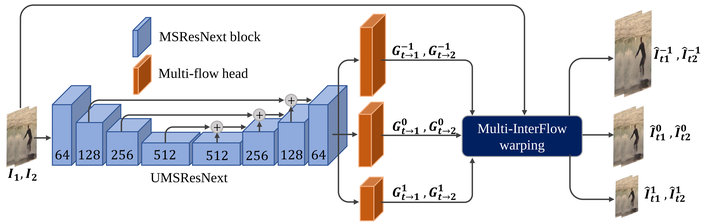
- Design **multi-branching** to combine advantages of both methods.
- Employ **3D CNN** on larger temporal window to enhance textural details.
- Use **spatio-temporal GAN** to enhance perceptual quality.

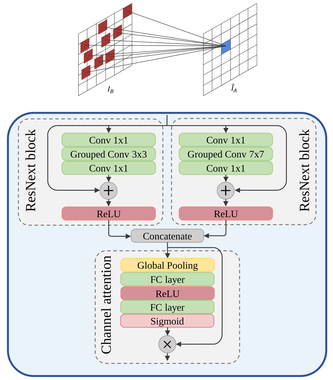
Proposed Method



Design Details

Multi-InterFlow Network:





Bidirectional Linear Flow Network: PWC-Net; $F_{1\to t} = 0.5F_{1\to 2}$ $F_{2\to t} = 0.5F_{2\to 1}$

Texture Enhancement Network: ResNet3D+channel attention with reduced feature dimensions.

Loss Functions:

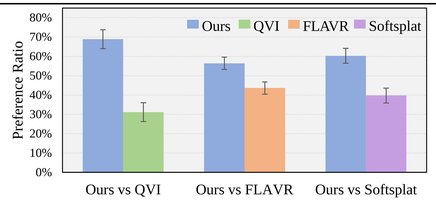
$$\mathcal{L}_{lap} = \sum_{s=1}^{S} 2^{s-1} \left\| L^s(I^{out}_t) - L^s(I^{gt}_t) \right\|_1 \qquad \mathcal{L}_p = \mathcal{L}_{lap} + \lambda \mathcal{L}_{adv} \qquad \mathcal{L}_{adv} = -\log(D(I^{out}_t, I_1, I_2))$$

Comparison with State of the Art

Note: each entry denotes PSNR/SSIM.

	UCF101	DAVIS	SNU-FILM Easy	SNU-FILM Medium	SNU-FILM Hard	SNU-FILM Extreme	VFITex
DVF	32.251/0.965	20.403/0.673	27.528/0.876	24.091/0.817	21.556/0.760	19.709/0.705	19.946/0.709
SuperSloMo	32.547/0.968	26.523/0.866	36.255/0.984	33.802/0.973	29.519/0.930	24.770/0.855	27.914/0.911
SepConv	32.524/0.968	26.441/0.853	39.894/0.990	35.264/0.976	29.620/0.926	24.653/0.851	27.635/0.907
DAIN	32.524/0.968	27.086/0.873	39.280/0.989	34.993/0.976	29.752/0.929	24.819/0.850	27.314/0.909
BMBC	32.729/0.969	26.835/0.869	39.809/0.990	35.437/0.978	29.942/0.933	24.715/0.856	27.337/0.904
AdaCoF	32.610/0.968	26.445/0.854	39.912/0.990	35.269/0.977	29.723/0.928	24.656/0.851	27.639/0.904
FeFlow	32.520/0.967	26.555/0.856	39.591/0.990	35.014/0.977	29.466/0.928	24.607/0.852	OOM
CDFI	32.653/0.968	26.471/0.857	39.881/0.990	35.224/0.977	29.660/0.929	24.645/0.854	27.576/0.906
CAIN	32.537/0.968	26.477/0.857	39.890/0.990	35.630/0.978	29.998/0.931	25.060/0.857	28.184/0.911
SoftSplat	32.835/0.969	27.582/0.881	40.165/0.991	36.017/0.979	30.604/0.937	25.436/0.864	28.813/0.924
EDSC	32.677/0.969	26.689/0.860	39.792/0.990	35.283/0.977	29.815/0.929	24.872/0.854	27.641/0.904
XVFI	32.224/0.966	26.565/0.863	38.849/0.989	34.497/0.975	29.381/0.929	24.677/0.855	27.759/0.909
QVI	32.668/0.967	27.483/0.883	36.648/0.985	34.637/0.978	30.614/0.947	25.426/0.866	28.819/0.926
FLAVR	33.389/0.971	27.450/0.873	40.135/0.990	35.988/0.979	30.541/0.937	25.188/0.860	28.487/0.915
ST-MFNet	33.384/0.970	28.287/0.895	40.775/0.992	37.111/0.985	31.698/0.951	25.810/0.874	29.175/0.929

	GoPro-4× PSNR (↑)	GoPro-4× SSIM (↑)	GoPro-8× PSNR (↑)	GoPro-8× SSIM (↑)
QVI	29.324	0.927	29.280	0.929
FLAVR	28.911	0.914	29.512	0.922
Softsplat	28.858	0.908	29.663	0.918
ST-MFNet	29.892	0.926	30.568	0.934

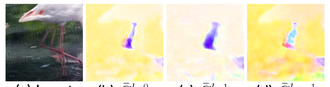


Ablation Study

	UCF101	DAVIS	VFITex
Ours-*w/o BLFNet*	33.218/0.970	27.767/0.881	28.498/0.915
Ours-*w/o MIFNet*	33.202/0.969	27.886/0.889	28.357/0.911
Ours-*w/o TENet*	32.895/0.970	27.484/0.880	28.241/0.910
Ours-*unet*	33.378/0.970	28.096/0.892	28.898/0.925
Ours-*w/o US*	33.371/0.970	28.155/0.893	28.977/0.925
Ours	33.384/0.970	28.287/0.895	29.175/0.929

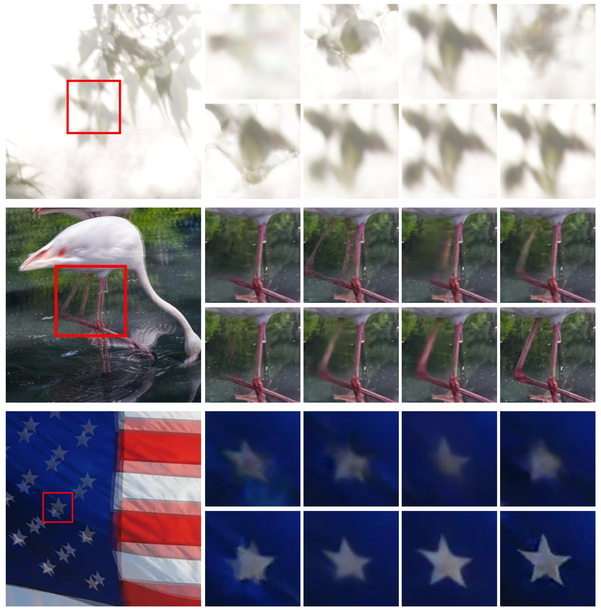
	UCF101	DAVIS	VFITex
Ours-$\mathcal{L}_{lap}$	0.036	0.125	0.216
TGAN	0.034	0.117	0.213
FIGAN	0.036	0.119	0.216
Ours-$\mathcal{L}_p$	0.033	0.116	0.210

Multi-InterFlow at 3 scales



(a) Input (b) $\bar{G}^{l=0}_{t\to1}$ (c) $\bar{G}^{l=1}_{t\to1}$ (d) $\bar{G}^{l=-1}_{t\to1}$

Qualitative Results



Overlaid frames

From top-left to bottom-right: BMBC, DAIN, Softsplat, FLAVR, QVI, Ours-$\mathcal{L}_{lap}$, Ours-$\mathcal{L}_p$, GT.