Perceptual Image Processing ALgorithms (PIPAL)

A Large-Scale Image Quality Assessment Dataset for Perceptual Image Restoration

Jinjin Gu[1], Haoming Cai[1,2], Haoyu Chen[1], Xiaoxing Ye[1],
Jimmy S[3]. Ren, Chao Dong[2]

[1] The Chinese University of Hong Kong, Shenzhen
[2] Shenzhen Institutes of Advanced Technology
[3] SenseTime Research

Image Restoration (IR) and Image Quality Assessment (IQA)

- **Image Restoration (IR)** aims at recovering a high-quality image from its degraded observation.
- **Image Quality Assessment (IQA)** methods were developed to measure the Perceptual Quality of images.
- **IQA methods** are widely used to evaluate **IR algorithms,** e.g., PSNR, SSIM and Perceptual Index (PI).

Perceptual Image Restoration

The invention of Generative Adversarial Networks (GANs) greatly improves the perceptual performance

Ground Truth

Less distortion
PSNR-oriented

Photo-realistic
GAN-based

Gap Between IQA Metric and Human Judgment

Increasing inconsistency between high numerical performances (PSNR, SSIM, PI, etc.) and perceptual performance.

Ground Truth
PSNR / SSIM

PSNR-oriented

GAN-based

Gap Between IQA Metric and Human Judgment

Before 2018, Evaluation Using PSNR/SSIM

Ground Truth
PSNR / SSIM

23.52 / 0.7056
Good in PSNR, SSIM

19.86 / 0.5530
Preferred by Human

Gap Between IQA Metric and Human Judgment

After 2018, Evaluation Using PI/NIQE

Ground Truth
PI / NIQE

3.80 / 6.47
Good in PI, NIQE

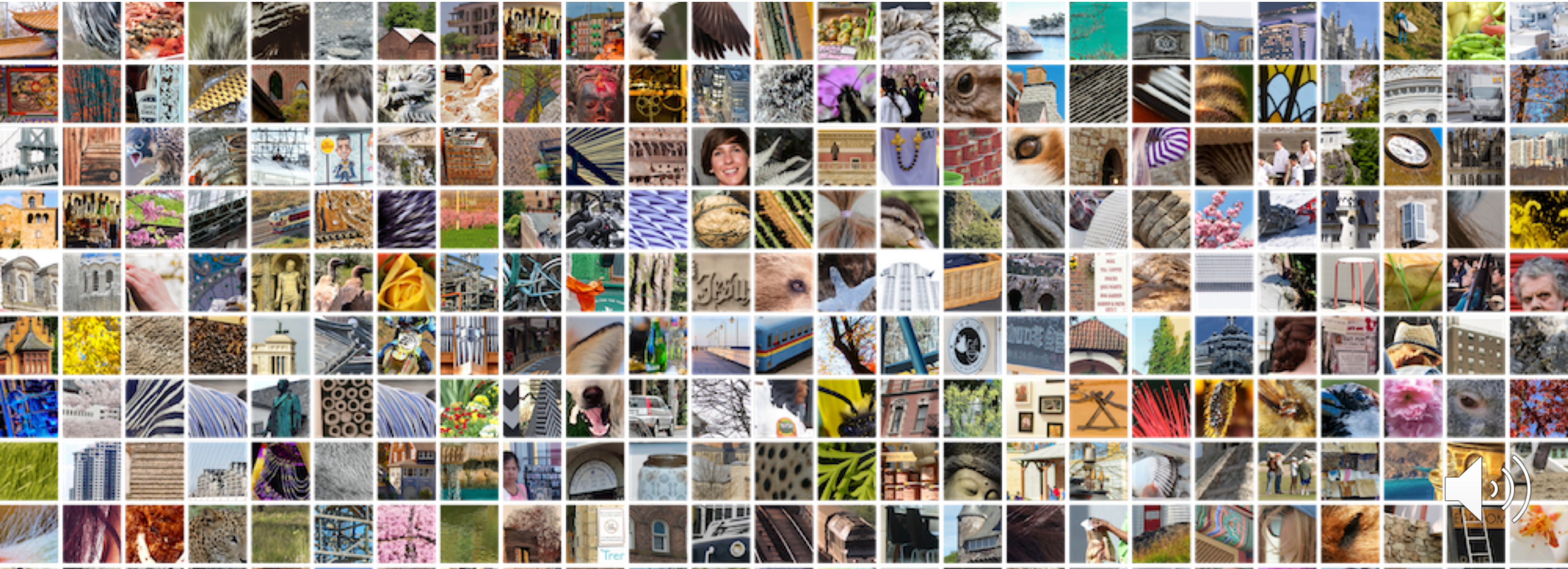
4.30 / 6.90
Preferred by Human

PI and NIQE are suggested in Blau, Y., & Michaeli, T. The perception-distortion tradeoff. CVPR 2018

- Can existing IQA methods objectively evaluate current Image Restoration algorithms, especially GAN-based algorithms?

- With the focus on beating benchmarks on the flawed IQA methods, are we getting better Image Restoration algorithms?

Perceptual Image Processing ALgorithms
PIPAL



PIPAL: Perceptual Image Processing ALgorithms

250
Reference Images

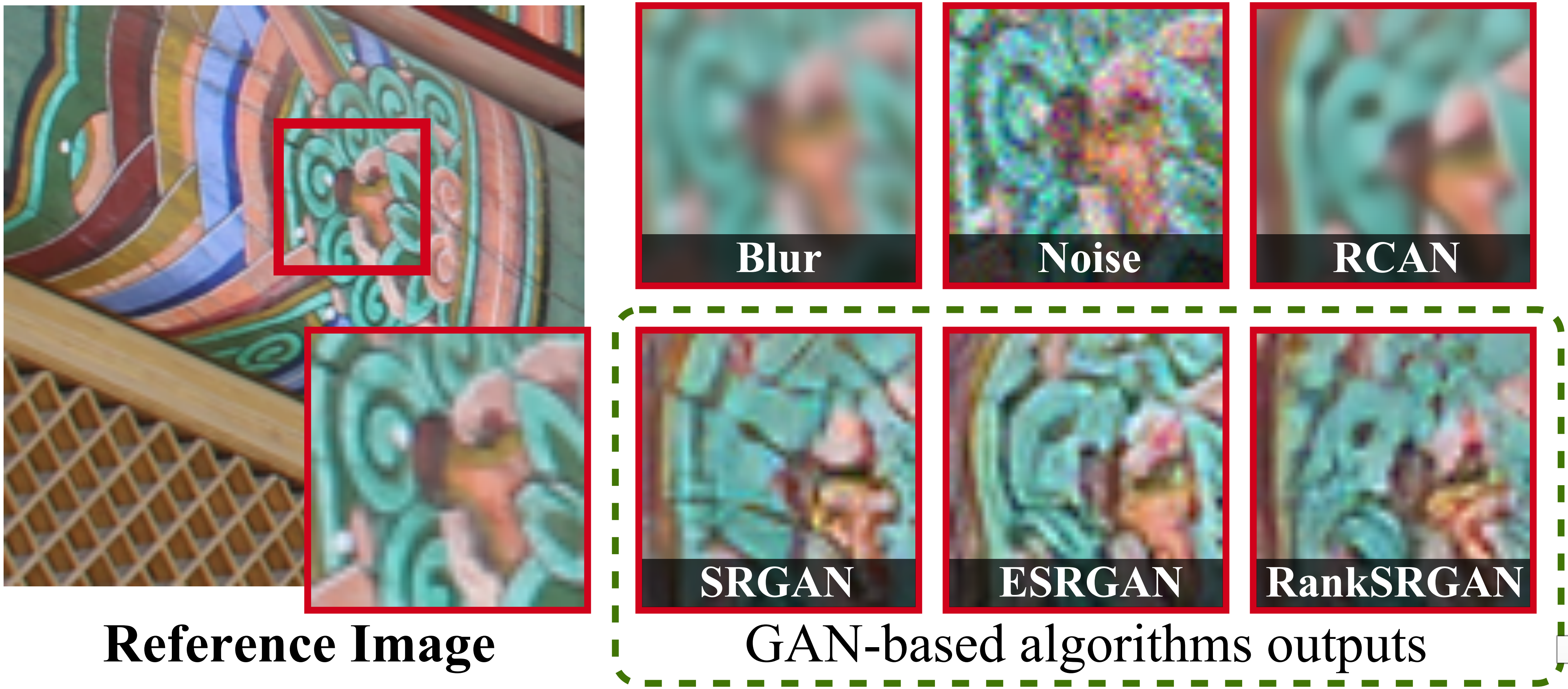
40
Distortion Types

29,000
Distortion Images

1,130,000
Human Ratings

PIPAL: Perceptual Image Processing ALgorithms

Novel GAN-based distortion



PIPAL: Perceptual Image Processing ALgorithms

Elo System:

- Possibility-based rating system, each image has an **Elo Score**.
- The difference of the Elo scores indicates the possibility of a user's preference.
- We update the Elo Score by **pairwise human judgments**.
- Extendible in the future

PIPAL: Perceptual Image Processing ALgorithms

Image A

Elo Score R_A

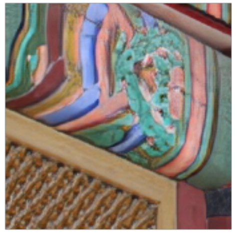


Image B

Elo Score R_B

PIPAL: Perceptual Image Processing ALgorithms

Image A
Elo Score R_A

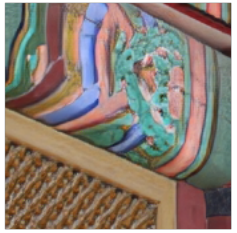


Image B
Elo Score R_B

The probability that one user prefer A than B

$$P_{A>B} = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

The probability that one user prefer B than A

$$P_{B>A} = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

PIPAL: Perceptual Image Processing ALgorithms

Image A

Elo Score R_A

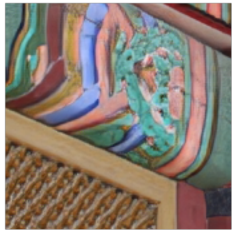


Image B

Elo Score R_B

The probability that one user prefer A than B

$$P_{A>B} = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

The probability that one user prefer B than A

$$P_{B>A} = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

User makes judgment

$$S_A = 1, S_B = 0$$

indicates user choose A,

$$S_A = 0, S_B = 1$$

indicates user choose B.

PIPAL: Perceptual Image Processing ALgorithms

Image A
Elo Score R_A

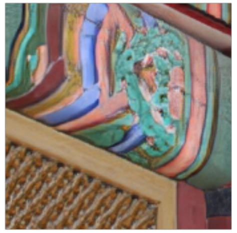


Image B
Elo Score R_B

The probability that one user prefer A than B

$$P_{A>B} = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

The probability that one user prefer B than A

$$P_{B>A} = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

User makes judgment

$$S_A = 1, S_B = 0$$

indicates user choose A,

$$S_A = 0, S_B = 1$$

indicates user choose B.

Update Elo Score

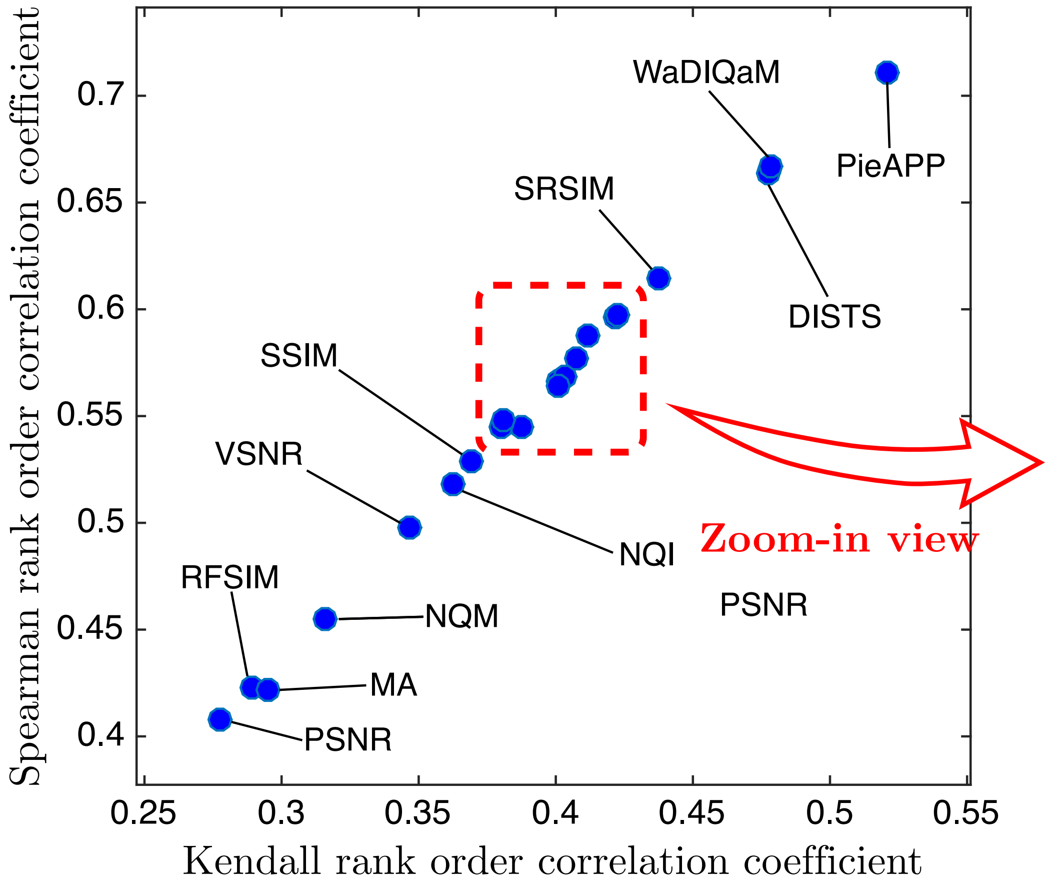
$$R_A^{new} = R_A + K(S_A - P_{A>B})$$

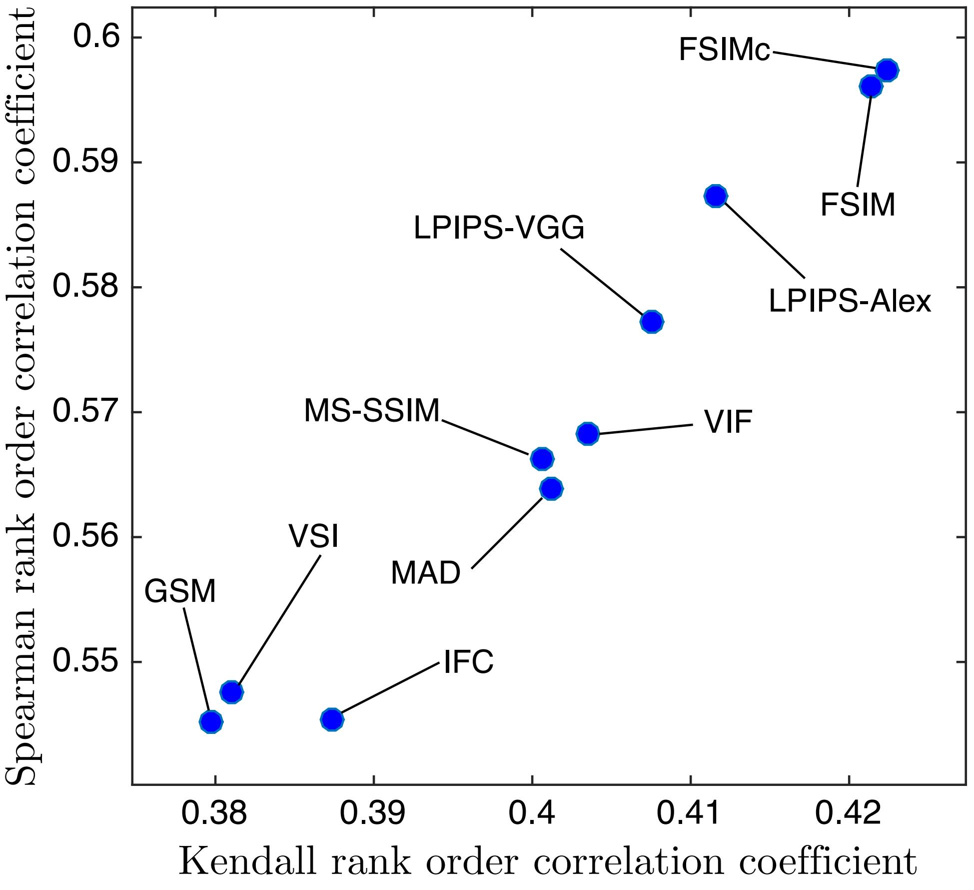
$$R_B^{new} = R_B + K(S_B - P_{B>A})$$

- Can existing IQA methods objectively evaluate current Image Restoration algorithms, especially GAN-based algorithms?
- With the focus on beating benchmarks on the flawed IQA methods, are we getting better Image Restoration algorithms?

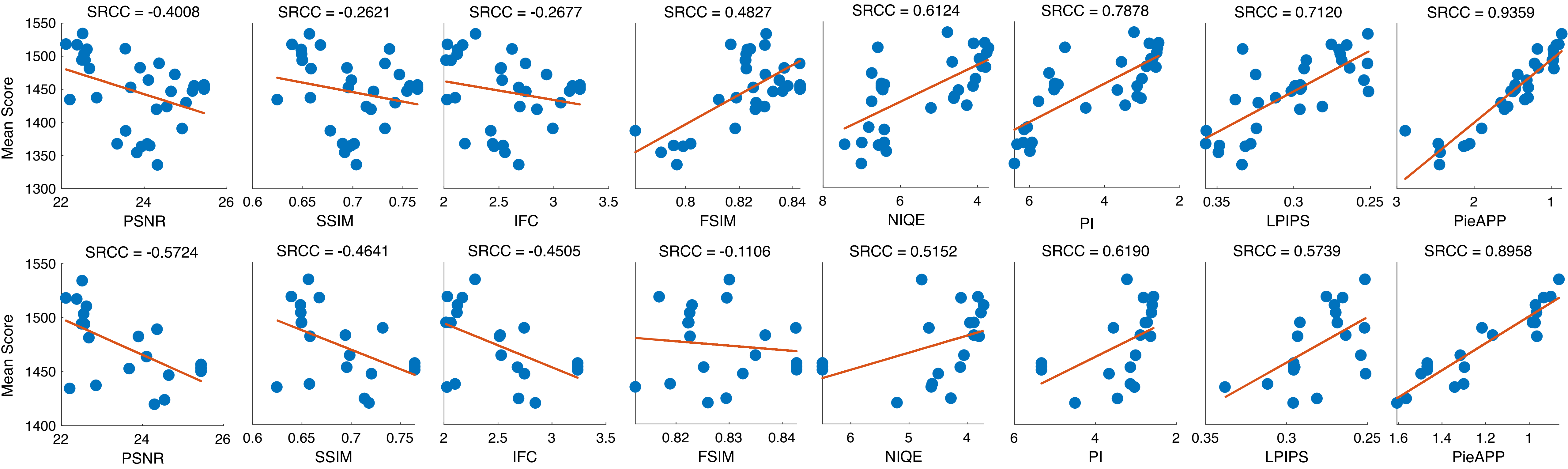
Benchmarking IQA methods

We collect 23 state-of-the-art IQA methods to build the benchmark.



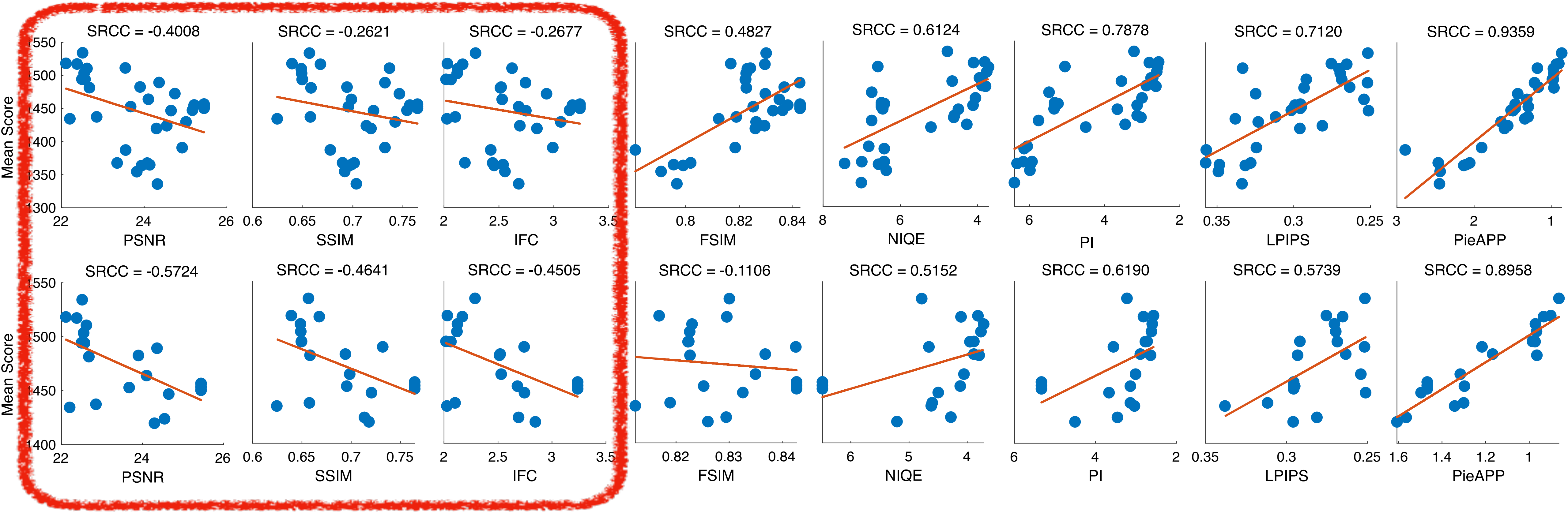


Benchmarking IQA methods



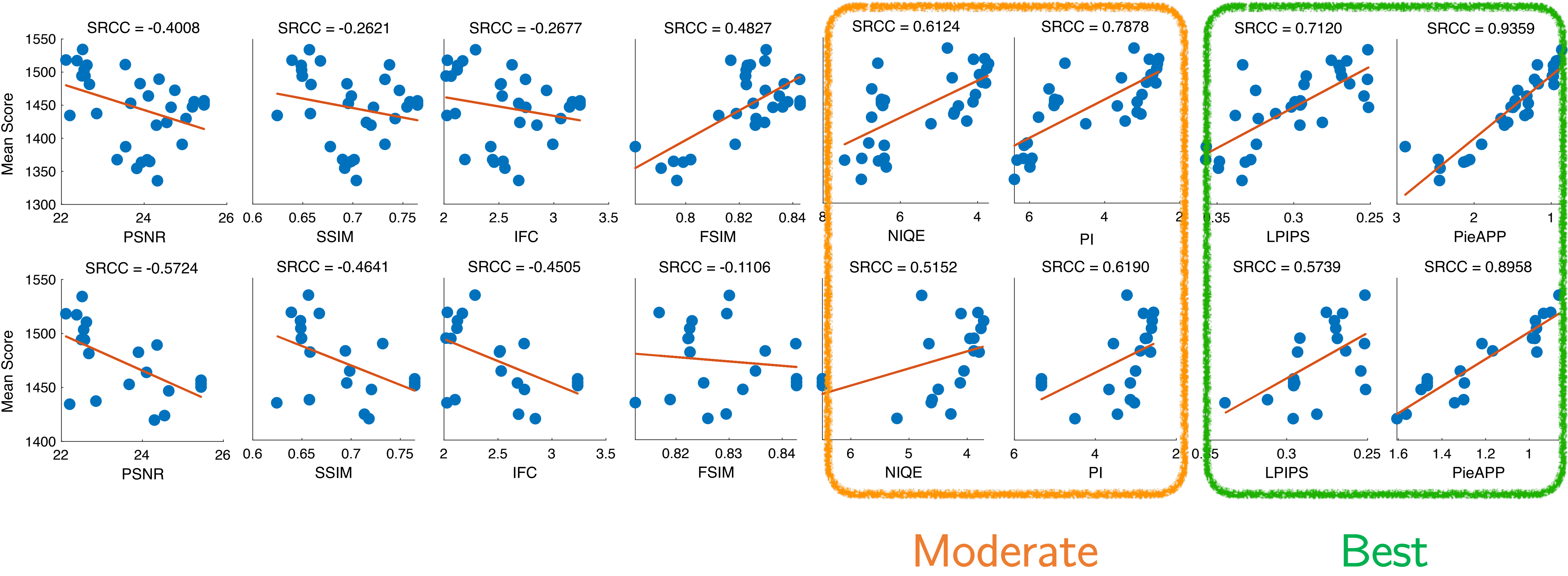
The first row shows the scatter plots of MOS score vs. IQA methods for all SR algorithms.
The second row gives scatter plots for GAN-based SR algorithms

Benchmarking IQA methods



Anti-correlated

Benchmarking IQA methods



Benchmarking IQA methods

- PIPAL poses challenges for IQA methods
- Evaluating Image Restoration algorithms only using existing IQA metrics is not appropriate
- Deep network based IQA methods (LPIPS, PieAPP, DISTS) perform better.

- Can existing IQA methods objectively evaluate current Image Restoration algorithms, especially GAN-based algorithms?

- With the focus on beating benchmarks on the flawed IQA methods, are we getting better Image Restoration algorithms?

Benchmarking Image Restoration Algorithms

We build benchmark for Image Restoration using algorithms in PIPAL.

Method	Year	PSNR ↑	NIQE ↓	PI ↓	MOS ↑
YY [52]	2013	23.35^{8}	6.4174^{8}	5.9344^{7}	1367.71^{8}
TSG [45]	2013	23.55^{7}	6.4163^{7}	6.1433^{10}	1387.24^{7}
A+ [46]	2014	23.82^{6}	6.3645^{5}	5.9897^{9}	1354.52^{12}
SRCNN [12]	2014	23.93^{5}	6.5657^{10}	5.9781^{8}	1363.68^{11}
FSRCNN [13]	2016	24.07^{4}	6.9985^{11}	6.1649^{11}	1367.49^{9}
VDSR [24]	2016	24.13^{3}	7.4436^{12}	6.3319^{12}	1364.90^{10}
EDSR [29]	2017	$\mathbf{25.17}^{2}$	6.4560^{9}	5.3463^{6}	1447.44^{6}
SRGAN [28]	2017	22.57^{10}	3.9527^{3}	2.7656^{3}	1494.14^{3}
RCAN [62]	2018	$\mathbf{25.21}^{1}$	6.4121^{6}	5.2430^{5}	1455.31^{5}
BOE [32]	2018	22.68^{9}	$\mathbf{3.7945}^{1}$	$\mathbf{2.6368}^{2}$	1481.51^{4}
ESRGAN [47]	2018	22.51^{11}	4.7821^{4}	3.2198^{4}	$\mathbf{1534.25}^{1}$
RankSRGAN [61]	2019	22.11^{12}	$\mathbf{3.8155}^{2}$	$\mathbf{2.5636}^{1}$	$\mathbf{1518.29}^{2}$

Benchmarking Image Restoration Algorithms

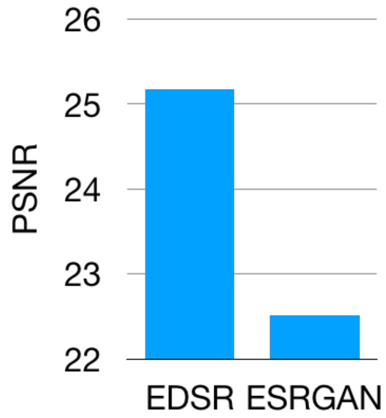
Method	Year	PSNR ↑	NIQE ↓	PI ↓	MOS ↑
YY [52]	2013	23.35^{8}	6.4174^{8}	5.9344^{7}	1367.71^{8}
TSG [45]	2013	23.55^{7}	6.4163^{7}	6.1433^{10}	1387.24^{7}
A+ [46]	2014	23.82^{6}	6.3645^{5}	5.9897^{9}	1354.52^{12}
SRCNN [12]	2014	23.93^{5}	6.5657^{10}	5.9781^{8}	1363.68^{11}
FSRCNN [13]	2016	24.07^{4}	6.9985^{11}	6.1649^{11}	1367.49^{9}
VDSR [24]	2016	24.13^{3}	7.4436^{12}	6.3319^{12}	1364.90^{10}
EDSR [29]	2017	$\mathbf{25.17}^{2}$	6.4560^{9}	5.3463^{6}	1447.44^{6}
SRGAN [28]	2017	22.57^{10}	3.9527^{3}	2.7656^{3}	1494.14^{3}
RCAN [62]	2018	$\mathbf{25.21}^{1}$	6.4121^{6}	5.2430^{5}	1455.31^{5}
BOE [32]	2018	22.68^{9}	$\mathbf{3.7945}^{1}$	$\mathbf{2.6368}^{2}$	1481.51^{4}
ESRGAN [47]	2018	22.51^{11}	4.7821^{4}	3.2198^{4}	$\mathbf{1534.25}^{1}$
RankSRGAN [61]	2019	22.11^{12}	$\mathbf{3.8155}^{2}$	$\mathbf{2.5636}^{1}$	$\mathbf{1518.29}^{2}$

Increase MOS by 90
in 4 years

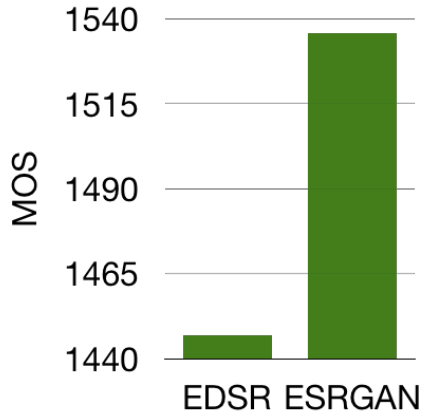
Benchmarking Image Restoration Algorithms

Method	Year	PSNR ↑	NIQE ↓	PI ↓	MOS ↑
YY [52]	2013	23.35^{8}	6.4174^{8}	5.9344^{7}	1367.71^{8}
TSG [45]	2013	23.55^{7}	6.4163^{7}	6.1433^{10}	1387.24^{7}
A+ [46]	2014	23.82^{6}	6.3645^{5}	5.9897^{9}	1354.52^{12}
SRCNN [12]	2014	23.93^{5}	6.5657^{10}	5.9781^{8}	1363.68^{11}
FSRCNN [13]	2016	24.07^{4}	6.9985^{11}	6.1649^{11}	1367.49^{9}
VDSR [24]	2016	24.13^{3}	7.4436^{12}	6.3319^{12}	1364.90^{10}
EDSR [29]	2017	$\mathbf{25.17}^{2}$	6.4560^{9}	5.3463^{6}	1447.44^{6}
SRGAN [28]	2017	22.57^{10}	3.9527^{3}	2.7656^{3}	1494.14^{3}
RCAN [62]	2018	$\mathbf{25.21}^{1}$	6.4121^{6}	5.2430^{5}	1455.31^{5}
BOE [32]	2018	22.68^{9}	$\mathbf{3.7945}^{1}$	$\mathbf{2.6368}^{2}$	1481.51^{4}
ESRGAN [47]	2018	22.51^{11}	4.7821^{4}	3.2198^{4}	$\mathbf{1534.25}^{1}$
RankSRGAN [61]	2019	22.11^{12}	$\mathbf{3.8155}^{2}$	$\mathbf{2.5636}^{1}$	$\mathbf{1518.29}^{2}$

Compare EDSR and ESRGAN



Higher is better

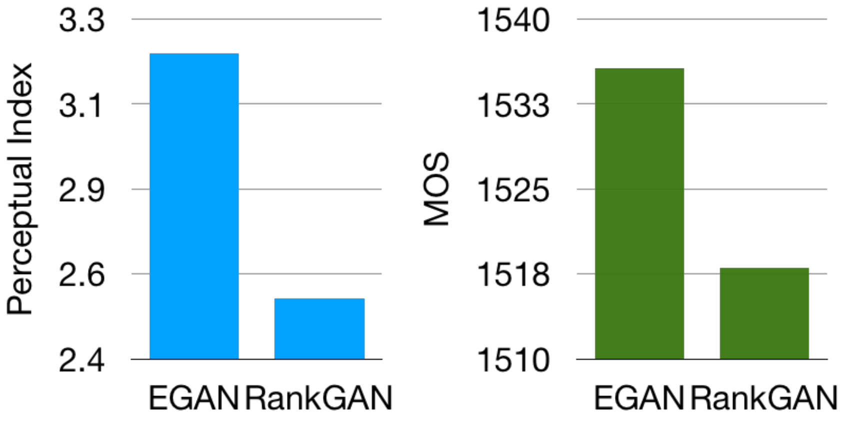


Higher is better

Benchmarking Image Restoration Algorithms

Method	Year	PSNR ↑	NIQE ↓	PI ↓	MOS ↑
YY [52]	2013	23.35^{8}	6.4174^{8}	5.9344^{7}	1367.71^{8}
TSG [45]	2013	23.55^{7}	6.4163^{7}	6.1433^{10}	1387.24^{7}
A+ [46]	2014	23.82^{6}	6.3645^{5}	5.9897^{9}	1354.52^{12}
SRCNN [12]	2014	23.93^{5}	6.5657^{10}	5.9781^{8}	1363.68^{11}
FSRCNN [13]	2016	24.07^{4}	6.9985^{11}	6.1649^{11}	1367.49^{9}
VDSR [24]	2016	24.13^{3}	7.4436^{12}	6.3319^{12}	1364.90^{10}
EDSR [29]	2017	$\mathbf{25.17}^{2}$	6.4560^{9}	5.3463^{6}	1447.44^{6}
SRGAN [28]	2017	22.57^{10}	3.9527^{3}	2.7656^{3}	1494.14^{3}
RCAN [62]	2018	$\mathbf{25.21}^{1}$	6.4121^{6}	5.2430^{5}	1455.31^{5}
BOE [32]	2018	22.68^{9}	$\mathbf{3.7945}^{1}$	$\mathbf{2.6368}^{2}$	1481.51^{4}
ESRGAN [47]	2018	22.51^{11}	4.7821^{4}	3.2198^{4}	$\mathbf{1534.25}^{1}$
RankSRGAN [61]	2019	22.11^{12}	$\mathbf{3.8155}^{2}$	$\mathbf{2.5636}^{1}$	$\mathbf{1518.29}^{2}$

Compare ESRGAN and RankSRGAN



Lower is better **Higher is better**

Benchmarking Super-Resolution Algorithms

- None of existing IQA methods is always effective in evaluation.
- Excessively optimizing performance on a specific IQA may cause a decrease in perceptual quality

Conclusion

- GAN-based Algorithms Pose challenges to IQA methods
- Existing IQA methods are inadequate in evaluating perceptual image restoration algorithms
- Deep network based IQA method show better performances.

Thank You

Project Page

X-Pixel Group

ECCV'20

ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU