北京交通大學
BEIJING JIAOTONG UNIVERSITY

数字媒体信息处理研究中心
Center of Digital Media Information Processing

Meeting of Paper Sharing

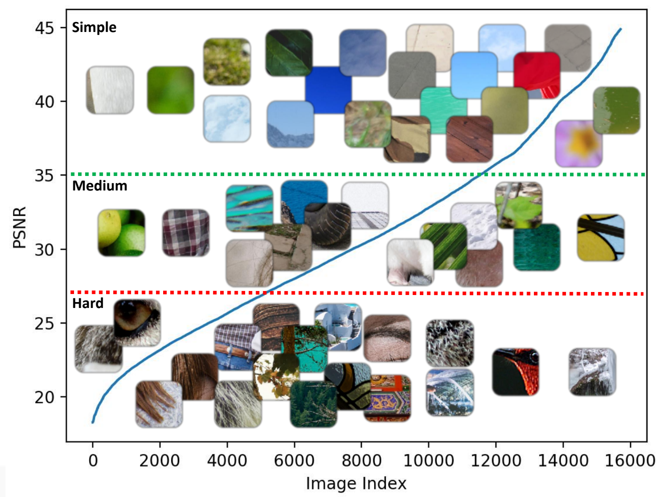
Restoration and Understanding of Visual Data

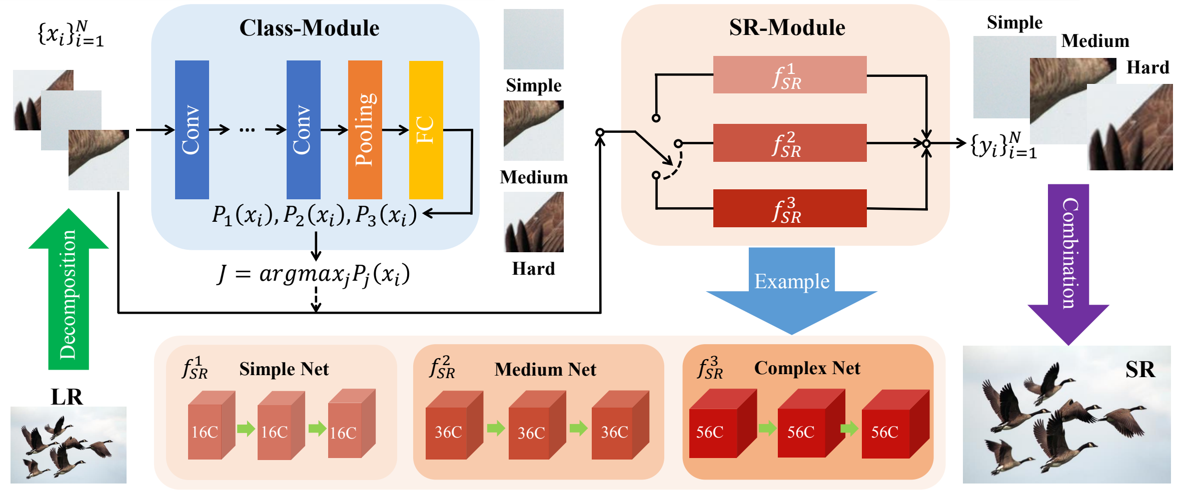
Qi Tang

2023/5/8

ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic

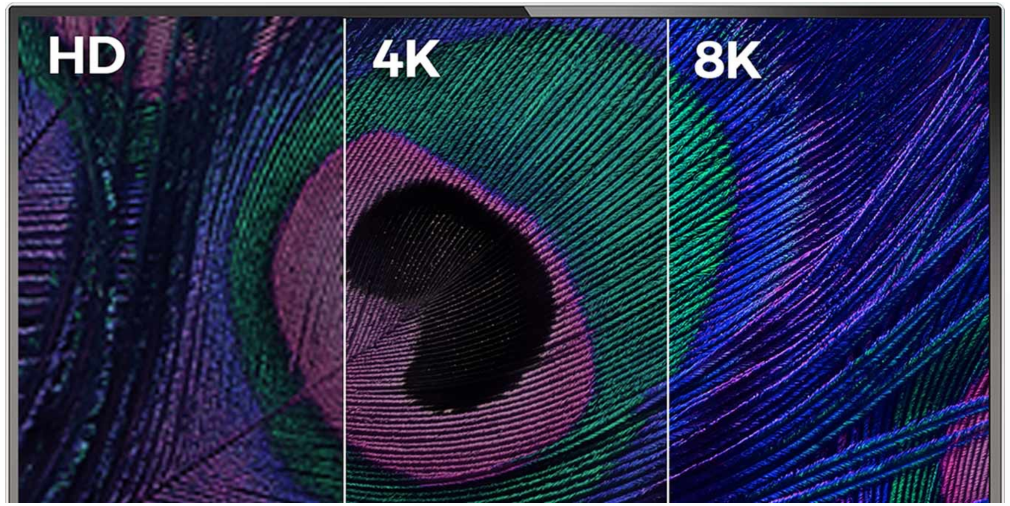
Kong[1,2] Hengyuan Zhao[1] Yu Qiao[1,3] Chao Dong[1,4] *
[1]Key Laboratory of Human-Machine Intelligence-Synergy Systems,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Shanghai AI Lab, Shanghai, China
[4]SIAT Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society





Restoration

Accelerate Super-Resolution

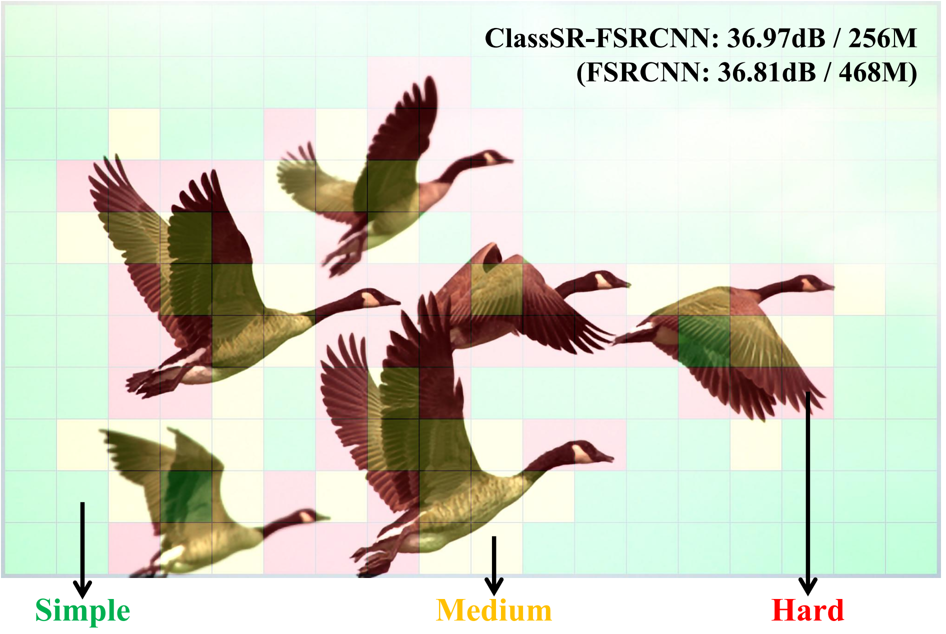


- The image/video resolution for smartphones and TV monitors has already reached 4K, or even 8K
- The memory and computational cost of methods built on CNNs will grow quadratically with the input size
- SR acceleration focus on proposing light-weight network structures

Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. In IEEE Conference on Computer Vision and Pattern Recognition. 12016-12025.

BEIJING JIAOTONG UNIVERSITY

Motivation

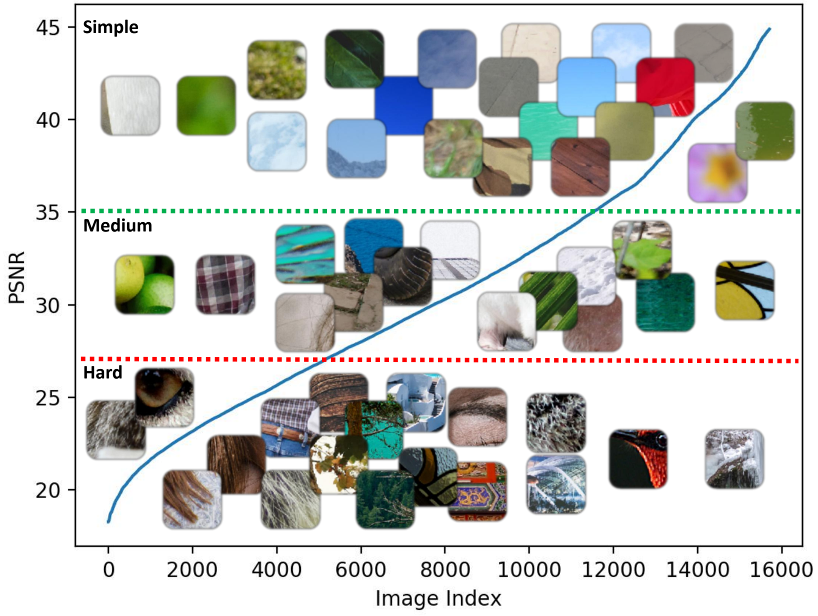


- Sub-images with high PSNR values are generally smooth, while the sub-images with low PSNR values contain complex textures

- Flat areas (color in light green) are processed with the simple network and the textures (color in red) are processed with the complex one

Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. In IEEE Conference on Computer Vision and Pattern Recognition. 12016-12025.

Motivation

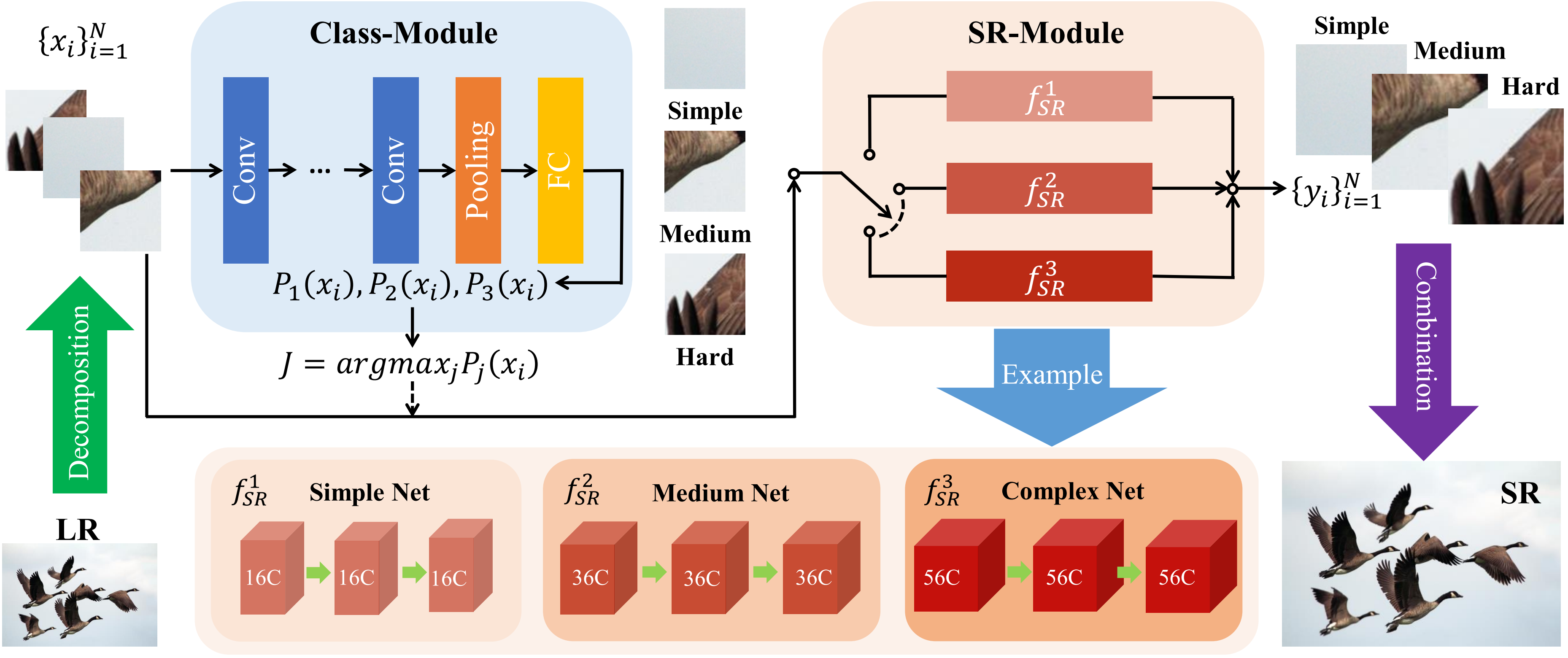


Model	FLOPs	Simple	Medium	Hard
FSRCNN (16)	141M	42.71dB	–	–
FSRCNN (36)	304M	–	29.62dB	–
FSRCNN (56)	468M	–	–	22.73dB
FSRCNN-O (56)	468M	42.70dB	29.69dB	22.71dB

- Sub-images with high PSNR values are generally smooth, while the sub-images with low PSNR values contain complex textures
- Flat areas (color in light green) are processed with the simple network and the textures (color in red) are processed with the complex one

Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. In IEEE Conference on Computer Vision and Pattern Recognition. 12016-12025.

ClassSR



$$L = w_1 \times L_1 + w_2 \times L_c + w_3 \times L_a \qquad L_c = -\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}|P_i(x) - P_j(x)|, \text{ s.t. } \sum_{i=1}^{M} P_i(x) = 1 \qquad L_a = \sum_{i=1}^{M}\left|\sum_{j=1}^{B} P_i(x_j) - \frac{B}{M}\right|$$

Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. In IEEE Conference on Computer Vision and Pattern Recognition. 12016-12025.

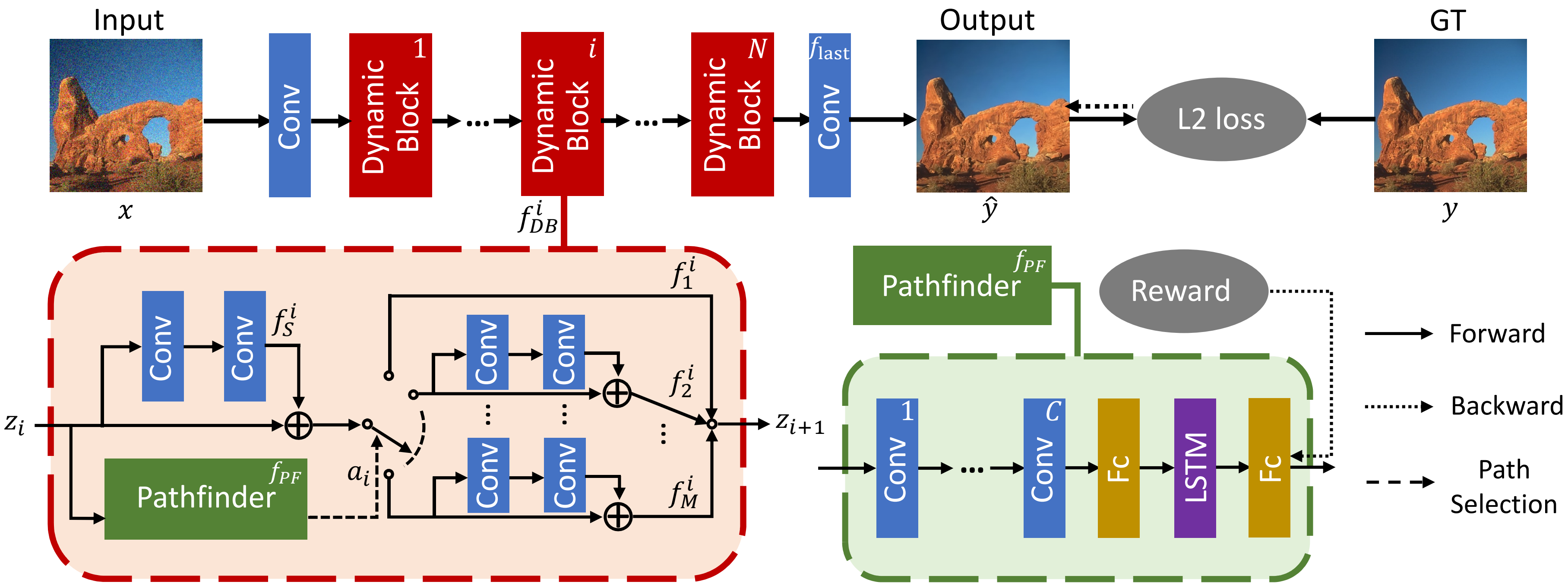
Experiments

Model	Parameters	Test2K	FLOPs	Test4K	FLOPs	Test8K	FLOPs
FSRCNN-O	25K	25.61dB	468M(100%)	26.90dB	468M(100%)	32.66dB	468M(100%)
ClassSR-FSRCNN	113K	25.61dB	311M(66%)	26.91dB	286M(61%)	32.73dB	238M(51%)
CARN-O	295K	25.95dB	1.15G(100%)	27.34dB	1.15G(100%)	33.18dB	1.15G(100%)
ClassSR-CARN	645K	26.01dB	814M(71%)	27.42dB	742M(64%)	33.24dB	608M(53%)
SRResNet-O	1.5M	26.19dB	5.20G(100%)	27.65dB	5.20G(100%)	33.50dB	5.20G(100%)
ClassSR-SRResNet	3.1M	26.20dB	3.62G(70%)	27.66dB	3.30G(63%)	33.50dB	2.70G(52%)
RCAN-O	15.6M	26.39dB	32.60G(100%)	27.89dB	32.60G(100%)	33.76dB	32.60G(100%)
ClassSR-RCAN	30.1M	26.39dB	21.22G(65%)	27.88dB	19.49G(60%)	33.73dB	16.36G(50%)

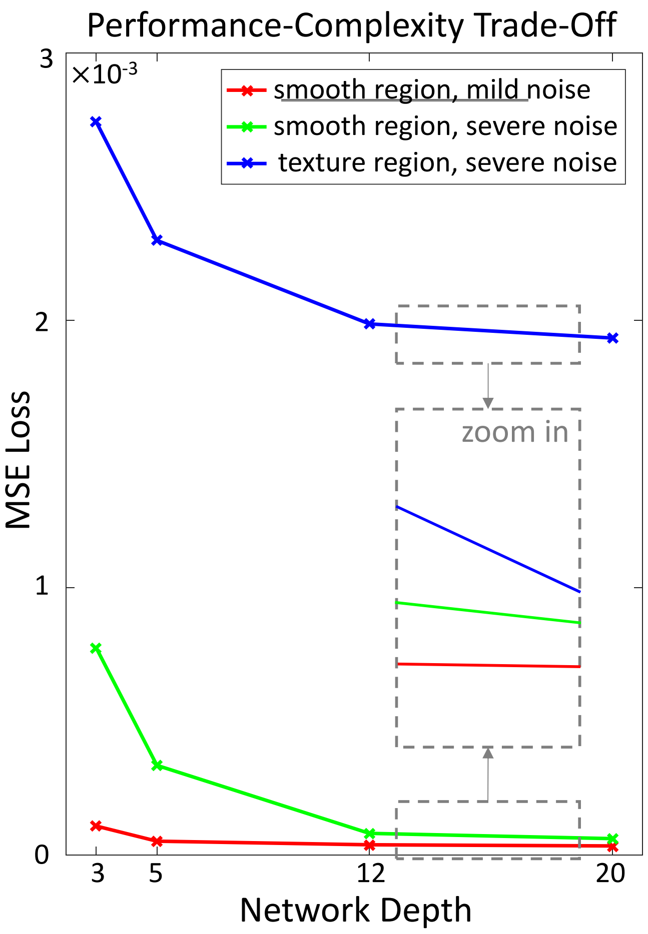
Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. 2021. ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. In IEEE Conference on Computer Vision and Pattern Recognition. 12016-12025.

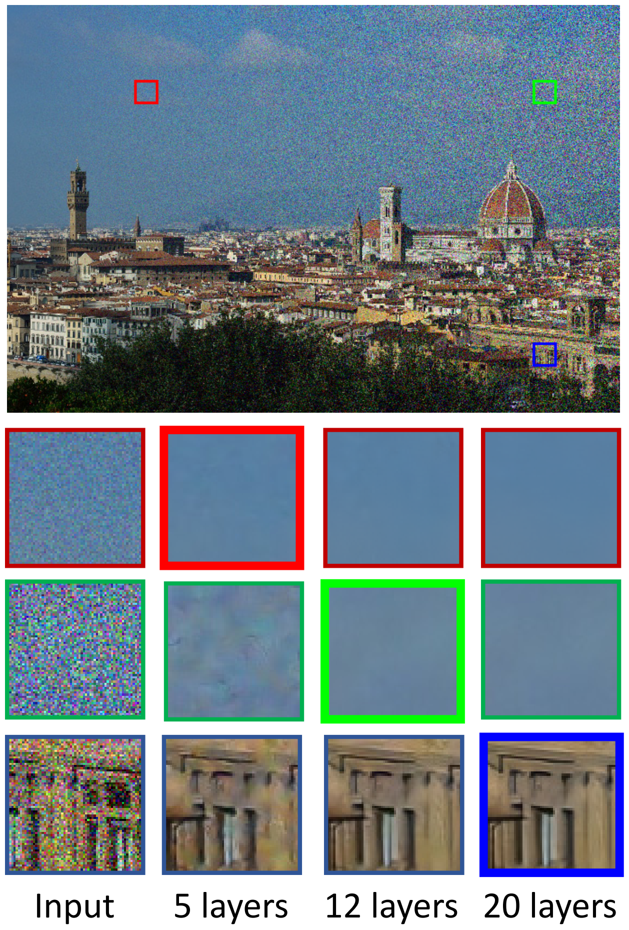
Path-Restore: Learning Network Path Selection for Image Restoration

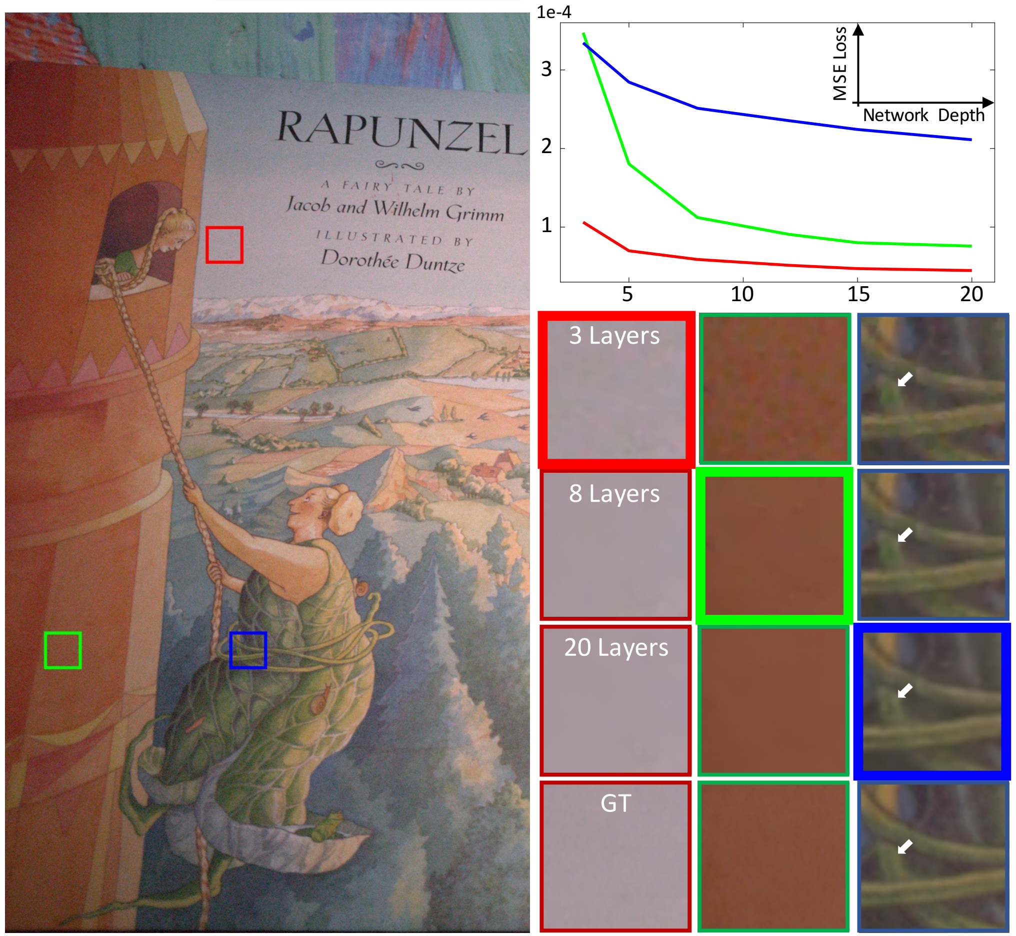
Ke Yu , Xintao Wang, Chao Dong , Xiaoou Tang, and Chen Change Loy



Accelerate Super-Resolution



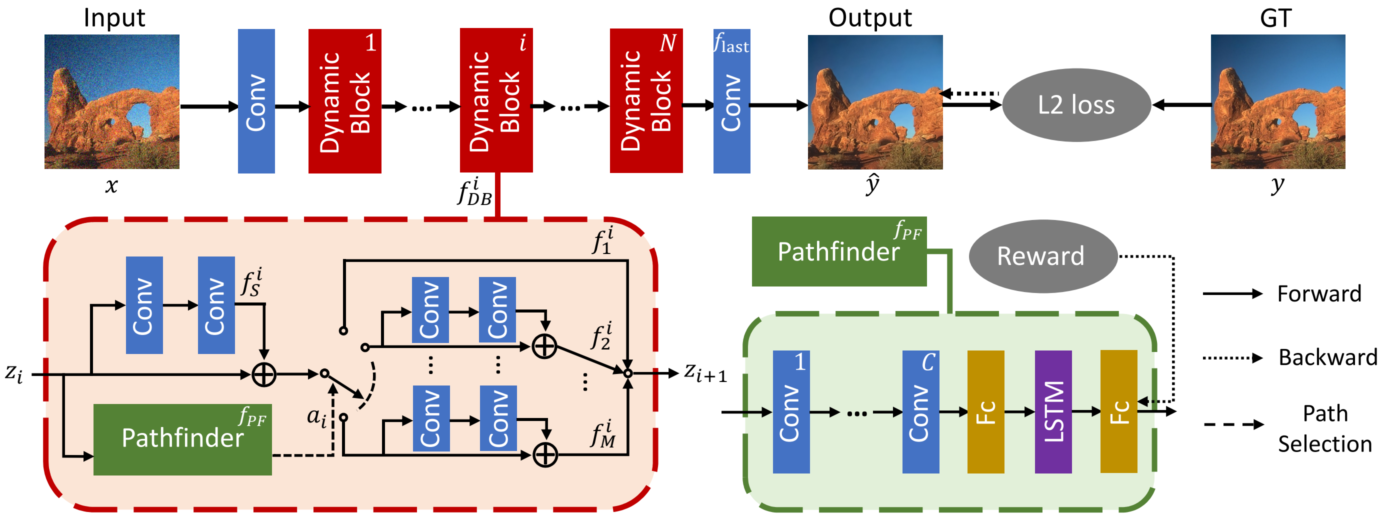


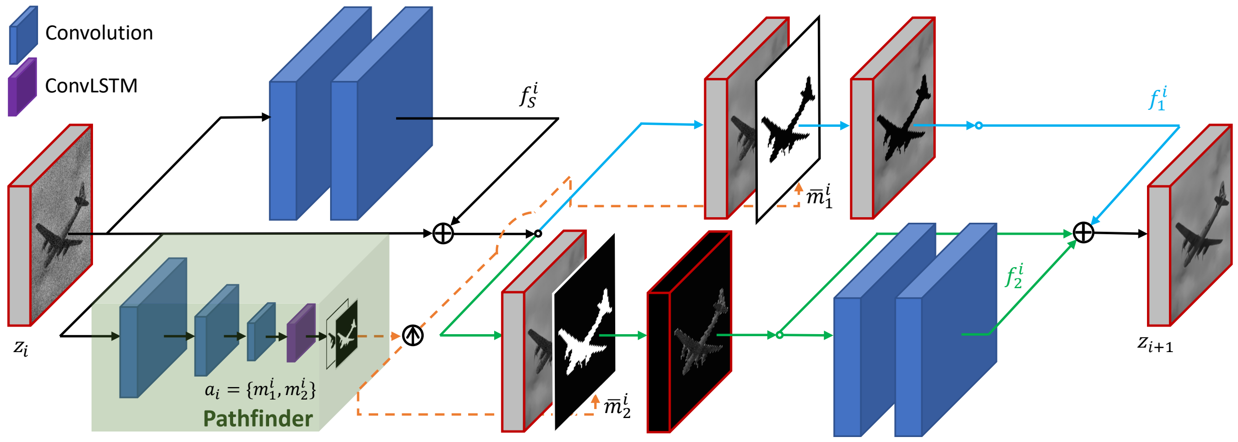


Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang and Chen Change Loy. 2022. Path-Restore: Learning Network Path Selection for Image Restoration. IEEE TPAMI 44, 10 (2022), 7078–7092.

Restoration

Accelerate Super-Resolution





$$z_{i+1} = f^i_{a_i}\left(f^i_S(z_i)\right)$$

$$r_i = \begin{cases} -p \times \left(1 - \mathbf{1}_{\{1\}}(a_i)\right), & 1 \le i < N \\ -p \times \left(1 - \mathbf{1}_{\{1\}}(a_i)\right) + d \times (-\Delta L_2), & i = N \end{cases}$$

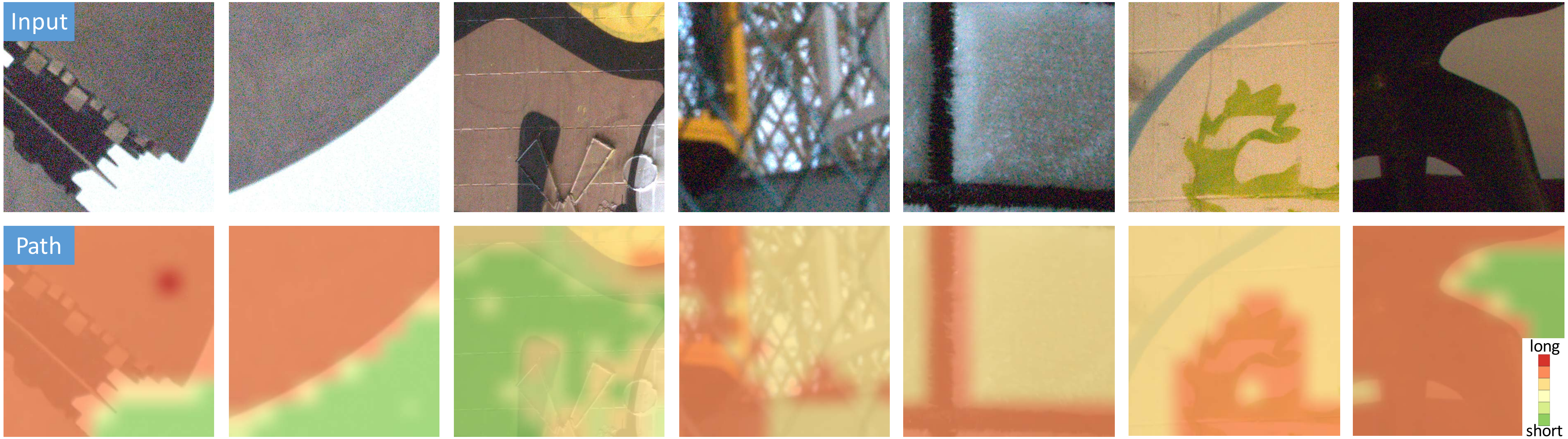
$$d = \begin{cases} L_d / L_0, & 0 \le L_d < L_0 \\ 1, & L_d \ge L_0 \end{cases}$$

$$z_{i+1} = \sum_{j=1}^{M} f^i_j\left(f^i_S(z_i) \odot \bar{m}^i_j\right)$$

Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang and Chen Change Loy. 2022. Path-Restore: Learning Network Path Selection for Image Restoration. IEEE TPAMI 44, 10 (2022), 7078–7092.

Restoration

Accelerate Super-Resolution

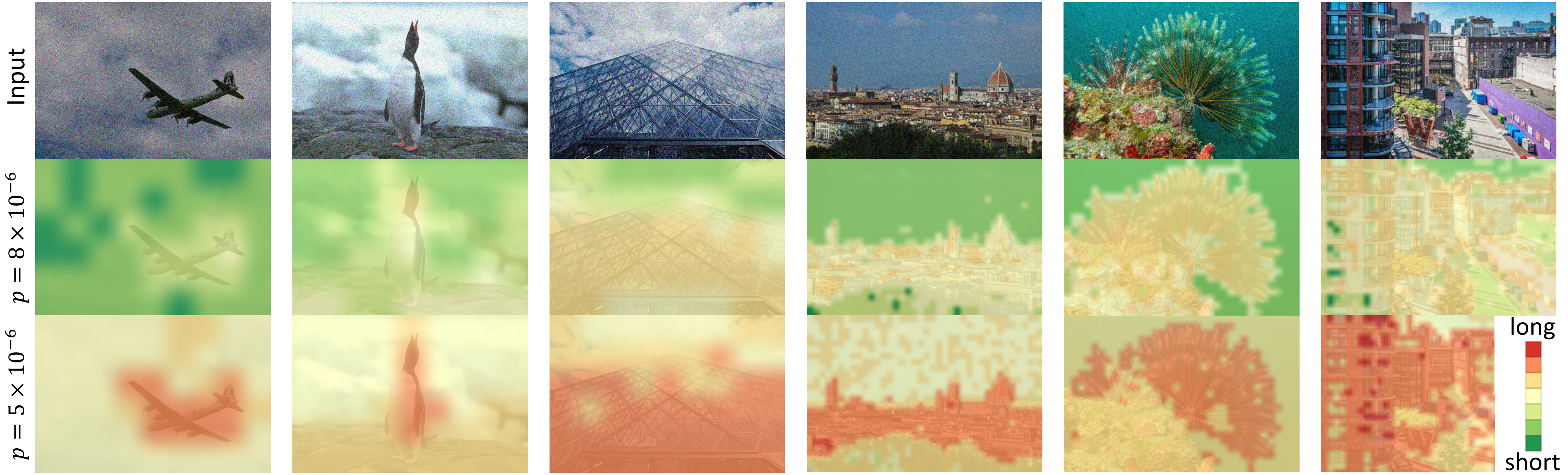


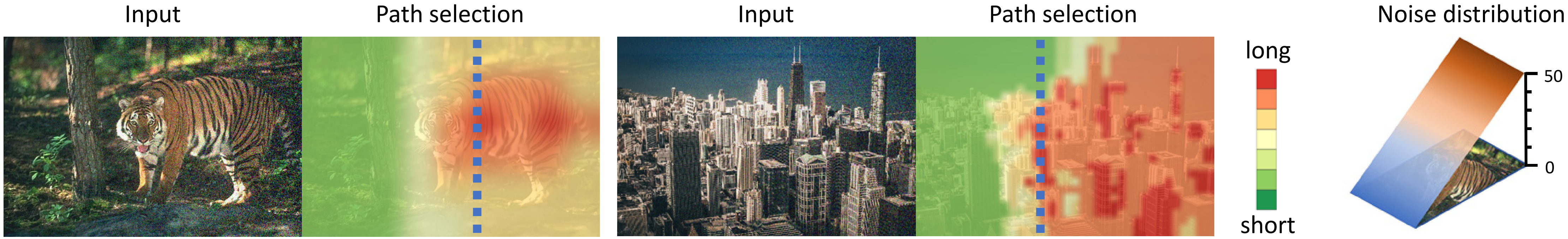
Dataset	CBSD68 [57]						DIV2K-T50 [58]					
Noise	uniform			spatially variant			uniform			spatially variant		
	σ=10	σ=50	FLOPs	linear	peaks	FLOPs	σ=10	σ=50	FLOPs	linear	peaks	FLOPs
DnCNN [23]	36.07	27.96	5.31G	31.17	31.15	5.31G	37.32	29.64	5.31G	32.82	32.64	5.31G
Path-Restore	36.04	27.96	**4.22G**	31.18	31.15	**4.22G**	37.26	29.64	**4.20G**	32.83	32.64	**4.17G**

The unit of FLOPs is Giga ($\times 10^9$). Path-Restore is consistently 25 percent faster (in terms of FLOPs) than DnCNN with comparable performance on different noise settings.

Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang and Chen Change Loy. 2022. Path-Restore: Learning Network Path Selection for Image Restoration. IEEE TPAMI 44, 10 (2022), 7078–7092.

Accelerate Super-Resolution

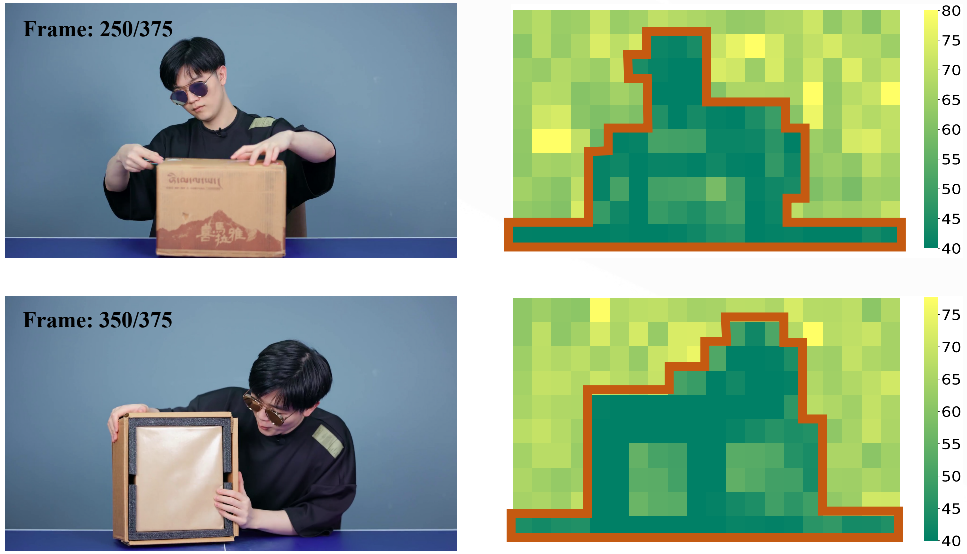


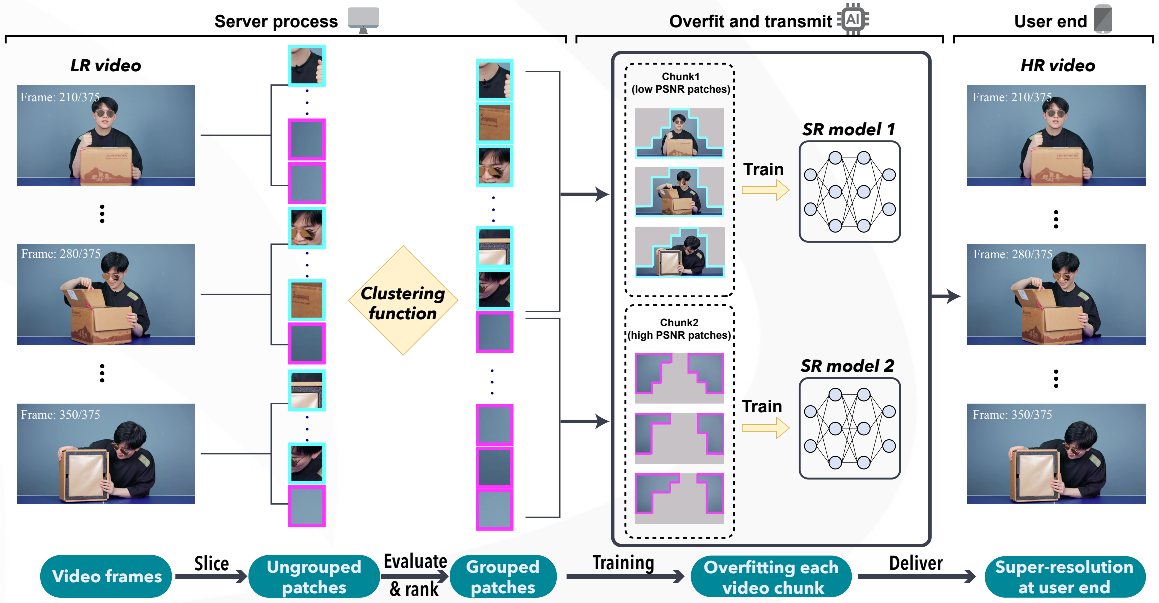


Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang and Chen Change Loy. 2022. Path-Restore: Learning Network Path Selection for Image Restoration. IEEE TPAMI 44, 10 (2022), 7078–7092.

Towards High-Quality and Efficient Video Super-Resolution via Spatial-Temporal Data Overfitting

Gen Li[1,†], Jie Ji[1,†], Minghai Qin[†], Wei Niu[2], Bin Ren[2], Fatemeh Afghah[1], Linke Guo[1], Xiaolong Ma[1]
[1]Clemson University [2]William & Mary





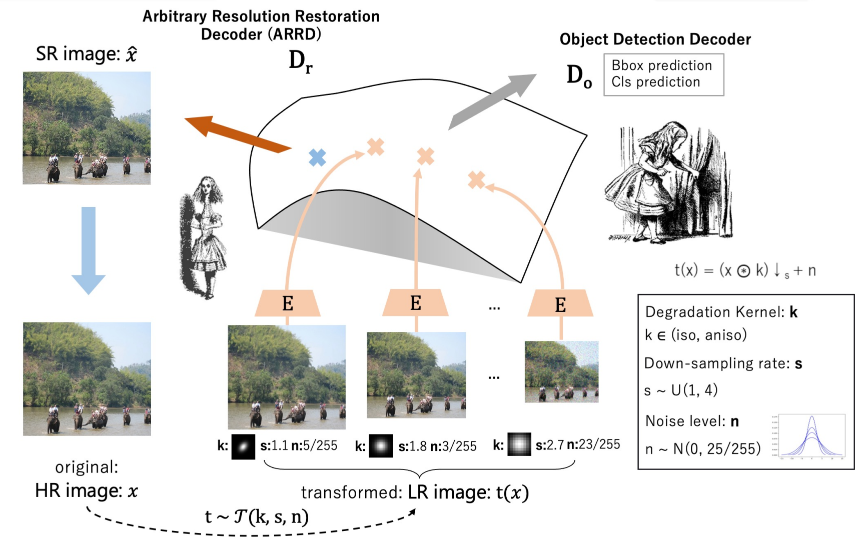
Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection

Ziteng Cui[1], Yingying Zhu[2], Lin Gu[3,4*], Guo-Jun Qi[5], Xiaoxiao Li[6], Renrui Zhang[7], Zenghui Zhang[1], and Tatsuya Harada[4,3]

[1] Shanghai Jiao Tong University [2] University of Texas at Arlington [3] RIKEN AIP

[4] The University of Tokyo [5] Laboratory for Machine Perception and Learning

[6] The University of British Columbia [7] Shanghai AI Laboratory



Understanding

Super-Resolution & Other Vision Tasks

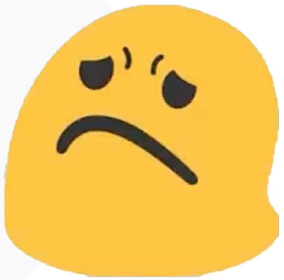
Low Resolution

Super-Resolution

High Resolution

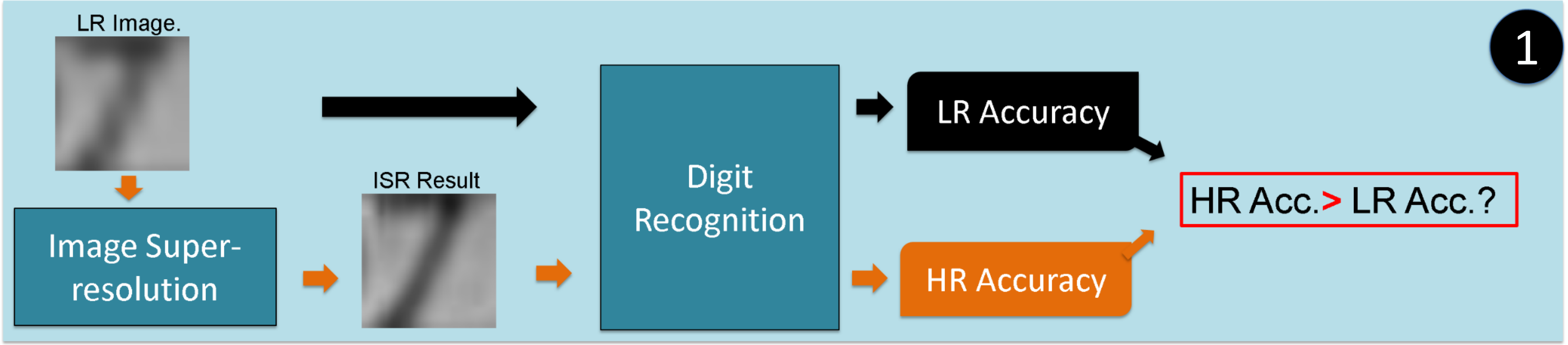
Methods merely evaluated perceptually

- Is Super-Resolution Helpful for Other Vision Tasks?
- How the usefulness correlate to perceptual quality?



Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

Helpful to Other Vision Task



- Same algorithm
- Two versions of input images

Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

Correlation with Perceptual Criteria

- Same super-resolved image
- Two evaluation methods: perceptual quality and usefulness

Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

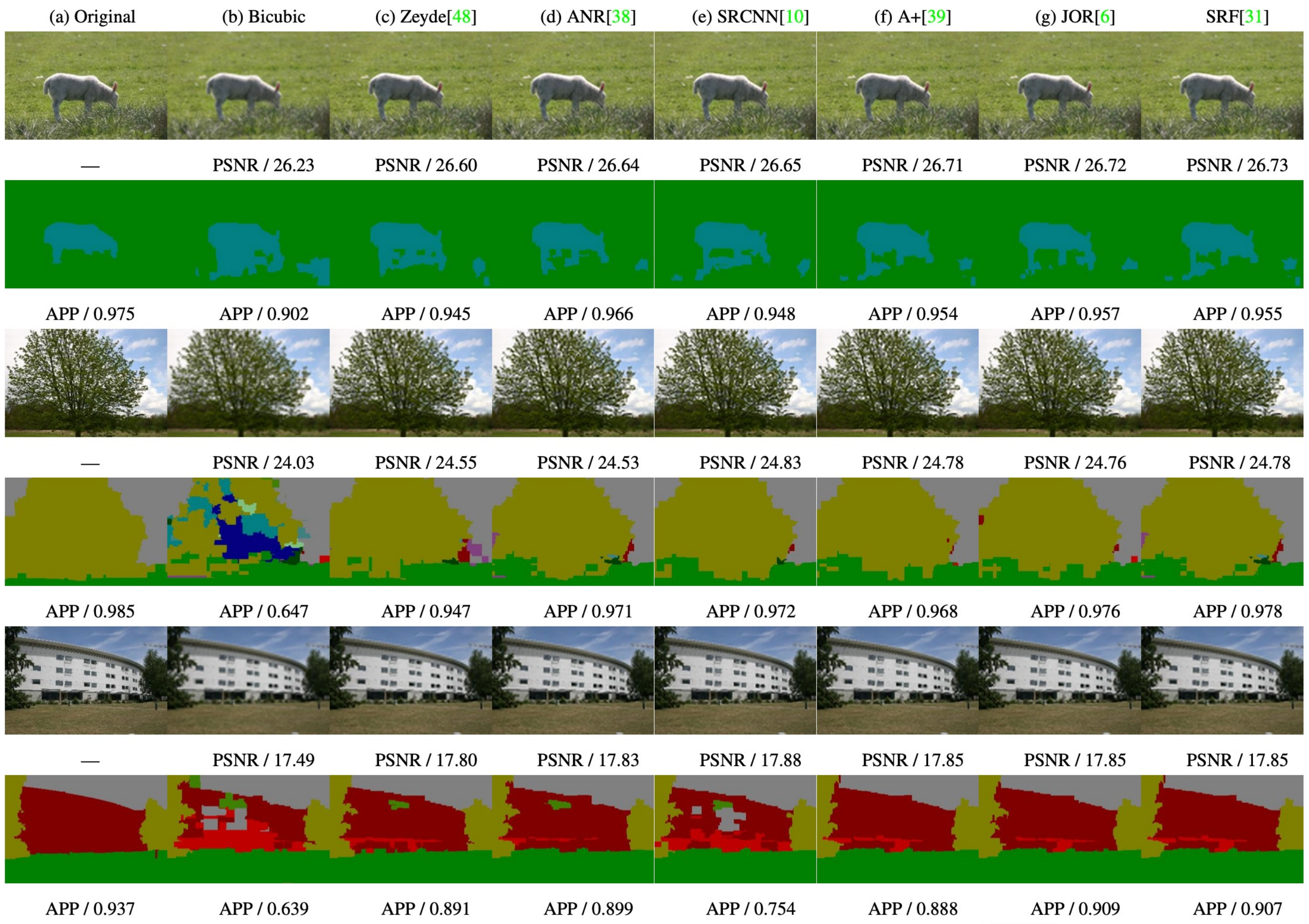
BEIJING JIAOTONG UNIVERSITY

Summary

- **Six** image super-resolution methods: Zeyde, ANR, A+, SRCNN, JOR, and SRF
- **Five** vision tasks: Boundary Detection, Semantic Image Segmentation, Digit Recognition, Scene Recognition, and Face Detection
- **Four** perceptual criteria: PSNR, SSIM, IFC, and NQM

Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

Semantic Segmentation

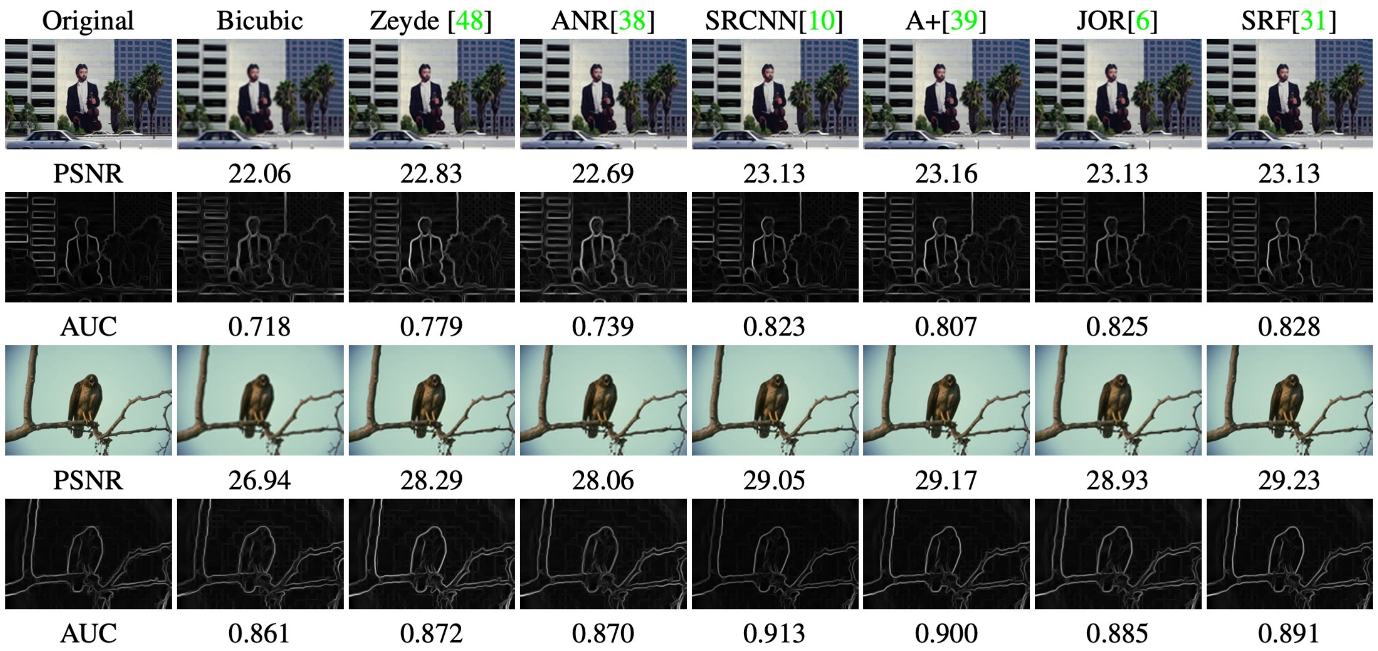


Examples for semantic image segmentation: super-resolved images with their PSNR values and the corresponding labeling results with their average precision over pixels (APP) are shown.

Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

Restoration and Understanding

Boundary Detection



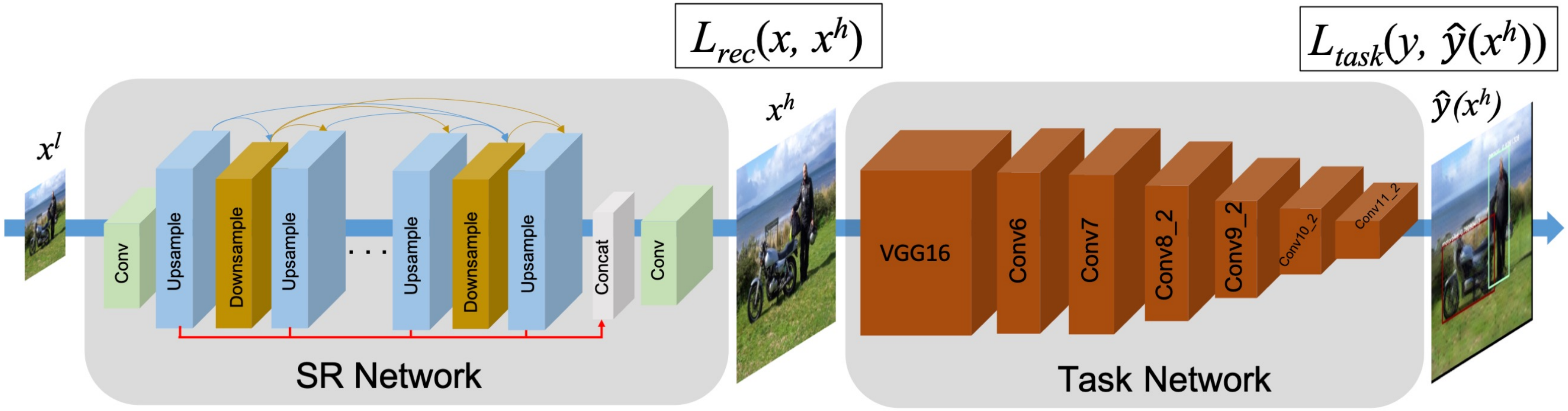
Super-resolved examples with their PSNR values and corresponding detected boundary maps by CBD with their AUC values.

Dengxin Dai, Yujian Wang, Yuhua Chen and Luc Van Gool. 2016. Is Image Super-Resolution Helpful for Other Vision Tasks? In IEEE Winter Conference on Applications of Computer Vision. 1-9.

Restoration and Understanding

Task Driven

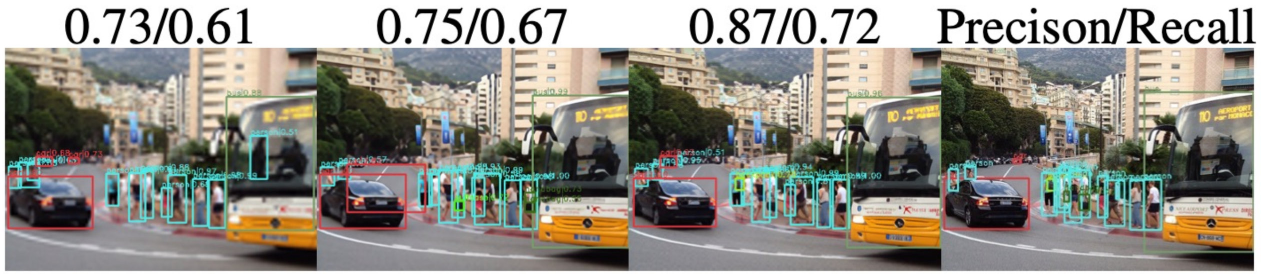
(a) HR | (b) LR | (c) Bicubic SR PSNR: 21.26 | (d) SR (no task) PSNR: 22.02 | (e) TDSR (proposed) PSNR: 21.54



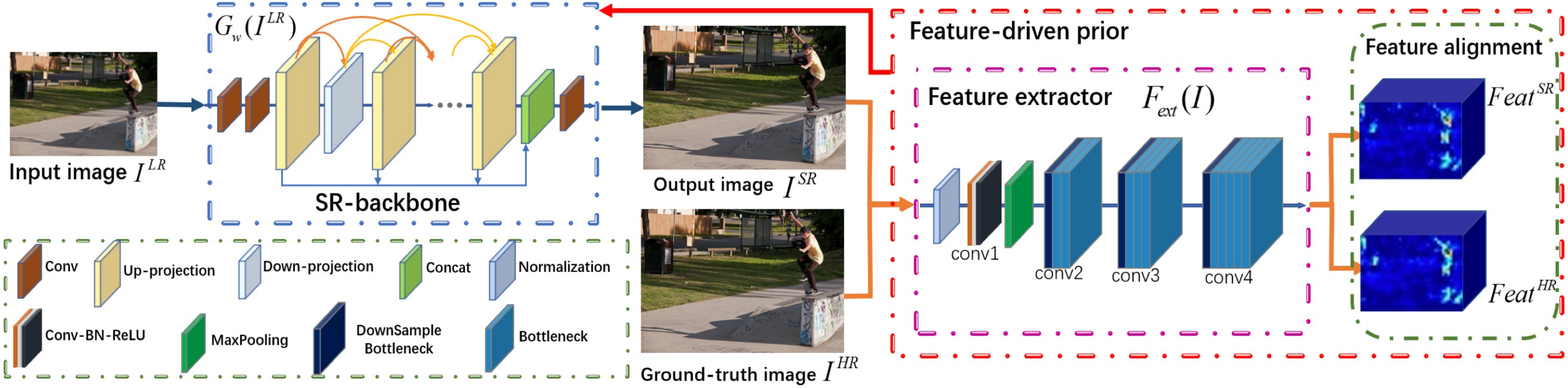
M. Haris, G. Shakhnarovich, N. Ukita, "Task-Driven Super Resolution: Object Detection in Low-resolution Images.", ICONIP2021.

Restoration and Understanding

Feature Driven



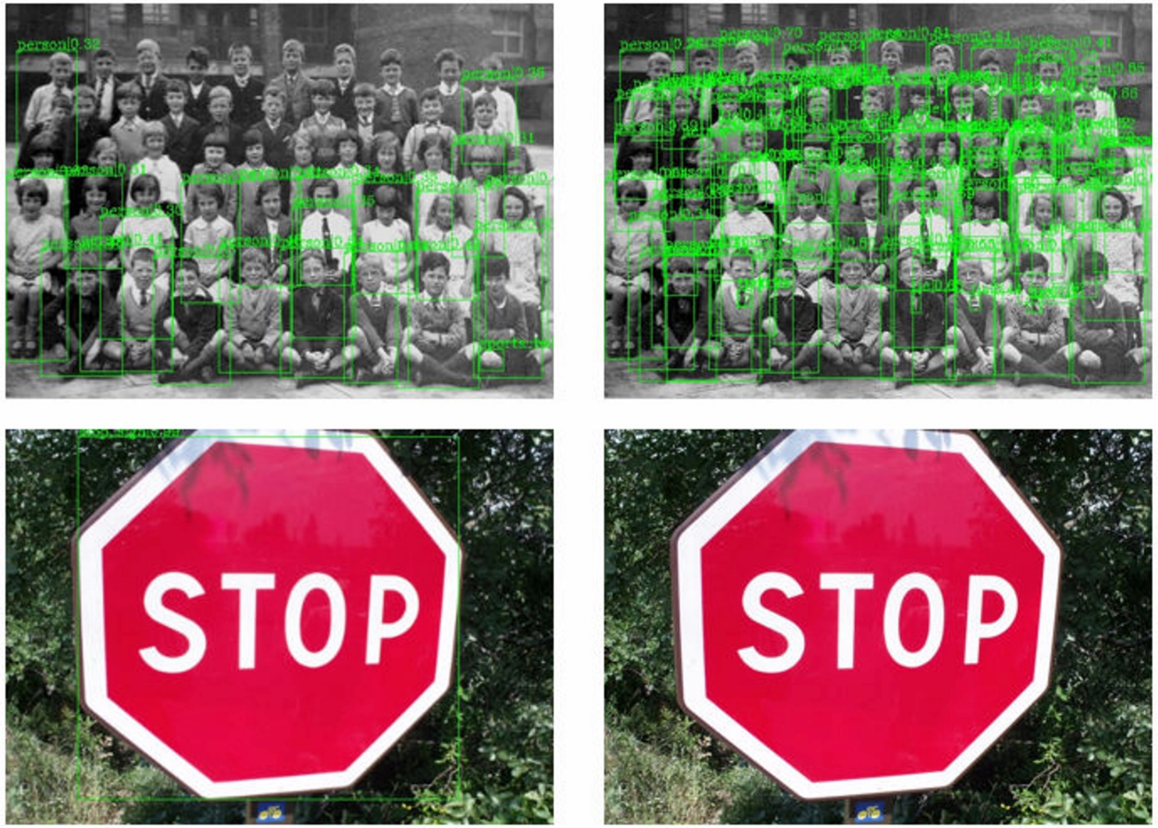
(a) Bicubic (b) D-DBPN (c) FDSR(ours) (d) HR(GT)



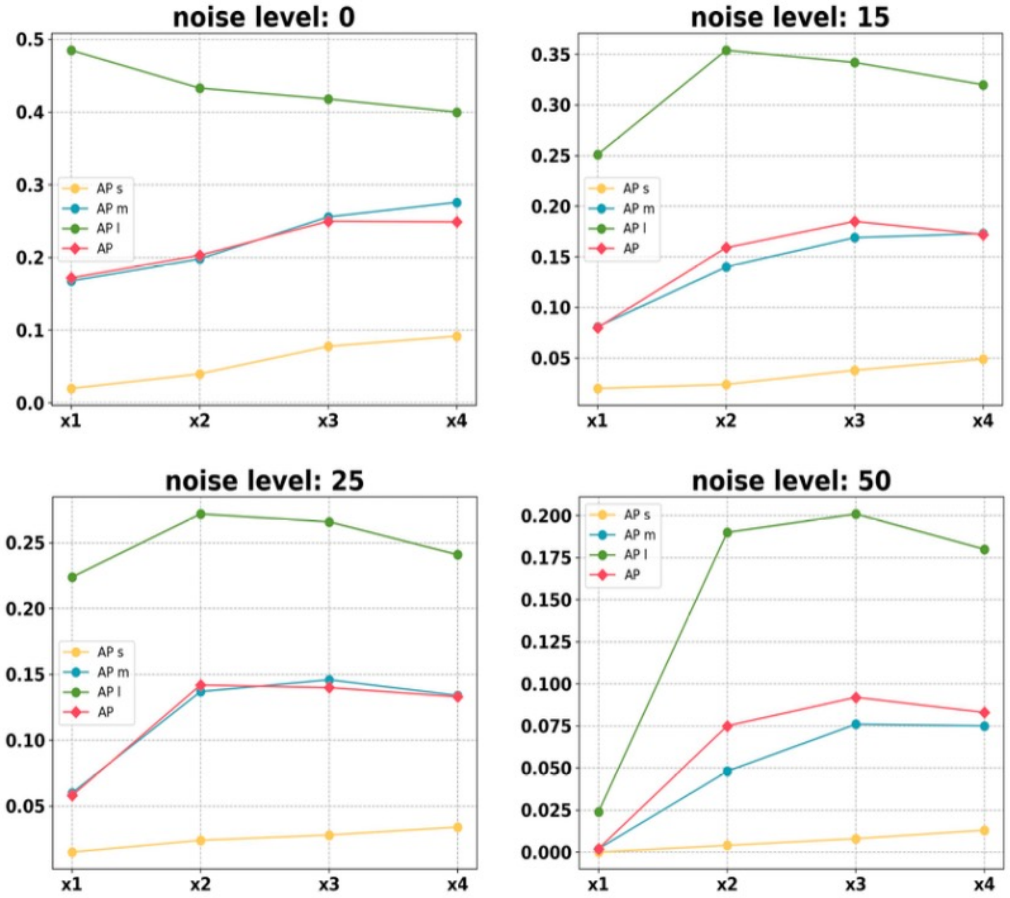
Bin Wang, Tao Lu and Yanduo Zhang. 2020. Feature-Driven Super-Resolution for Object Detection. In IEEE International Conference on Control, Robotics and Cybernetics. 211-215.

Restoration and Understanding

Equivariant Representation

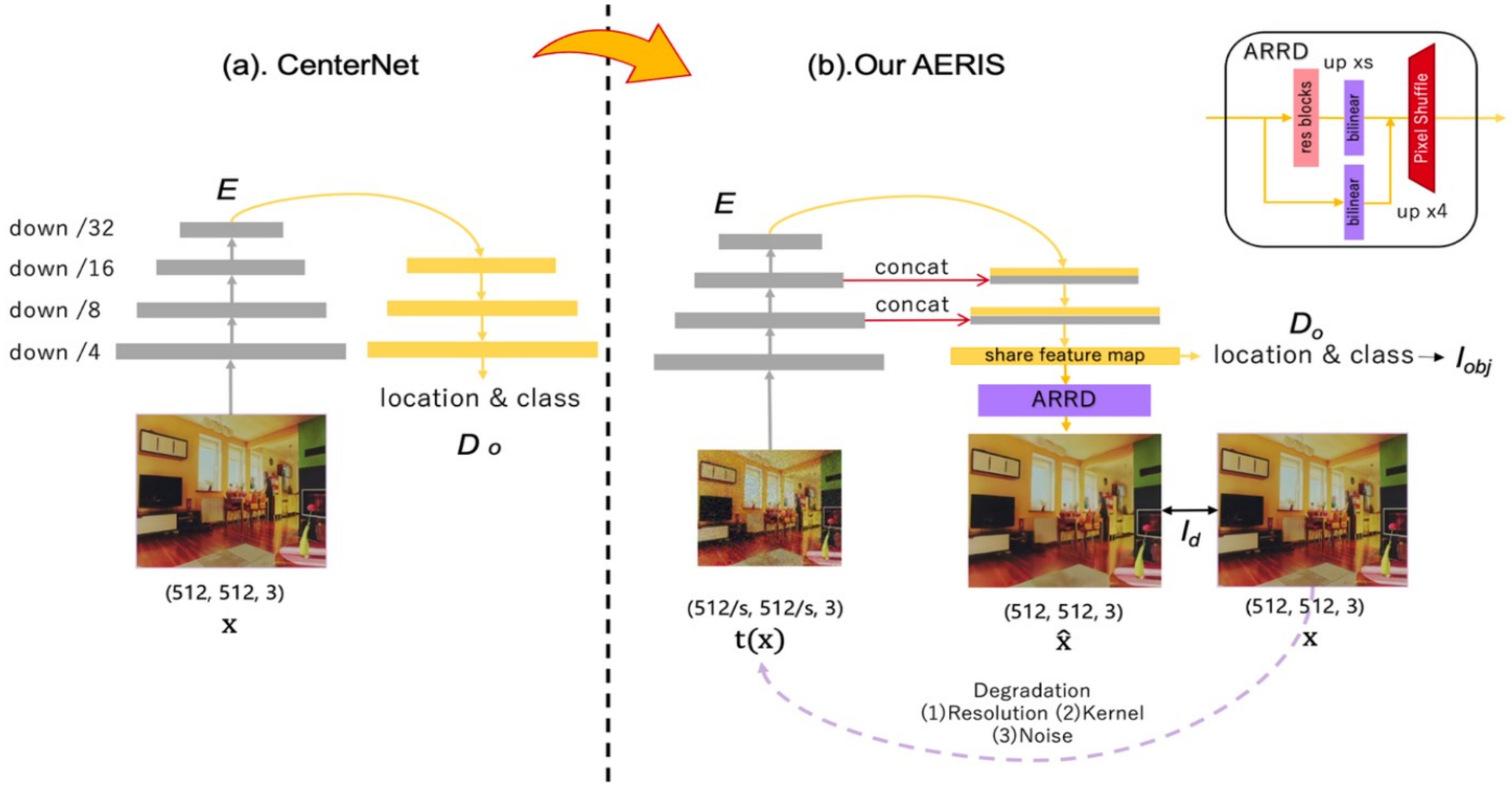


resolution: (667, 400)   resolution: (2666, 1600)



Cui, Ziteng, Ying J. Zhu, Lin Gu, Guo-Jun Qi, Xiaoxiao Li, Renrui Zhang, Zenghui Zhang and Tatsuya Harada. 2022. "Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection." European Conference on Computer Vision.

Equivariant Representation



Cui, Ziteng, Ying J. Zhu, Lin Gu, Guo-Jun Qi, Xiaoxiao Li, Renrui Zhang, Zenghui Zhang and Tatsuya Harada. 2022. "Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection." European Conference on Computer Vision.

Equivariant Representation

Algorithm 1 AERIS Algorithm Pipeline

(1). Data Generation:

B: batch size, C: channel, H: image height, W: image width

inputs: HR image $x = (B, C, H, W)$, down-sample factor $s \sim (1.0, 4.0)$

outputs: degraded LR image $t(x) = (B, C, \frac{H}{s}, \frac{W}{s})$

  for i in range(B): **do**

    (1). Convolution with blur kernel k

    (2). Down-sampling with rate s

    (3). Add noise n

  end for

(2). Training:

inputs: Degraded LR image $t(x) = (B, C, \frac{H}{s}, \frac{W}{s})$

outputs: detection output, estimated SR image $\hat{x}$

encoding:

$t(x) \xrightarrow{E} E(t(x))$

decoding:

data restoration decoding: $\hat{x} = D_r(E(t(x)))$

detection decoding: detection results $= D_o(E(t(x)))$

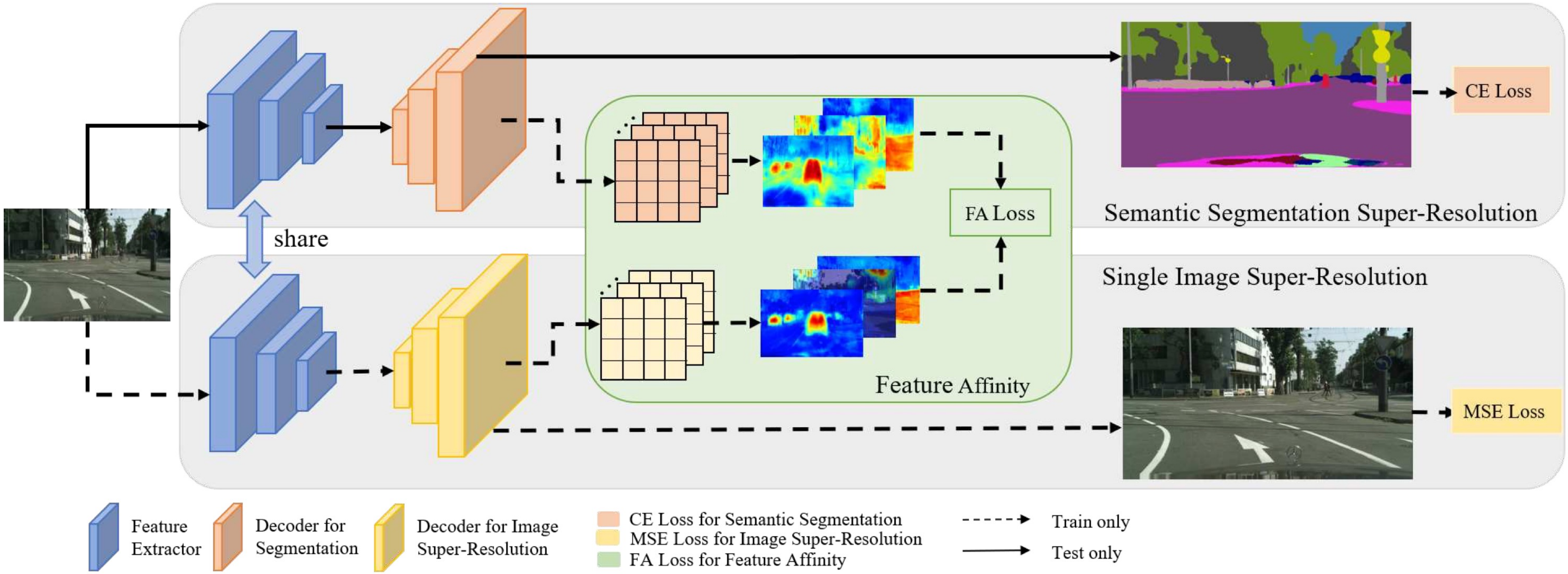
Cui, Ziteng, Ying J. Zhu, Lin Gu, Guo-Jun Qi, Xiaoxiao Li, Renrui Zhang, Zenghui Zhang and Tatsuya Harada. 2022. "Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection." European Conference on Computer Vision.

Restoration and Understanding

Equivariant Representation

Cui, Ziteng, Ying J. Zhu, Lin Gu, Guo-Jun Qi, Xiaoxiao Li, Renrui Zhang, Zenghui Zhang and Tatsuya Harada. 2022. "Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection." European Conference on Computer Vision.

Parallel Structure

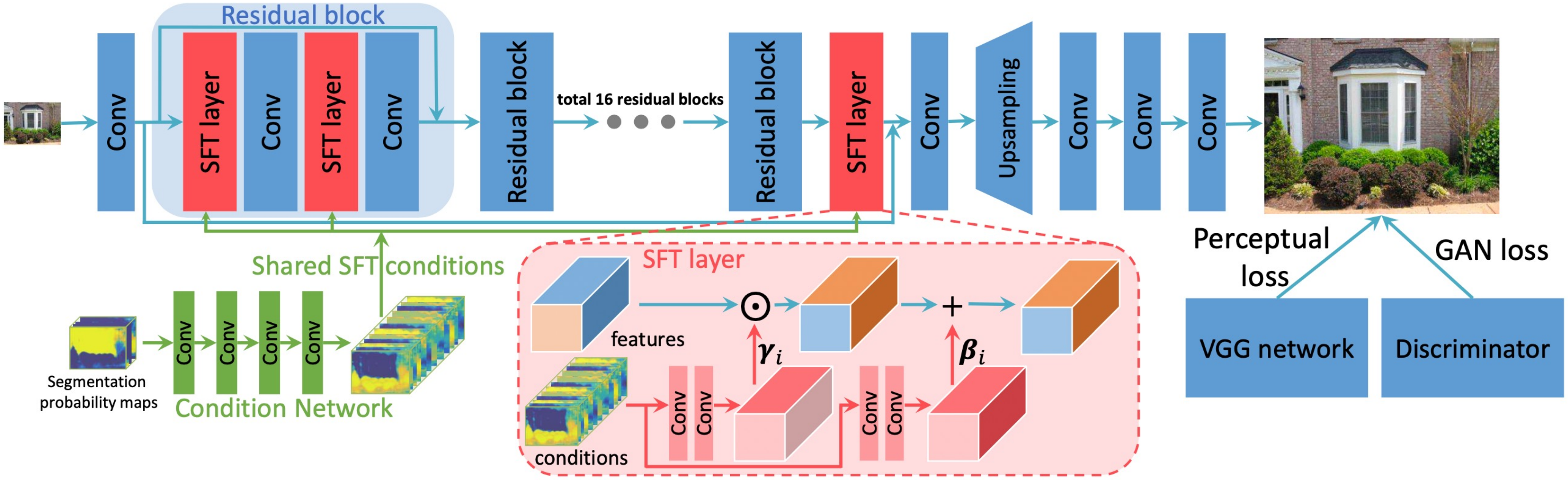


The overview of DSRL framework, which includes three parts: Semantic Segmentation Super-Resolution (SSSR) branch, Single Image Super-Resolution (SISR) branch, and Feature Affinity (FA) module. The encoder is shared between the SSSR branch and the SISR branch.

Li Wang, Dong Li, Yousong Zhu, Lu Tian, and Yi Shan. 2020. Dual Super-Resolution Learning for Semantic Segmentation. In IEEE Conference on Computer Vision and Pattern Recognition.

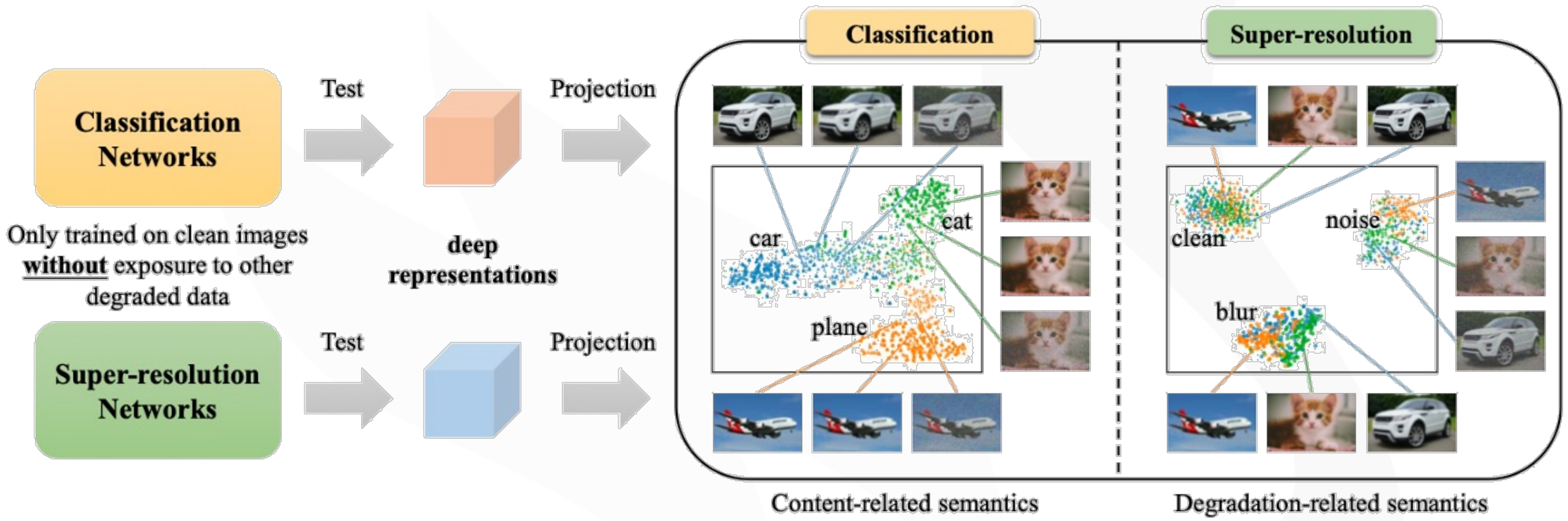
Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform

Xintao Wang[1] Ke Yu[1] Chao Dong[2] Chen Change Loy[1]
[1]CUHK - SenseTime Joint Lab, The Chinese University of Hong Kong, [2]SenseTime Research



BEIJING JIAOTONG UNIVERSITY

Semantic

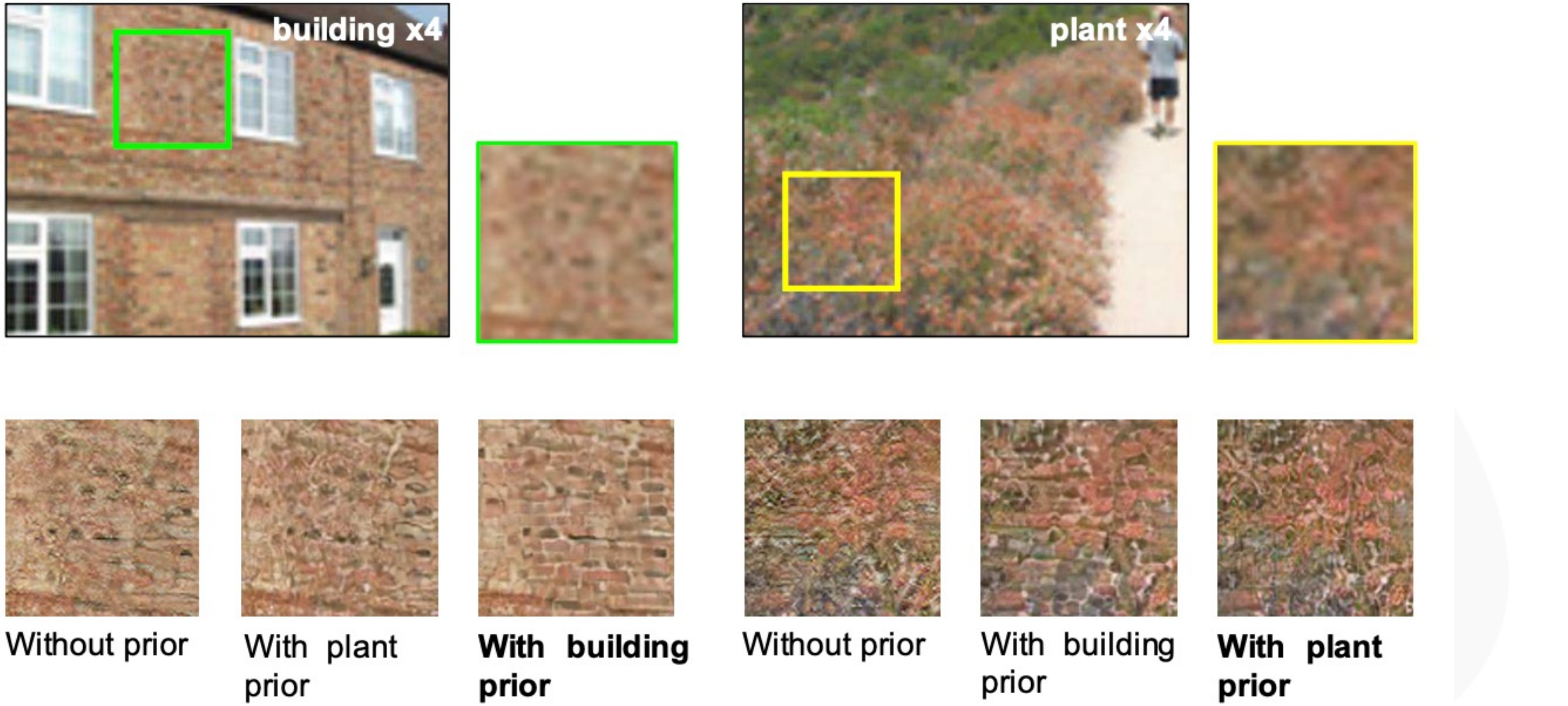


Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

BEIJING JIAOTONG UNIVERSITY

Semantic

The extracted building and plant patches from two low- resolution images look very similar. Using adversarial loss and perceptual loss without prior could add details that are not faithful to the underlying class.



Xiangtao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In IEEE Conference on Computer Vision and Pattern Recognition.

Restoration and Understanding

Semantic

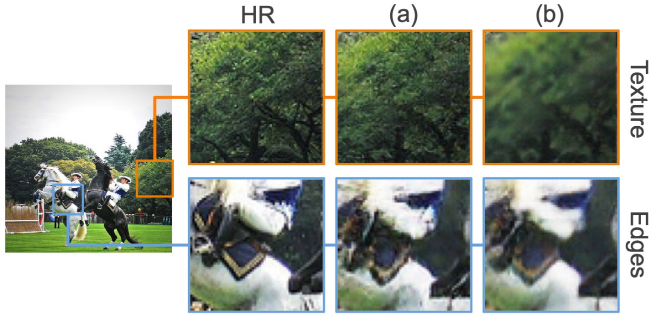
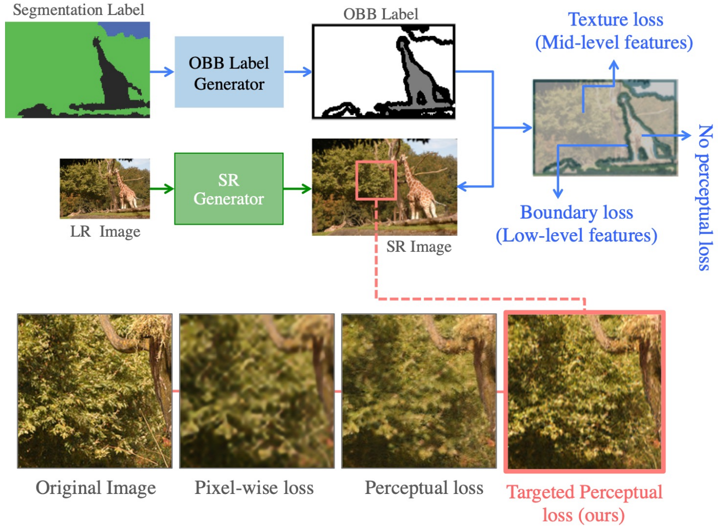
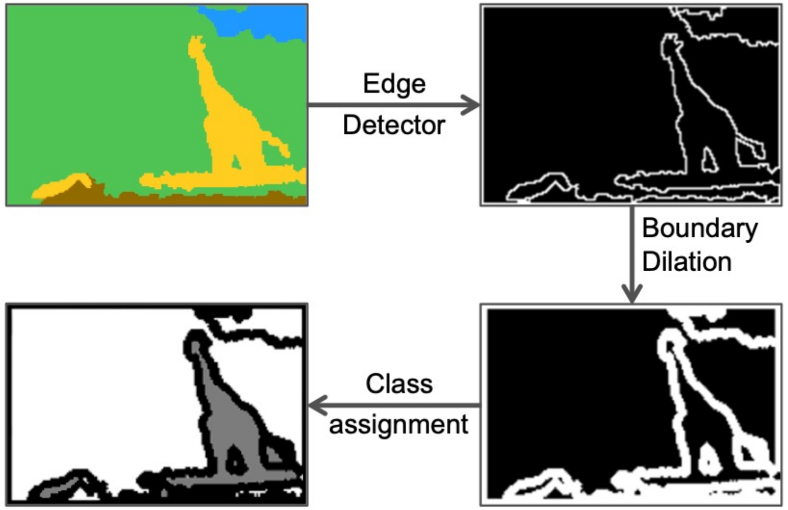
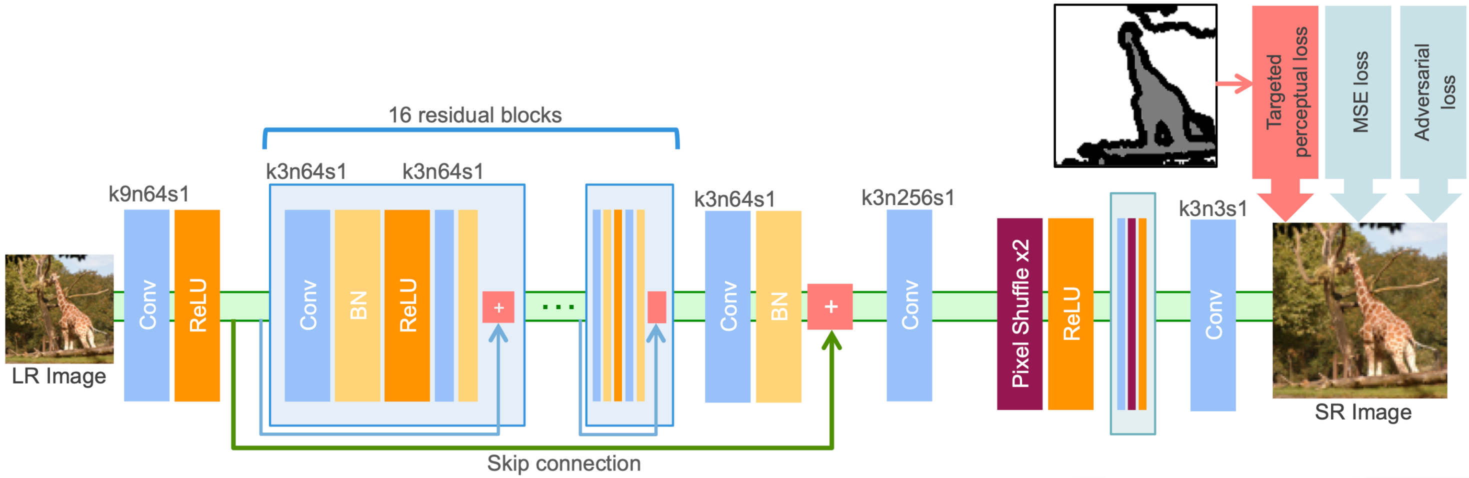


Figure 2. The effect of choosing different CNN layers to estimate the perceptual loss on different regions of an image, e.g., edges and textures: (a) using a deeper convolutional layer (mid-level features), ReLU 4-1 of VGG-16 [29] and, (b) using an early convolutional layer (low-level features), ReLU 1-2 of the VGG-16 network.

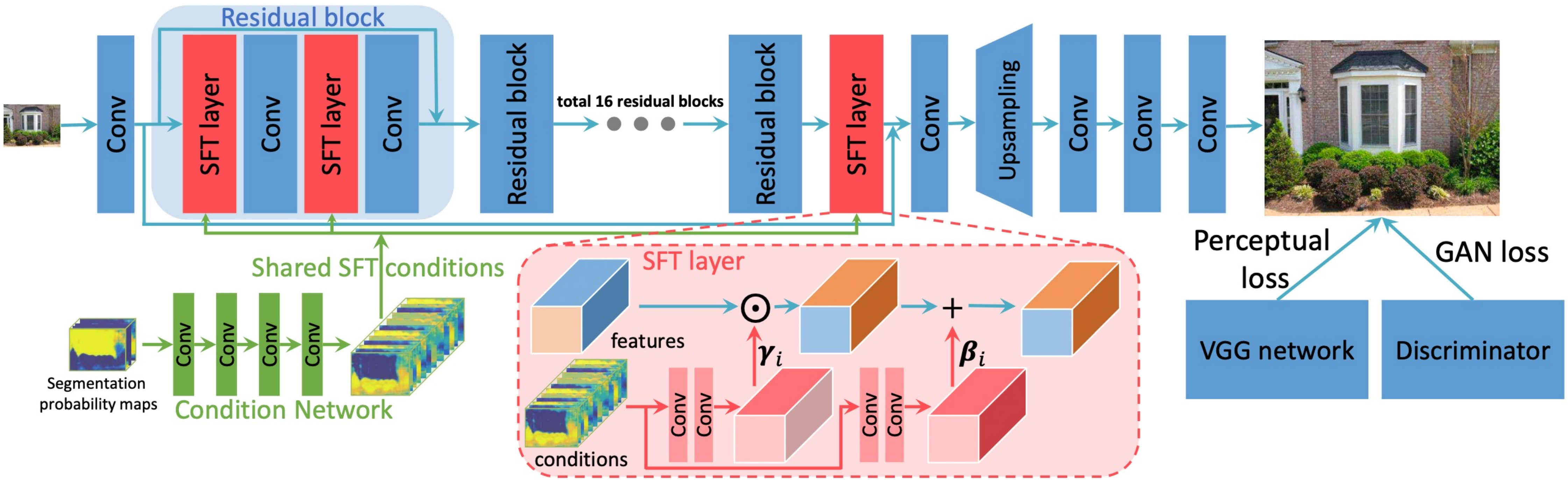






Mohammad Saeed Rad, Behzad Bozorgtabar, Urs-Viktor Marti, Max Basler, Hazım Kemal Ekenel, and Jean-Philippe Thiran. 2019. SROBB: Targeted perceptual loss for single image super-resolution. In IEEE International Conference on Computer Vision. 2710-2719 .

Semantic

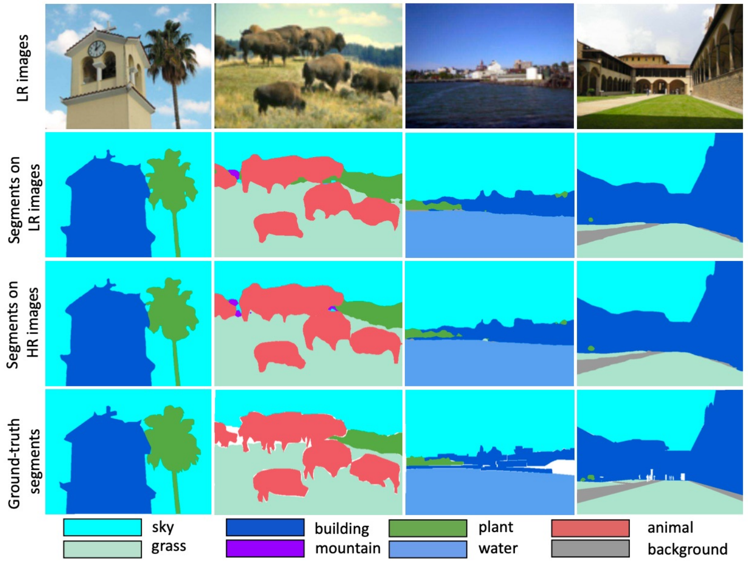


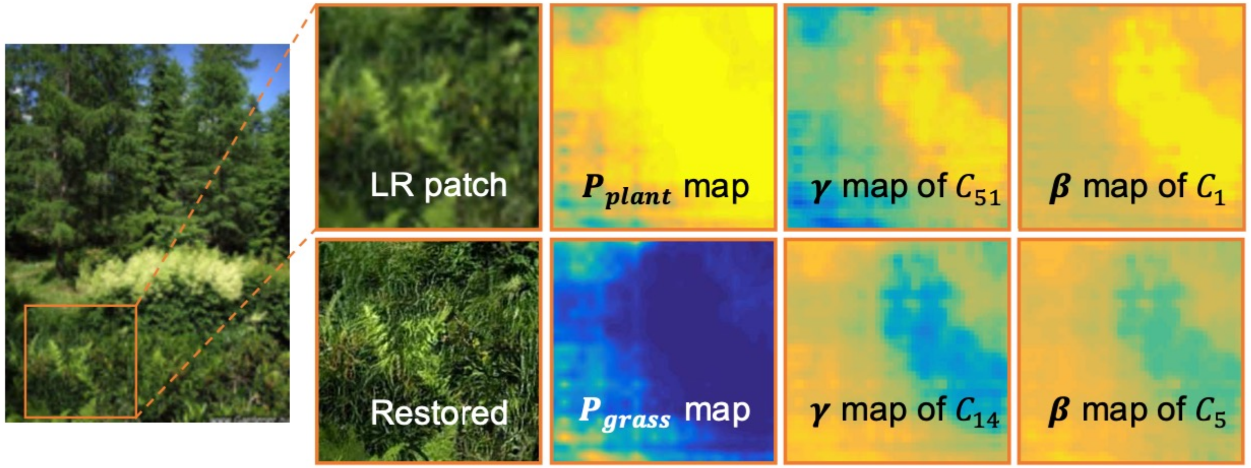
$$\hat{\boldsymbol{y}} = G_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{\gamma}, \boldsymbol{\beta}), \quad (\boldsymbol{\gamma}, \boldsymbol{\beta}) = \mathcal{M}(\Psi)$$

$$\mathrm{SFT}(\boldsymbol{F} \mid \boldsymbol{\gamma}, \boldsymbol{\beta}) = \boldsymbol{\gamma} \odot \boldsymbol{F} + \boldsymbol{\beta}$$

Xiangtao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In IEEE Conference on Computer Vision and Pattern Recognition.

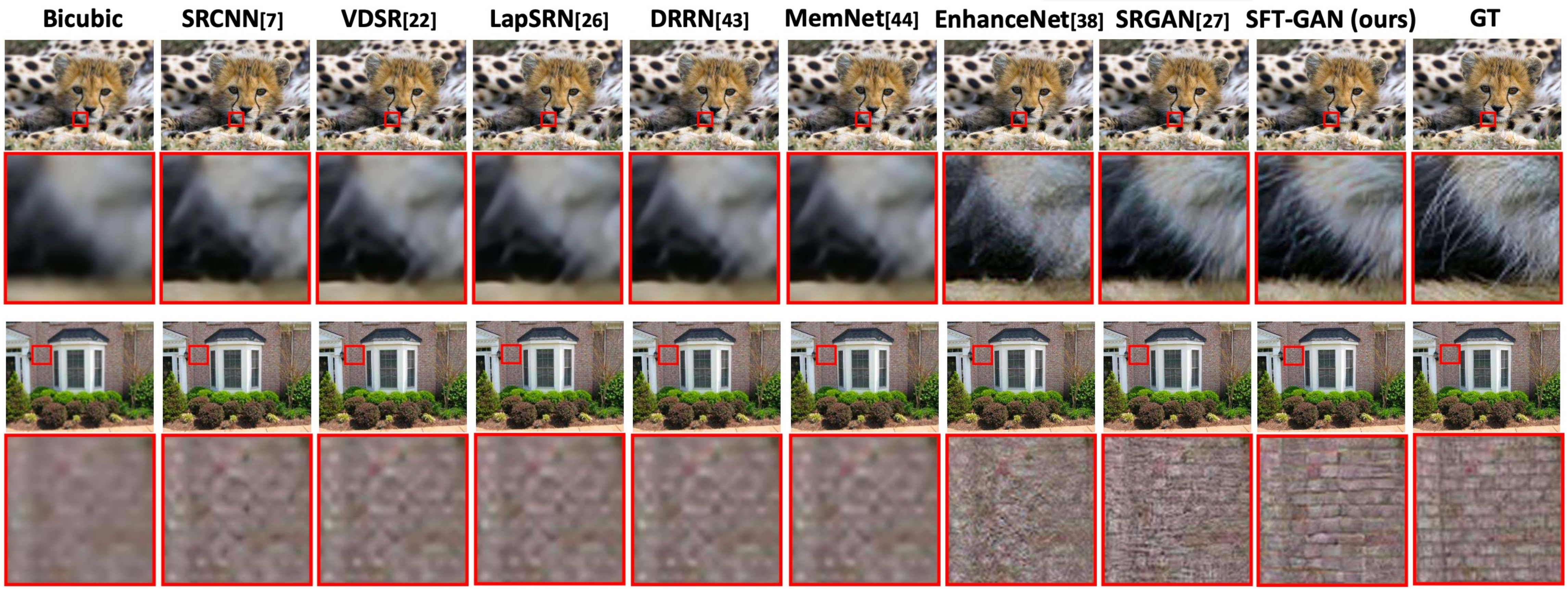
Semantic





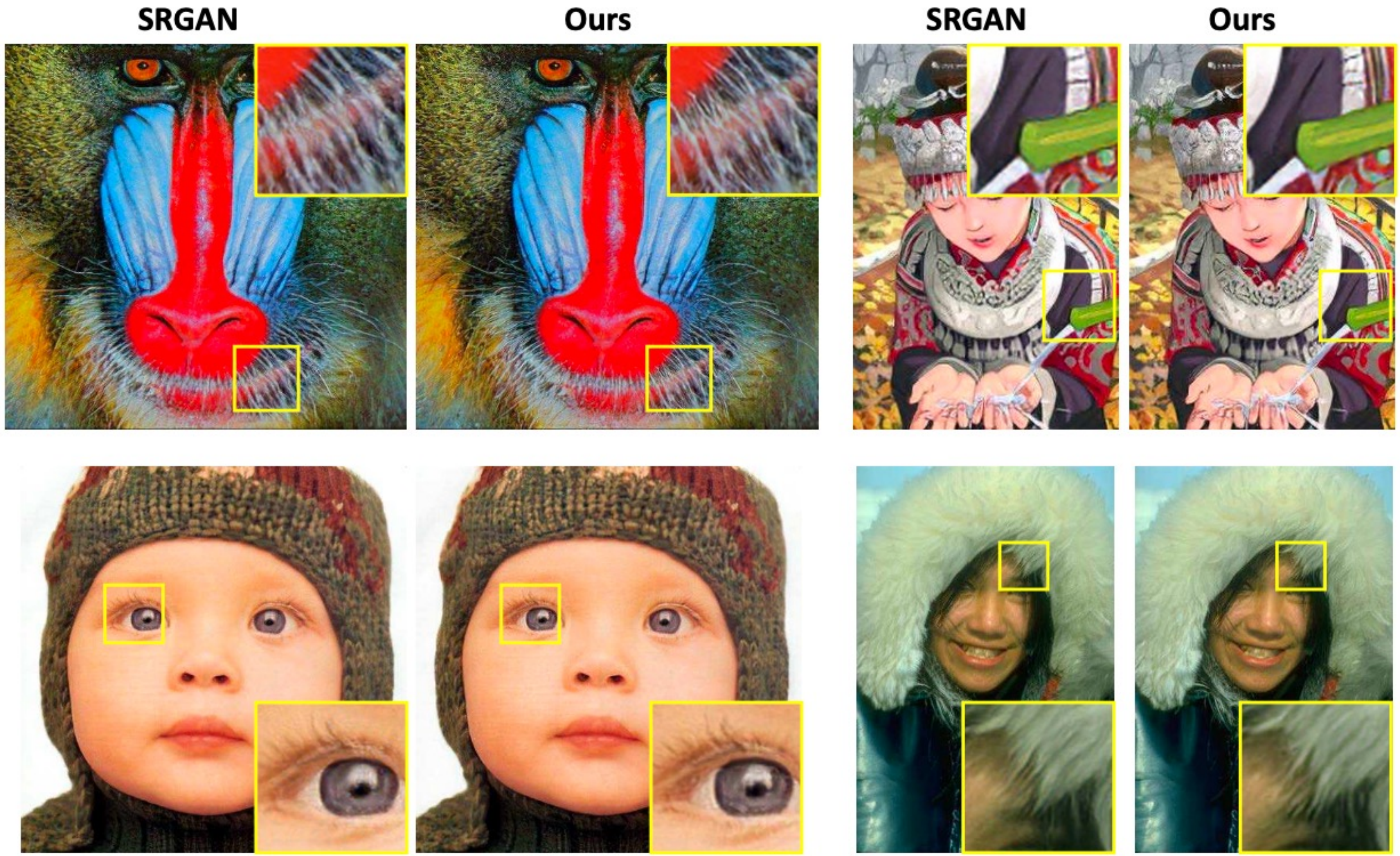
Xiangtao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In IEEE Conference on Computer Vision and Pattern Recognition.

Semantic



Xiangtao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In IEEE Conference on Computer Vision and Pattern Recognition.

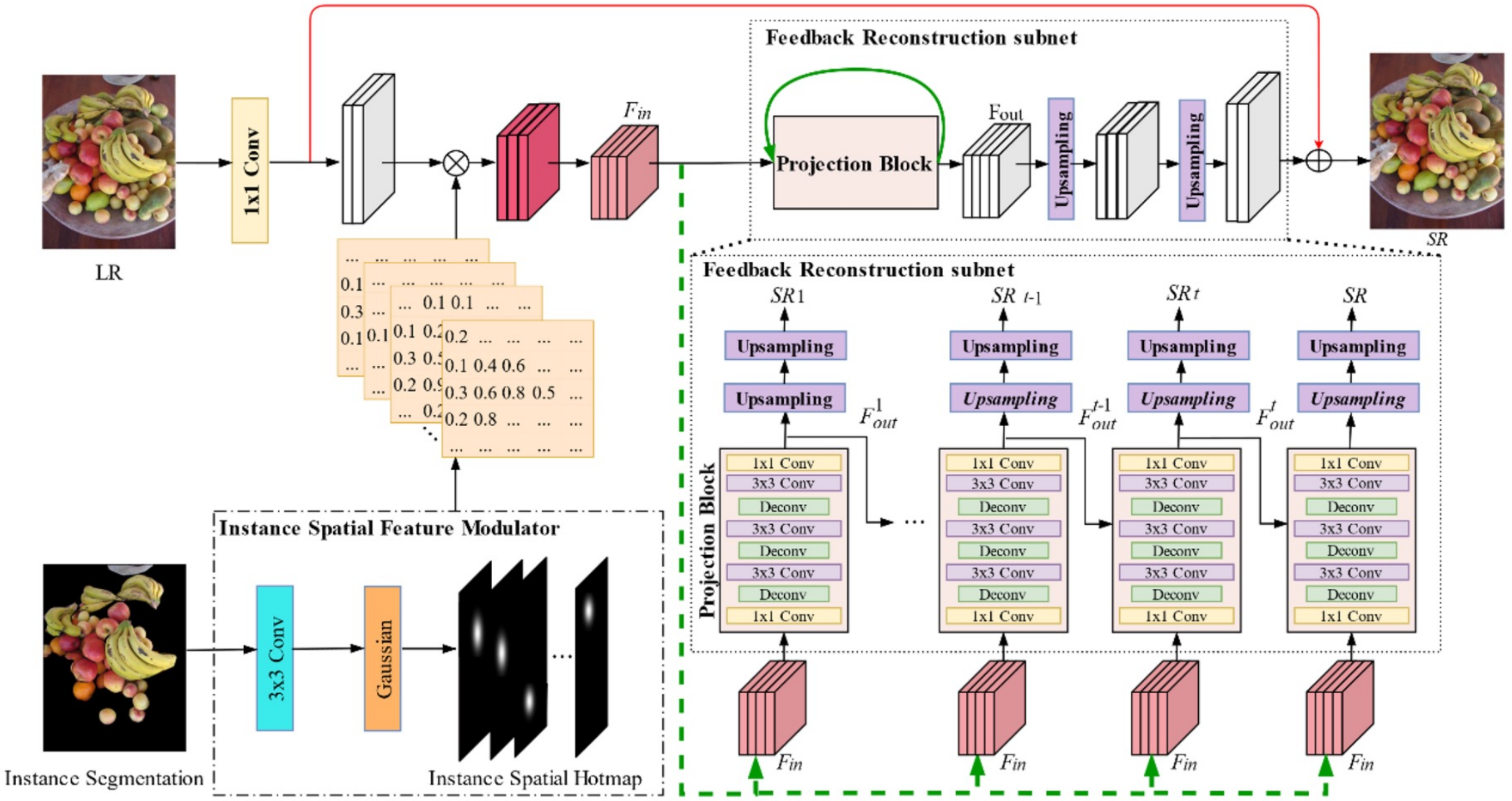
Semantic



When facing with other scenes or the absence of segmentation probability maps, our model degenerates itself as SR- GAN and produces comparative results with SRGAN.

Xiangtao Wang, Ke Yu, Chao Dong, and Chen Change Loy. 2018. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In IEEE Conference on Computer Vision and Pattern Recognition.

Instance



Lihua Fu, Hanxu Jiang, Huixian Wu, Shaoxing Yan, Junxiang Wang, and Dan Wang. 2022. Image super-resolution reconstruction based on instance spatial feature modulation and feedback mechanism. Applied Intelligence. 1–15

Thank you