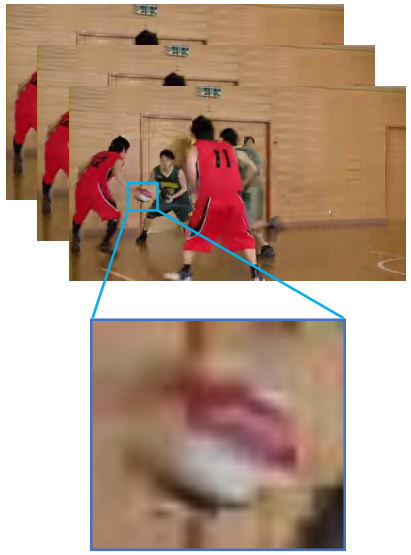
NTIRE 2022 Challenge on Super-Resolution and Quality Enhancement of Compressed Video

Ren Yang and Radu Timofte

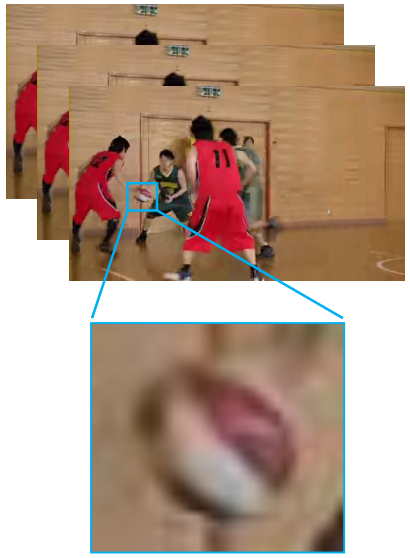
VOYAGE 81

Background

video compression
(H.264, HEVC, VVC, etc.)



Compressed video
(quality loss)



After enhancement

Compressed video
Low resolution
(poor quality)

Enhanced video
High resolution

Challenge

Track 1: Quality Enhancement

Compression: HM 16.20, LDP configuration, QP = 37

```
path_to_HM16.20/bin/TAppEncoderStatic
-c path_to_HM16.20/cfg/encoder_lowdelay_P_main.cfg
-c path_to_HM16.20/cfg/per-sequence/BasketballDrill.cfg
-i xxx.yuv -q 37 -wdt (width) -hgt (height) -f (frame_num) -fr (frame_rate)
-b xxx.mkv
```

Evaluation: **PSNR (used for ranking)**, MS-SSIM

Track 2/3: Super-Resolution and Quality Enhancement (Track 2: x2 SR, Track 3: x4 SR)

Downsampling:

Downsample the videos by x2 bicubic (Track 2) or x4 bicubic (Track 3) using the commands below. We define `w` and `h` as the width and height after downsampling. In Track 1, `w = width, h = height`. In Track 2, `w = width/2, h = height/2`. In Track 3, `w = width/4, h = height/4`

```
ffmpeg -pix_fmt yuv420p -s (width)x(height) -i xxx.yuv -vf scale=(w)x(h):flags=bicubic xxx.yuv
```

Compression: HM 16.20, LDP configuration, QP = 37

Evaluation: **PSNR (used for ranking)**, MS-SSIM

LDV 2.0 Dataset

LDV 2.0 (335 videos) = LDV (240 videos) + 95 videos

- Collecting

The videos in LDV 2.0 are collected from YouTube. To ensure the high quality, we only collect the videos with 4K resolution, and without obvious compression artifacts. All source videos used for our LDV 2.0 dataset have the licence of *Creative Commons Attribution licence (reuse allowed).*

- Downscaling

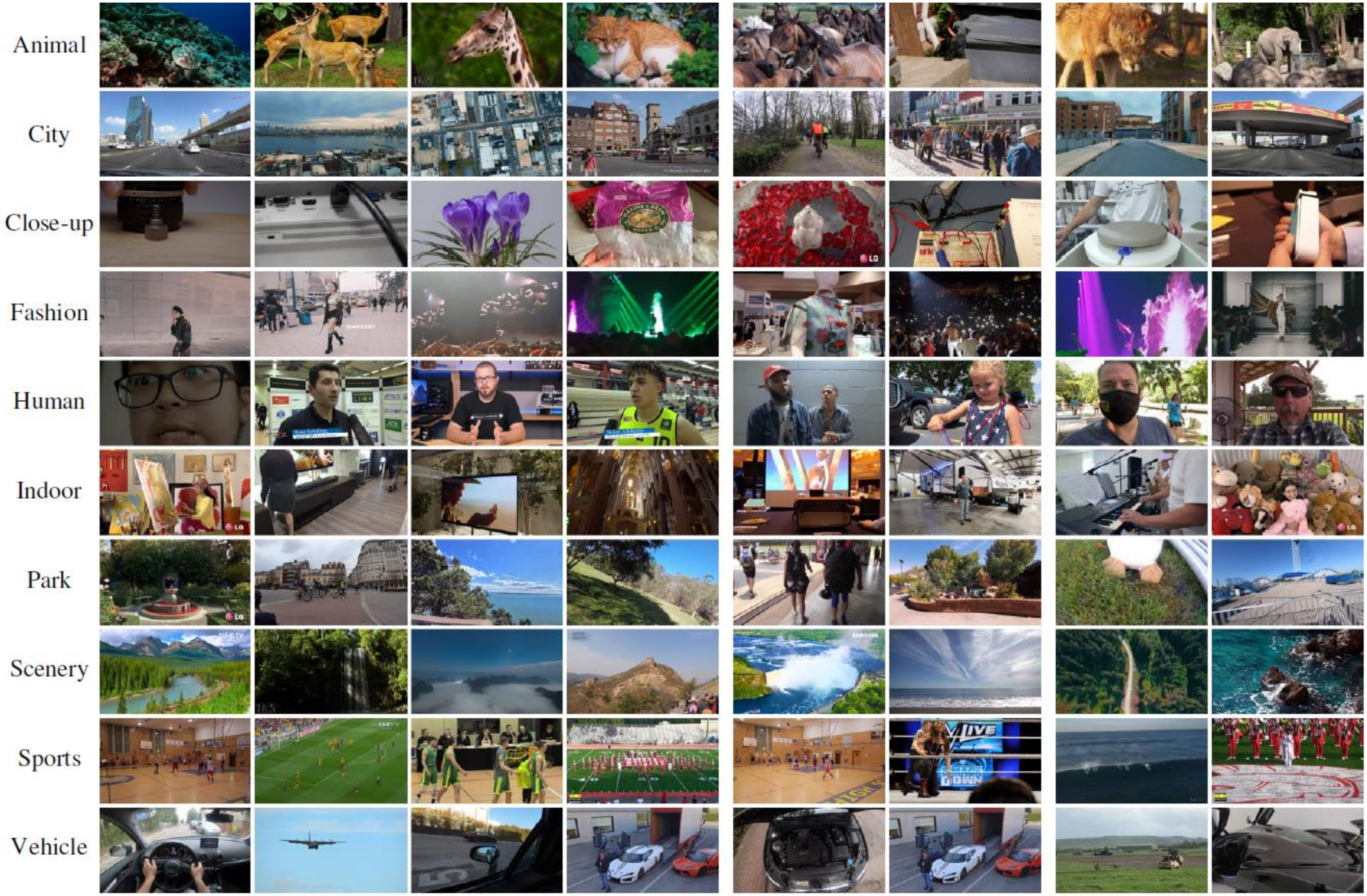
The videos are downscaled to remove compression artifacts, and the width and height of each video are cropped to the multiples of 8, due to the requirement of the HEVC test model (HM). We follow the standard datasets, *e.g.*, JCT-VC, to convert videos to the format of YUV 4:2:0.

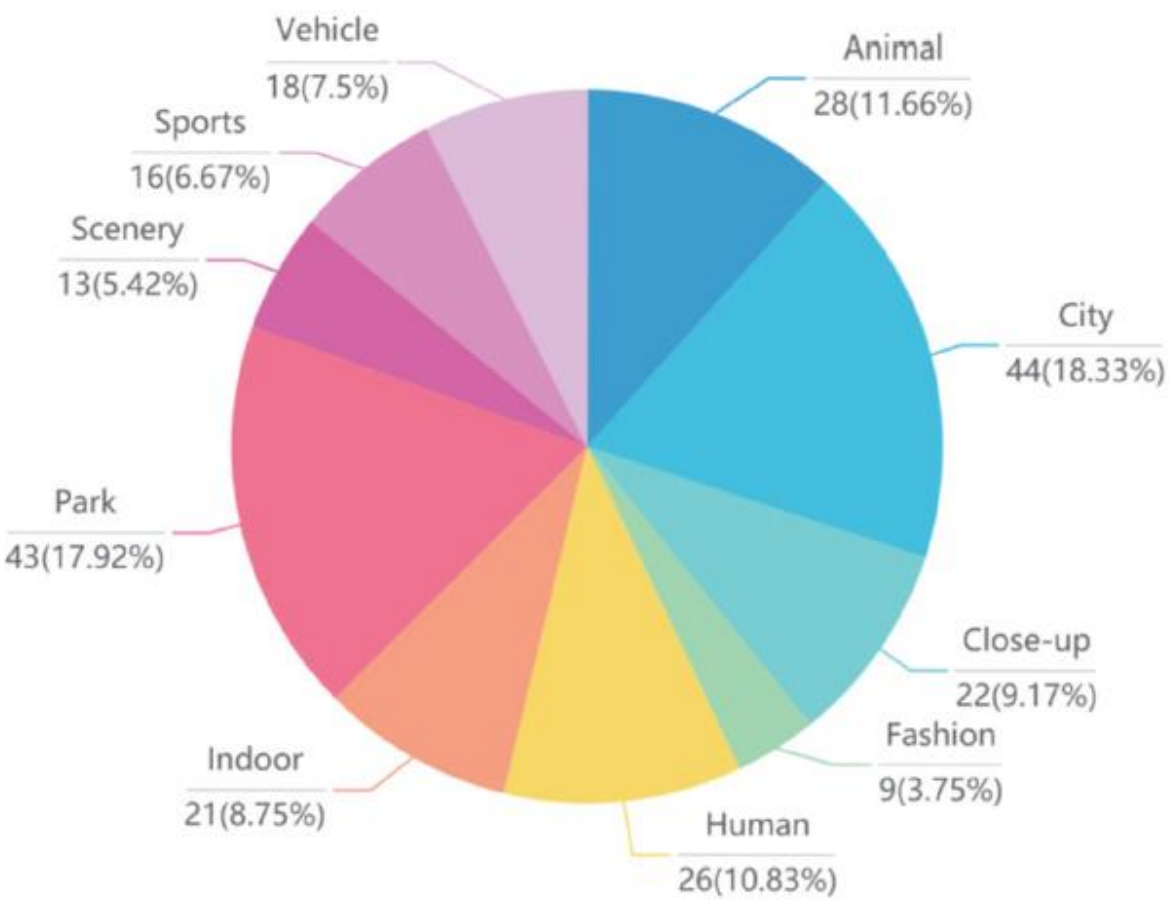
- Diversity

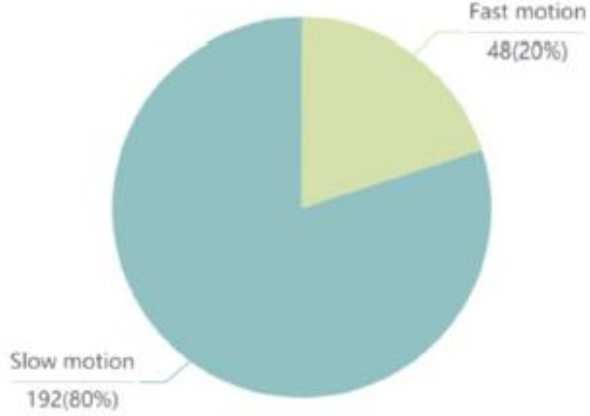
The LDV 2.0 dataset contains 10 categories of scenes, i.e., animal, city, close-up, fashion, human, indoor, park, scenery, sports and vehicle, and with different kinds of motion and large range of frame-rates.

- Partition

Training set: LDV dataset (240 videos)
Track 1 validation: 15 videos
Track 1 test: 15 videos
Track 2 validation: 15 videos
Track 2 test: 15 videos
Track 3 validation: 15 videos
Track 3 test: 15 videos

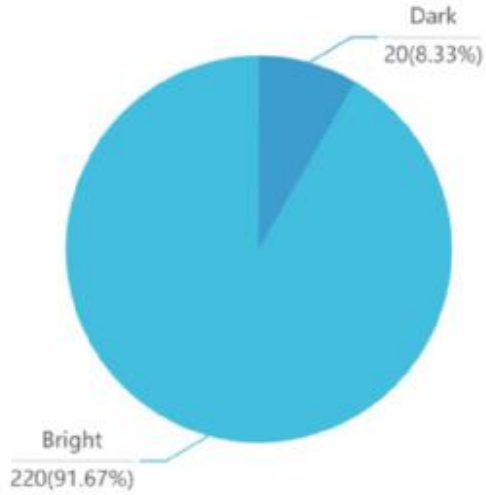

Animal
City
Close-up
Fashion
Human
Indoor
Park
Scenery
Sports
Vehicle


Vehicle
18(7.5%)
Animal
28(11.66%)
Sports
16(6.67%)
Scenery
13(5.42%)
City
44(18.33%)
Park
43(17.92%)
Close-up
22(9.17%)
Fashion
9(3.75%)
Indoor
21(8.75%)
Human
26(10.83%)


Fast motion
48(20%)
Slow motion
192(80%)


High frame-rate
68(28.33%)
Low frame-rate
172(71.67%)


Shaky
75(31.25%)
Steady
165(68.75%)


Dark
20(8.33%)
Bright
220(91.67%)

Results: Track 1

Table 1. Results of Track 1 (quality enhancement). Blue indicates the state-of-the-art method.

Team	PSNR (dB)															
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	Ave.
TaoMC2	34.20	32.71	31.59	37.12	**32.75**	30.81	**30.79**	**34.58**	**32.13**	**37.13**	**27.35**	**24.42**	**30.60**	**33.38**	31.54	**32.07**
GY-Lab	**34.23**	**32.90**	**31.61**	**37.32**	32.70	**30.83**	30.77	34.55	32.06	36.87	27.34	24.28	30.58	33.34	**31.56**	32.06
HIT&ACE	34.07	32.68	31.47	37.10	32.60	30.73	30.70	34.38	31.98	36.69	27.27	24.38	30.54	33.20	31.40	31.94
BOE-IOT-AIBD	33.96	32.58	31.45	36.90	32.45	30.63	30.54	34.25	31.91	36.51	27.16	24.05	30.47	33.12	31.33	31.82
OCL-VCE	33.81	32.54	31.30	36.82	32.28	30.59	30.47	34.07	31.75	36.28	27.08	24.00	30.37	33.01	31.25	31.71
BasicVSR++ [10]	33.73	32.42	31.22	36.75	32.16	30.57	30.41	33.99	31.68	36.20	27.06	23.93	30.24	32.94	31.20	31.63
OREO	33.64	32.38	31.16	36.80	32.08	30.55	30.37	34.02	31.63	36.12	27.02	23.98	30.30	32.91	31.10	31.60
UESTC+XJU CV	33.57	32.33	31.04	36.60	31.97	30.46	30.24	33.76	31.47	35.96	26.94	23.77	30.19	32.77	31.05	31.47
AVRT	33.19	31.19	30.14	35.59	31.37	29.87	29.88	33.15	31.05	35.79	26.81	23.61	29.39	32.32	29.77	30.88
Compressed video	32.43	30.18*	29.05	34.31	30.49	28.99	28.84	31.80	29.86	34.65	26.30	22.87	28.42	31.21	29.03	29.90

* The frames with MSE = 0 are excluded when calculating the average PSNR.

Results: Track 2

Table 2. Results of Track 2 (quality enhancement and ×2 super-resolution)

Team	PSNR (dB)															
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	Ave.
TaoMC2	**27.71**	**24.11**	**26.53**	**31.30**	**31.84**	**27.11**	**25.13**	**24.20**	**28.76**	**32.16**	**26.39**	**27.83**	**26.27**	**24.73**	**29.17**	**27.55**
GY-Lab	27.56	23.89	26.47	31.26	31.69	27.03	25.05	24.12	28.61	32.15	26.34	27.71	26.08	24.61	**29.17**	27.45
HIT&ACE	27.43	23.79	26.38	31.13	31.42	26.96	24.94	23.91	28.38	32.01	26.21	27.61	25.95	24.47	29.06	27.31
ZX_VIP	27.46	23.77	26.42	31.03	31.50	26.83	24.91	24.00	28.46	31.69	26.23	27.61	25.97	24.55	29.00	27.30
Trick collector	27.44	23.90	26.40	31.02	31.38	26.94	25.00	22.28	28.34	31.86	26.17	27.58	26.03	24.49	29.01	27.19
HMSR	27.18	23.54	26.26	30.85	31.22	26.69	24.68	23.84	28.28	31.61	26.13	27.45	25.67	24.34	28.68	27.09
TBE	26.87	23.23	26.12	30.53	30.76	26.36	24.36	23.65	27.92	31.27	25.93	27.23	25.28	24.07	28.34	26.80
AVRT	26.58	22.88	25.92	30.21	30.39	26.23	24.01	23.43	27.68	30.77	25.77	26.94	24.70	23.68	27.95	26.48
Bicubic ×2	25.55	22.03	25.53	29.12	29.33	25.20	23.17	22.67	26.54	29.65	25.12	26.25	23.92	22.76	26.83	25.58

Results: Track 3

Table 3. Results of Track 3 (quality enhancement and ×4 super-resolution). Blue indicates the state-of-the-art method.

Team	PSNR (dB)															
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	Ave.
GY-Lab	27.13	21.22	**25.15**	**24.78**	**26.80**	24.39	21.93	**24.32**	**28.05**	22.33	22.24	**24.06**	20.88	**25.66**	24.49	**24.23**
TaoMC2	27.20	**21.24**	25.14	24.71	26.72	24.19	21.98	24.31	28.04	**22.35**	22.24	**24.06**	20.93	25.63	24.49	24.22
NoahTerminalCV	**28.40**	20.84	24.61	23.91	26.44	**25.66**	**22.06**	23.83	26.79	21.86	**23.81**	23.75	**21.56**	24.90	**24.61**	24.20
HIT&ACE	27.06	21.10	24.99	24.49	26.33	23.93	21.83	24.18	27.74	22.18	22.13	23.95	20.75	25.48	24.30	24.03
XPixel	27.02	21.10	24.96	24.58	26.43	24.04	21.80	24.15	27.48	22.15	22.14	23.98	20.81	25.42	24.26	24.02
Trick collector	26.75	21.15	25.01	24.01	26.23	23.85	21.04	24.16	27.66	22.12	22.10	23.88	20.69	25.38	24.26	23.88
HyperPixel	26.72	20.87	24.68	23.82	25.92	23.49	21.46	23.90	27.16	21.90	22.04	23.66	20.51	25.12	23.88	23.68
CVStars	26.75	20.88	24.72	23.66	25.90	23.71	21.52	23.81	26.94	21.86	22.16	23.77	20.50	25.00	23.90	23.67
AVRT	26.63	20.75	24.39	23.58	25.72	23.48	21.39	23.69	26.64	21.76	22.07	23.54	20.41	24.98	23.70	23.52
StarRay	26.46	20.66	24.17	22.97	24.15	23.22	21.19	23.37	26.01	21.49	22.10	23.41	20.22	24.39	23.48	23.15
CDVSR [13]*	26.20	20.64	24.03	22.61	24.90	23.02	21.02	23.24	25.88	21.35	21.97	23.18	20.13	23.89	23.40	23.03
Modern_SR	26.38	20.49	23.90	22.76	24.81	22.93	21.09	23.13	25.56	21.29	21.92	23.27	20.14	24.18	23.31	23.01
TUK-IKLAB	26.19	20.55	23.74	22.37	24.17	22.39	20.80	22.84	25.10	21.07	21.94	23.05	19.89	23.35	23.12	22.70
Bicubic ×4	26.04	20.52	23.53	22.14	24.24	22.06	20.58	22.63	24.80	20.89	21.91	22.92	19.78	23.32	23.06	22.56

* The CDVSR [13] method only enhances the Y channel and upsamples the U and V channels by the bicubic algorithm.

Results: time complexity, hardware, test strategies and training data

Table 4. The time complexity, hardware, test strategies and training data of the proposed methods (reported by the participants).

Team	Running time (s) per frame			Hardware	Ensemble / Fusion	Extra training data
	Track 1	Track 2	Track 3			
TaoMC2	44.1	44.1	13.0	Tesla V100	Flip/rotation x8, two models	870 videos from YouTube [28]
GY-Lab	6.9	4.6	11.5	Tesla V100	Spatial-temporal ensemble and several models	REDS [49], Vimeo90K [69], YouTube [28]
HIT&ACE	17.30	10.97	17.40	Tesla V100	Flip/rotation x8, two models	540 samples from YouTube [28]
NoahTerminalCV	-	-	150	Tesla V100	Flip/rotation x8, five networks	90,000 videos from YouTube [28]
BOE-IOT-AIBD	1.61	-	-	Tesla V100	Flip/rotation x8	-
ZX_VIP	-	12	-	Tesla V100	Flip/rotation x8	REDS
OCL-VCE	28.72	-	-	Tesla T4	Flip/rotation x8	-
Trick collector	-	2.56	3.2	Tesla A100	Flip/rotation x6/x8, model voting	REDS [49]
XPixel	-	-	13.02	Tesla A100	Flip/rotation x8	REDS [49], Vimeo90K [69] and 2174 clips
OREO	19.4			Tesla A40	Flip/rotation x8	-
HMSR		14.36		Tesla A100	Flip/rotation x8	1274 additional from Youtube [28]
UESTC+XJU CV	0.16	-	-	GeForce RTX 3090	-	-
TBE	-	0.90	-	Tesla V100	-	91 videos
HyperPixel	-	-	0.44	Tesla V100	Flip/rotation x8	-
CVStars	-	-	10	Tesla V100	Flip/rotation x8, epoch-ensemble	-
AVRT	27	8	2	Tesla A100	Flip/rotation x4	202 videos
StarRay	-	-	4.0	GeForce RTX 2080 Ti	Two models with different loss	-
Modern_SR	-	-	0.86	GeForce RTX 3080	-	-
TUK-IKLAB	-	-	$\leq$1.0	GeForce RTX 3090	-	-

Methods

Preliminary: BasicVSR++ (Chan et al., in CVPR 2022, Winner of NTIRE 2021)

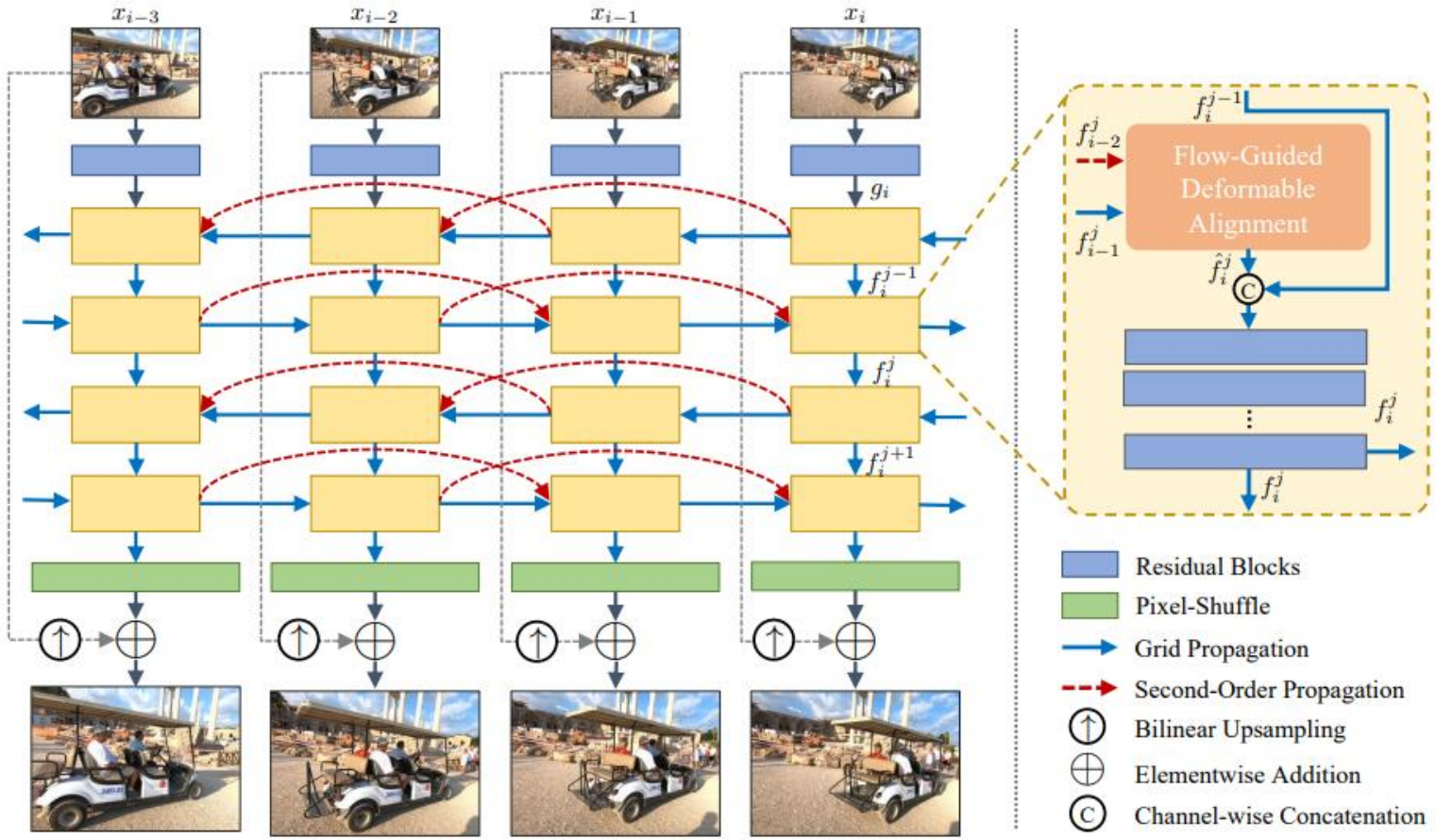


Figure 2. **An Overview of BasicVSR++.** BasicVSR++ consists of two redesigns to improve propagation and alignment in its predecessor BasicVSR [3]. For propagation, we introduce grid propagation (blue solid lines) to refine features bidirectionally. In addition, second-order connection (red dotted lines) is adopted to improve the robustness of propagation. Within each propagation branch, flow-guided deformable alignment is proposed to increase the offset diversity while overcoming the offset overflow problem.

Methods

NTIRE 2022 Winner: TaoMC2 Team

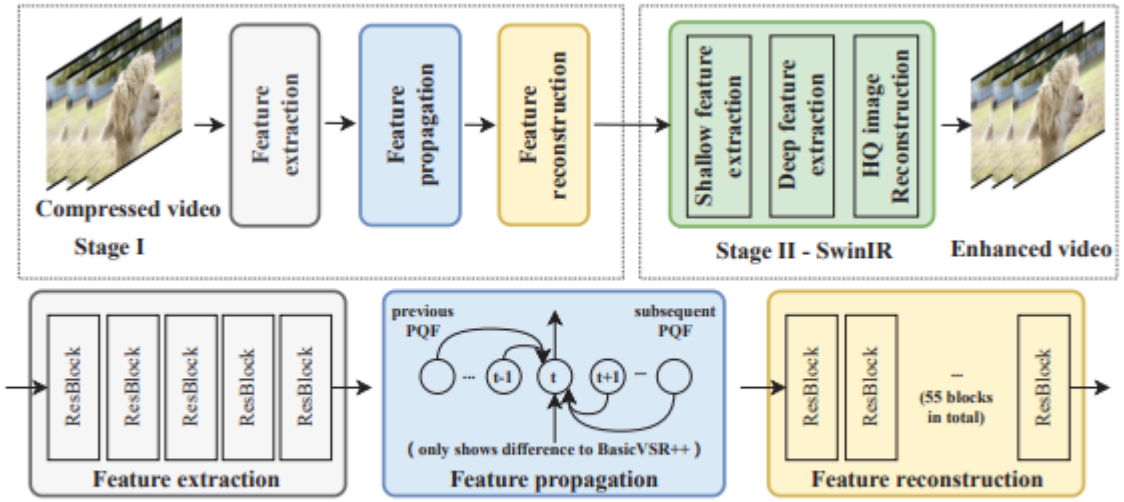


Figure 1. The two-stage approach proposed by the TaoMC2 Team.

The authors will present their method in the following.

Paper

Project page

Methods

NTIRE 2022 Winner: GY-Lab Team

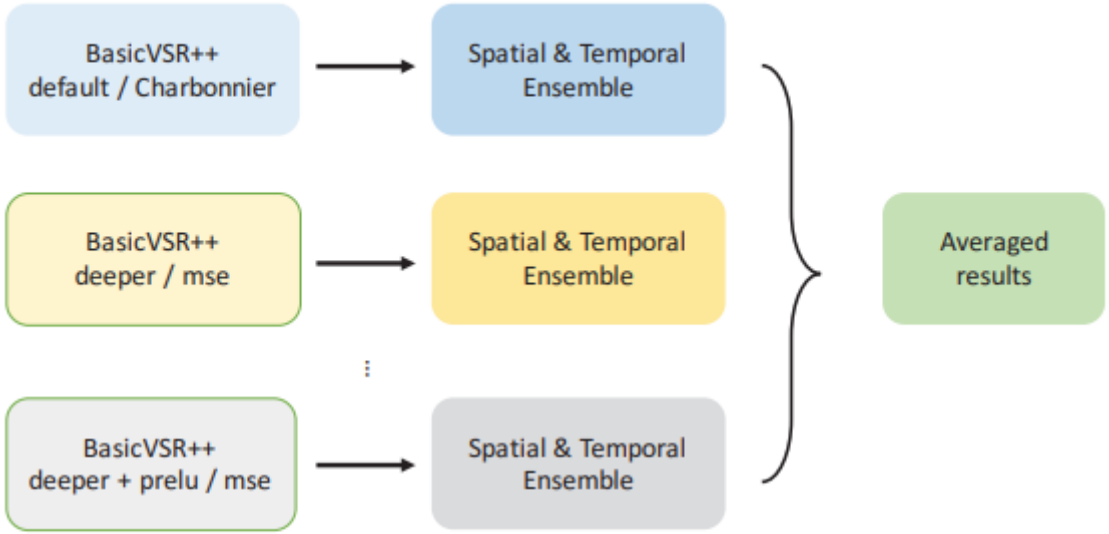


Figure 3. The pipeline of the method proposed by the GY-Lab Team in all three tracks. Models under different settings (structure, loss functions) are first trained independently. They apply both spatial and temporal ensemble on each model and average the outputs of all models as the final result.

Report

Yang, Ren, Radu Timofte, et al. "NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1221-1238. 2022.

Report

LDV 2.0 Dataset

Challenge homepage

Cite:

```
@inproceedings{yang2022ntire,
  title={NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results},
  author={Yang, Ren and Timofte, Radu and others},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops},
  pages={1221--1238},
  year={2022}
}
```

Congratulations to Award Winners
Thanks to all participants!

Ren Yang and Radu Timofte

VOYAGE 81