MMEditing

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Improving Video Super-Resolution with Enhanced Propagation and Alignment

Chen Change Loy

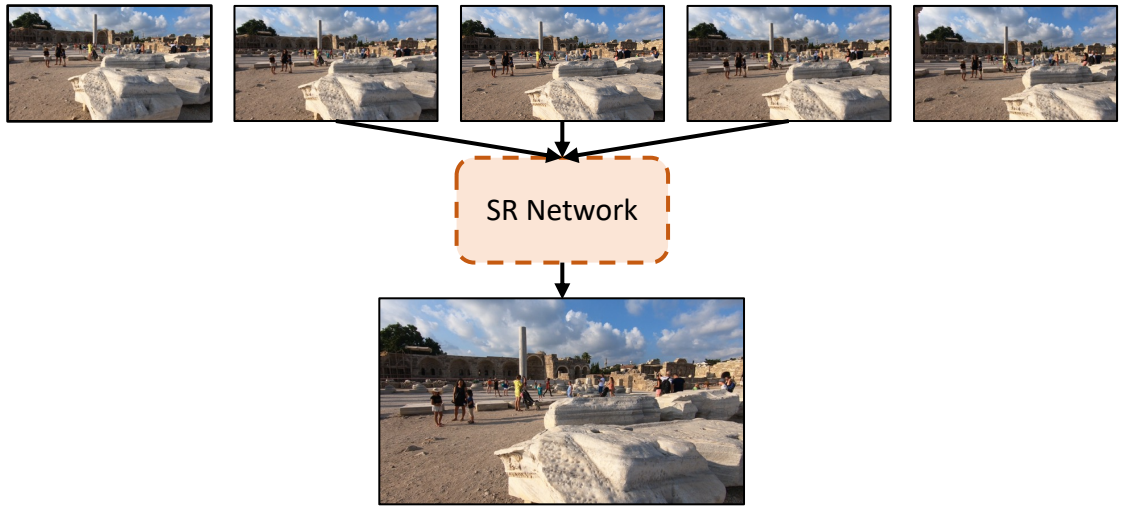
MMLab@NTU, Nanyang Technological University

Video Super-Resolution

Challenges:

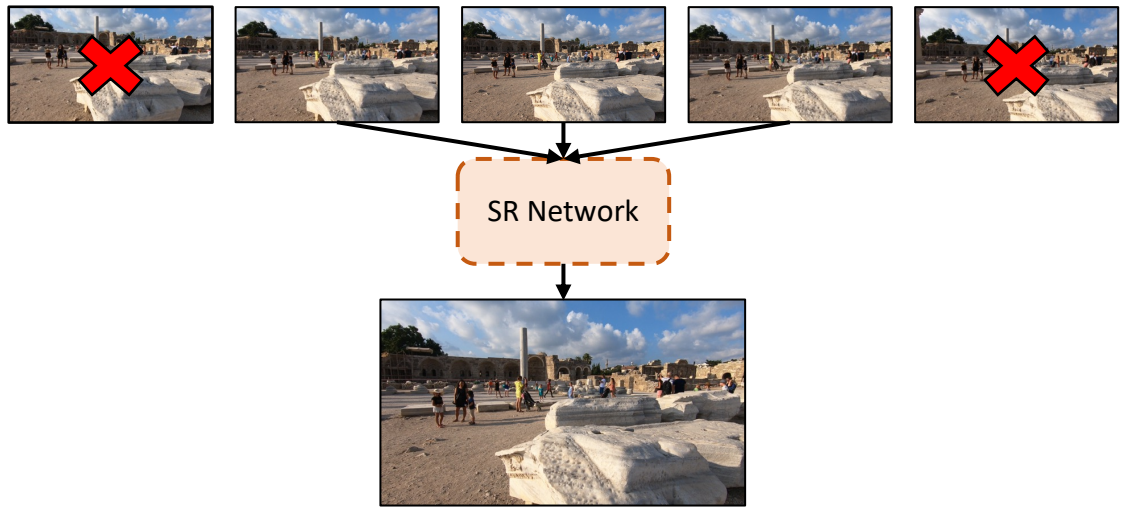
Aggregate information from highly-related but misaligned frames

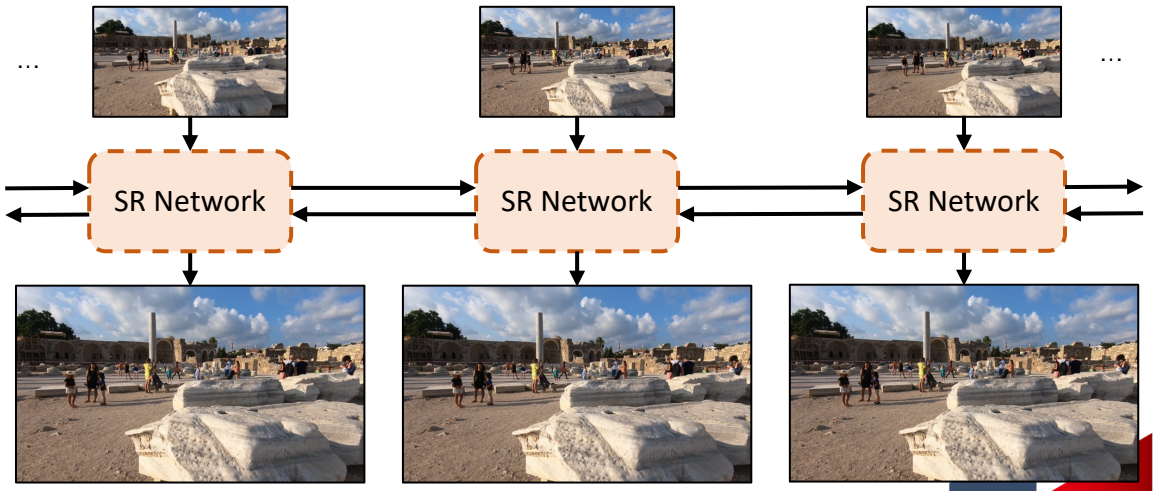
Diverse and complex degradations



Existing Work

- Sliding-window Framework
  - Restore frames using information within a local window
- Recurrent Framework
  - Propagate the information across the whole video sequence





Our Solutions

2021

BasicVSR
CVPR 2021

A simple backbone with good tradeoff in performance and efficiency

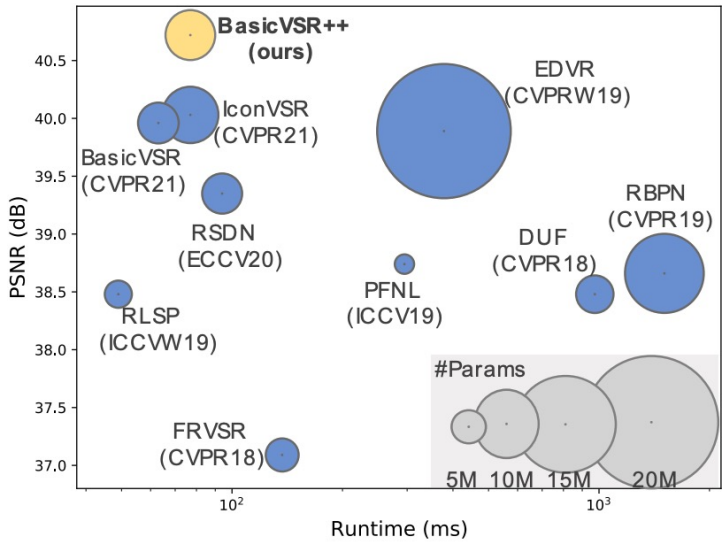
Effective propagation and alignment are essential in video super-resolution

2022

BasicVSR++
CVPR 2022, 2021 NTIRE Champions

Second-Order Grid Propagation
Effective aggregation of video information

Flow-Guided Deformable Alignment
Flexible alignment with motion guidance



RealBasicVSR
CVPR 2022

Increase the resolution of videos containing unknown degradations

Balance between detail synthesis and artifact suppression

Improve training efficiency

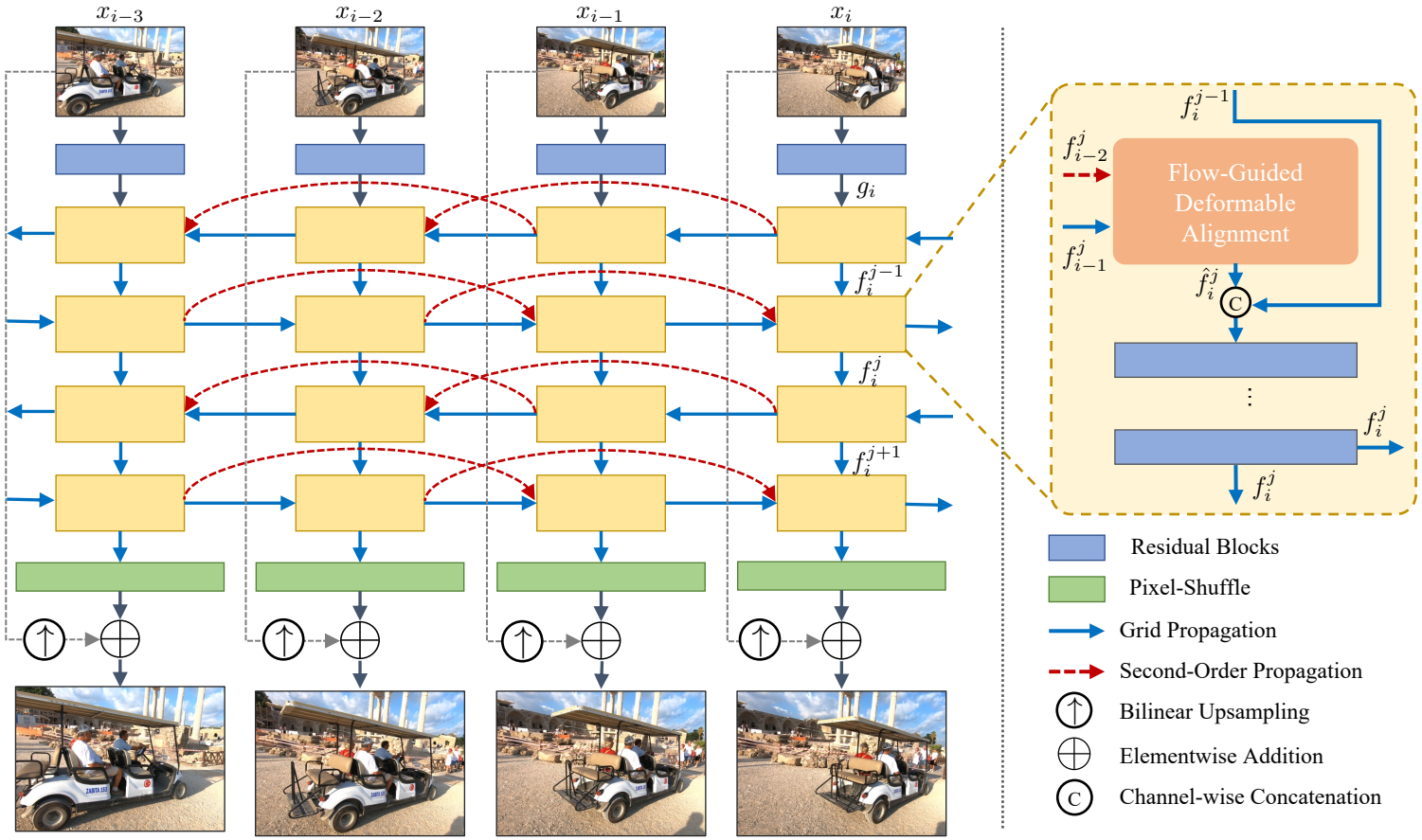
MMEditing

BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment

Kelvin C.K. Chan Shangchen Zhou Xiangyu Xu Chen Change Loy

MMLab@NTU, Nanyang Technological University

Overview of BasicVSR++



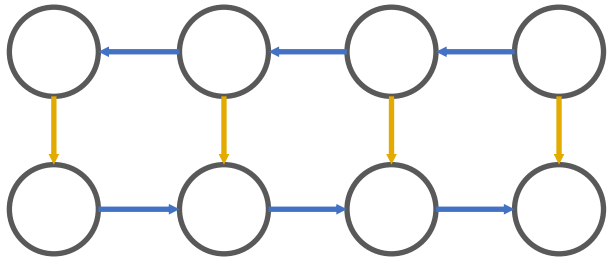
- Residual blocks to extract shallow features
- Second-Order Grid Propagation to effectively distribute information back-and-forth
- Flow-Guided Deformable Alignment for flexible alignment

Second-Order Grid Propagation

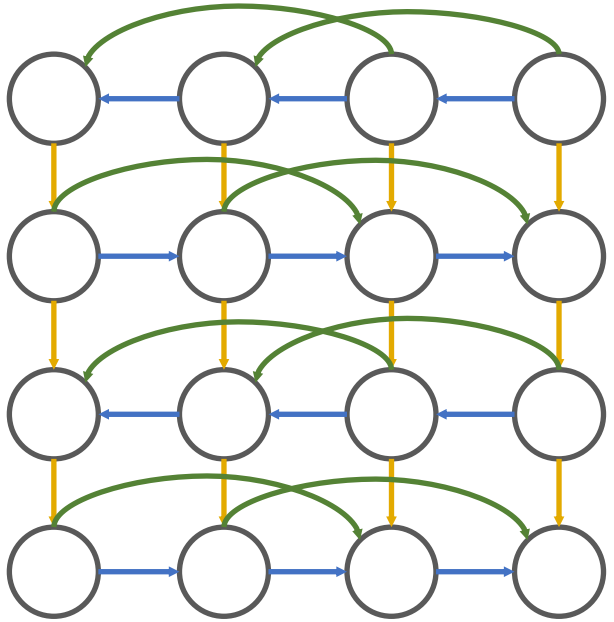
- Second-Order Propagation
  - Propagate information further
  - Aggregate information from multiple frames, improving robustness

- Grid Propagation
  - Propagation is beneficial to restoration
  - Features are propagated back-and-forth
  - Repeated refinement through propagation

- The more effective propagation helps to save parameters
  - BasicVSR++ surpasses BasicVSR by a significant 0.82 dB in PSNR with similar number of parameter



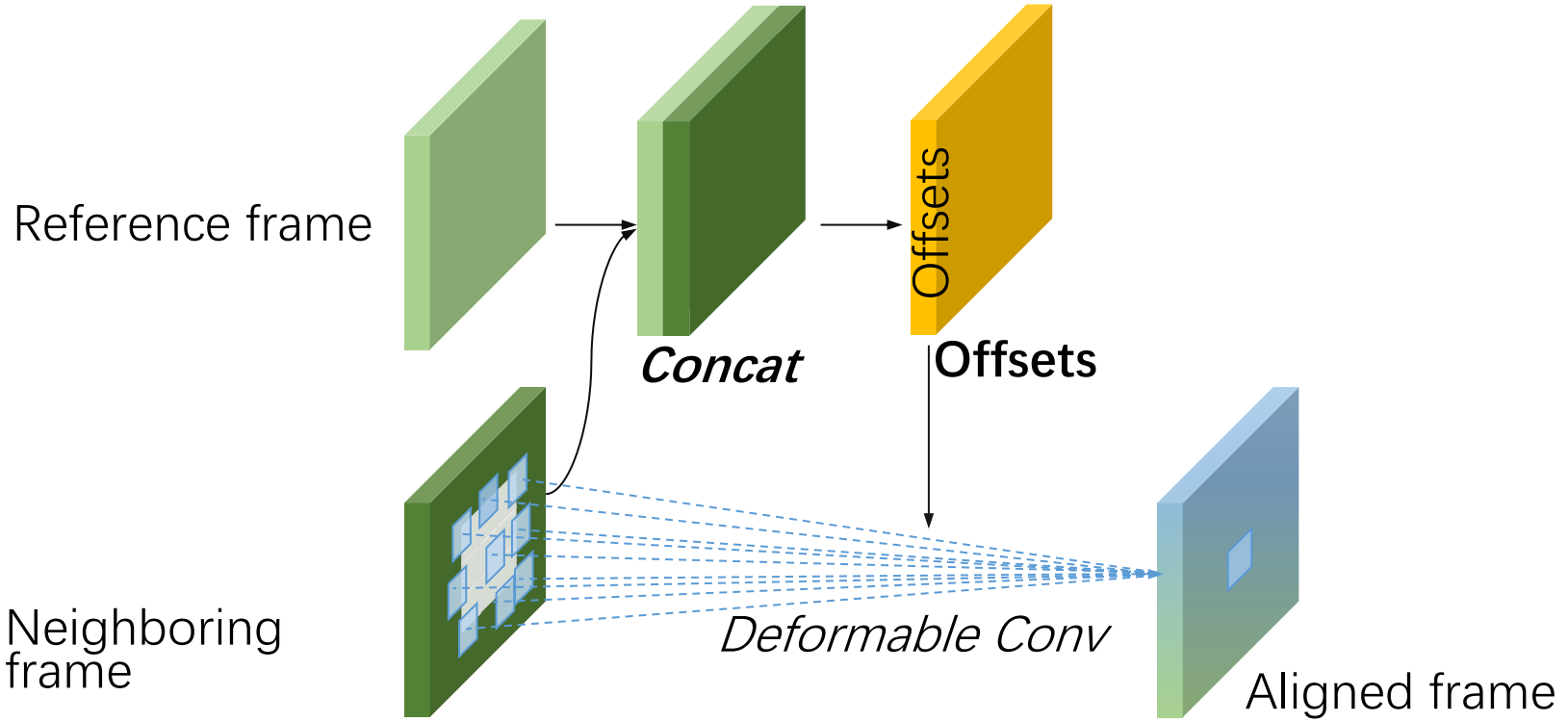
BasicVSR



BasicVSR++

Flow-Guided Deformable Alignment

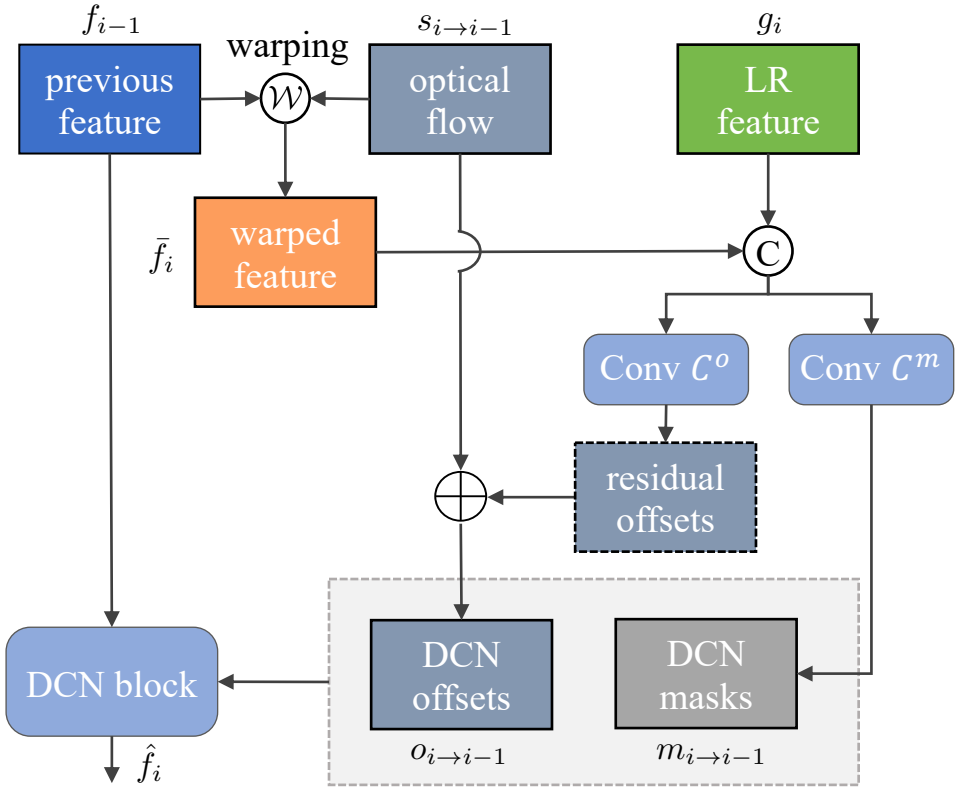
Deformable convolution for alignment - conventional method (e.g., in TDAN and EDVR)



[2] Chan et al., Understanding Deformable Alignment in Video Super-Resolution , AAAI, 2021

Flow-Guided Deformable Alignment

- Motivations
  - Deformable alignment is effective but unstable
  - DCN offsets are highly related to optical flow [2]
- Our approach:
  - Use optical flow as base offset
  - Learn residual offsets for DCN
- Advantages:
  - Feature pre-alignment eases offset learning
  - Learning only small deviation stabilizes training
  - Modulation masks act as attention, providing flexibility



[2] Chan et al., Understanding Deformable Alignment in Video Super-Resolution , AAAI, 2021

Ablations

- Each element brings >0.1 dB improvement
  - Notably, our flow-guided deformable alignment leads to 0.46 dB gain

- An improvement of 0.91 dB over the baseline

	(A)	(B)	(C)	**BasicVSR++**
Flow-Guided Deform. Align.		✓	✓	✓
Second-Order Propagation			✓	✓
Grid Propagation				✓
PSNR (dB)	31.48	31.94	32.08	32.39

Benefits of second-order grid propagation

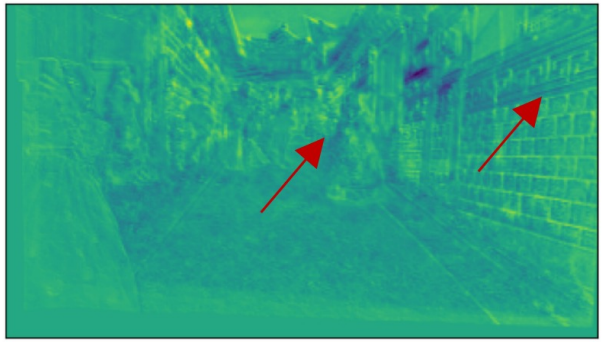
(a) Second-Order Propagation

(b) Grid Propagation

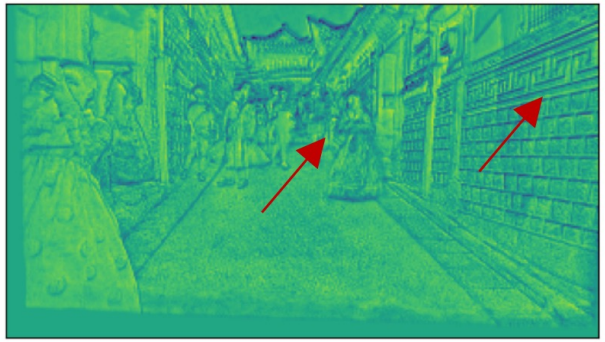
Benefits of flow-guided deformable alignment

Reference image

Neighboring image



Aligned by optical flow



Aligned by flow-guided deformable alignment


Low-Resolution Input
Output of BasicVSR++

Extension to Video Restoration

- Our original BasicVSR++ works for super-resolution
  - Input size is smaller than output size

- We introduce a simple extension
  - Downsample with strided convolutions if input size equals output size
  - The downsampling factor is selected based on needs
    - Larger -> faster
    - Smaller -> better performance

Experimental Results

- Significantly outperform existing works by a large margin
  - 1 dB improvement on Vid4 over IconVSR
  - NTIRE 2021 Champion

			BI degradation			BD degradation		
	Params (M)	Runtime (ms)	REDS4 [19]	Vimeo-90K-T [33]	Vid4 [18]	UDM10 [35]	Vimeo-90K-T [33]	Vid4 [18]
Bicubic	-	-	26.14/0.7292	31.32/0.8684	23.78/0.6347	28.47/0.8253	31.30/0.8687	21.80/0.5246
VESPCN [1]	-	-	-	-	25.35/0.7557	-	-	-
SPMC [25]	-	-	-	-	25.88/0.7752	-	-	-
TOFlow [33]	-	-	27.98/0.7990	33.08/0.9054	25.89/0.7651	36.26/0.9438	34.62/0.9212	-
FRVSR [22]	5.1	137	-	-	-	37.09/0.9522	35.64/0.9319	26.69/0.8103
DUF [14]	5.8	974	28.63/0.8251	-	-	38.48/0.9605	36.87/0.9447	27.38/0.8329
RBPN [8]	12.2	1507	30.09/0.8590	37.07/0.9435	27.12/0.8180	38.66/0.9596	37.20/0.9458	-
EDVR-M [29]	3.3	118	30.53/0.8699	37.09/0.9446	27.10/0.8186	39.40/0.9663	37.33/0.9484	27.45/0.8406
EDVR [29]	20.6	378	31.09/0.8800	37.61/0.9489	27.35/0.8264	39.89/0.9686	37.81/0.9523	27.85/0.8503
PFNL [35]	3.0	295	29.63/0.8502	36.14/0.9363	26.73/0.8029	38.74/0.9627	-	27.16/0.8355
MuCAN [16]	-	-	30.88/0.8750	37.32/0.9465	-	-	-	-
TGA [12]	5.8	-	-	-	-	-	37.59/0.9516	27.63/0.8423
RLSP [7]	4.2	49	-	-	-	38.48/0.9606	36.49/0.9403	27.48/0.8388
RSDN [11]	6.2	94	-	-	-	39.35/0.9653	37.23/0.9471	27.92/0.8505
RRN [13]	3.4	45	-	-	-	38.96/0.9644	-	27.69/0.8488
BasicVSR [3]	6.3	63	31.42/0.8909	37.18/0.9450	27.24/0.8251	39.96/0.9694	37.53/0.9498	27.96/0.8553
IconVSR [3]	8.7	70	31.67/0.8948	37.47/0.9476	27.39/0.8279	40.03/0.9694	37.84/0.9524	28.04/0.8570
VSR-Tran [2]	32.6	4312	31.06/0.8815	37.71/0.9494	27.36/0.8258	-	-	-
BasicVSR++	7.3	77	**32.39/0.9069**	**37.79/0.9500**	**27.79/0.8400**	**40.72/0.9722**	**38.21/0.9550**	**29.04/0.8753**

Experimental Results

- Deblurring
  - 1.48 and 1.81 dB improvements on DVD over ARVo when downsampling the input 4x and 2x, respectively

	EDVR [30]	Tao *et al.* [26]	Su *et al.* [24]	DBLRNet [35]	STFAN [36]	Xiang *et al.* [32]	TSP [22]	Suin *et al.* [25]	ARVo [15]	**BasicVSR++$_{\downarrow 4}$**	**BasicVSR++$_{\downarrow 2}$**
PSNR	28.51	29.98	30.01	30.08	31.15	31.68	32.13	32.53	32.80	34.28	**34.61**
SSIM	0.864	0.884	0.888	0.885	0.905	0.916	0.827	0.947	0.935	0.951	**0.954**

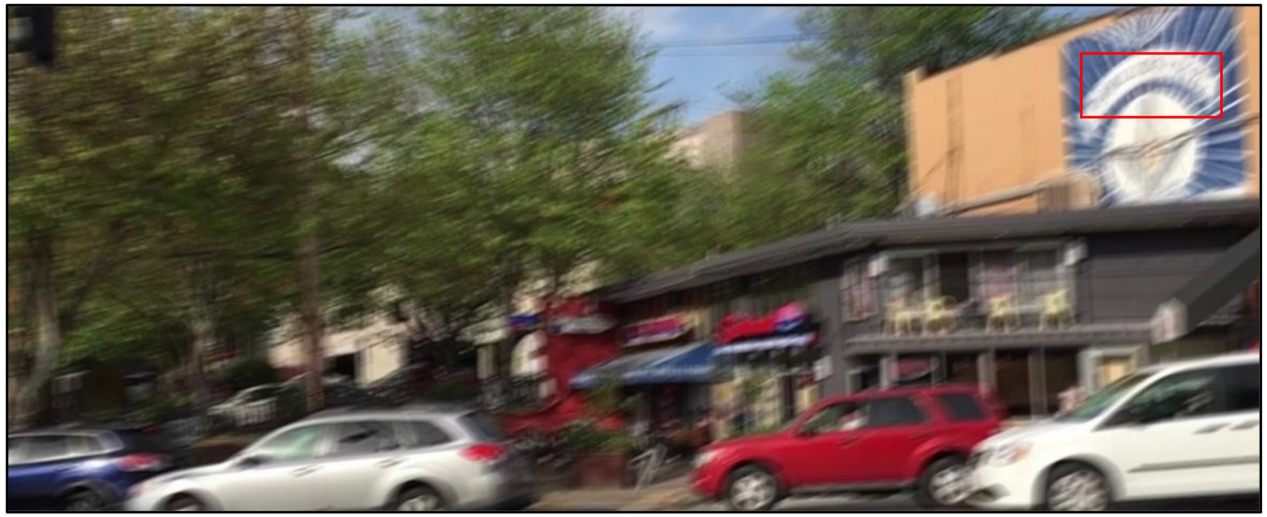
- Denoising
  - 0.62 and 1.97 dB on DAVIS over PaCNet when downsampling the input 4x and 2x, respectively

	VBM4D [19]	VNLB [1]	DVDnet [27]	FastDVDnet [28]	VNLNet [9]	PaCNet [29]	**BasicVSR++$_{\downarrow 4}$**	**BasicVSR++$_{\downarrow 2}$**	Δ
σ=10	37.58/-	38.85/-	38.13/0.9657	38.71/0.9672	39.56/0.9707	39.97/0.9713	40.13/0.9754	**40.97/0.9786**	1.00/0.0073
σ=20	33.88/-	35.68/-	35.70/0.9422	35.77/0.9405	36.53/0.9464	37.10/0.9470	37.41/0.9598	**38.58/0.9666**	1.48/0.0196
σ=30	31.65/-	33.73/-	34.08/0.9188	34.04/0.9167	-	35.07/0.9211	35.74/0.9457	**37.14/0.9560**	2.07/0.0349
σ=40	30.05/-	32.32/-	32.86/0.8962	32.82/0.8949	33.32/0.8996	33.57/0.8969	34.49/0.9321	**36.06/0.9459**	2.49/0.0490
σ=50	28.80/-	31.13/-	31.85/0.8745	31.86/0.8747	-	32.39/0.8743	33.45/0.9179	**35.18/0.9358**	2.79/0.0615
Average	32.39/-	34.34/-	34.52/0.9195	31.64/0.9188	-	35.62/0.9221	36.24/0.9462	**37.59/0.9566**	1.97/0.0345

- Compressed Video Enhancement
  - 2 champions in NTIRE 2021, with 4x downsampling

[3] Suin et al., Gated spatio-temporal attention-guided video deblurring, CVPR, 2021
[4] Vaksman et al., Patch Craft: Video denoising by deep modeling and patch matching, ICCV, 2021

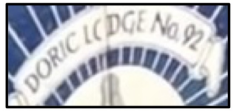
Experimental Results



Input

STFAN

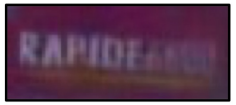
TSP



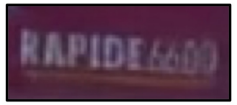
BasicVSR++ (ours)

GT

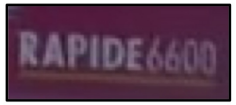
Input



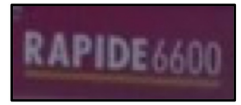
VBM4D



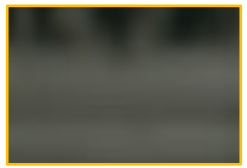
FastDVDnet



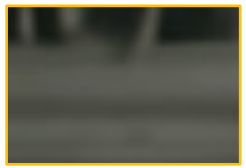
BasicVSR++ (ours)



GT



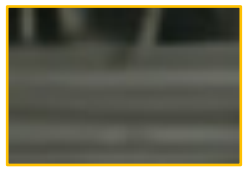
Bicubic



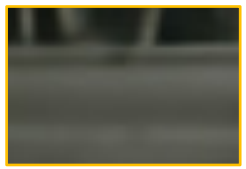
RBPN

EDVR-M

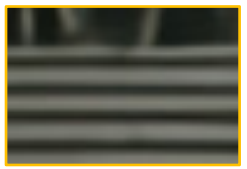
EDVR



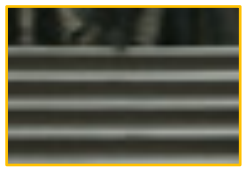
BasicVSR



IconVSR



BasicVSR++ (ours)



GT

Summary

- We build upon BasicVSR and proposes BasicVSR++
  - Second-Order Grid Propagation – Effective aggregation of video information
  - Flow-Guided Deformable Alignment – Flexible alignment with motion guidance
- Further propose an extension to various video restoration tasks
  - Downsample with strided convolution for efficiency
  - Downsampling factor is determined by the user for speed-performance tradeoff
- State-of-the-art performance on
  - Video super-resolution
  - Video deblurring
  - Video denoising
  - Compressed video enhancement

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Investigating Tradeoffs in Real-World Video Super-Resolution

Kelvin C.K. Chan Shangchen Zhou Xiangyu Xu Chen Change Loy

S-Lab, Nanyang Technological University

Introduction

- Recurrent networks are effective in synthetic settings (e.g., bicubic), but not necessarily useful in real-world videos
- Training recurrent networks for real-world video super-resolution requires prohibitive resources
- No diverse datasets for evaluation

Image Pre-Cleaning

- In non-blind settings, using more frames => better
- In real-world settings, artifacts are amplified with more frames
- Networks are unable to distinguish high-frequency details from artifacts
  - Artifacts are exaggerated with longer propagation

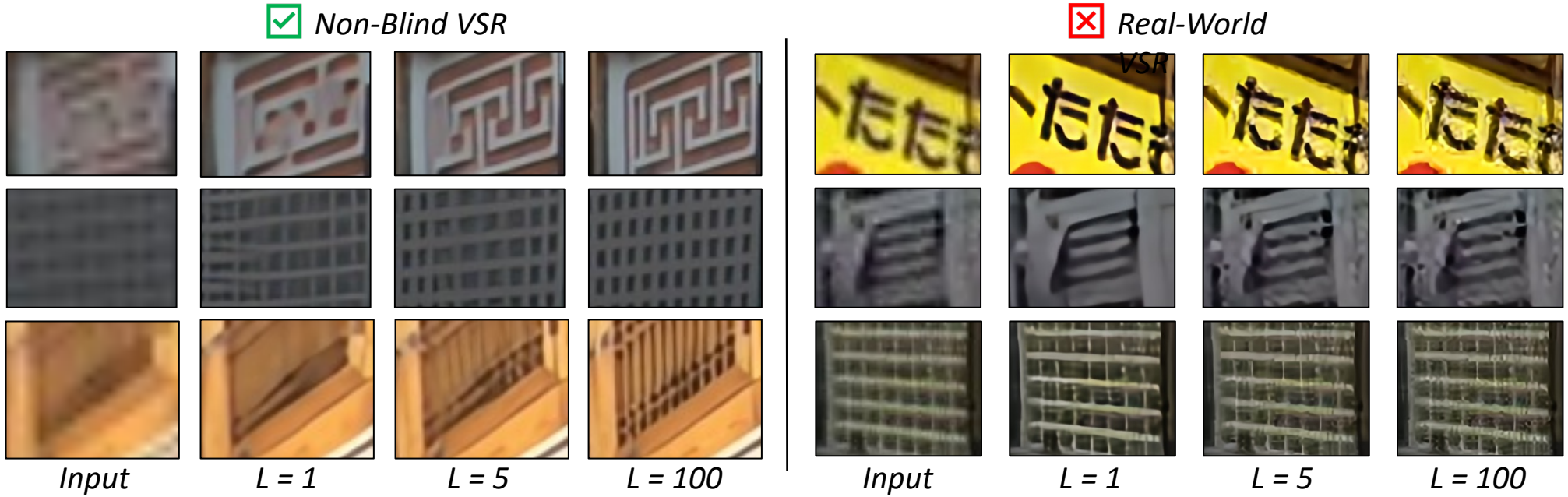


Image Pre-Cleaning

- Reduce the artifacts before propagation
- Two-stage: Cleaning + Details Synthesis

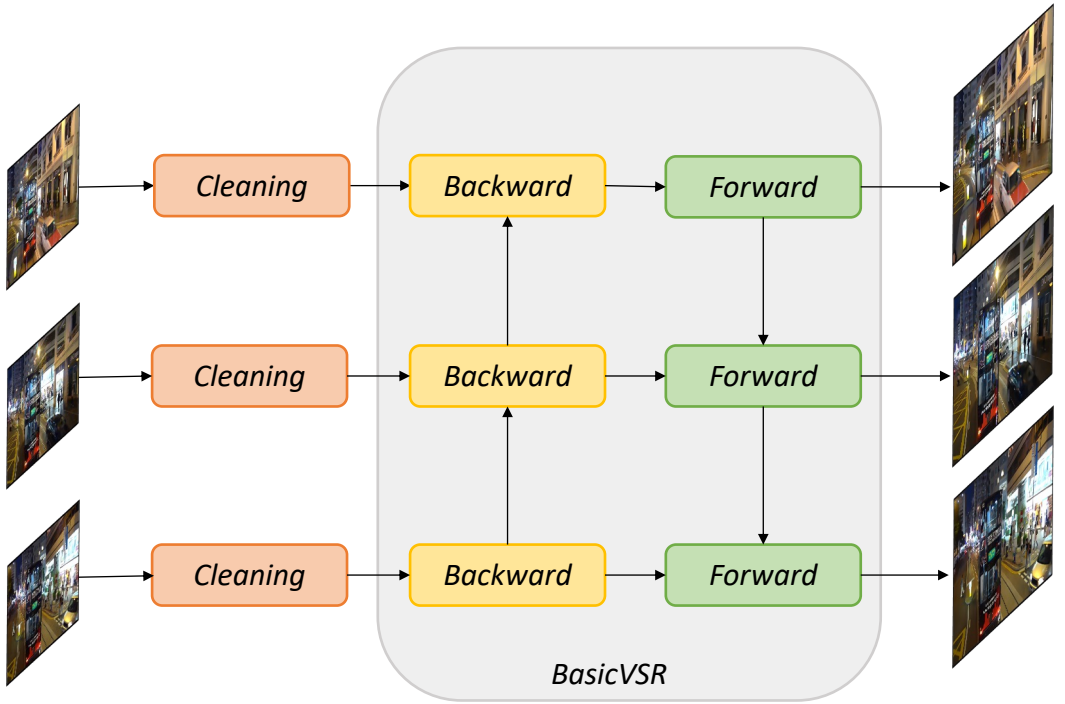
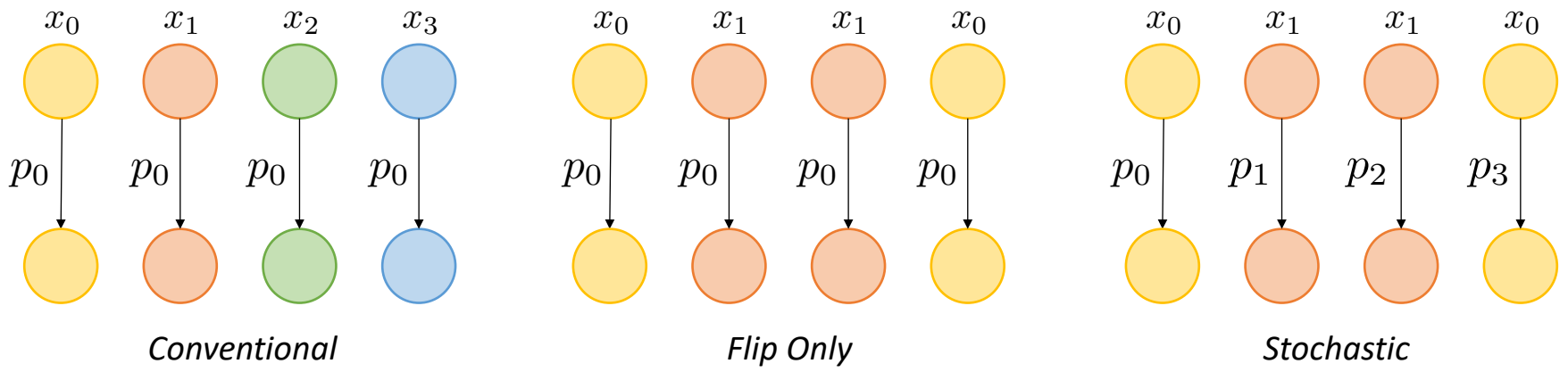


Image Pre-Cleaning

- A cleaning loss is used to guide the cleaning module
  - Using a low-resolution ground-truth
  - Simple MSE loss suffices
  - Essential for removing artifacts

Stochastic Degradation Scheme

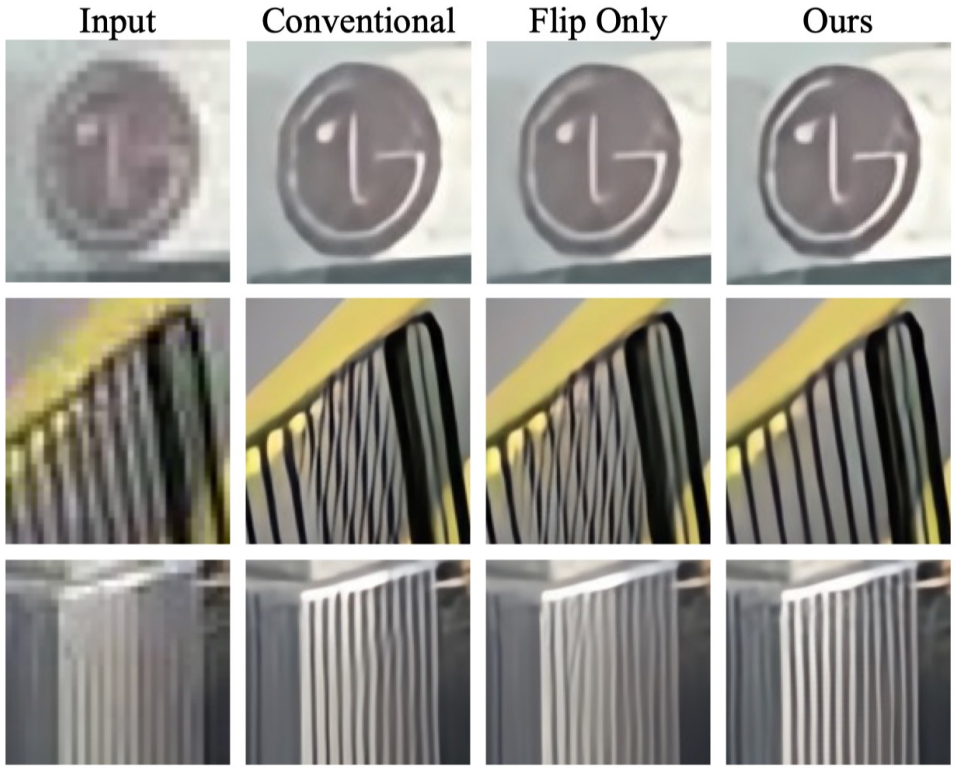
- Training recurrent networks require long sequences
- Training real-world models need larger batch size to stabilize training
- Our Approach:
  - Flip the sequence: Load half of the frames -> same length but smaller CPU load
  - Random-walk degradations: Increase degradation diversity



Conventional *Flip Only* *Stochastic*

Stochastic Degradation Scheme

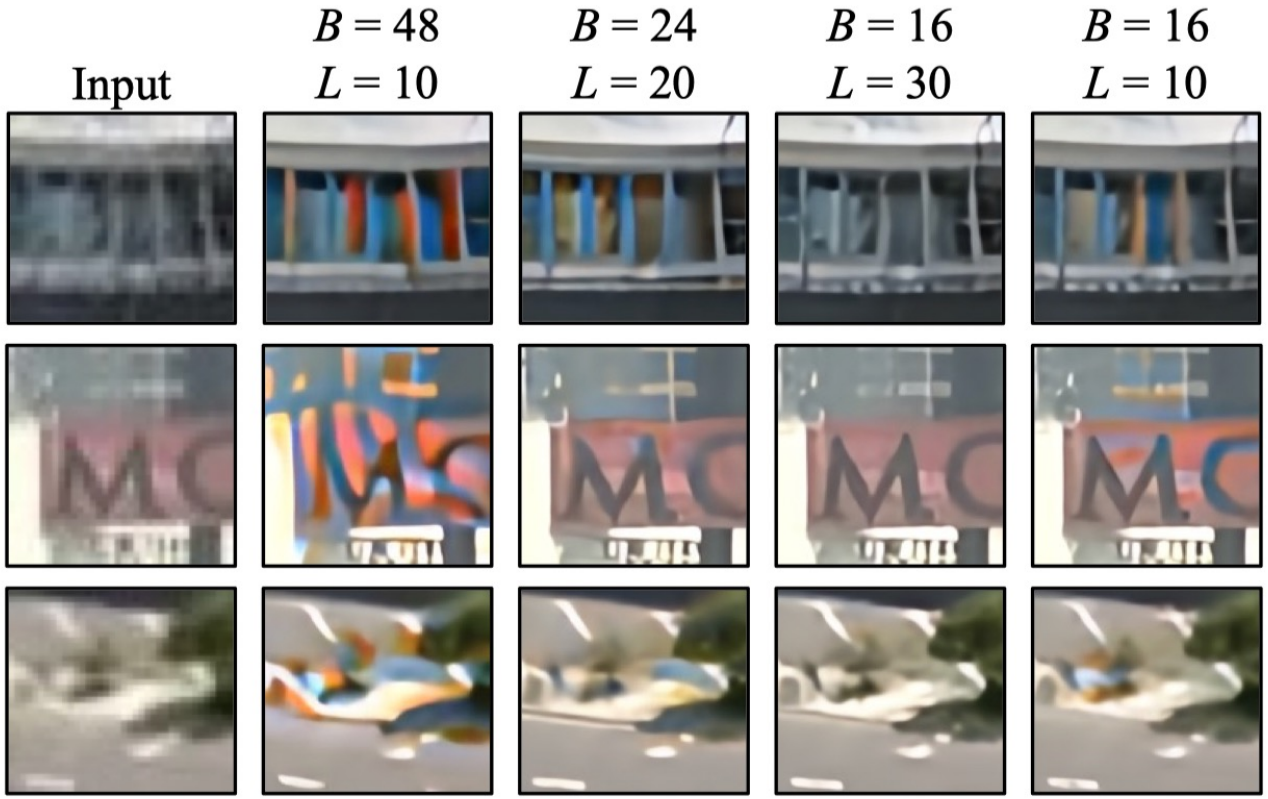
- *40% faster training speed without sacrificing output quality*



	Time per iteration ↓	NIQE ↓
Conventional Scheme	~2.5s	4.7191
Flip Only	~1.5s	4.6926
Stochastic Degradation	~1.5s	4.6836

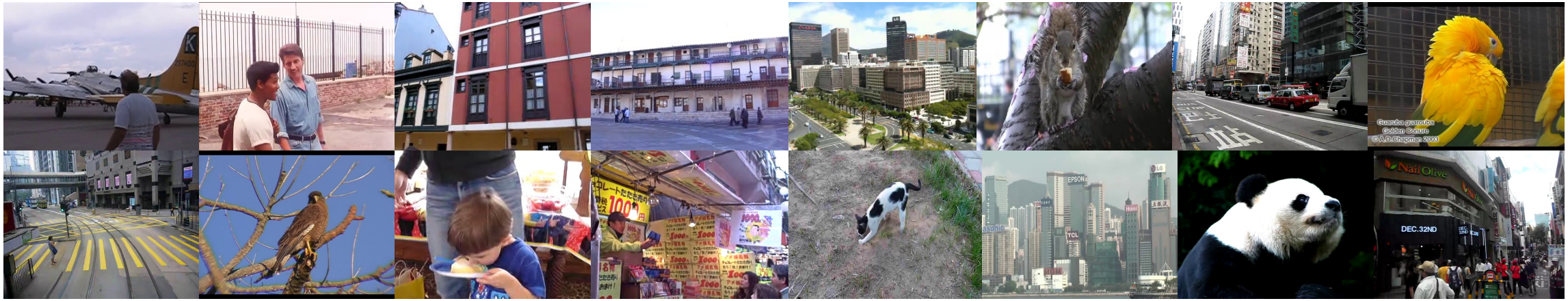
Batch Size vs. Sequence Length

- Go for longer sequences rather than larger batch size if resources are limited



VideoLQ Dataset

- Test dataset consists of 50 low-quality videos
- Extracted from video-hosting sites (e.g., YouTube, Flickr)
- Evaluate both performance and generalizability



Experimental Results

- Only RealBasicVSR is able to restore the fine details
  - Aggregate information from other frames through long-term propagation

Experimental Results

- RealBasicVSR outperforms existing works in non-reference metrics

	Bicubic	BasicVSR++ [6]	RealVSR [41]	DAN [27]	DBVSR [33]	BSRGAN [44]	Real-ESRGAN [36]	RealSR [18]	**RealBasicVSR**
Params (M)	-	7.3	**2.7**	4.3	25.5	16.7	16.7	16.7	6.3
Runtime (ms)	-	77	1082	185	239	149	149	149	**63**
NRQM [28] ↑	2.8016	3.5646	2.4958	3.3346	3.4097	5.7172	5.7108	5.6187	**6.0477**
NIQE [31] ↓	8.0049	6.3662	8.0606	7.1230	6.7866	4.2460	4.2091	4.1482	**3.7662**
PI [1] ↓	7.6017	6.4008	7.7824	6.8942	6.6885	4.2644	4.2492	4.2648	**3.8593**
BRISQUE [30] ↓	54.899	50.841	54.988	51.563	50.936	30.213	32.103	30.542	**29.030**


RealBasicVSR
Input

Summary

- Unlike non-blind VSR, real-world VSR induces various tradeoffs
  - Detail synthesis vs. Artifact exaggeration
  - Performance vs. Training time
  - Batch size vs. Sequence Length
- We study the tradeoffs and provide respective discussion
  - With image pre-cleaning, RealBasicVSR restores details with amplifying artifacts
  - Stochastic degradation reduces training time by 40%
  - Longer length is preferred when resources are limited
- VideoLQ dataset for evaluation

Code

BasicVSR++

BasicVSR++ (extension)

RealBasicVSR

MMLab@NTU

Mobile Intelligent Photography & Imaging Workshop 2022

MIPI Workshop in conjunction with ECCV 2022, Tel-Aviv, Israel (Live on Zoom)

October, 2022

Start Challenge

http://mipi-challenge.org/

Five challenge tracks that emphasize the integration of novel image sensors and image algorithms

Top industry and academic speakers

Michael S. Brown
Professor of York University
Director of Samsung AI Center

Peyman Milanfar
Principal Scientist of Google Research

Mohit Gupta
Assistant Professor of University of Wisconsin-Madison

Wolfgang Heidrich
Professor of KAUST Visual Computing Center

Tomoo Mitsunaga
Manager of Sony Europe B. V.

The workshop's main focus is on MIPI, emphasizing the integration of novel image sensors and imaging algorithms. Together with the workshop, we organize a few exciting challenges and invite renowned researchers from both industry and academia to share their insights and recent work. Our challenge includes five tracks:

- **RGB+ToF Depth Completion** uses sparse, noisy ToF depth measurements with RGB images to obtain a complete depth map.
- **Quad-Bayer Re-mosaic** converts Quad-Bayer RAW data into Bayer format so that it can be processed with standard ISPs.
- **RGBW Sensor Re-mosaic** converts RGBW RAW data into Bayer format so that it can be processed with standard ISPs.
- **RGBW Sensor Fusion** fuses Bayer data and a monochrome channel data into Bayer format to increase SNR and spatial resolution.
- **Under-display Camera Image Restoration** improves the visual quality of image captured by a new imaging system equipped under-display camera.

Unlike previous workshops that focus on image or video manipulation, restoration and enhancement, or the efficient designs of AI models for mobile devices, the central theme of our workshop encompasses new sensors and imaging systems, which are the indispensable foundation for mobile intelligent photography and imaging. As the first workshop of this kind, MIPI aims to organize a dedicated workshop so that we can solicit relevant solutions and attract a focused group from both academia and industry for fruitful discussions.

Datasets and Submission

Datasets are available at the Codalab site of each challenge track. Submissions to all phases will be done through the CodaLab site. Please register to the site and refer to the instructions on how to download the datasets and submit your results. The evaluation metrics of each track will be introduced in the respective site.

Link to Codalab: **RGB+ToF Depth Completion** – **Quad-Bayer Re-mosaic** – **RGBW Sensor Re-mosaic** – **RGBW Sensor Fusion** – **Under-display Camera Image Restoration**

Awards and Prizes

The winner teams of each track will receive a certificate and be awarded a cash prize. Challenge participants with the most successful and innovative methods will be invited to present at the MIPI workshop. The cash prize for each track:

- First place: USD 1000
- Second place: USD 800
- Third place: USD 500

Important Dates

Challenge

Workshop

Event	Date (Always 11:59 PM Pacific Time)
Site online	Apr 08, 2022
Release of training data and validation data	May 15, 2022
Validation server online	May 15, 2022
Release of test data, test server online	Jun 25, 2022
Test results submission deadline, test server closed	Jul 20, 2022
Fact sheets submission deadline	Jul 20, 2022
Final test and rating results release to participants	Jul 30, 2022

http://mipi-challenge.org/

Important Dates

Challenge

Workshop

Event	Date (Always 11:59 PM Pacific Time)
Site online	Apr 08, 2022
CMT online	Apr 25, 2022
Paper submission deadline	Aug 08, 2022
Notification to authors	Aug 18, 2022
Camera ready deadline	Aug 22, 2022
Workshop date	TBA

http://mipi-challenge.org/