ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

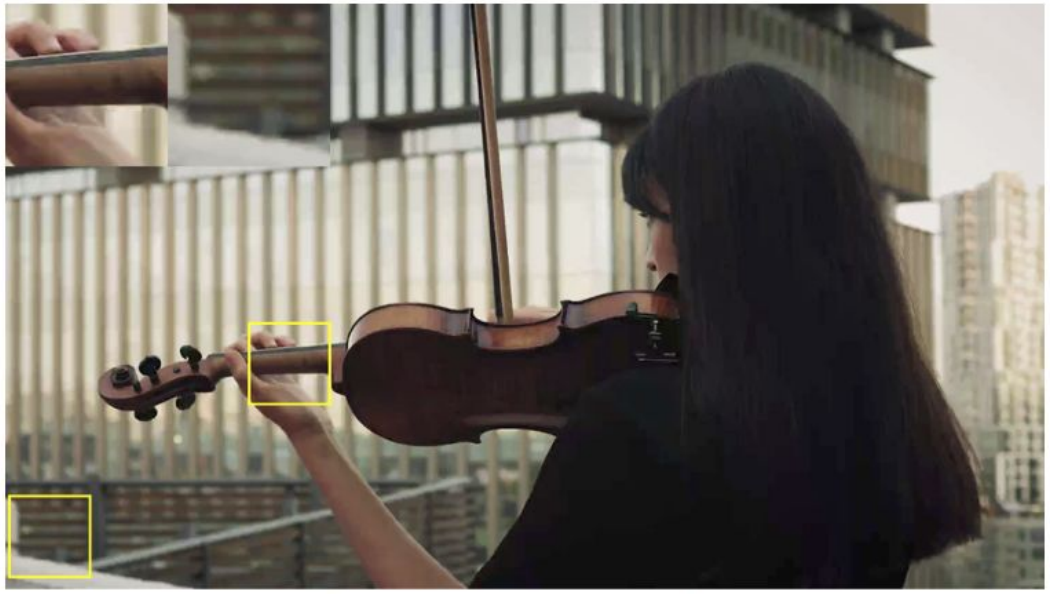
CVL Computer Vision Lab

NTIRE 2020 Challenge on Video Quality Mapping: Methods and Results

Dario Fuoli, Zhiwu Huang, Martin Danelljan, Radu Timofte and et. al.,

Computer Vision Lab @ ETH Zurich

Track 1: Supervised Video Quality Mapping



(a) More compressed

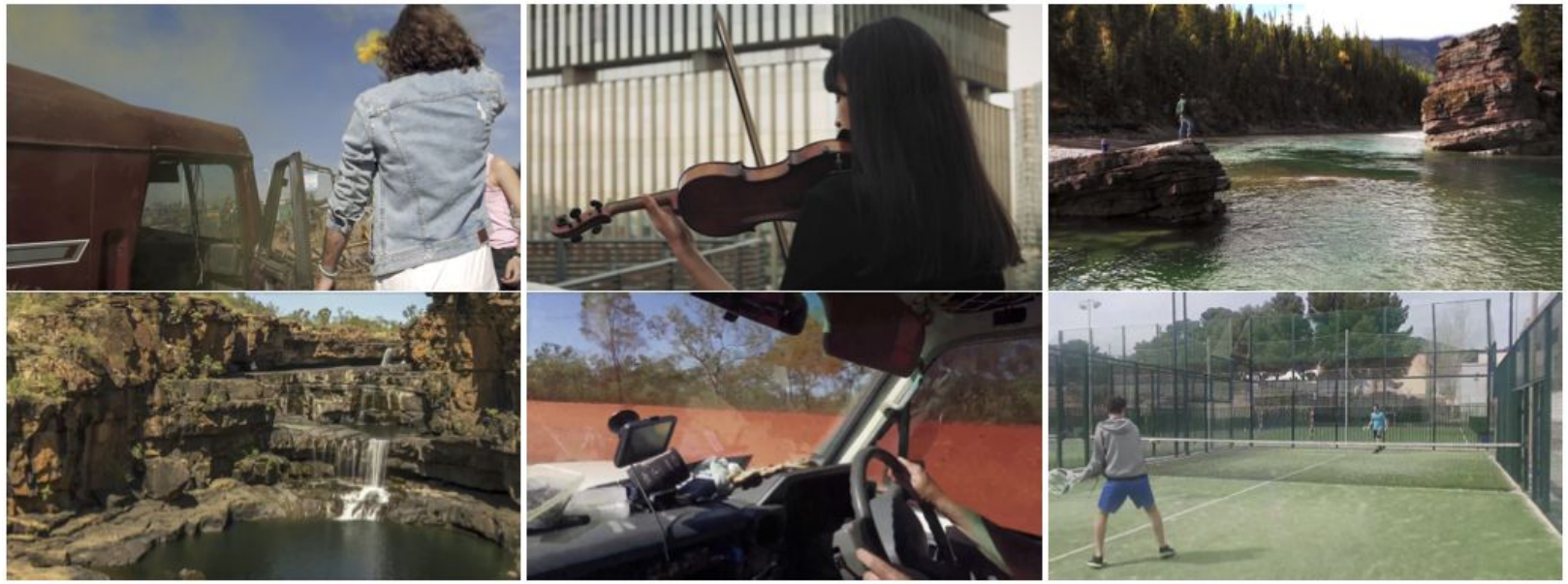
(b) Less compressed

- Mapping from more compressed to less compressed (H.264 compression)
- Pairs of well-aligned videos

Track 1: Dataset

IntVid Dataset [1]

- Internet videos
- 12 diverse scenarios
- H.264 compression
- 60 pairs of less and more compressed for training
- 30 pairs for validation
- 30 pairs for test



[1] Sohyeong Kim, Guanju Li, Dario Fuoli, Martin Danelljan, Zhiwu Huang, Shuhang Gu and Radu Timofte. The Vid3oC and IntVID Datasets for Video Super Resolution and Quality Mapping. In ICCV 2019 Workshops.

Track 1: Evaluation protocol

- Validation Phase
  - available validation source
  - no access to validation target
  - feedback (e.g., PSNR, SSIM etc.) from online server on CodaLab
  - maximum 10 submissions per day, maximum 20 submissions in total
- Test phase
  - no online feedback
  - final submission
    - full set of resulting frames
    - meta information such as author list, institution, method description

Track 1: Competition Results

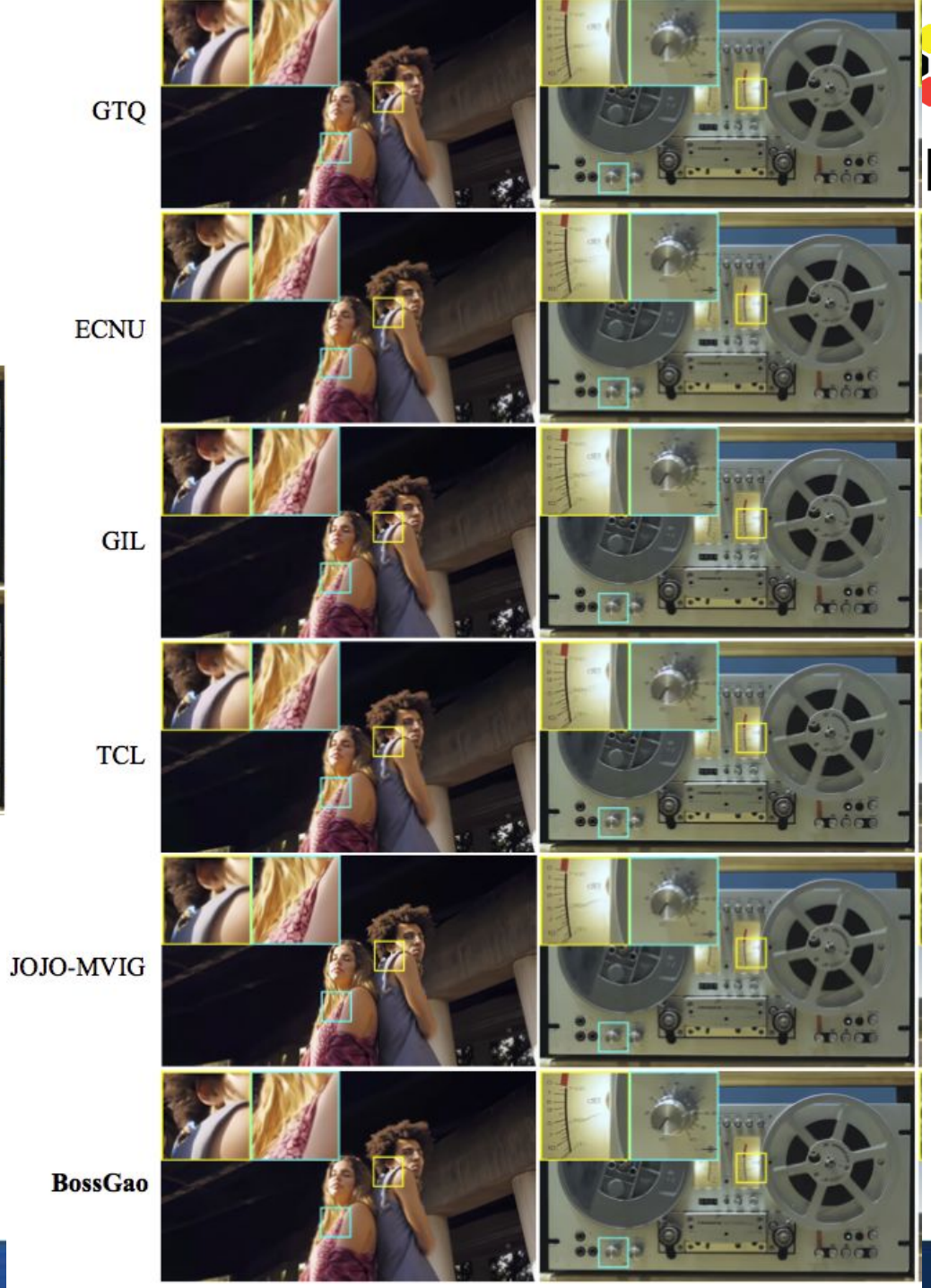
	Method	↑PSNR	↑SSIM	↓LPIPS	TrainingReq	TrainingTime	TestReq	TestTime	Parameters	ExtraData
Participants	BossGao	**32.419**	**0.905**	0.177	8×V100	5-10d	1×V100	4s	n/a	No
	JOJO-MVIG	32.167	0.901	0.182	2×1080Ti	≈ 4d	1×1080Ti	2.07s	≈22.75M	No
	GTQ	32.126	0.900	0.187	2×2080Ti	≈ 5d	1×2080Ti	9.74s	19.76M	No
	ECNU	31.719	0.896	0.198	2×1080Ti	2-3d	1×1080Ti	1.1s	n/a	No
	TCL	31.701	0.897	0.193	2×1080Ti	≈ 3d	1×1080Ti	25s	≈8.92M	No
	GIL	31.579	0.894	0.195	1×970Ti	≈ 6d	1×970Ti	11.37s	3.60M	No
	7-th team	30.598	0.878	**0.176**	n/a	4d	n/a	0.5s	≈7.92M	Yes
	No processing	30.553	0.877	**0.176**						

Bold: best, Underline: second and third best
TrainingTime: days, TestTime: seconds per frame

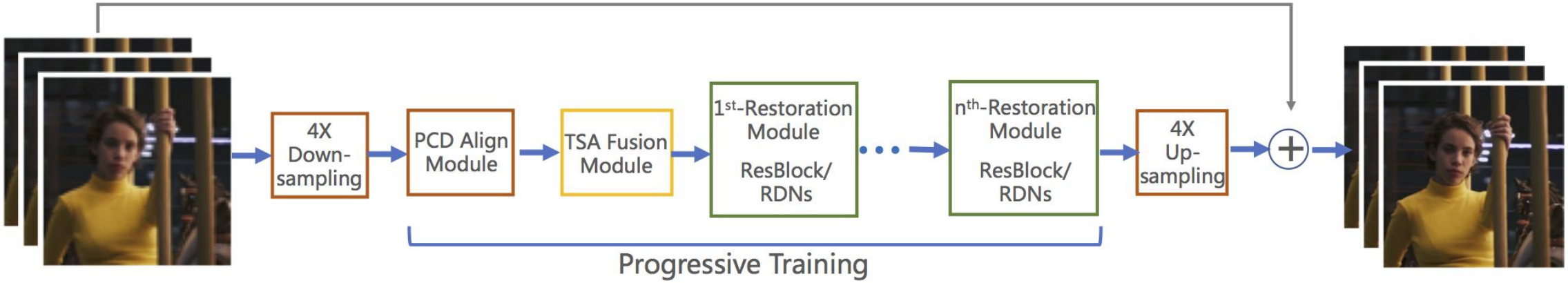
Winner: BossGao (Cao Li, Dongliang He, Fu Li, Shilei Wen, Baidu Inc., China)
Runner-up: JOJO-MVIG (Muchen Li, Siyao Li, Bo Pang, Cewu Lu, Shanghai Jiao Tong University, China)
Second Runner-up: GTQ (Hua Wang, Longcun Jin, Dewei Su, South China University of Technology, China)

[10] LPIPS: Zhang et al., The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In CVPR 2018.

Track 1: Visual Results



Track 1: Winner Team's Method - BossGao



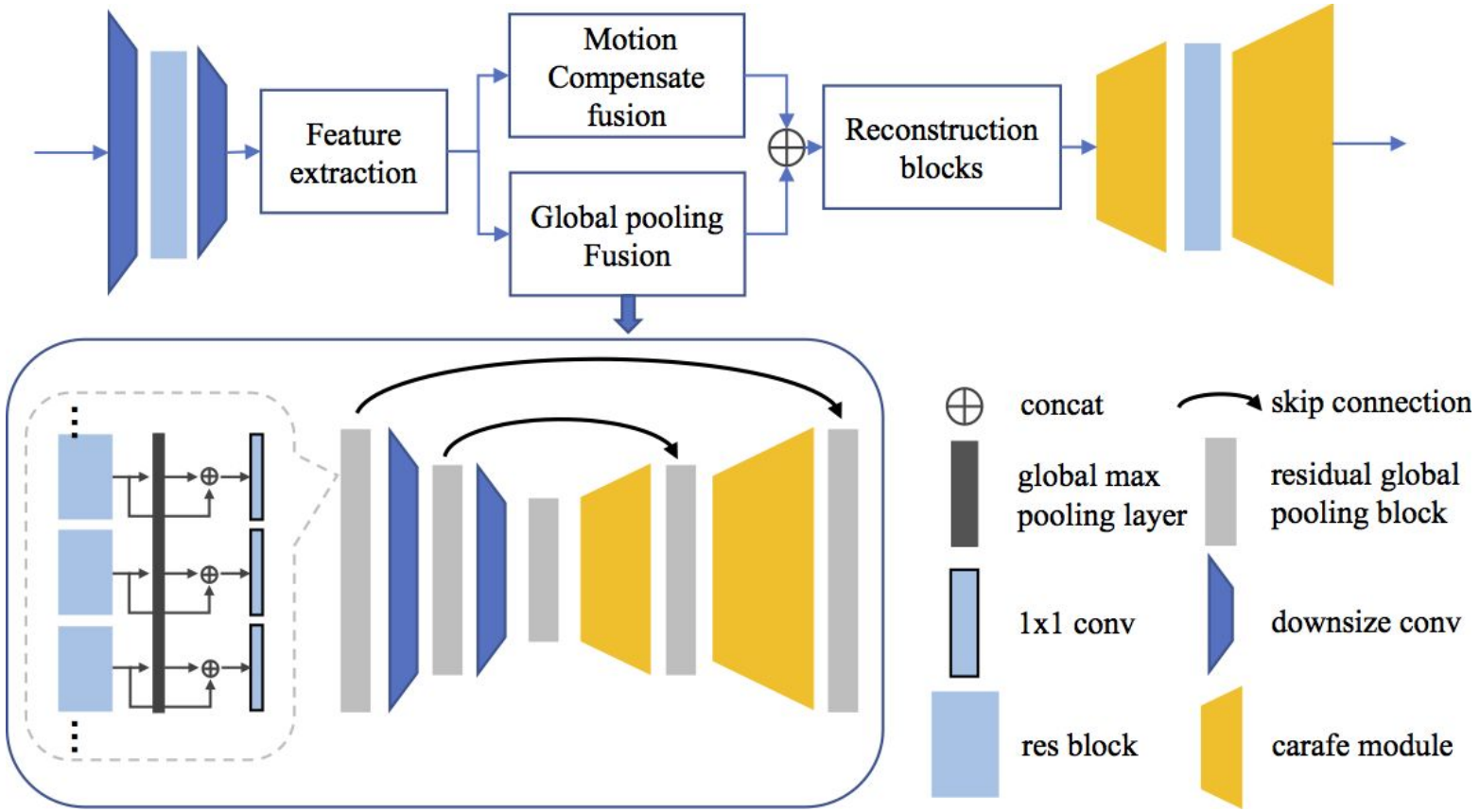
- Pyramid, Cascading and Deformable convolution [2]
- Temporal and Spatial Attention [2]
- ResBlock/ Residual dense block [3]
- Progressive training on the suggested modules

[2] Wang et al., EDVR: Video Restoration with Enhanced Deformable Convolutional Networks. In CVPR 2019 Workshops.

[3] Zhang et al., Residual Dense Network for Image Super-Resolution. In CVPR 2018.

Track 1: Runner-Up Team's Method - JOJO-MVIG

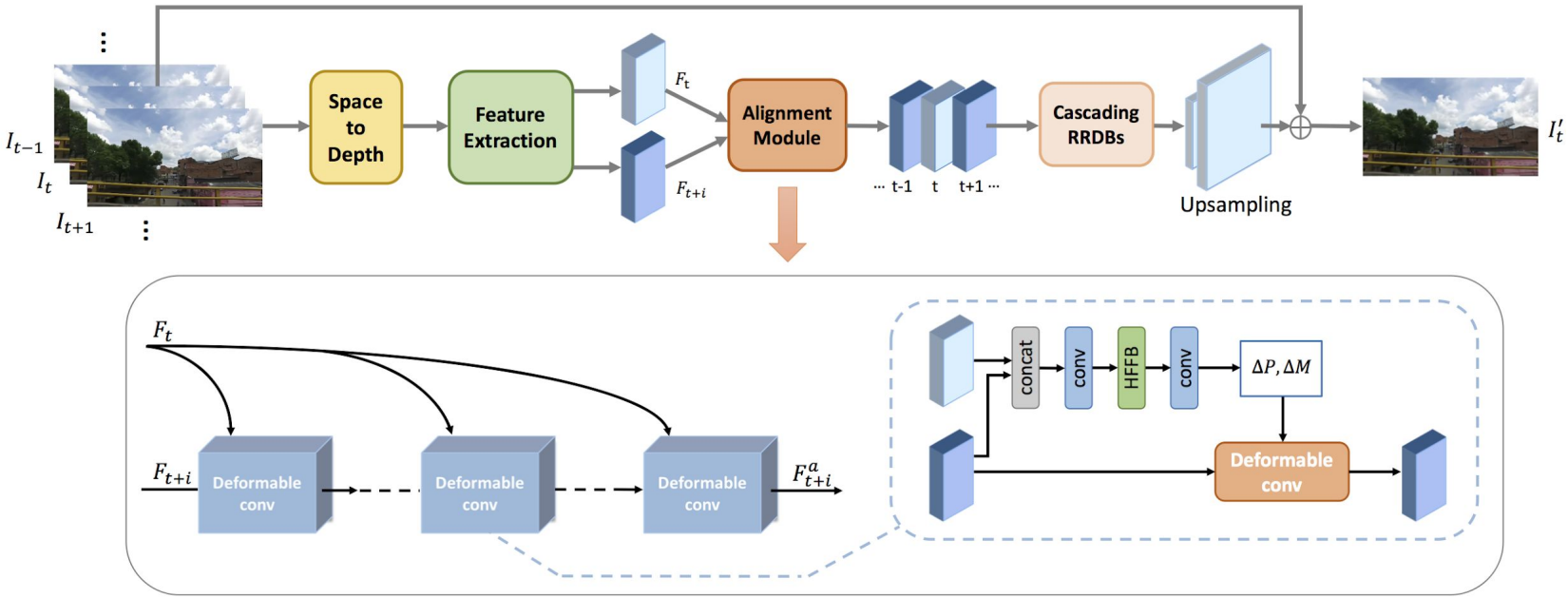
- Feature extraction module [2]
- Two spatio-temporal fusion paths [2]
- Reconstruction module [2]



[2] Wang et al., EDVR: Video Restoration with Enhanced Deformable Convolutional Networks. In CVPR 2019 Workshops.

Track 1: Second Runner-up Team's method - GTQ

- Space to depth shuffling operation
- Feature extraction
- Alignment module using deformable convolutions [4] and hierarchical feature fusion blocks [5]
- Residual dense blocks [3]



[3] Zhang et al., Residual Dense Network for Image Super-Resolution. In CVPR 2018.
[4] Zhu et al., Deformable ConvNets v2: More Deformable, Better Results. In CVPR 2019.
[5] Hui et al., Progressive Perception-Oriented Network for Single Image Super-Resolution. arXiv 2019.

Track 2: Weakly-Supervised Video Quality Mapping

(a) ZED captured

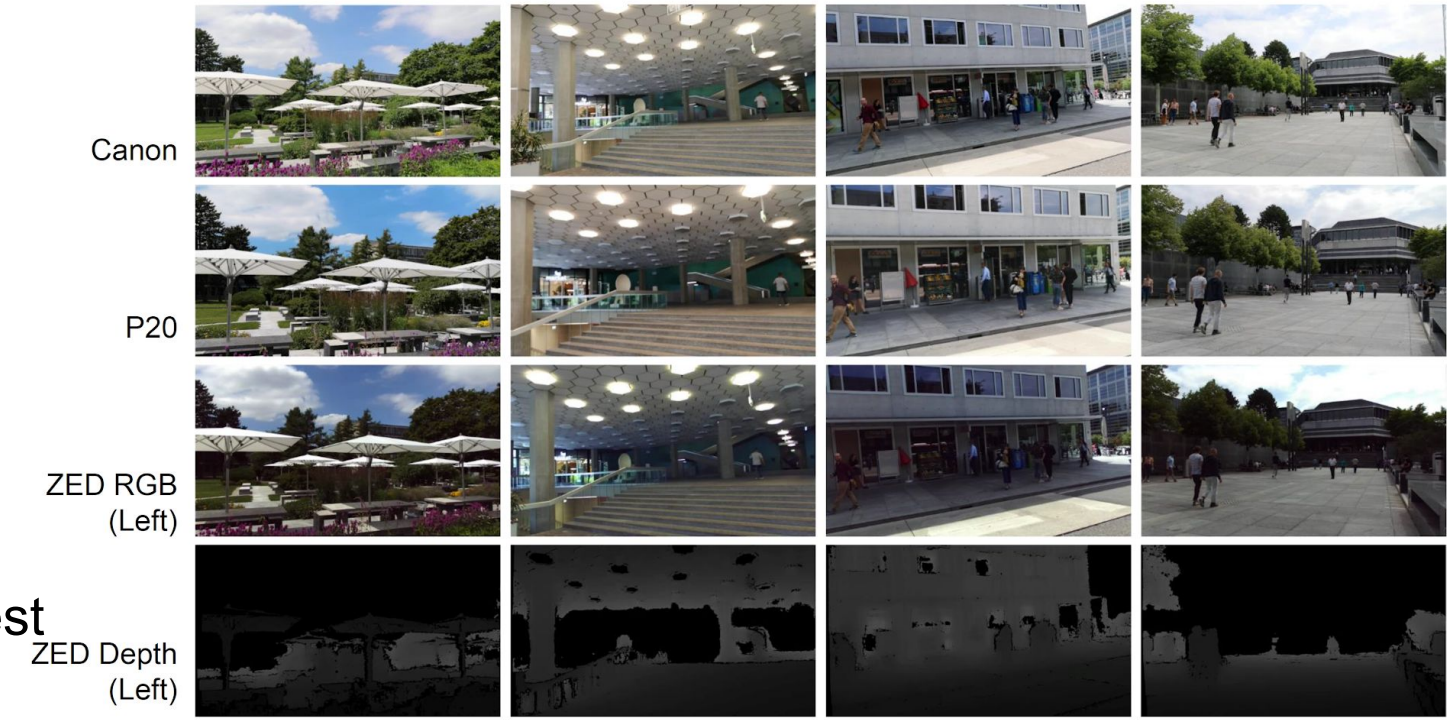
(b) Canon captured

- Mapping from ZED camera captured videos to Canon cameral captured videos
- Pairs of weakly-aligned videos

Track 2: Dataset

Vid3oC Dataset [1]

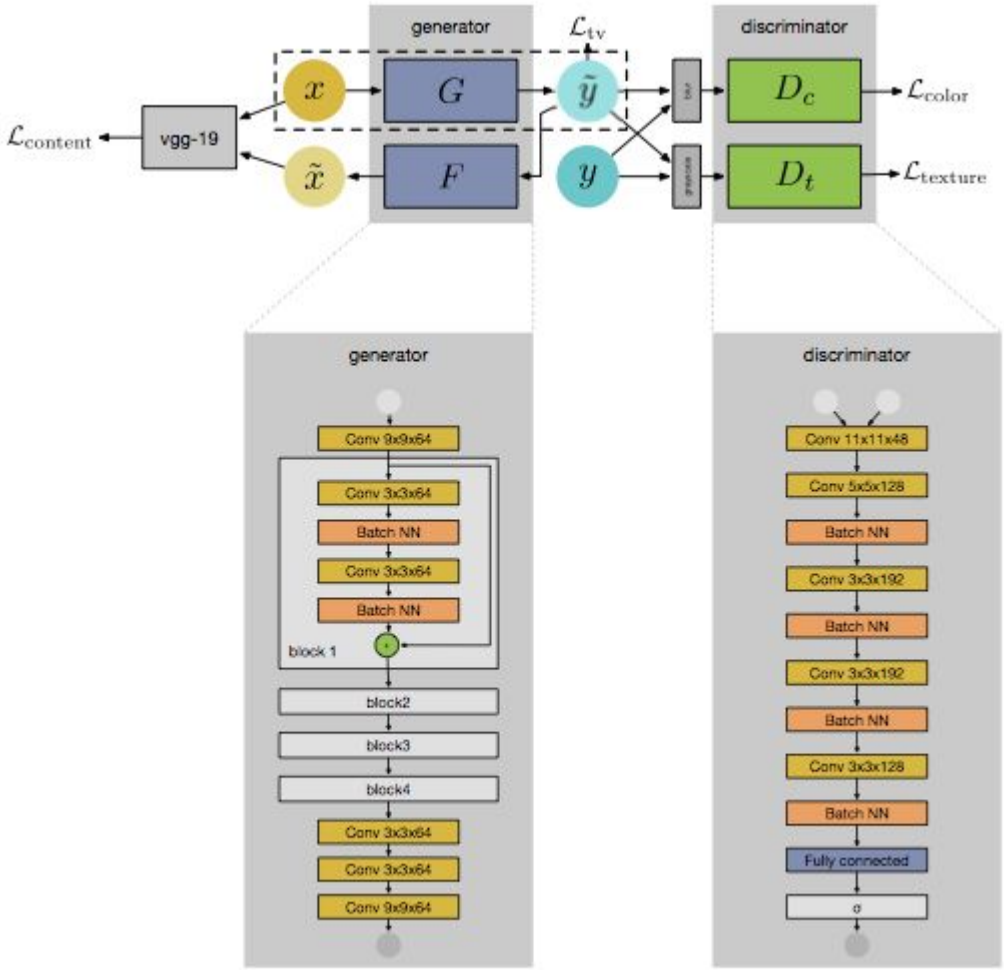
- **Canon 5D Mark IV** high quality DSLR
- **ZED** stereo camera
- 50 weakly-paired videos for training
- 16 weak-pairs for validation
- 16 weak-pairs for test



[1] Sohyeong Kim, Guanju Li, Dario Fuoli, Martin Danelljan, Zhiwu Huang, Shuhang Gu and Radu Timofte. The Vid3oC and IntVID Datasets for Video Super Resolution and Quality Mapping. In ICCV 2019 Workshops.

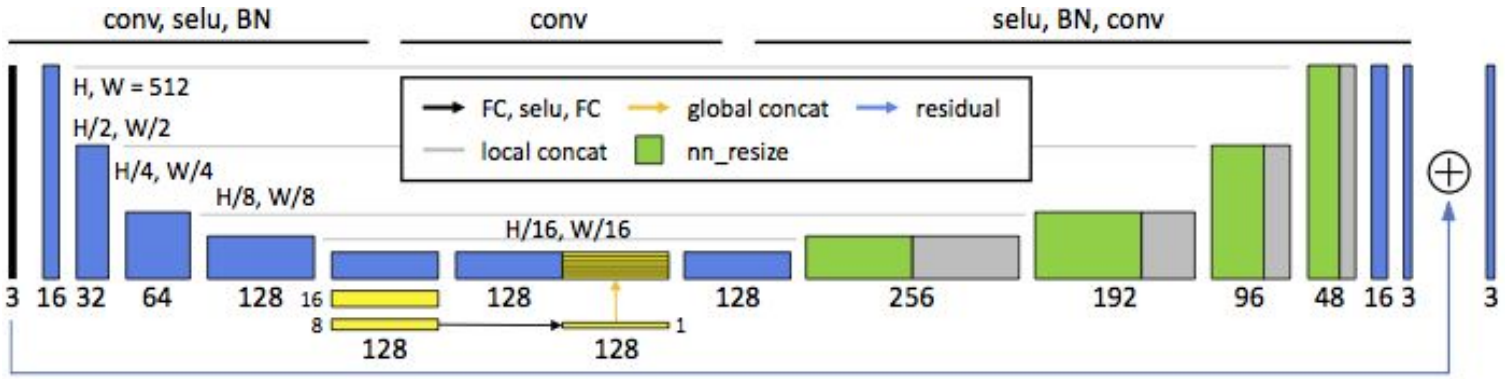
Track 2: Competing Method1 - WESPE [6]

- Generative and inverse generative network
  - fully-convolutional residual CNNs
- Two discriminators
  - color-based
  - Texture based
- Adversarial training
  - JS divergence-based min-max loss



[6] Ignatov et al., WESPE: weakly supervised photo enhancer for digital cameras, In CVPRW 2018.

Track 2: Competing Method2 - DPE [7]



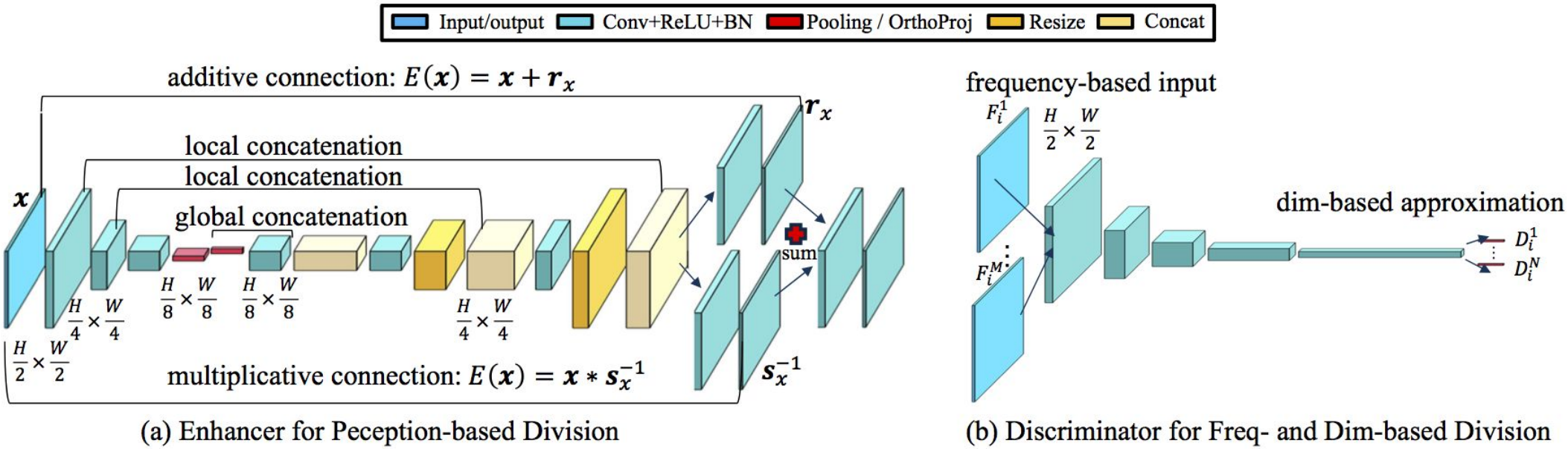
(a) generator

(b) discriminator

- Generator using U-Net with global concatenation
- Discriminator
- Adversarial training
  - Adaptive Wasserstein distance based min-max loss

[7] Chen et al., Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs, *In CVPR 2018.*

Track 2: Competing Method3- DACAL [8]



(a) Enhancer for Peception-based Division

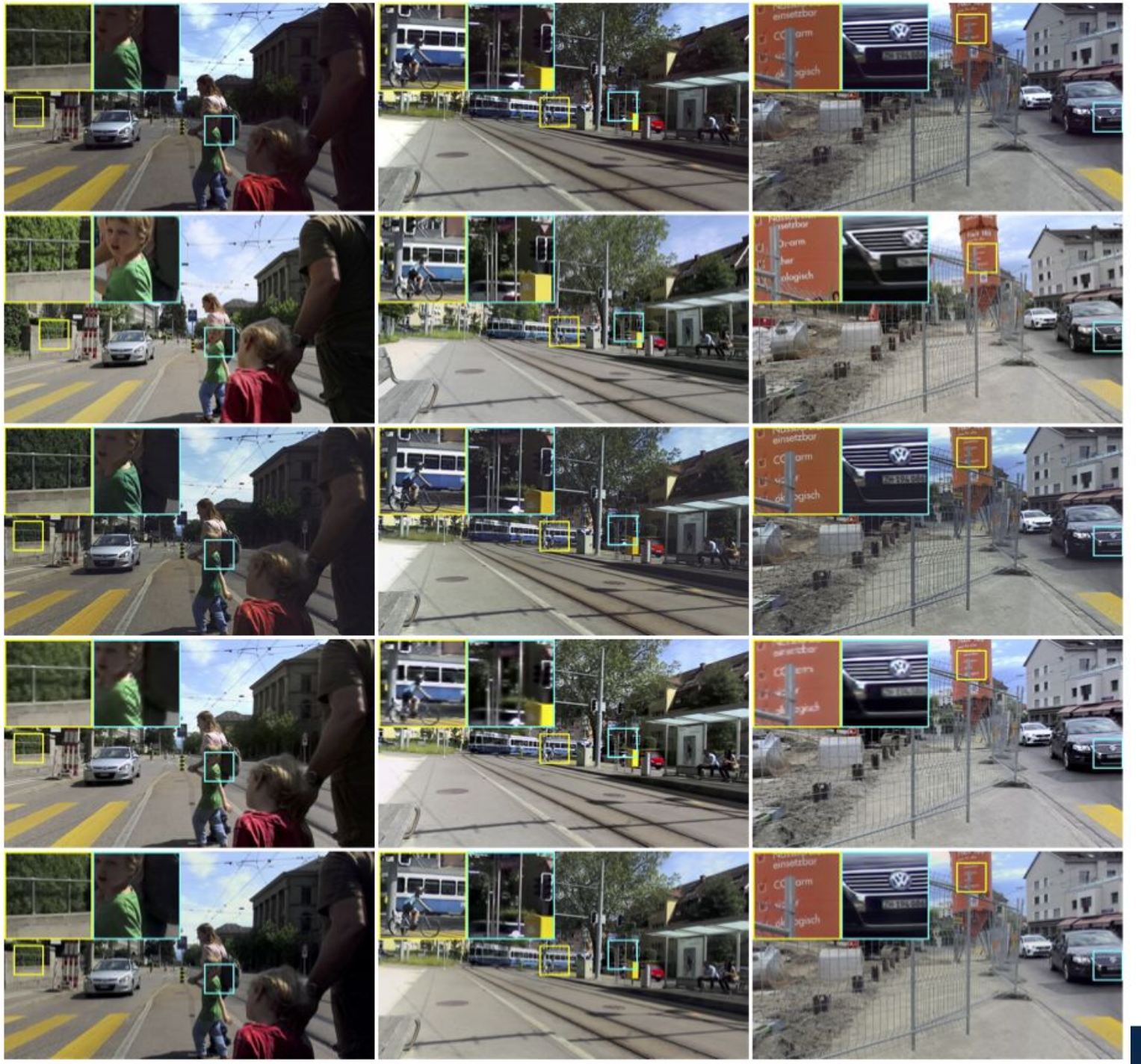
(b) Discriminator for Freq- and Dim-based Division

- Perception-based generator
- Frequency-based discriminator
- Dim-based min-max loss (Divide-and-Conquer Adversarial Learning): Adaptive version of sliced Wasserstein GAN [9]

[8] Huang et al., Divide-and-Conquer Adversarial Learning for High-resolution Image and Video Enhancement, arXiv 2019
[9] Wu et al., Sliced Wasserstein Generative Modeling, In CVPR 2019.

Track 2: Results

	Source	WESPE	DPE	DACAL
LPIPS↓	0.590	0.584	0.631	**0.565**



Errata: As the visual results are obtained on validation set, the camera-ready wrongly used the test set for reference to compute LPIPS. We have corrected the LPIPS score in our latest arXiv paper.

[10] LPIPS: Zhang et al., The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In CVPR 2018.