ACM multimedia

中国科学技术大学
University of Science and Technology of China

Space-Time Video Super-Resolution Using Temporal Profiles

Zeyu Xiao, Zhiwei Xiong, Xueyang Fu, Dong Liu, Zheng-Jun Zha
University of Science and Technology of China
zwxiong@ustc.edu.cn

VIDAR
Visual Information Discovery And Recovery

NEL-BITA
National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application

Introduction

Space-Time Video Super-Resolution

- Video spatial SR
- Video frame interpolation (video temporal SR)
- Recover a **high-frame-rate** and **high-resolution** video from its low-frame-rate and low-resolution observation simultaneously

Introduction

Space-Time Video Super-Resolution

low-frame-rate & low-resolution

high-frame-rate & high-resolution

Introduction

Applications

- Film making
- UHD displays
- Video replay in sports games

Related Work

Video Frame Interpolation (VFI) and Video Super-Resolution (VSR)

- VFI → Recover unseen intermediate frames to up-convert frame rate
- VSR → Improve the resolution of frames by exploring temporal information
- Cascade VFI and VSR in a two-stage manner
  - accumulated errors
  - unexpected artifacts
  - intra-related spatial-temporal information cannot be fully exploited

Related Work

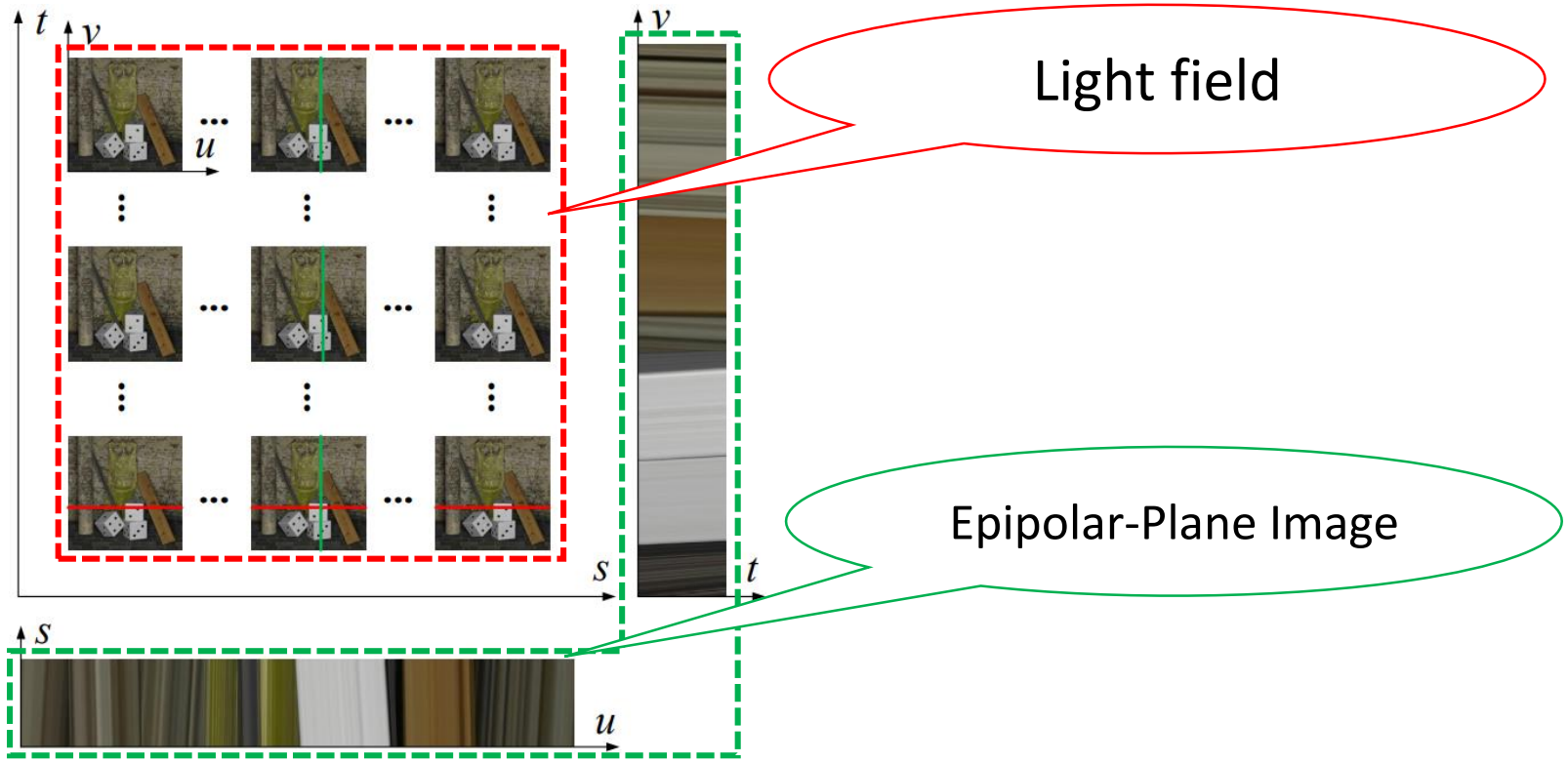
Space-Time Video Super-Resolution

- Zooming Slow-Mo
  - one-stage deep-learning-based framework
  - frame feature temporal interpolation, deformable ConvLSTM and frame reconstruction modules
  - limited temporal context and unrealistic artifacts

Related Work

Epipolar-Plane Image (EPI) and Temporal Profile (TP)

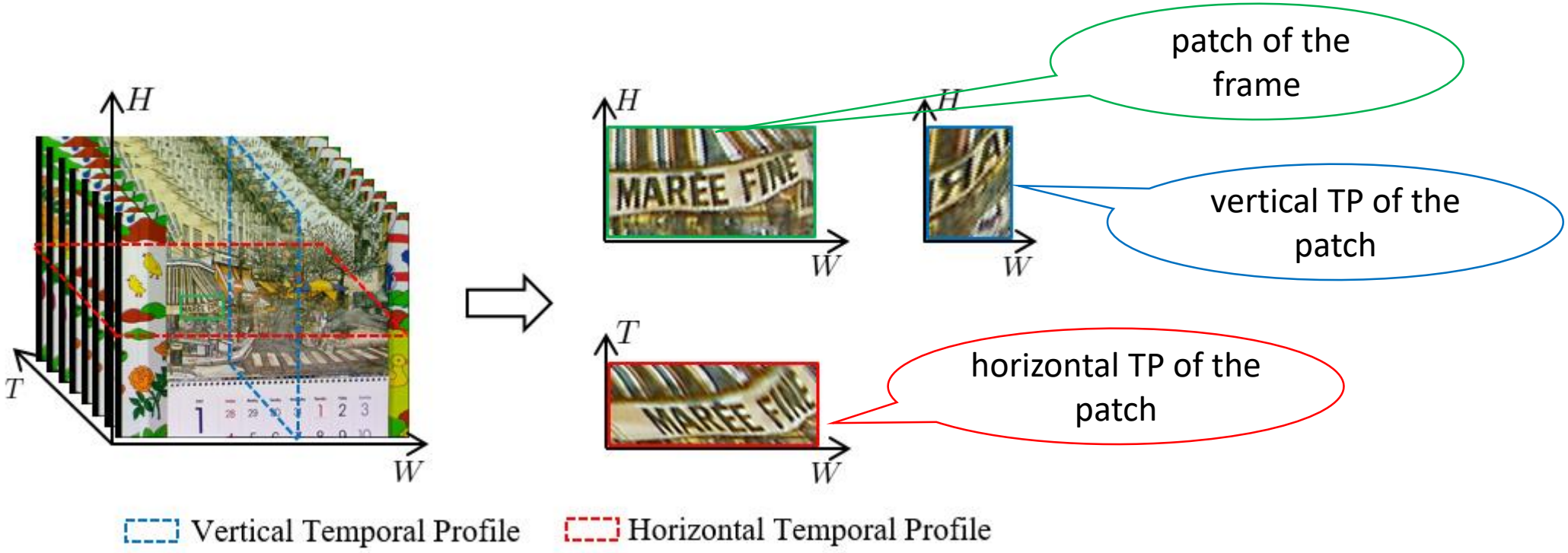
- EPI → Slice contains spatial-angular information in light field
- TP → Slice contains spatial-temporal information in video frames



Related Work

Epipolar-Plane Image (EPI) and Temporal Profile (TP)

- EPI → Slice contains spatial-angular information in light field
- TP → Slice contains spatial-temporal information in video frames



Motivation

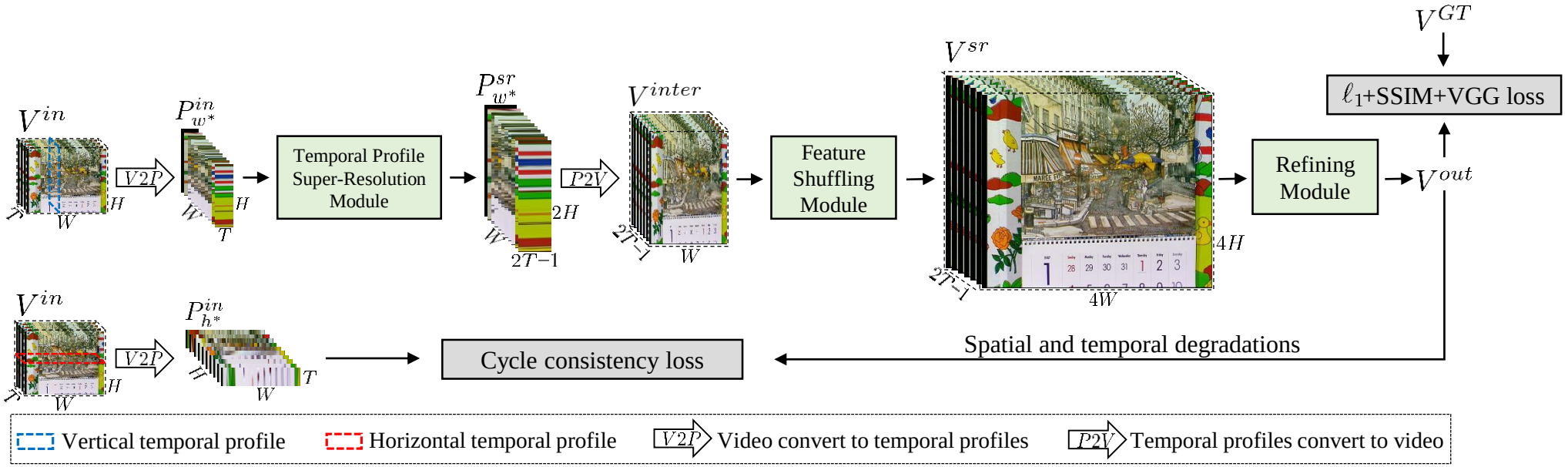
Observation

- Horizontal and vertical TPs maintain similar structures to those in the spatial domain

Benefition

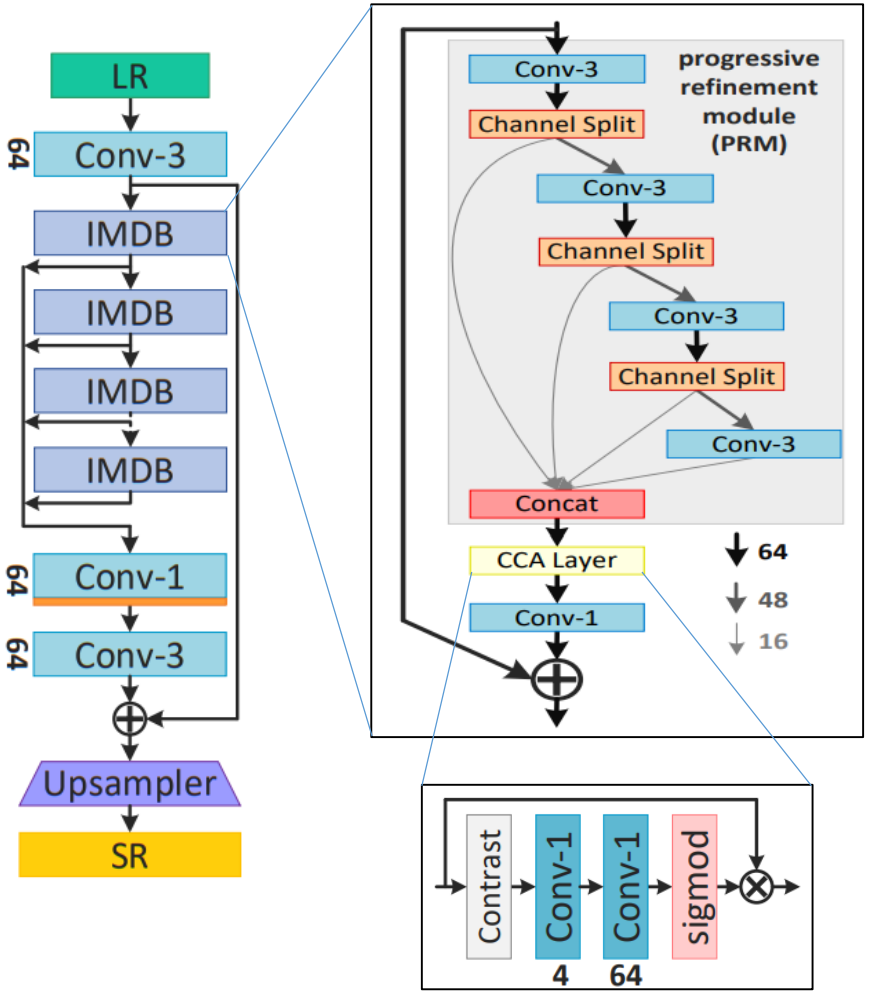
- STVSR can be modeled as a learning-based restoration task focusing on the specific 2D structure of TPs
- TPs contain both space and time dimensions, spatial-temporal correlation can be better exploited
- Longer-term temporal context can be integrated by TPs in a more flexible way

Proposed Method



Proposed Method

Temporal Profile Super-Resolution Module



Loss Function

$$\mathcal{L}_{\mathrm{TPSRM}}\left(\Theta_{tpsrm}\right)=\frac{1}{W_{train}}\sum_{w^*=1}^{W_{train}}\left\|N_{\mathrm{TPSRM}}\left(P_{w^*}^{in}\right)-P_{w^*}^{de}\right\|_1$$

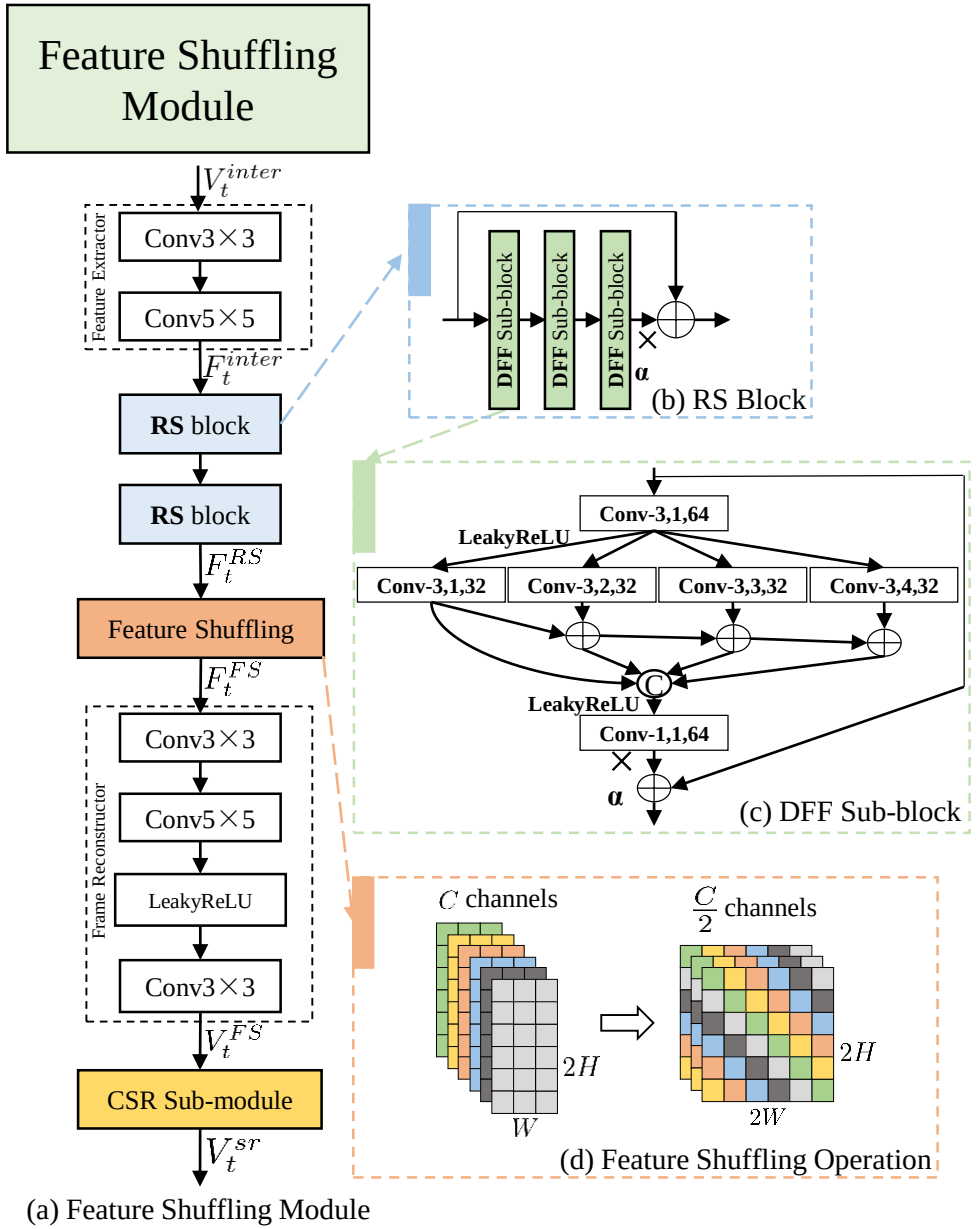
$\left\{P_{w^*}^{in},P_{w^*}^{de}\right\}_{w^*=1}^{W_{train}}$: training set

W_{train} : total number of extracted TPs from training video clips

$P_{w^*}^{de}\in\mathbb{R}^{2H\times(2T-1)}$: vertical TPs converted from GT training videos with degradation

Θ_{tpsrm} : learnable parameter set

Proposed Method



(a) Feature Shuffling Module

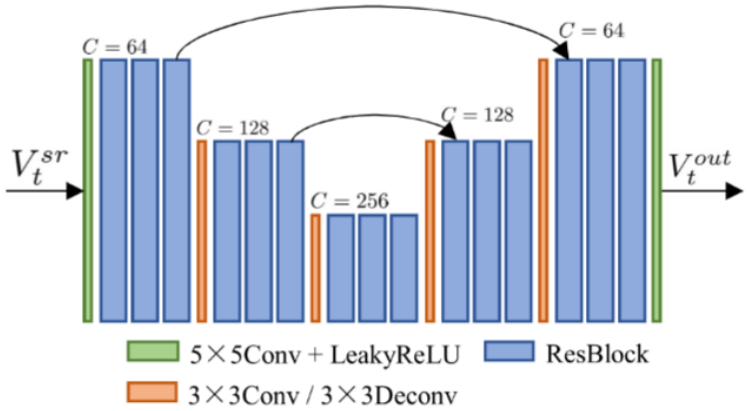
Loss Function

$$\mathcal{L}_{\mathrm{FSM}}\left(\Theta_{fsm}\right)=\frac{1}{T_{train}}\sum_{t=1}^{T_{train}}\left\|N_{\mathrm{FSM}}\left(V_t^{inter}\right)-V_t^{GT}\right\|_1$$

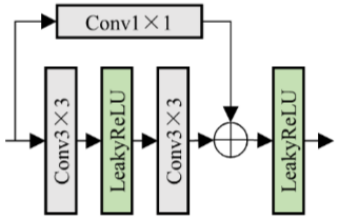
$\{\mathrm{V}_t^{inter},\mathrm{V}_t^{GT}\}_{t=1}^{T_{train}}$:	training set
T_{train}:	total number of frames used for training
$\mathrm{V}^{inter}\in\mathbb{R}^{W\times 2H\times(2T-1)}$:	video converted from SRed TPs
$V_t^{GT}\in\mathbb{R}^{4W\times 4H}$:	GT HFR and HR training video frames
Θ_{fsm} :	learnable parameter set

Proposed Method

Refining Module



(a) Refine Module



(b) ResBlock

Loss Function

$$\mathcal{L}_{\mathrm{RM}}\left(\Theta_{rm}\right)=\mathcal{L}_{\mathrm{RM}}^{\ell_1}+\lambda_1\mathcal{L}_{\mathrm{RM}}^{SSIM}+\lambda_2\mathcal{L}_{\mathrm{RM}}^{VGG}+\lambda_3\mathcal{L}_{\mathrm{RM}}^{Cycle}$$

$$\begin{cases}\mathcal{L}_{\mathrm{RM}}^{\ell_1}=\frac{1}{T_{train}}\sum_{t=1}^{T_{train}}\left\|N_{\mathrm{RM}}\left(V_t^{sr}\right)-V_t^{GT}\right\|_1\\ \mathcal{L}_{\mathrm{RM}}^{SSIM}=\frac{1}{T_{train}}\sum_{t=1}^{T_{train}}\left(1-\mathrm{SSIM}\left(N_{\mathrm{RM}}\left(V_t^{sr}\right),V_t^{GT}\right)\right)\\ \mathcal{L}_{\mathrm{RM}}^{VGG}=\frac{1}{T_{train}}\sum_{\mathrm{t}=1}^{T_{train}}\sum_{j=1,3,5}\left\|\phi_j\left(N_{\mathrm{RM}}\left(V_t^{sr}\right)\right)-\phi_j\left(V_t^{GT}\right)\right\|_2^2\\ \mathcal{L}_{\mathrm{RM}}^{Cycle}=\frac{1}{H_{train}}\sum_{h^*=1}^{H_{train}}\left\|P_{h^*}^{de}-P_{h^*}^{in}\right\|_1\end{cases}$$

$$V_t^{sr}=N_{\mathrm{FSM}}\left(V_t^{inter}\right),t=1,\ldots,2T-1$$

ϕ_j : j-th layer of the VGG-19network

$P_{h^*}^{de}\in\mathbb{R}^{W\times T}$: horizontal TPs converted from the reconstructed video after degradation

$P_{h^*}^{in}\in\mathbb{R}^{W\times T}$: horizontal TPs converted from the input video

H_{train} : total number of TPs used for training

Θ_{rm} : learnable parameter set

$\lambda_1,\lambda_2,\lambda_3$: weighting factors

Experiments

Comparison to State-of-the-art

Method		Vimeo90-Slow			Vimeo90K-Medium			Vimeo90K-Fast			Vid4			UCF101		
VFI (×2)	VSR (×4)	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
SepConv	IMDN	31.75	0.8851	7.6781	33.13	0.8986	7.7814	34.31	0.9177	8.5542	24.87	0.7150	6.3421	29.10	0.8790	7.6561
SepConv	SAN	32.12	0.8966	7.1001	33.59	0.9125	7.4623	34.97	0.9194	8.4790	24.93	0.7240	5.8864	29.80	0.8896	7.3087
SepConv	EDVR	32.97	0.9110	7.0023	34.25	0.9240	7.4016	35.51	0.9253	8.4753	25.93	0.7792	5.7024	30.19	0.8994	7.3915
DAIN	IMDN	31.84	0.8878	7.1319	33.39	0.9073	7.5839	34.74	0.9182	8.4278	24.93	0.7197	6.1853	29.57	0.8882	7.2996
DAIN	SAN	32.26	0.8993	7.0546	33.82	0.9249	7.4468	35.27	0.9244	8.4775	25.14	0.7301	5.7853	30.13	0.8990	7.3214
DAIN	EDVR	33.21	0.9126	7.0638	34.73	0.9283	7.3923	35.71	0.9307	8.4696	26.12	0.7856	5.6243	30.54	0.9001	7.4961
VSR (×4)	VFI (×2)	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
IMDN	SepConv	32.01	0.8867	7.6661	33.22	0.9016	7.6521	34.50	0.9181	8.5417	24.88	0.7155	6.3336	29.12	0.8801	7.4421
IMDN	DAIN	32.27	0.8916	6.9916	33.73	0.9167	7.1657	35.15	0.9206	8.4121	24.99	0.7227	6.2116	29.79	0.8901	7.2246
SAN	SepConv	32.32	0.9006	6.9912	33.73	0.9154	7.3151	35.33	0.9233	8.4221	25.01	0.7313	5.8714	29.92	0.8898	7.3151
SAN	DAIN	32.56	0.9113	6.8954	34.12	0.9284	7.4315	35.47	0.9246	8.3876	25.26	0.7515	6.1654	30.31	0.8992	7.2157
End-to-end Framework		PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
Zooming Slow-Mo		33.29	0.9127	6.9397	35.24	0.9347	7.3461	36.43	0.9337	8.4093	26.30	0.7975	5.6203	30.90	0.9095	7.2914
Ours		33.40	0.9217	6.1725	35.55	0.9358	6.3704	36.29	0.9322	7.1320	26.50	0.8182	5.4762	31.18	0.9119	6.8464

Experiments

Comparison to State-of-the-art

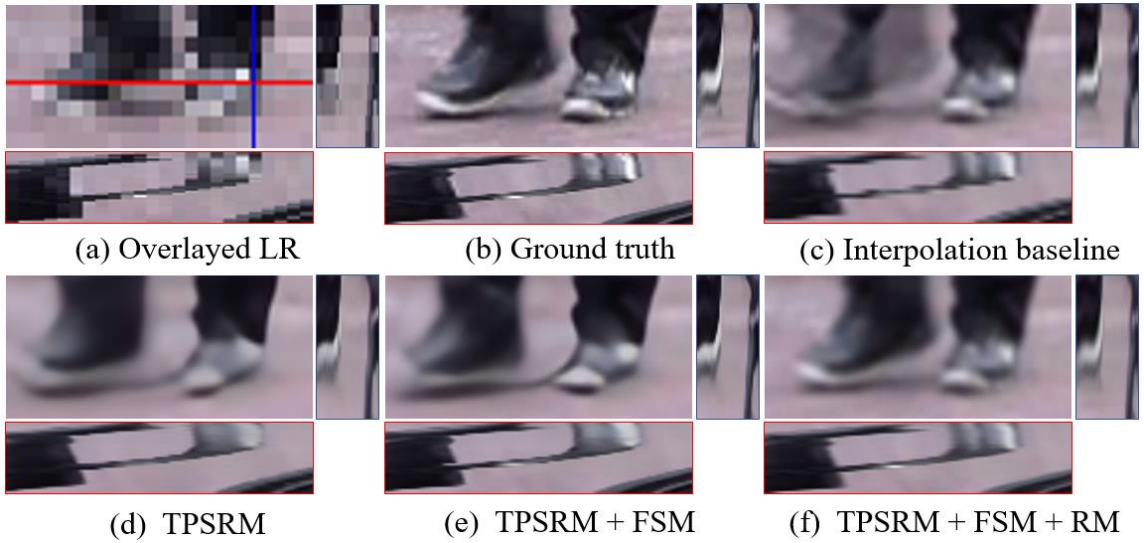
Experiments

Model parameters and average inference time on the Vid4 dataset with 1080Ti

Method	Parameters (Million)	Average Inference Time (s/frame)
DAIN [1] + EDVR [36]	24.0+20.7	0.8940
Zooming Slow-Mo [39]	11.10	0.1995
Ours	**7.53**	**0.1328**

Experiments

Ablation Study: Investigation of different modules



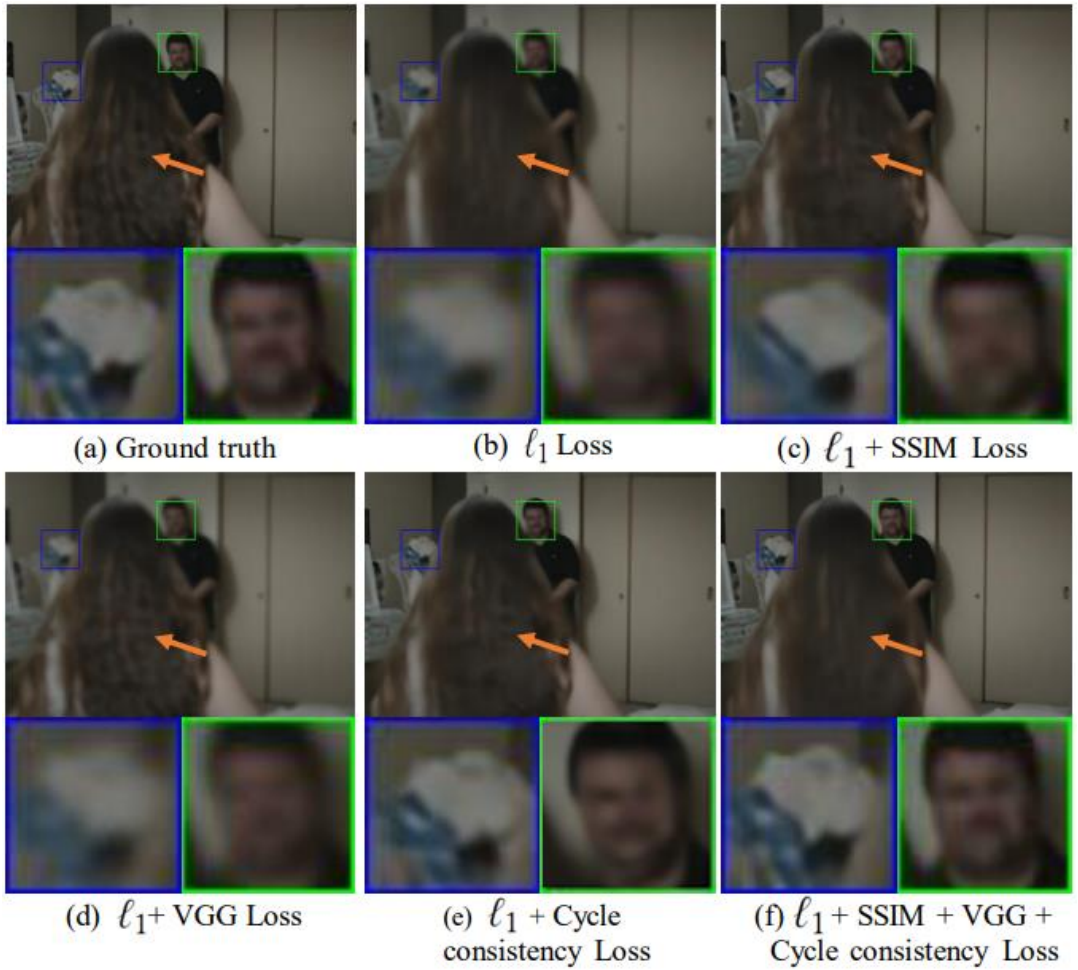
(a) Overlayed LR (b) Ground truth (c) Interpolation baseline

(d) TPSRM (e) TPSRM + FSM (f) TPSRM + FSM + RM

Method	TPSRM	FSM	RM	PSNR↑	SSIM↑
(a)	✗	✗	✗	24.62	0.7626
(b)	✓	✗	✗	25.41	0.7743
(c)	✓	✓	✗	25.97	0.7976
(d)	✓	✓	✓	**26.50**	**0.8182**

Experiments

Ablation Study: Investigation of different modules



(a) Ground truth (b) ℓ_1 Loss (c) ℓ_1 + SSIM Loss

(d) ℓ_1+ VGG Loss (e) ℓ_1 + Cycle consistency Loss (f) ℓ_1 + SSIM + VGG + Cycle consistency Loss

Loss functions setting	PSNR↑	SSIM↑	NIQE↓
ℓ_1	33.37	0.9177	7.1346
ℓ_1 + SSIM	33.39	0.9217	6.5150
ℓ_1 + VGG	33.37	0.9212	6.1064
ℓ_1 + Cycle	33.40	0.9216	6.4165
ℓ_1 + SSIM + VGG + Cycle	33.40	0.9217	6.1725

Application

Old Movie Restoration

(a) Zooming Slow-Mo (b) Ours

Limitations

(a) Overlayed LR (b) Ground truth

(c) Zooming Slow-Mo (d) Ours

ACM multimedia

中国科学技术大学
University of Science and Technology of China

Thanks for your listening!

Project page
http://home.ustc.edu.cn/~zeyuxiao/mm2020/TPVSR.html

VIDAR
Visual Information Discovery And Recovery

NEL-BITA
National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application