Toward Real World Image Processing Applications: Challenges and Opportunities.

JINJIN GU
School of Electrical and Information Engineering,
The University of Sydney.
May. 2021

Research Outline

Real World Image Restoration

Apply advanced image processing technologies in real world applications.

Interpreting Image Generation

Understanding latent space of GANs and make use of GANs' powerful representation.

Applying Deep Signal Processing Technologies in Industrial

Empowering existing industrial facilities with advanced signal processing technologies.

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

Our product
Vivo X23 Smart Camera Solution

Application Goals

- Super-Resolution (Zoom in)
- Denoising
- Multi-Frame Processing

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

Our product
Vivo X23 Smart Camera Solution

Challenges

- Unknown Downsampling for SR
- Unknown Noise for Denoising
- Mixture Problem of Denoising and SR
- Unaligned Multiple Frames
- Time constraint (4K Res Image, Less than 300ms)
- No Powerful GPU

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

Our product
Vivo X23 Smart Camera Solution

Challenges

- **Unknown Downsampling for SR**
- Unknown Noise for Denoising
- Mixture Problem of Denoising and SR
- Unaligned Multiple Frames
- Time constraint (4K Res Image, Less than 300ms)
- No Powerful GPU

Background for Image Restoration

Super-Resolution

$$I_{LR} = (I_{HR} \odot k) \downarrow + n$$

Low-Res High-Res Blur kernel Noise

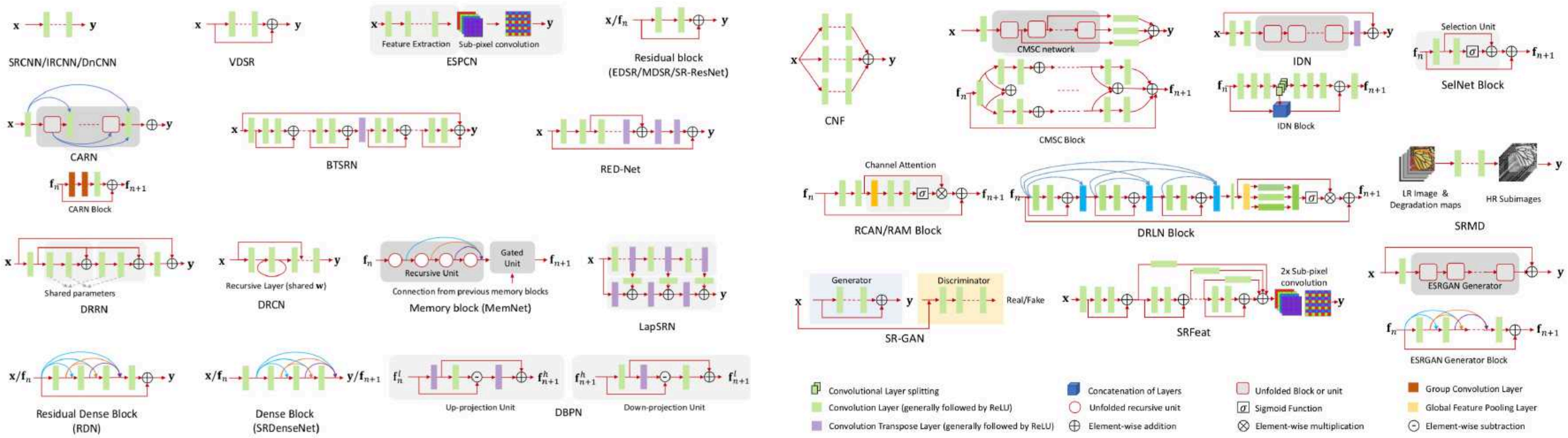
Denoising

$$I_{Noisy} = I_{Clean} + n$$

Noisy Image Clean Image Noise

Image Restoration: "Undo" these degradations

Deep Networks for Image Restoration



[Anwar, S., Khan, S., & Barnes, N. (2019). A Deep Journey into Super-resolution: A survey. *arXiv preprint arXiv:1904.07523*.]

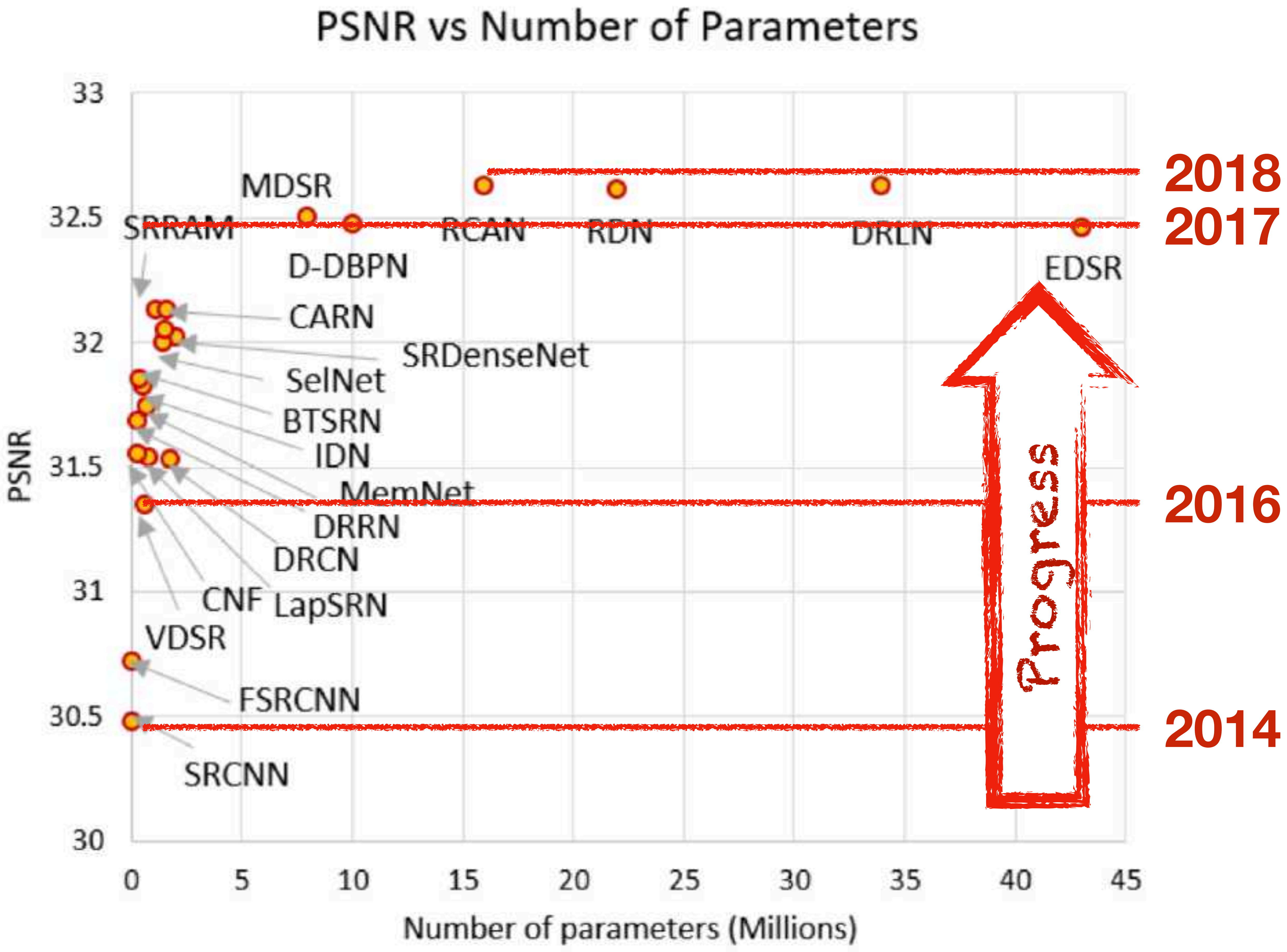
Deep Networks for Image Restoration

Great Progress Made by Deep Networks!

2014 state-of-the-art: 30.5db

2018 state-of-the-art: 32.5db

2db of progress in 4 years!



[Anwar, S., Khan, S., & Barnes, N. (2019). A Deep Journey into Super-resolution: A survey. arXiv preprint arXiv:1904.07523.]

Current Image Restoration Research

Great Progress (Under Controlled Problem Settings)

Super-Resolution

$$I_{LR} = (I_{HR} \odot k) \downarrow + n$$

Low-Res High-Res Blur kernel Noise

Fixed blur kernel!

No noise term!

Denoising

$$I_{Noisy} = I_{Clean} + n$$

Noisy Image Clean Image Noise

Focus on Gaussian Noise!

Current Image Restoration Research

Great Progress (Under Controlled Problem Settings)

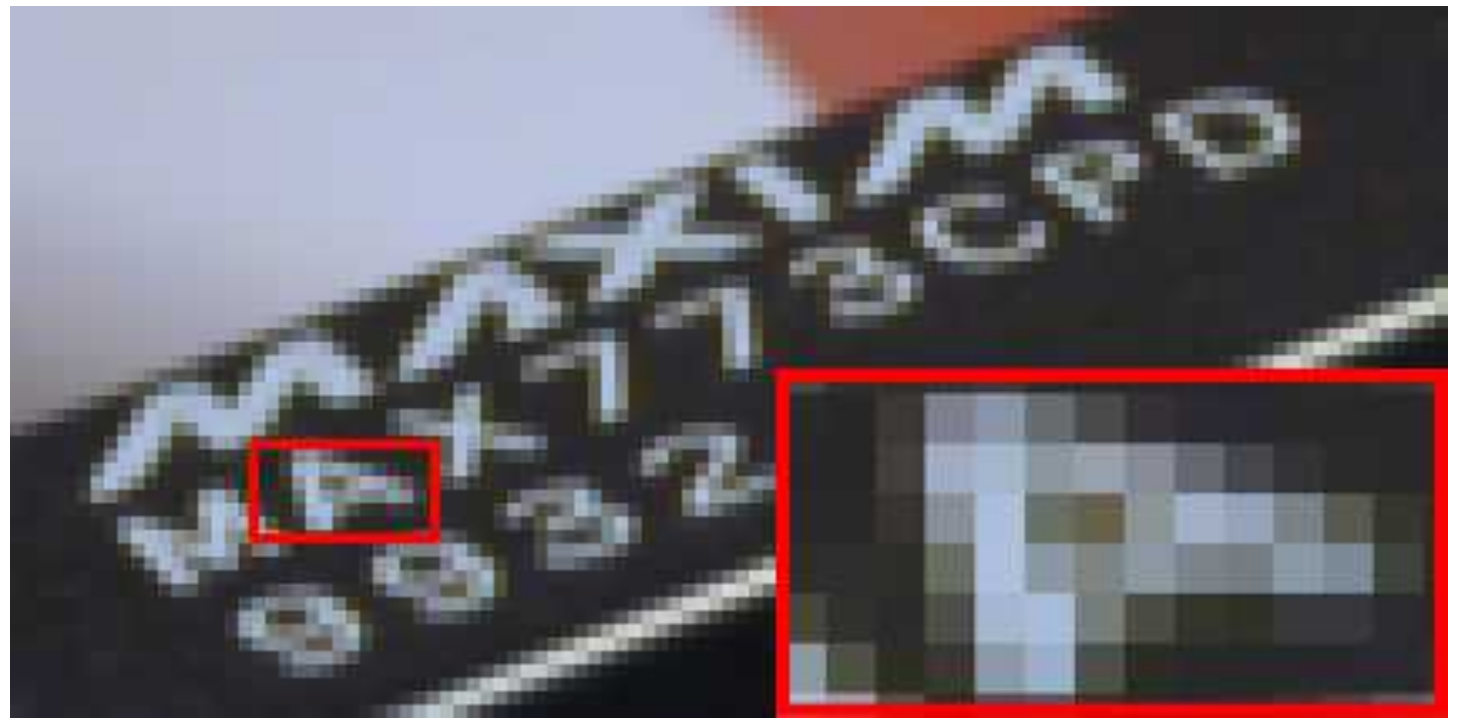
Super-Resolution

$$I_{LR} = (I_{HR} \odot k) \downarrow + n$$

Low-Res High-Res Blur kernel Noise

Fixed blur kernel!

No noise term!



Real Low-Res Image

EDSR (STOA 2017)

Blurry Effect?

Blind Super-Resolution with Iterative Kernel Correction

Jinjin Gu, Hannan Lu, Wangmeng Zuo, Chao Dong, CVPR 2019

Super-Resolution

$$I_{LR} = (I_{HR} \odot k) \downarrow + n$$

Low-Res High-Res Blur kernel Noise

The Blur Kernel is Unknown in Blind SR and Real World Applications

Low-Res Input

Non-Blind SR

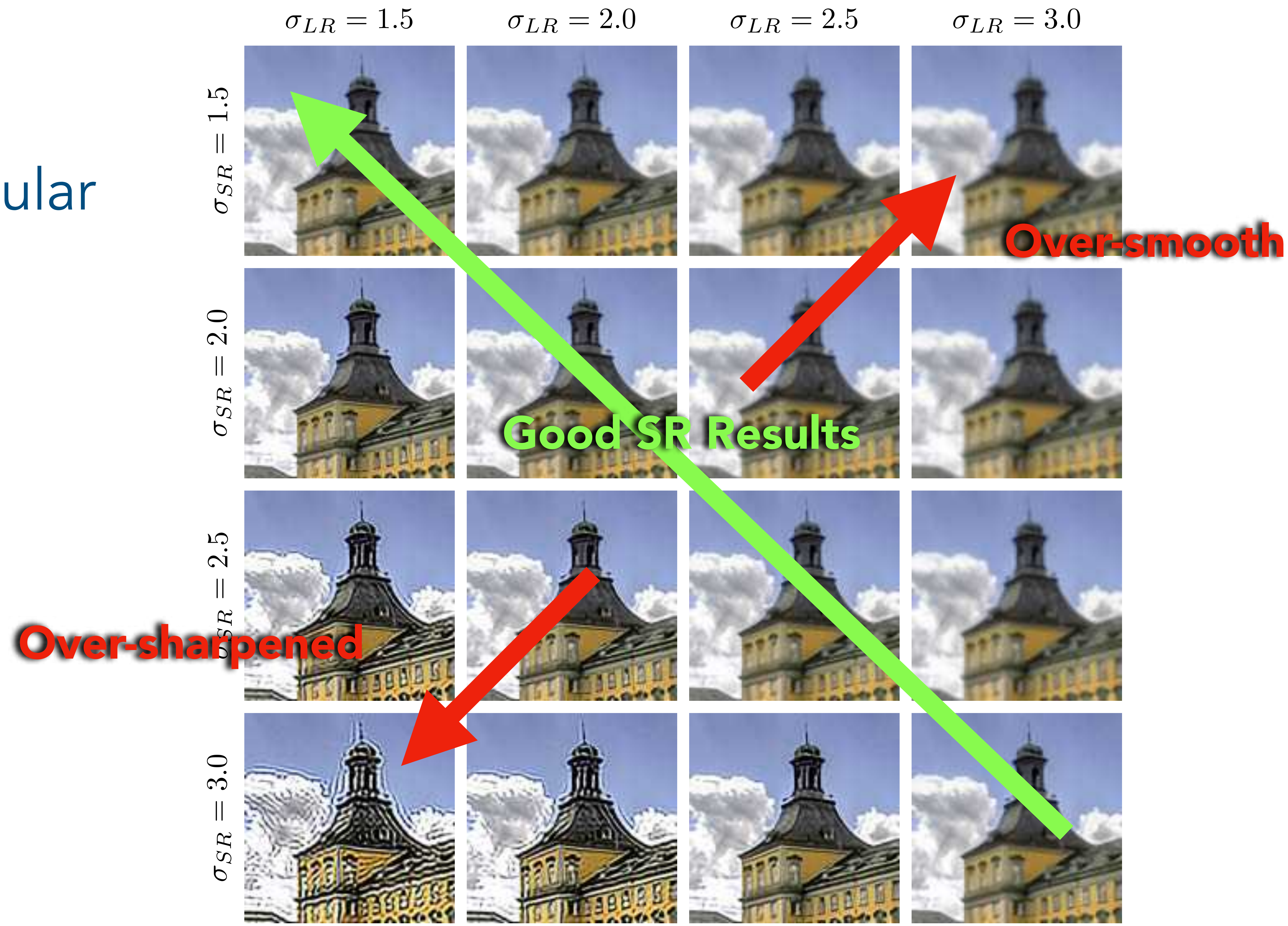
Ours

LONG BEACH
CALIFORNIA
June 16-20, 2019

What if the kernel is incorrect during SR?

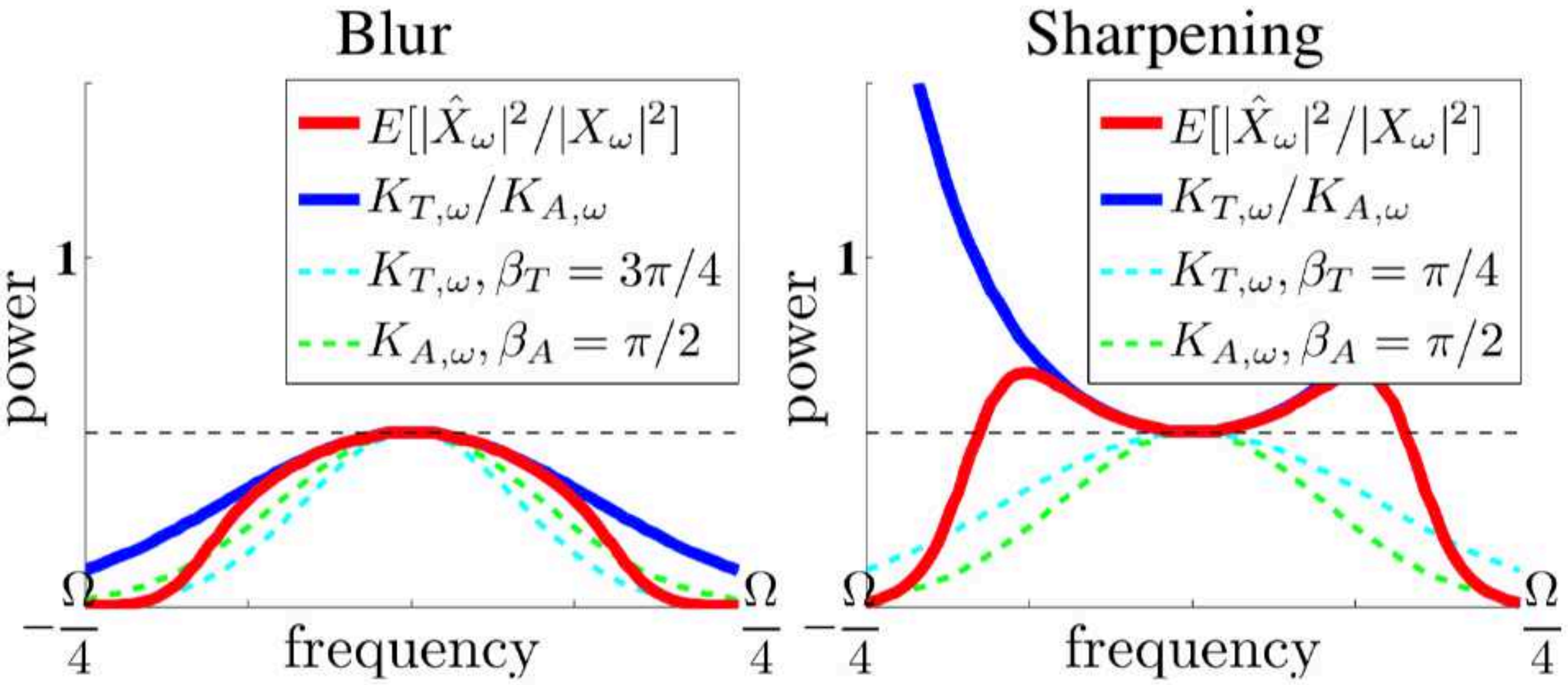
Kernel mismatch brings regular artifacts:

- Over-smooth
- Over-sharpened



Why Regular Artifacts?

Theoretical Analysis Using MAP and Simple Gradient Prior

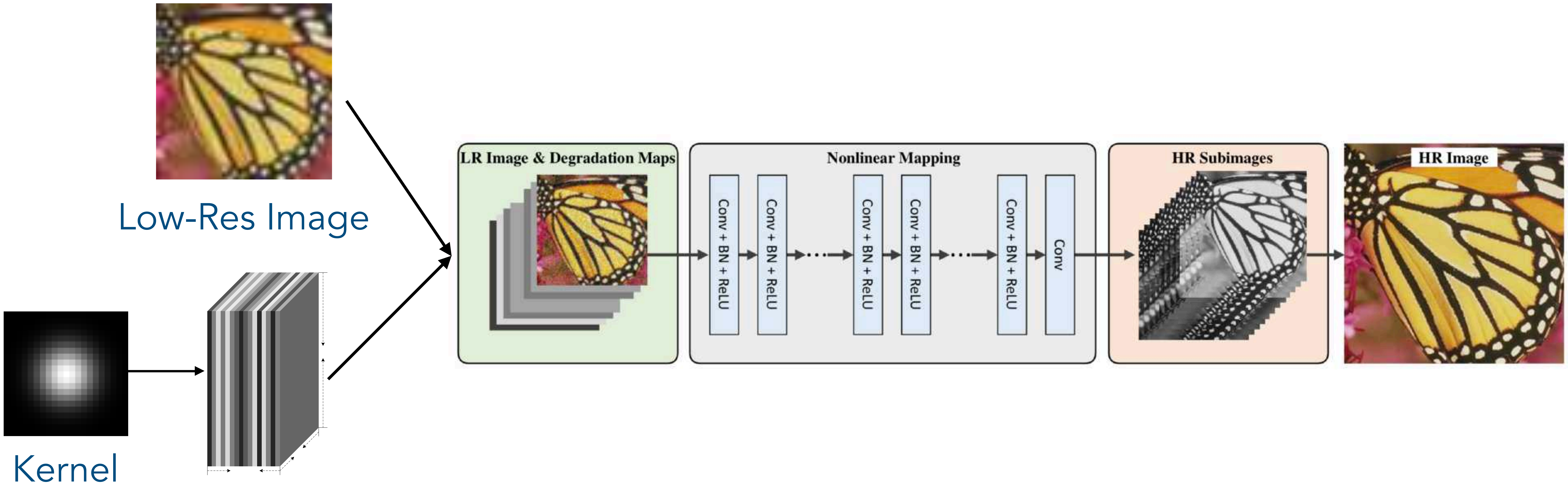


On the Left: yields a Gaussian blur filter; On the Right: gives a sharpening filter.

[Efrat, N., Glasner, D., Apartsin, A., Nadler, B., & Levin, A. Accurate blur models vs. image priors in single image super-resolution. ICCV 2013]

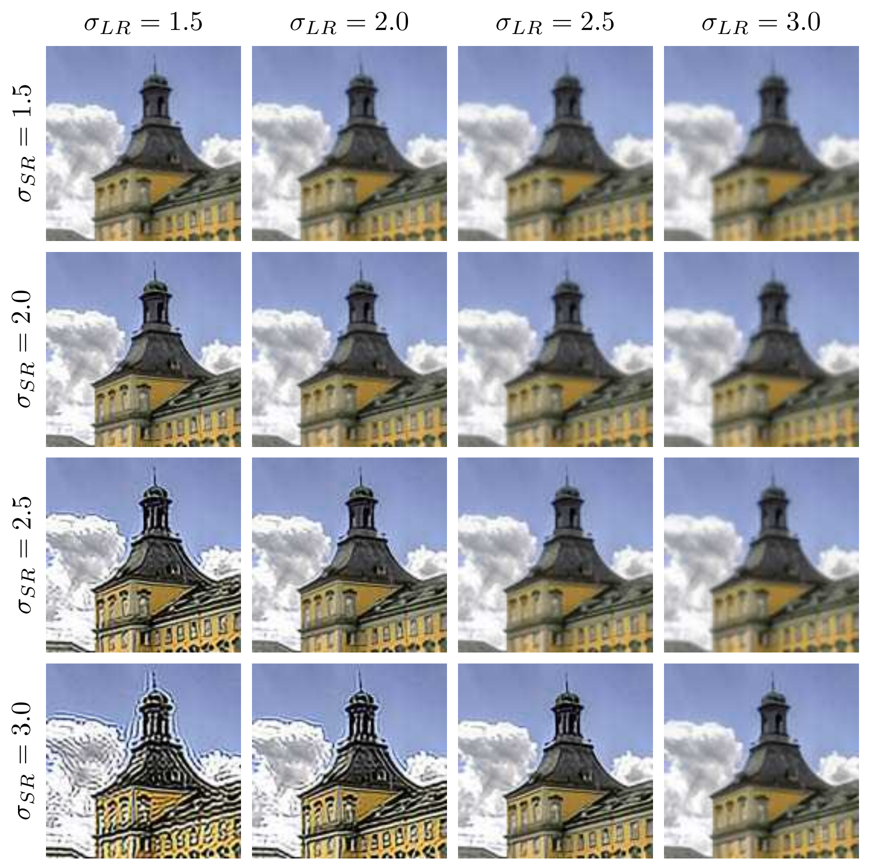
Learning Single Network for Multiple Kernels

Super-resolve images according to given blur kernel (SRMD)

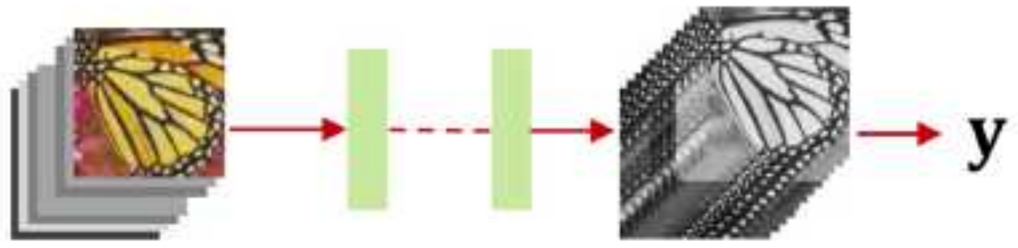


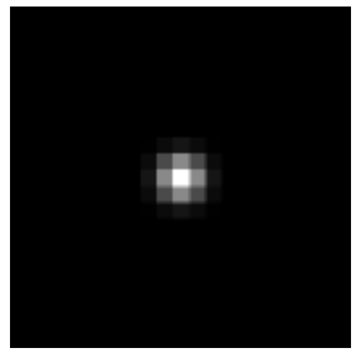
[Zhang, K., Zuo, W., & Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. CVPR 2018.]

Try Kernels in Real Applications



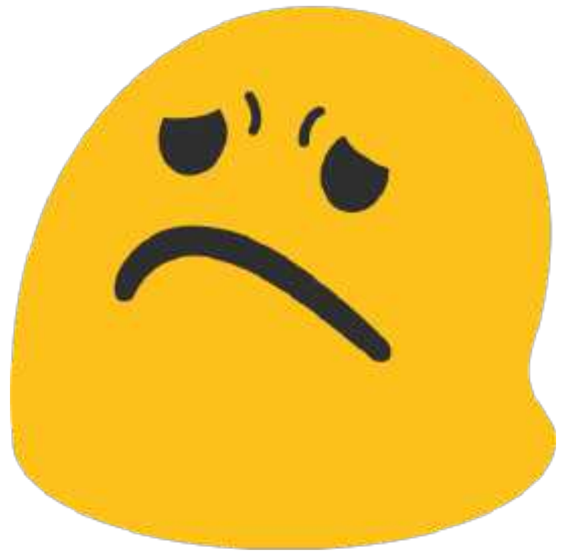
Low-Res Image





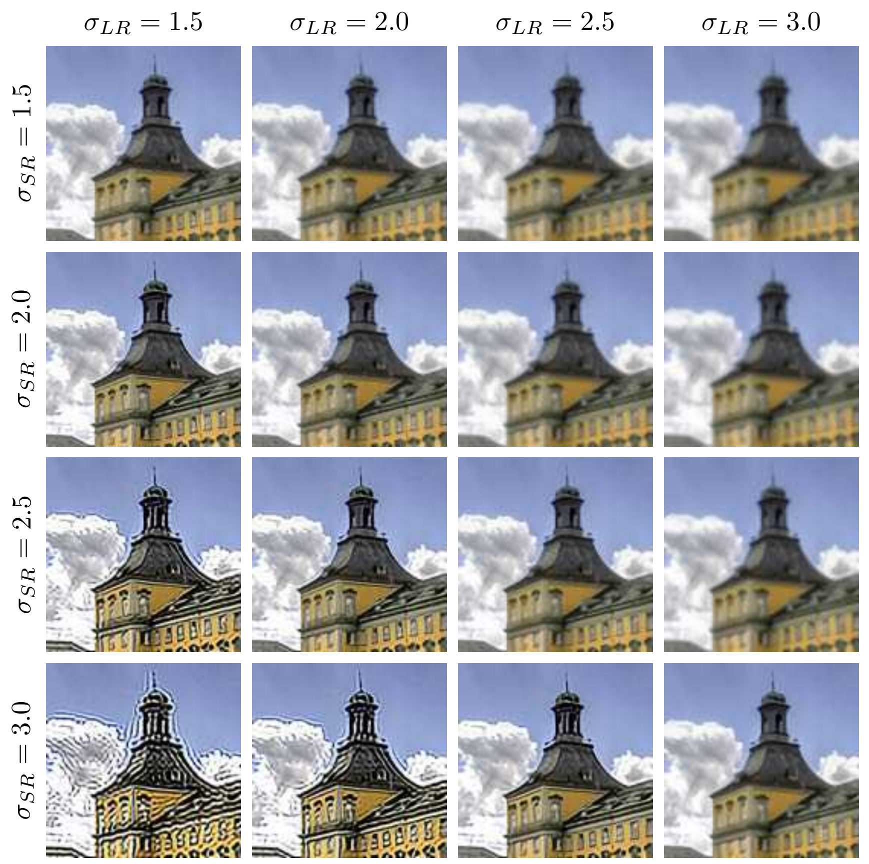
Kernel

SR Image

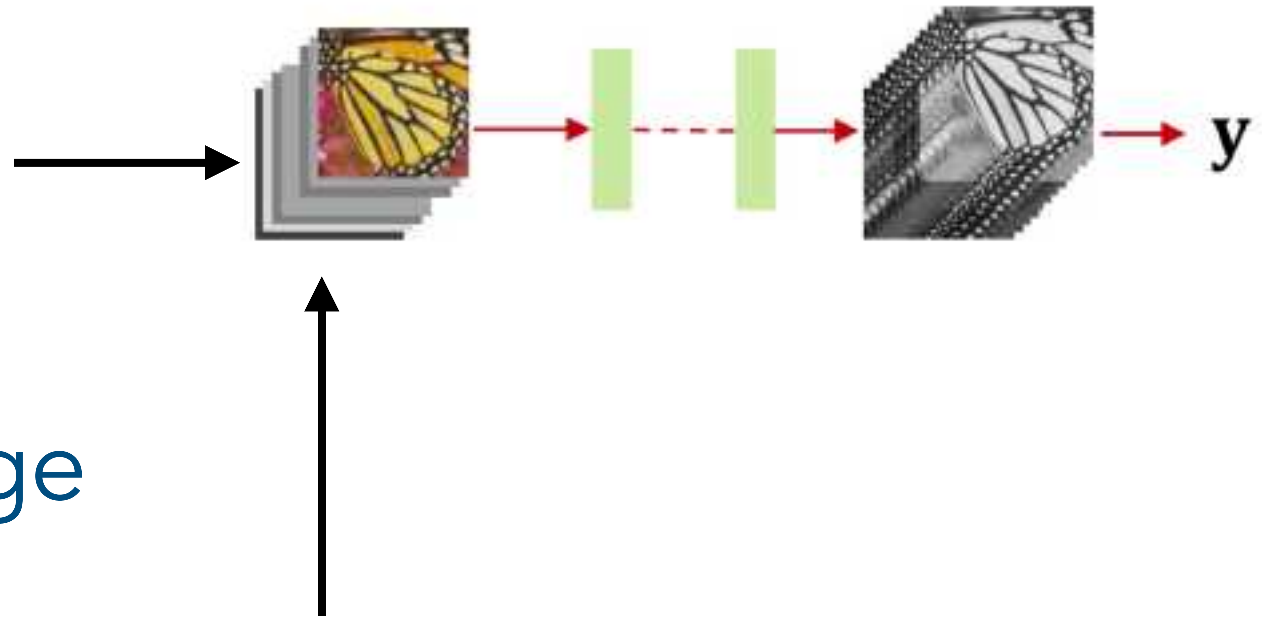


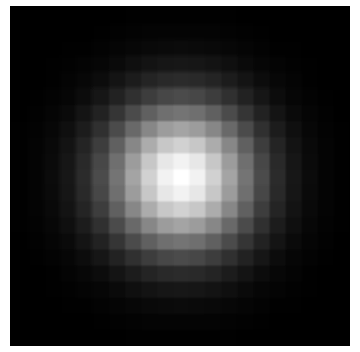
User: Over-smooth.
Try a Wider Kernel

Try Kernels in Real Applications



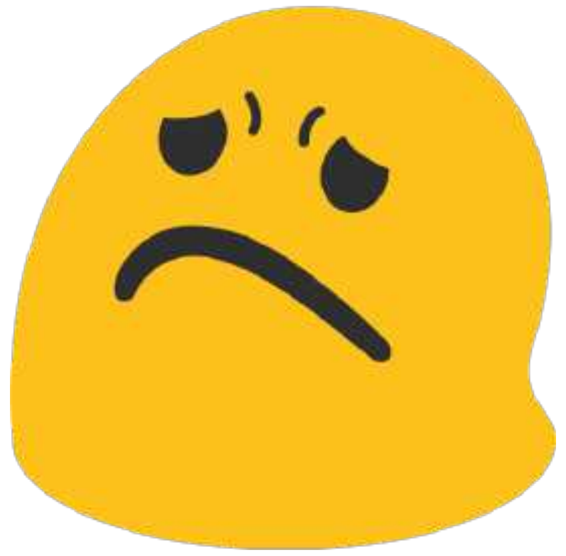
Low-Res Image





Kernel

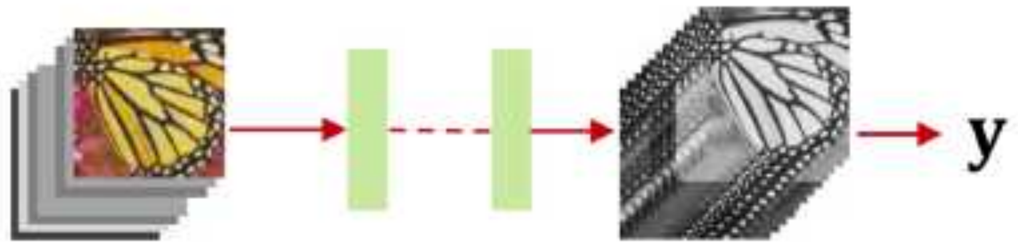
SR Image

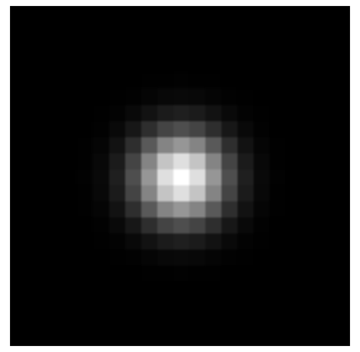


User: Over-sharpened Artifacts.
Try a Sharper kernel

Try Kernels in Real Applications

Low-Res Image





Kernel

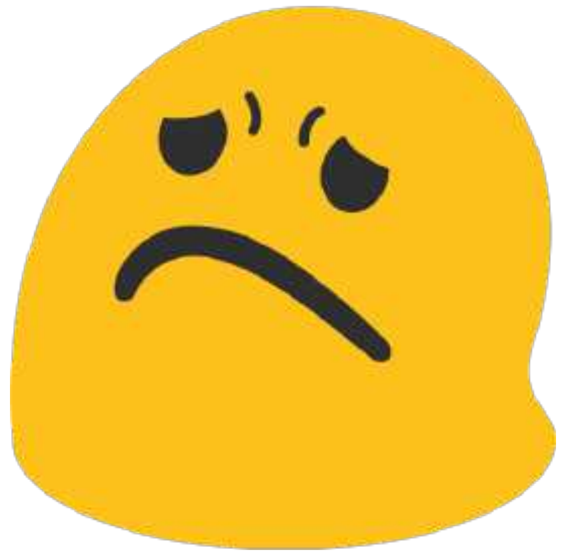
SR Image

User: Good One, Use this Kernel!

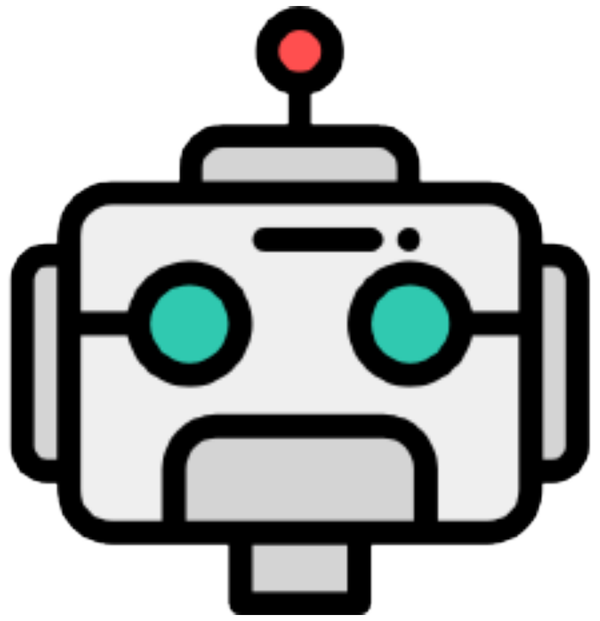
Estimate Kernels Automatically

In real applications, how to try kernels for thousands of images?

Can we train a Network to do this for us?

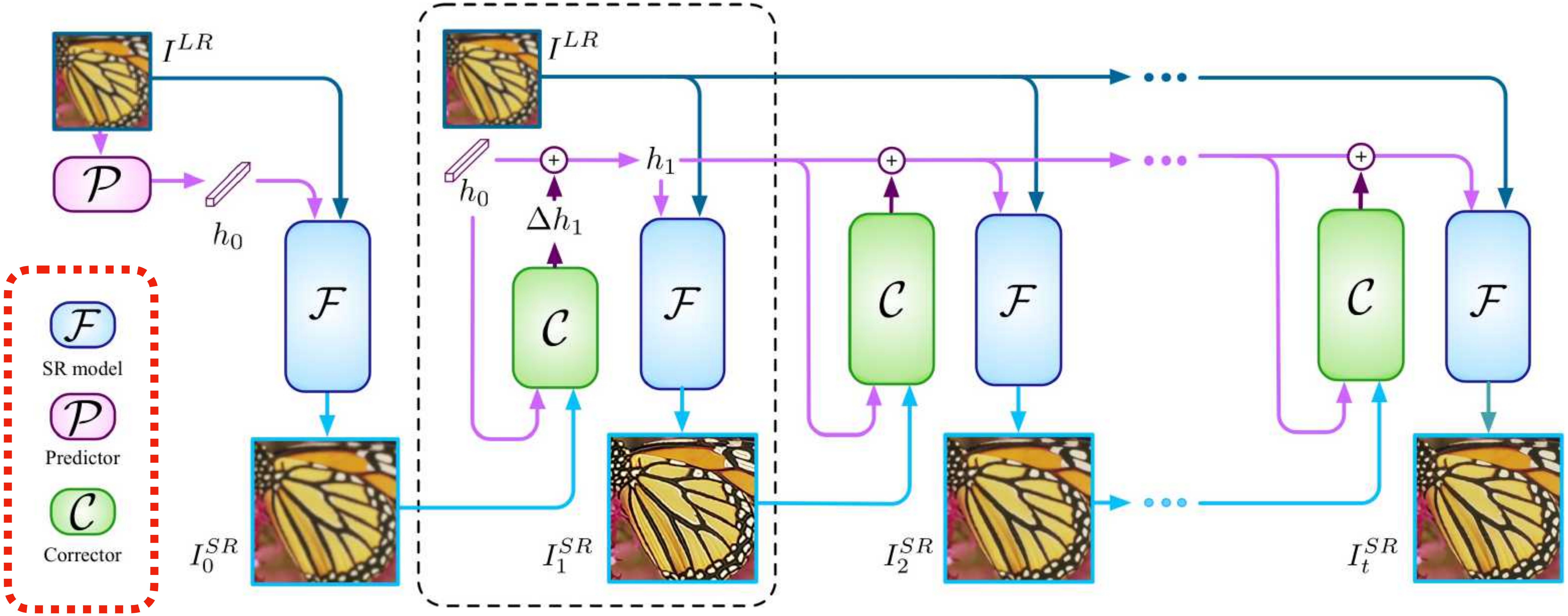


User: Search Kernels for every images?



Train an Algorithm to do it.

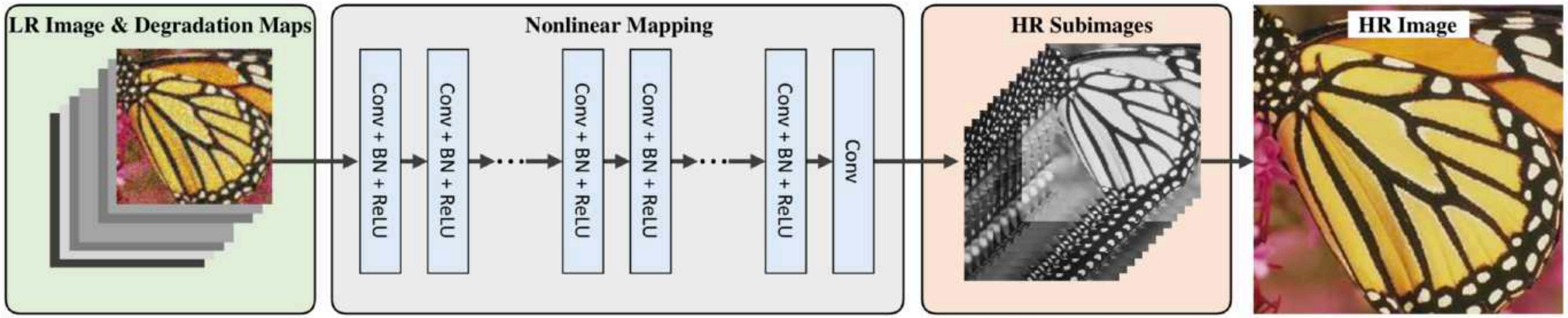
Iterative Kernel Correction Framework



Three components

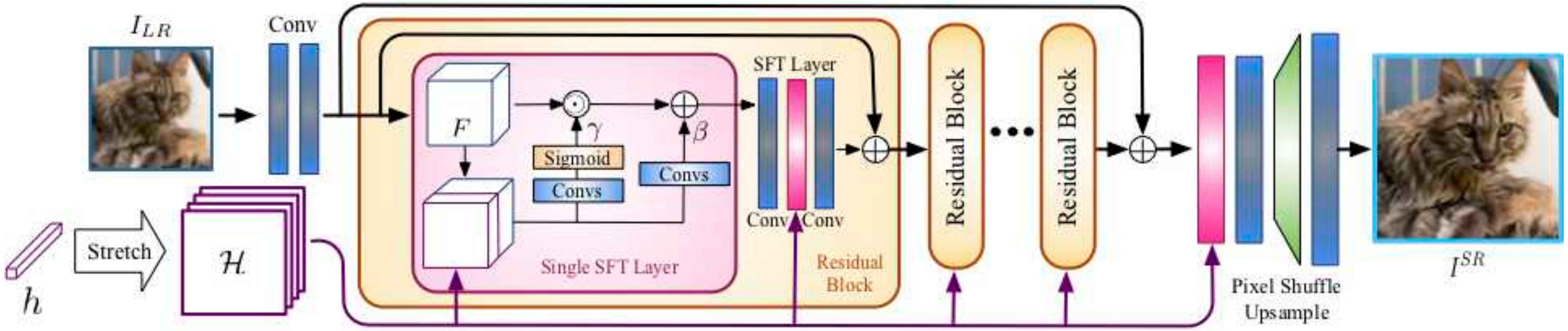
New SR Model for Multiple Degradations

SRMD, CVPR 2018

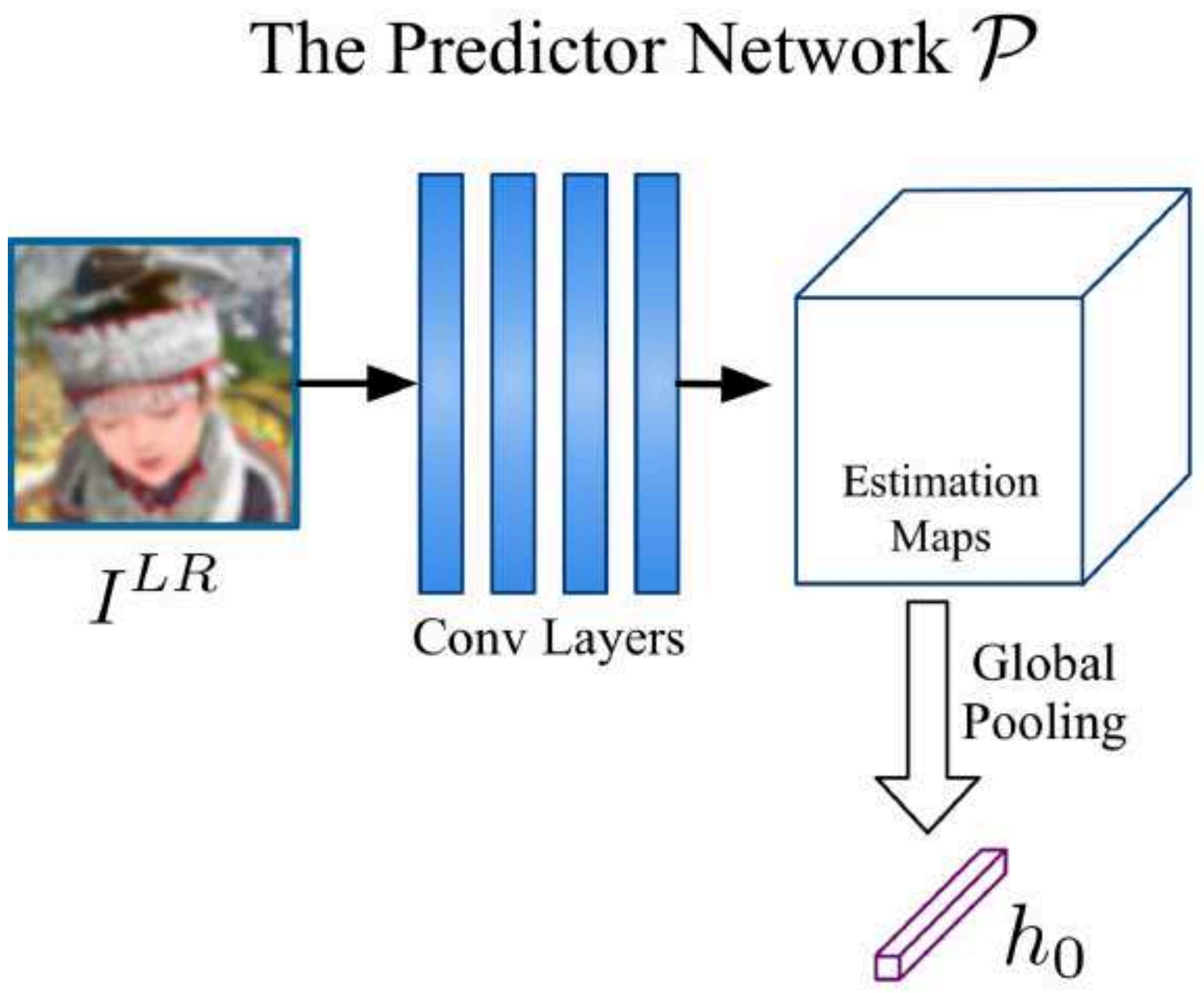


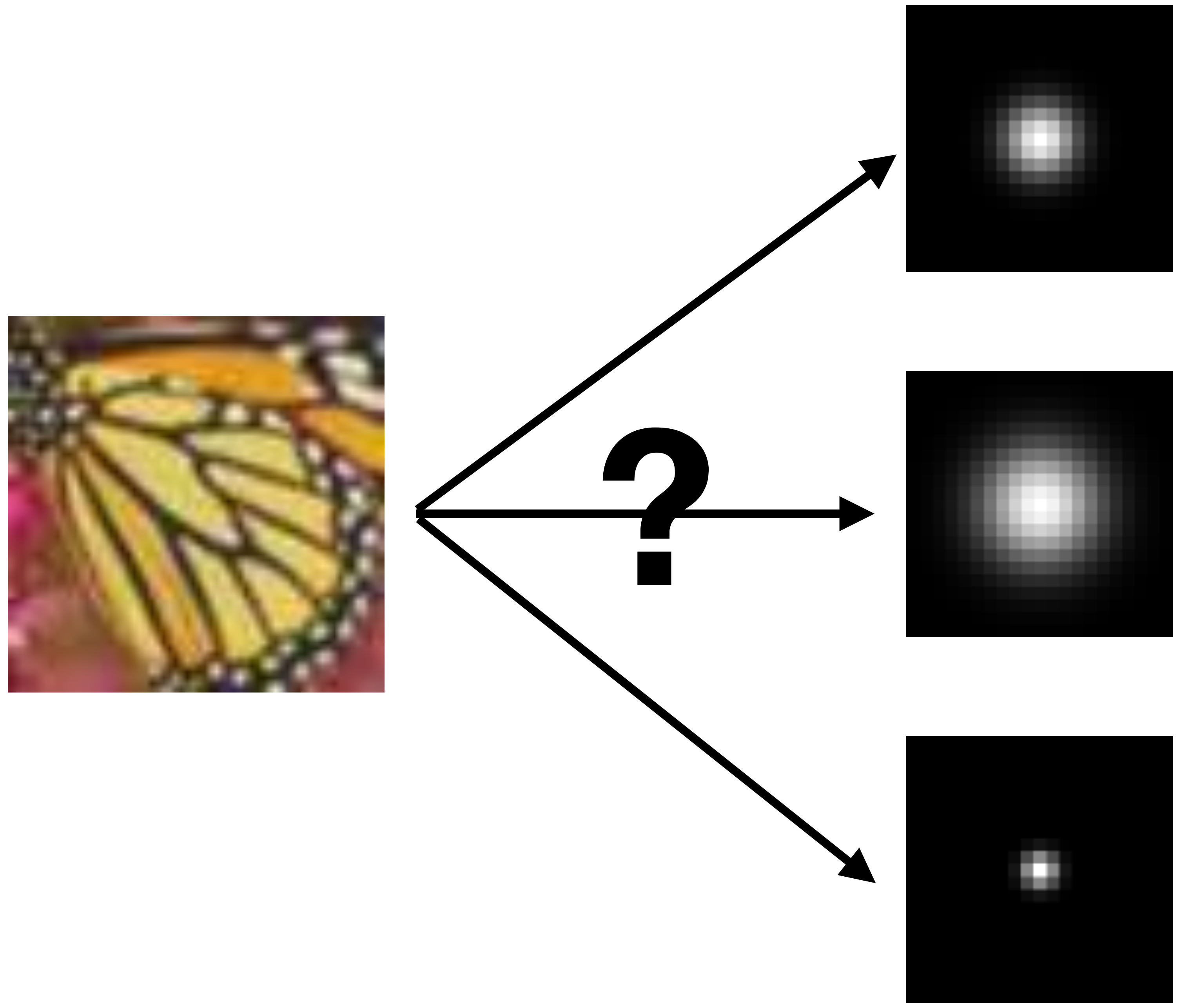
[Zhang, K., Zuo, W., & Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. CVPR 2018.]

Our SFTMD, CVPR 2019



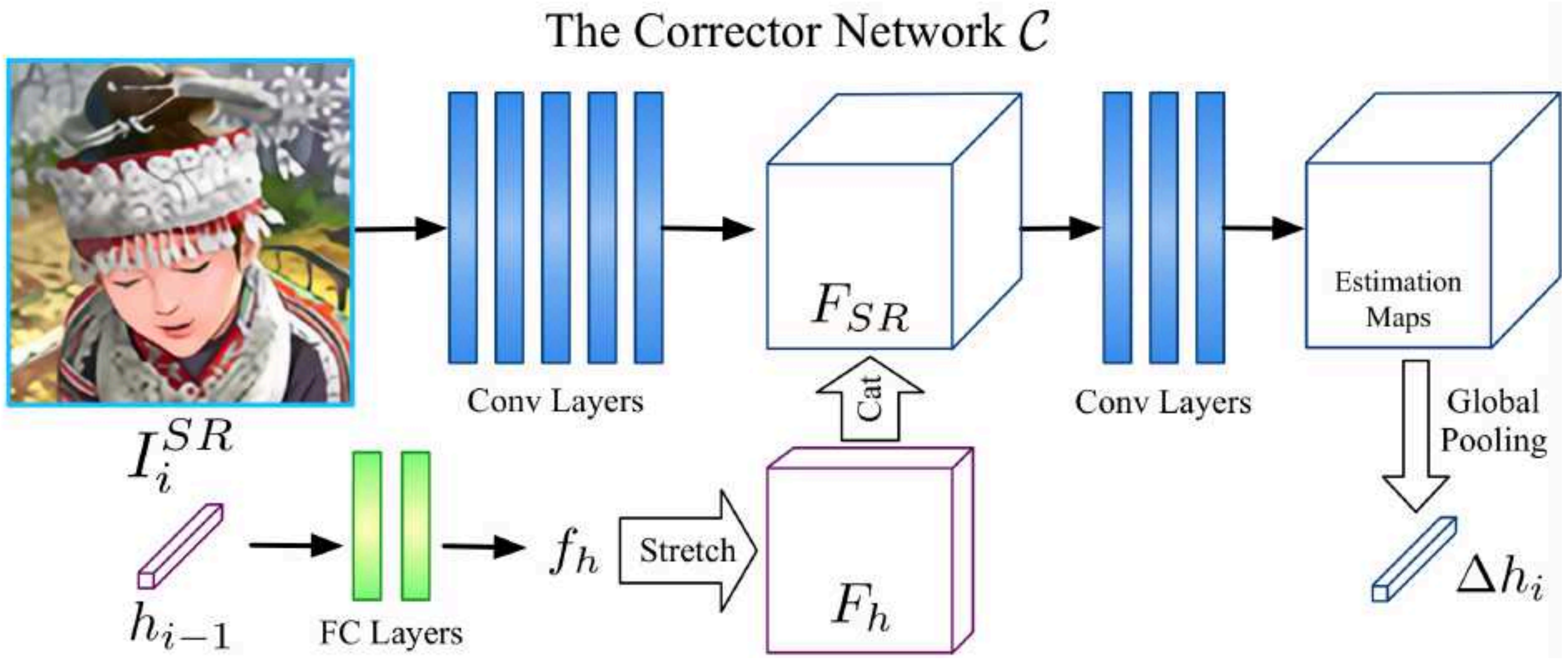
Predictor in Iterative Correction



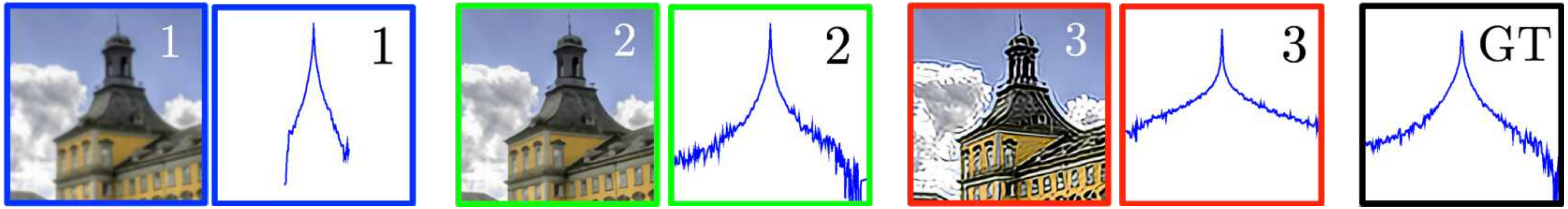


ill-posed Problem!

Corrector in Iterative Correction

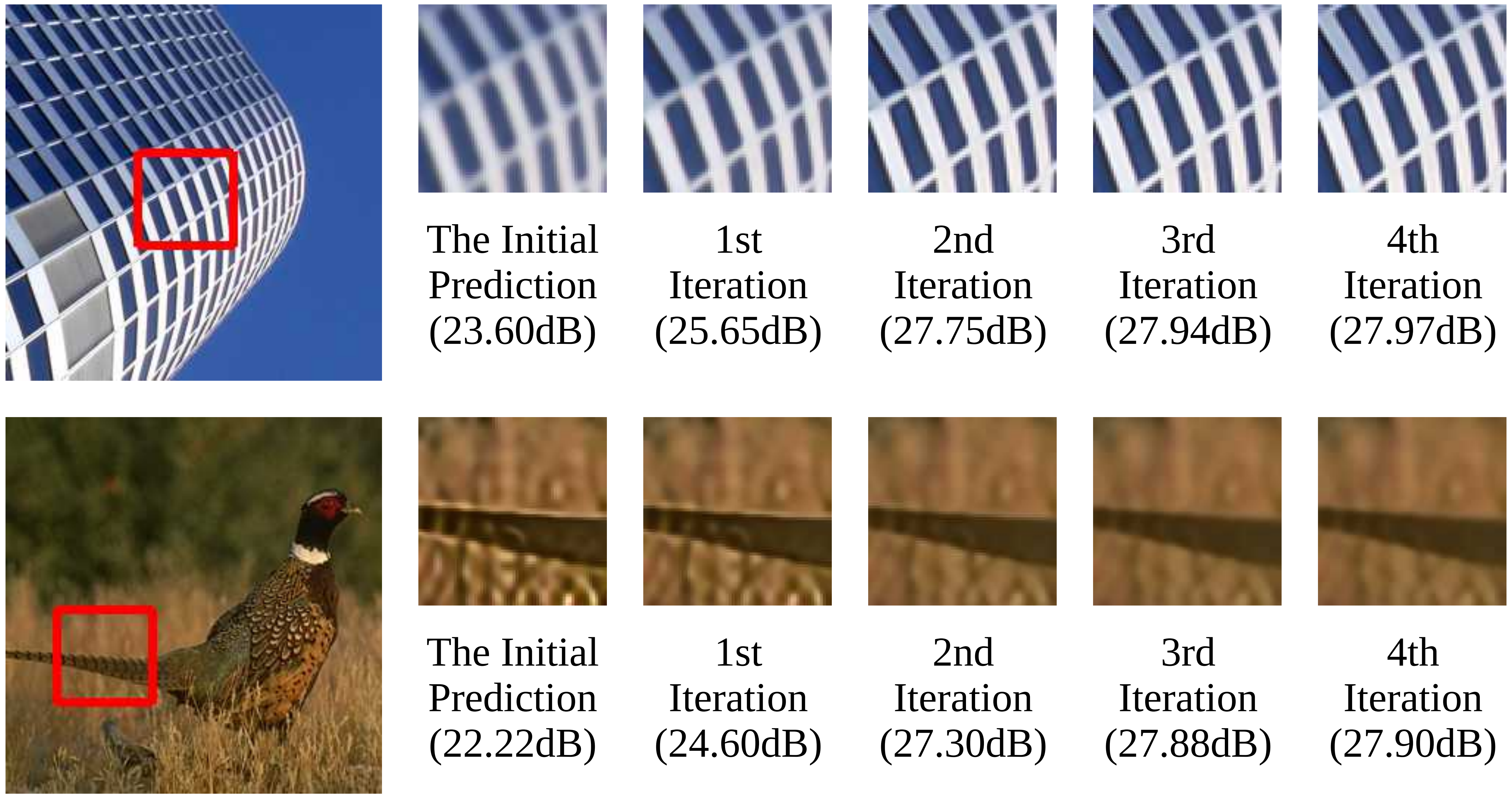


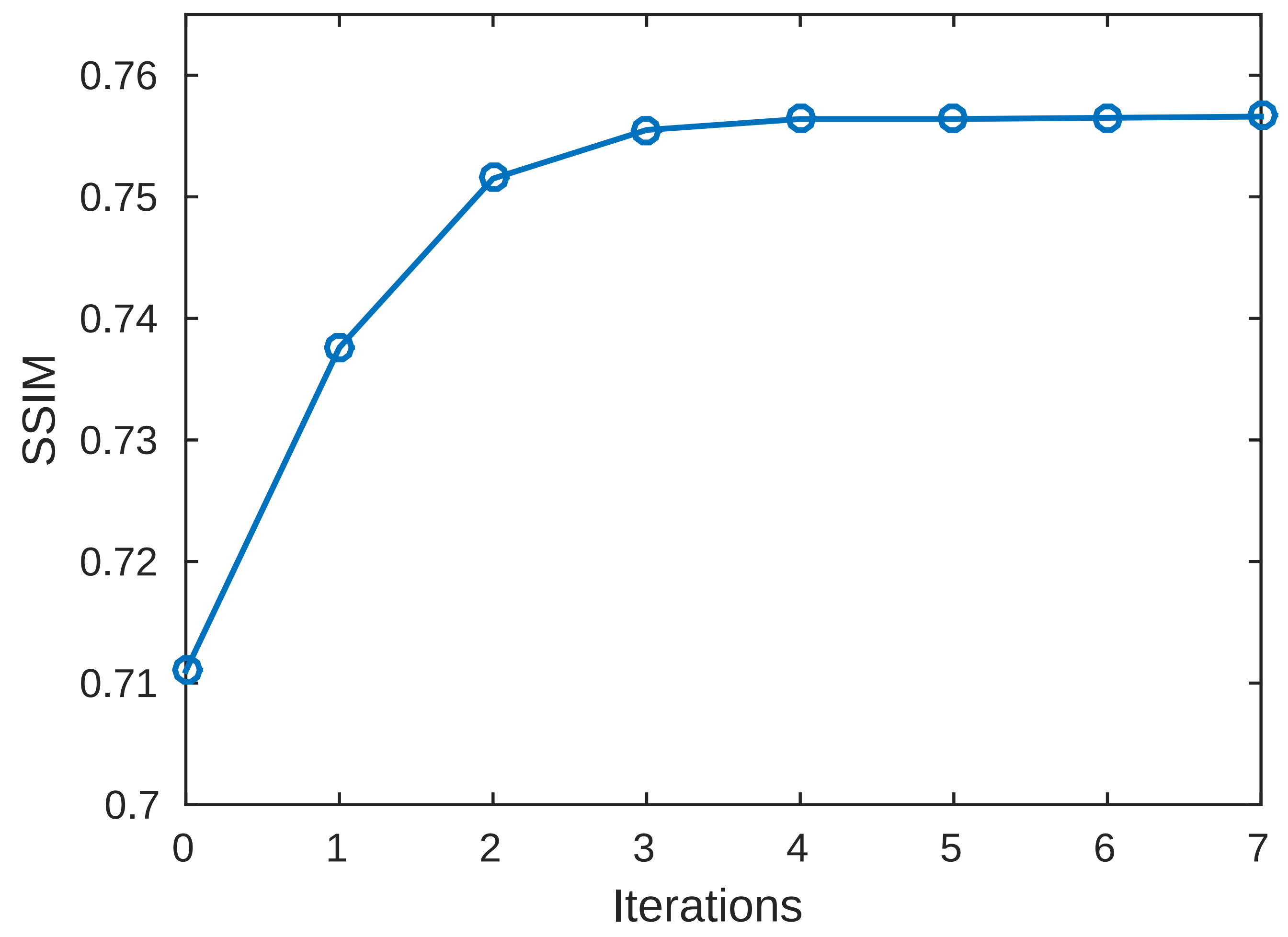
Corrector learns the prior of SR results with correct kernel.

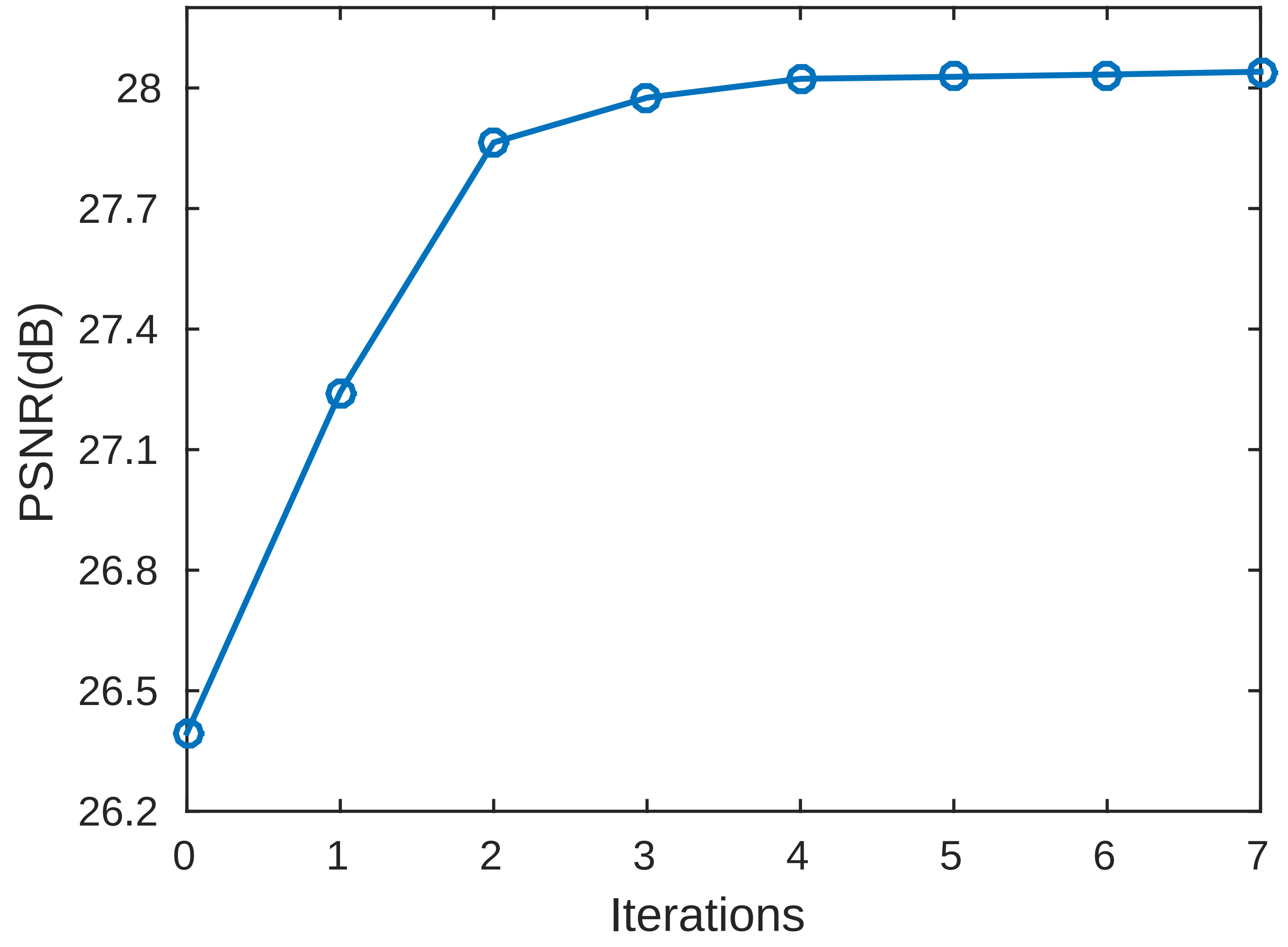


Different images and their gradient distributions.

Blind SR results for each step







Qualitative and quantitative results for different correction steps

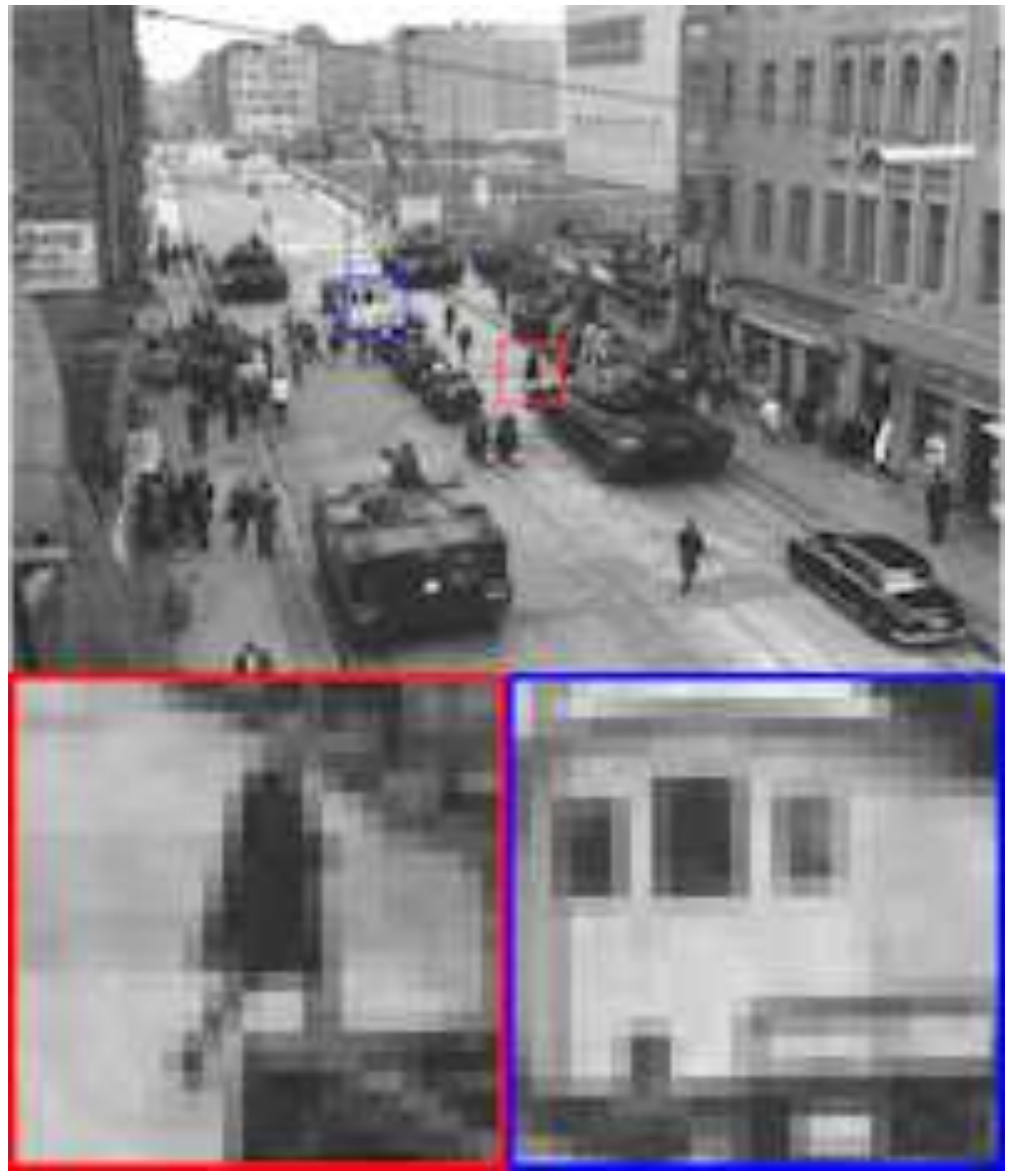
Comparison Blind SR Methods

Synthetic Data $I_{LR} = (I_{HR} \odot k) \downarrow + n$

A+ EDSR ZSSR IKC *w/t* Correction Ours

Comparison Blind SR Methods

Real Image



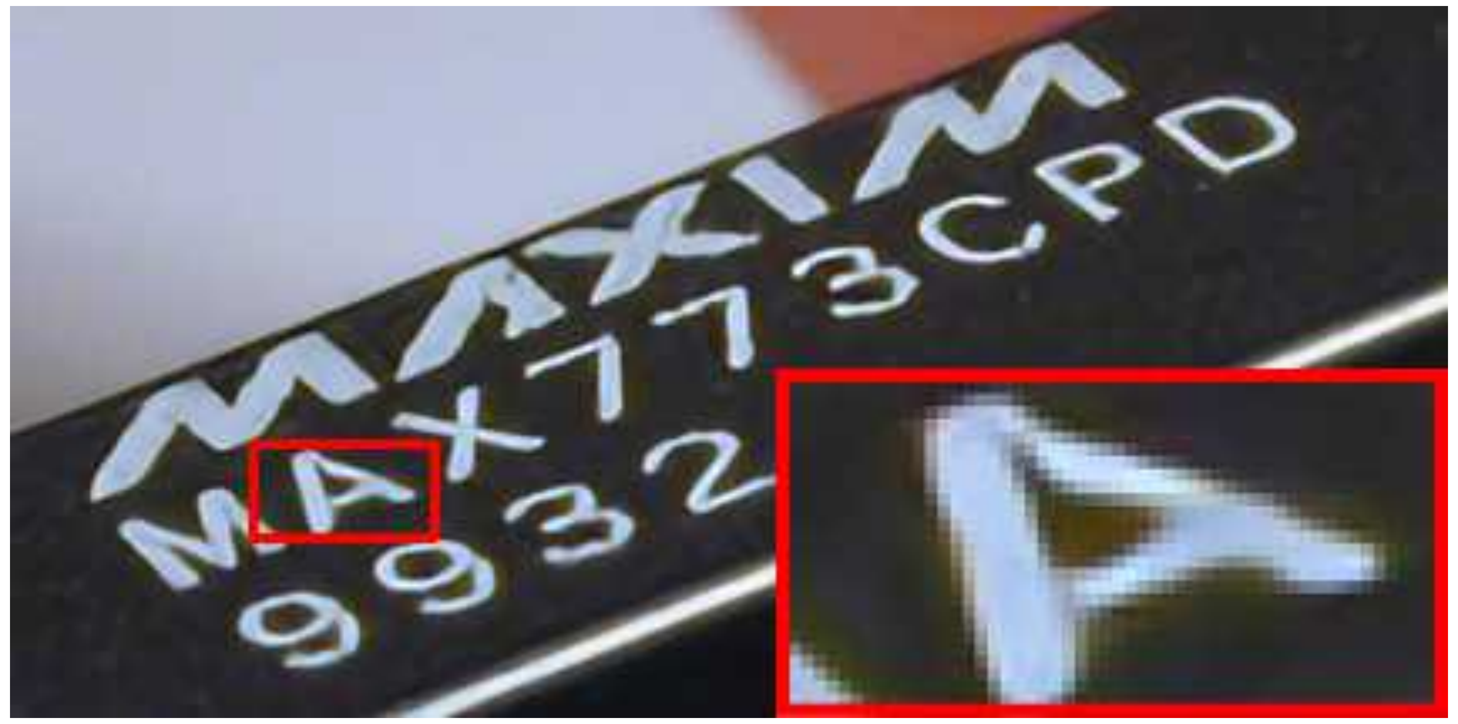
Low-Res Image A+ ZSSR Ours

Comparison IKC with Hand-craft Kernel

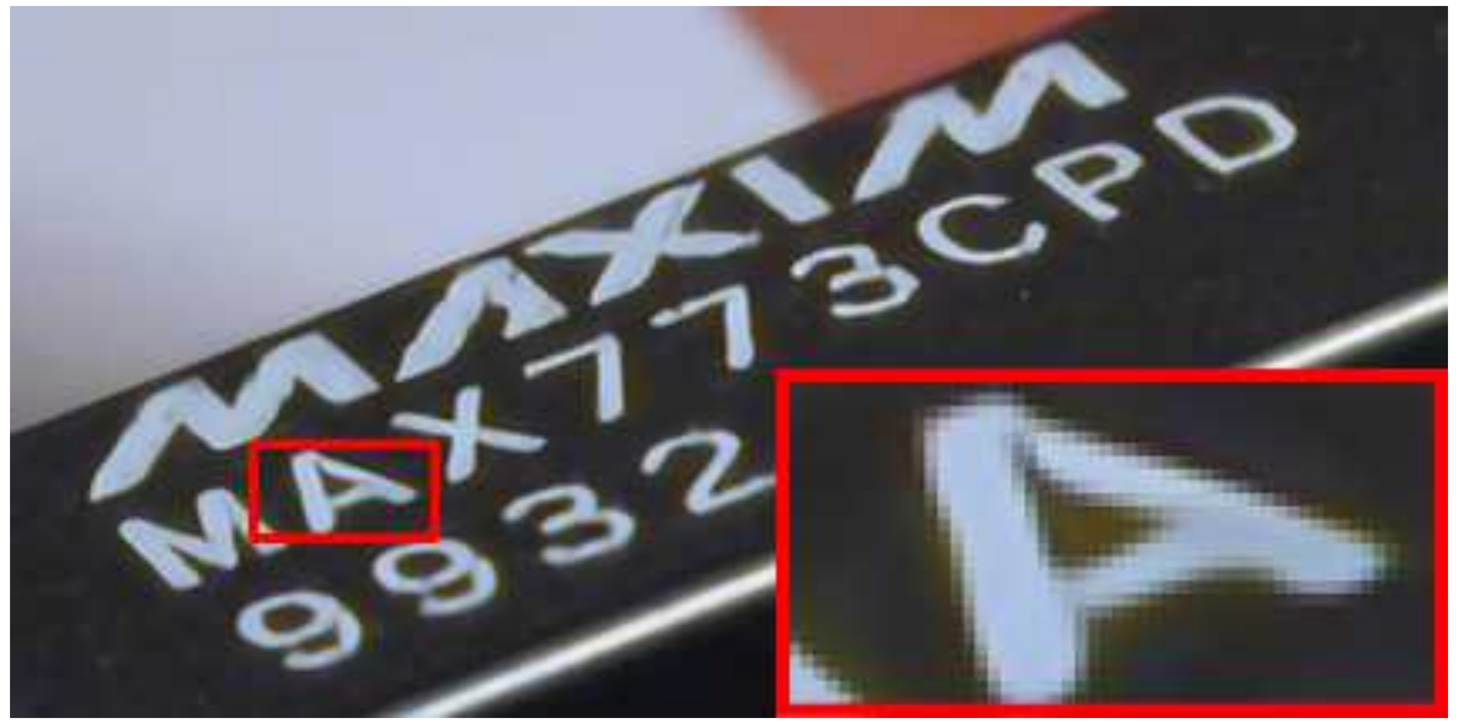
Real Image

Low-Res Image

EDSR



Hand-craft + SRMD



Ours

Blind Super-Resolution with Iterative Kernel Correction

Jinjin Gu, Hannan Lu, Wangmeng Zuo, Chao Dong, CVPR 2019

Super-Resolution

$$I_{LR} = (I_{HR} \odot k) \downarrow + n$$

Low-Res High-Res Blur kernel Noise

The Blur Kernel is Unknown in Blind SR and Real World Applications

Low-Res Input Non-Blind SR Ours

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

Our product
Vivo X23 Smart Camera Solution

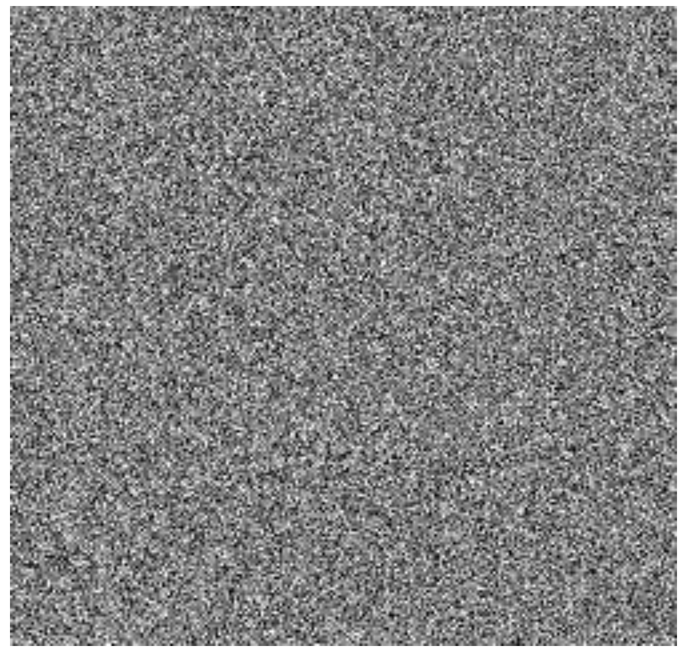
Challenges

- Unknown Downsampling for SR
- **Unknown Noise for Denoising**
- Mixture Problem of Denoising and SR
- Unaligned Multiple Frames
- Time constraint (4K Res Image, Less than 300ms)
- No Powerful GPU

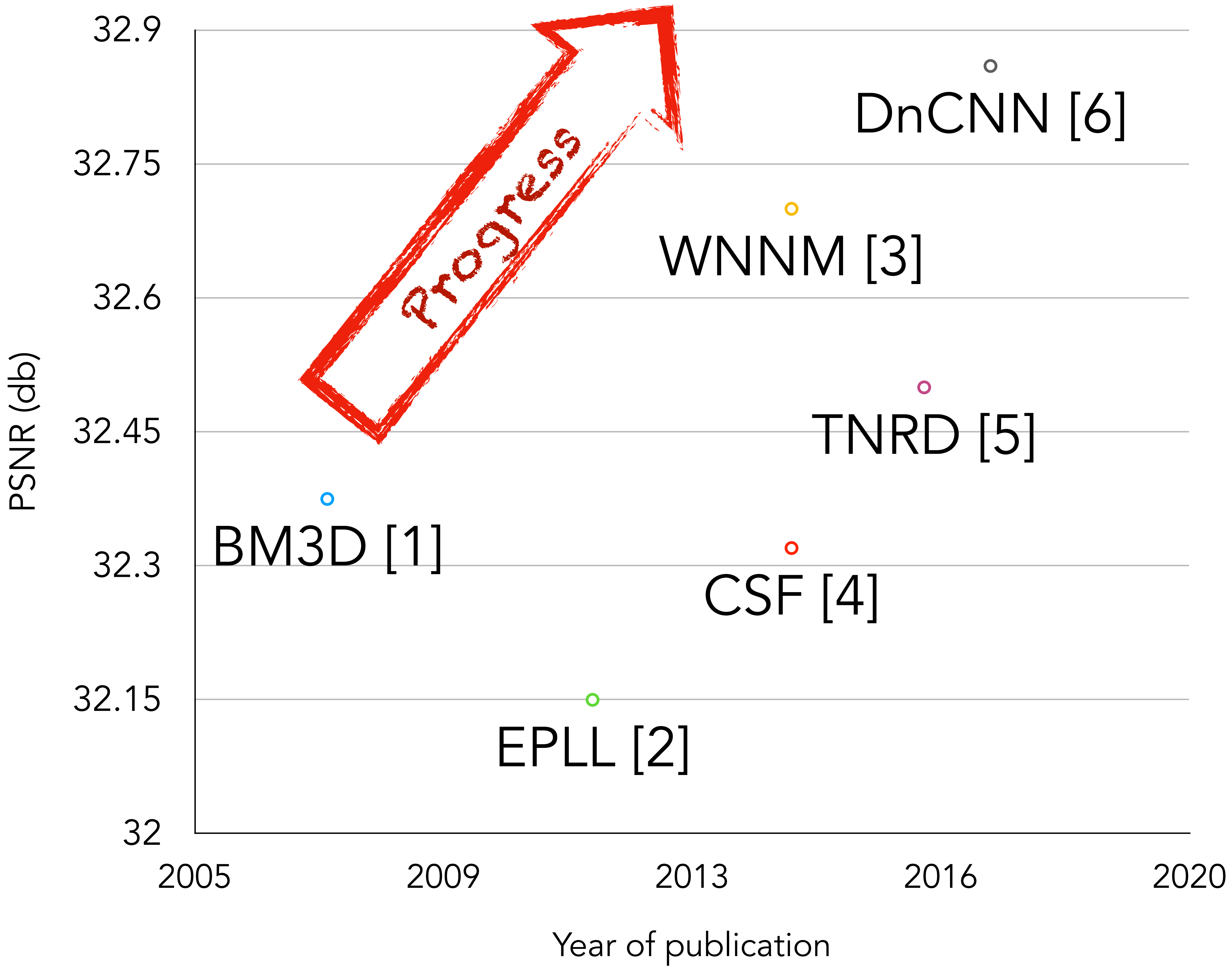
Denoising Progress: Synthetic Data (Gaussian)

Clean image

+



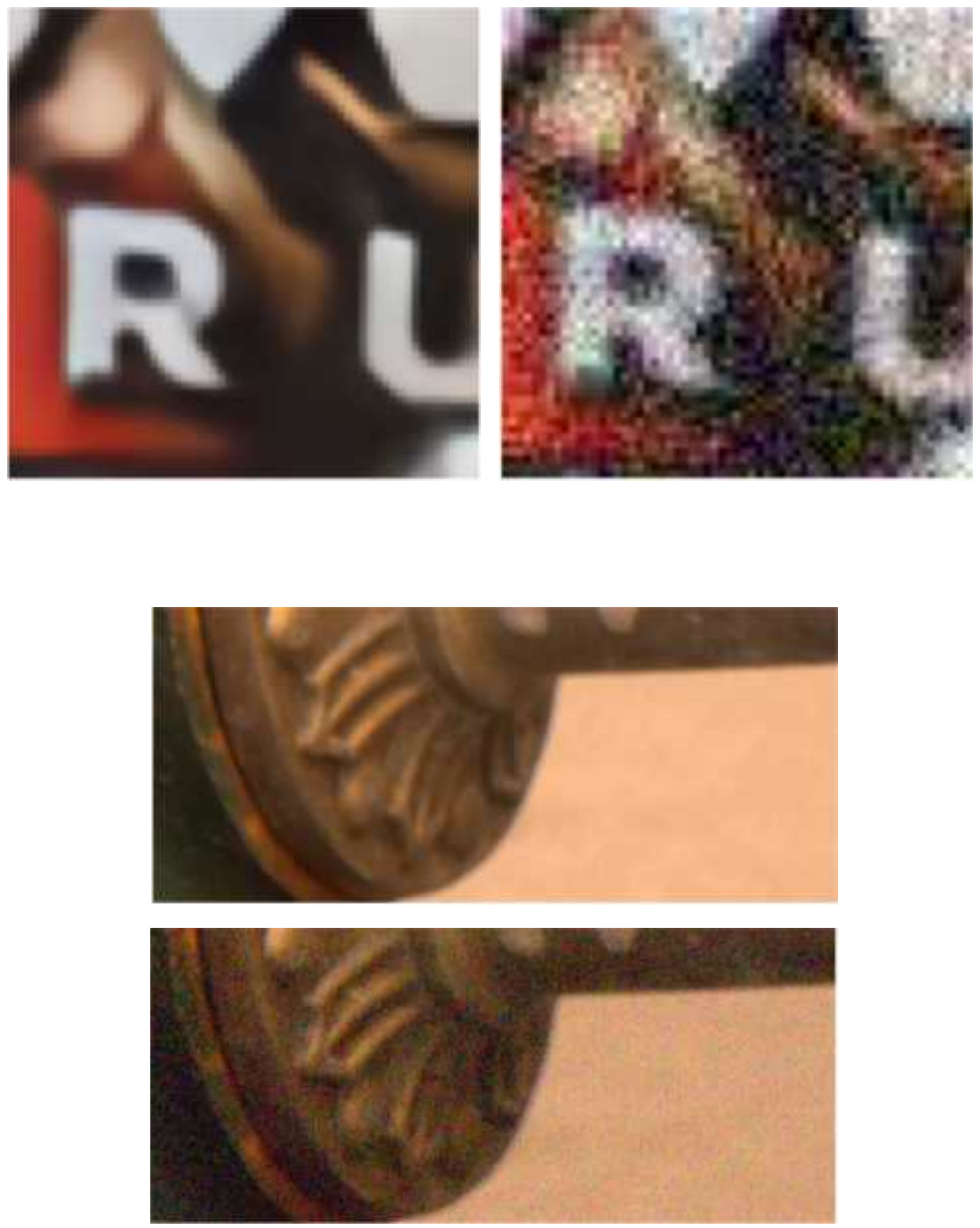
Gaussian white noise

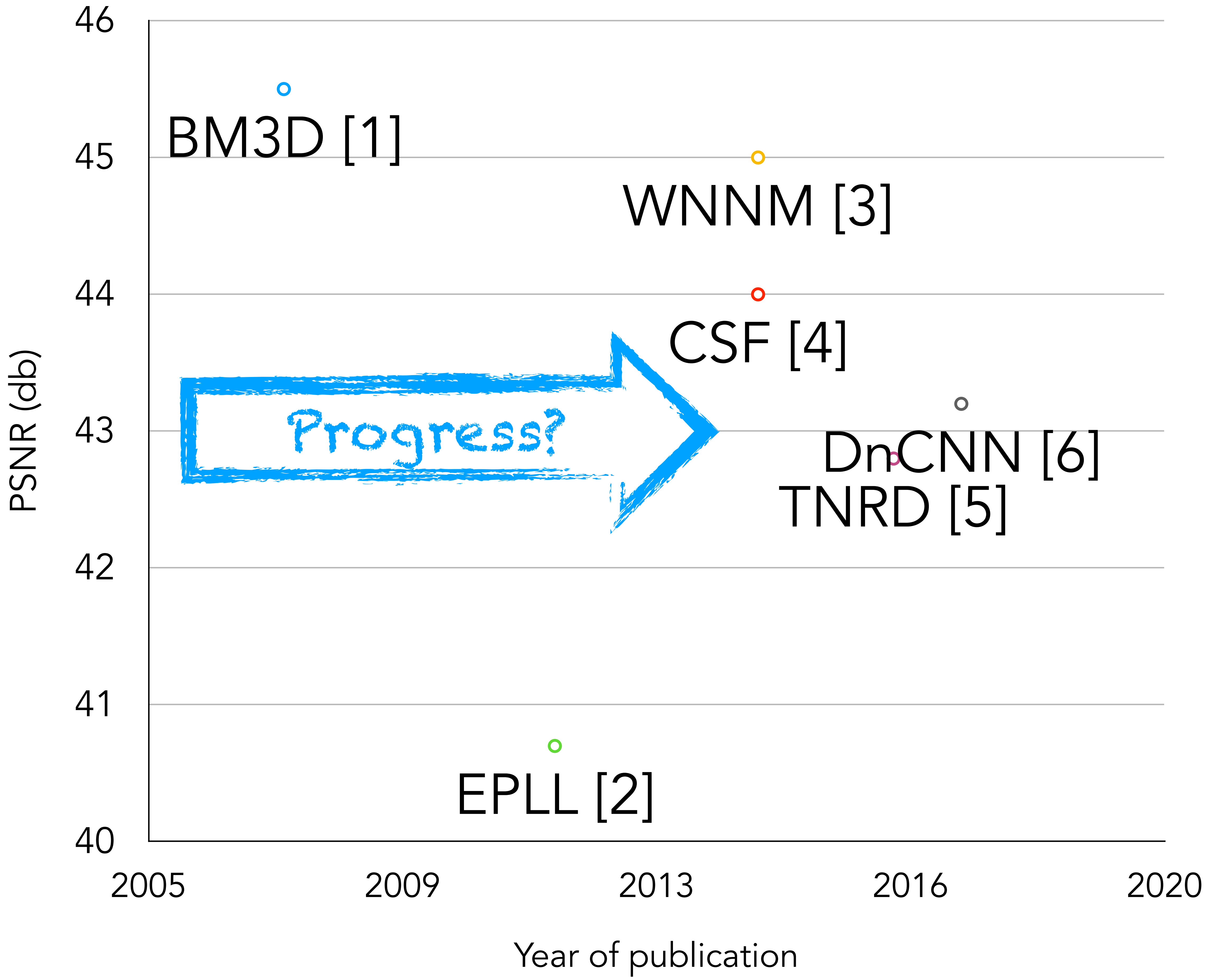


[1] Dabov et al., TIP 2007
[2] Zoran & Weiss, ICCV 2011
[3] Gu et al., CVPR 2014
[4] Schmidt & Roth, CVPR 2014
[5] Chen & Pock, TPAMI 2016
[6] Zhang et al., TIP 2017

[Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, 2017]

Denoising Progress: Real Camera Data





[1] Dabov et al., TIP 2007
[2] Zoran & Weiss, ICCV 2011
[3] Gu et al., CVPR 2014
[4] Schmidt & Roth, CVPR 2014
[5] Chen & Pock, TPAMI 2016
[6] Zhang et al., TIP 2017

[Plotz, T., & Roth, S. Benchmarking denoising algorithms with real photographs. CVPR 2017]

Possible Solutions

We have roughly two ways to go:

1. More realistic synthesize data.
2. Collecting real data pairs.

[Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., & Barron, J. T. Unprocessing images for learned raw denoising. CVPR 2019]

[Cai, J., Zeng, H., Yong, H., Cao, Z., & Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. ICCV 2019.]

Engineering Solution: Learning From Real Image Pairs

- Collect data pairs using real camera.

- Not only us, many coinstantaneous works also apply this strategy
  - *Chen, C., Xiong, Z., Tian, X., Zha, Z. J., & Wu, F. Camera Lens Super-Resolution. CVPR 2019.*
  - *Zhang, X., Chen, Q., Ng, R., & Koltun, V. Zoom to Learn, Learn to Zoom. CVPR 2019*
  - *Cai, J., Zeng, H., Yong, H., Cao, Z., & Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. ICCV 2019.*

Our darkroom

Engineering Solution: Learning From Real Image Pairs

Advantages:

- It works for real cameras.
- It can solve blind SR and blind noise problem at same time.

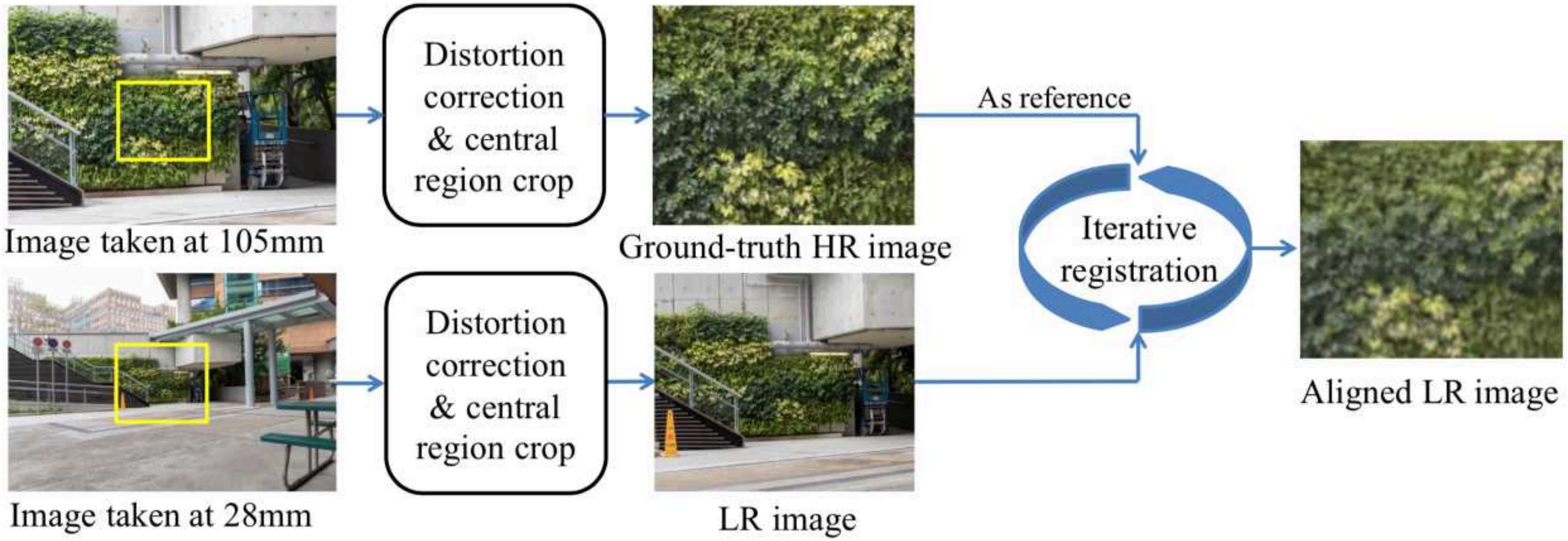
Disadvantages:

- Expensive in data collection. Many dirty works. Training with little data.
- Data can not reuse, need to collect new data for new camera.
- Not insightful in academic.

Real Image Super-Resolution

- RealSR Dataset

- NTIRE Real SR Challenge



[Cai, J., Zeng, H., Yong, H., Cao, Z., & Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. ICCV 2019.]

Expensive in data collection = Lack of training data

In our project:

- 200 image pairs for training.
- 15 images for testing

In NTIRE Real SR Challenge:

- 60 images for training
- 20 images for validation
- 20 images for testing

Resulting in Overfitting in training!

— If using PSNR as metric

Suppressing Model Overfitting for Image Super-Resolution Networks

Ruicheng Feng, Jinjin Gu, Yu Qiao, Chao Dong, CVPRW 2019, Spotlight Talk

- Real SR image pairs can not be simply synthesized.
- We are the first to study the overfitting problem in Super-Resolution.

Real LR Image

Real HR Image

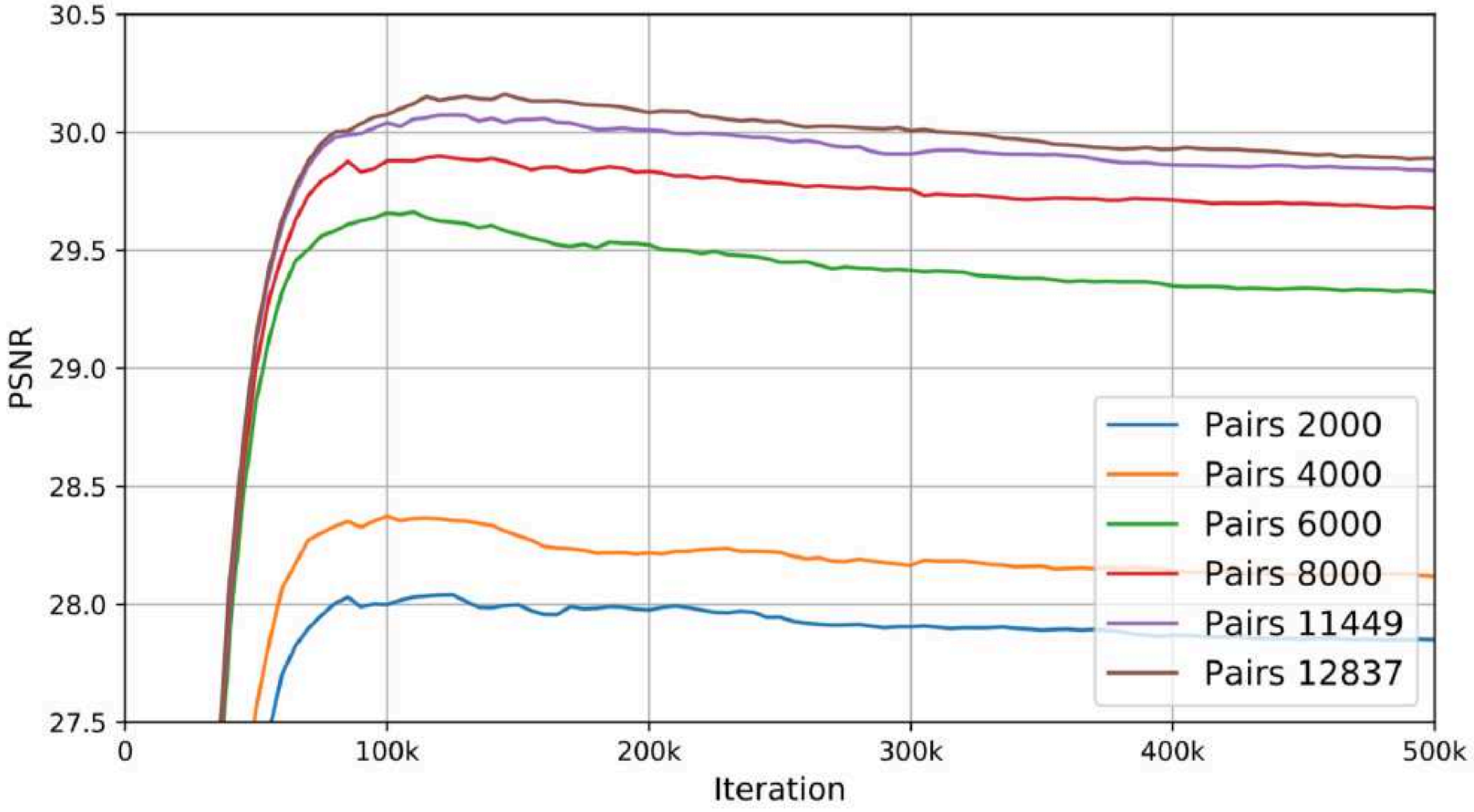
NTIRE 2019

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Overfitting in Super-Resolution

Data volume

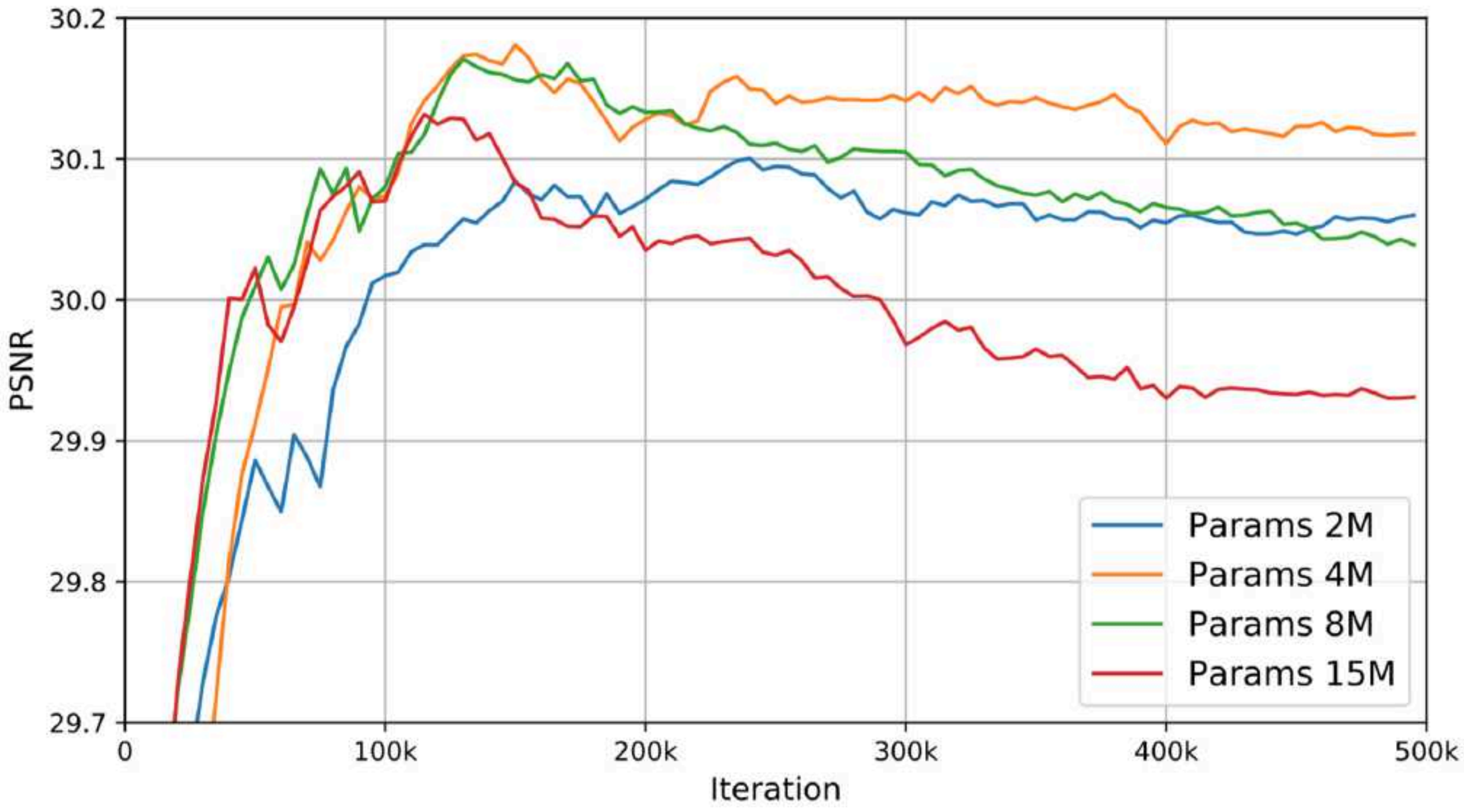
- Train large model (26M parameters) with different size of dataset.
- Increasing amounts of training data will lead to better performance



Overfitting in Super-Resolution

Model complexity

- Train models of different amounts of parameters.
- Larger models suffer from severe overfitting problem



Problem Formulation

We propose two simple yet effective strategies to suppress overfitting.

- Mixup
- Data synthesis

$$\hat{\theta} = \arg\min_{\theta} \sum_{(x,y)\in(\hat{X},\hat{Y})} L(f_\theta(x), y)$$

$$\Downarrow$$

$$\hat{\theta} = \arg\min_{\theta} \underbrace{\sum_{(x,y)\in(\hat{X},\hat{Y})} L(f_\theta(x), y)}_{\text{Observation set}} + \underbrace{\sum_{(x,y)\in(X',Y')} L(f_\theta(x), y)}_{\text{Augmentation set}}$$

Mixup Trick for SR

Randomly draw two samples (x_i, y_i) and (x_j, y_j) from $(\hat{x}, \hat{y})$.

$$x' = \lambda x_i + (1 - \lambda) x_j$$
$$y' = \lambda y_i + (1 - \lambda) y_j$$

Two random HR images　　Mixup HR Image

Two random LR images　　Mixup LR Image

Mixup Trick for SR

Randomly draw two samples (x_i, y_i) and (x_j, y_j) from $(\hat{x}, \hat{y})$.

$$x' = \lambda x_i + (1 - \lambda) x_j$$
$$y' = \lambda y_i + (1 - \lambda) y_j$$

Mixup HR Image

Mixup LR Image

(x', y') follows same degradation model!

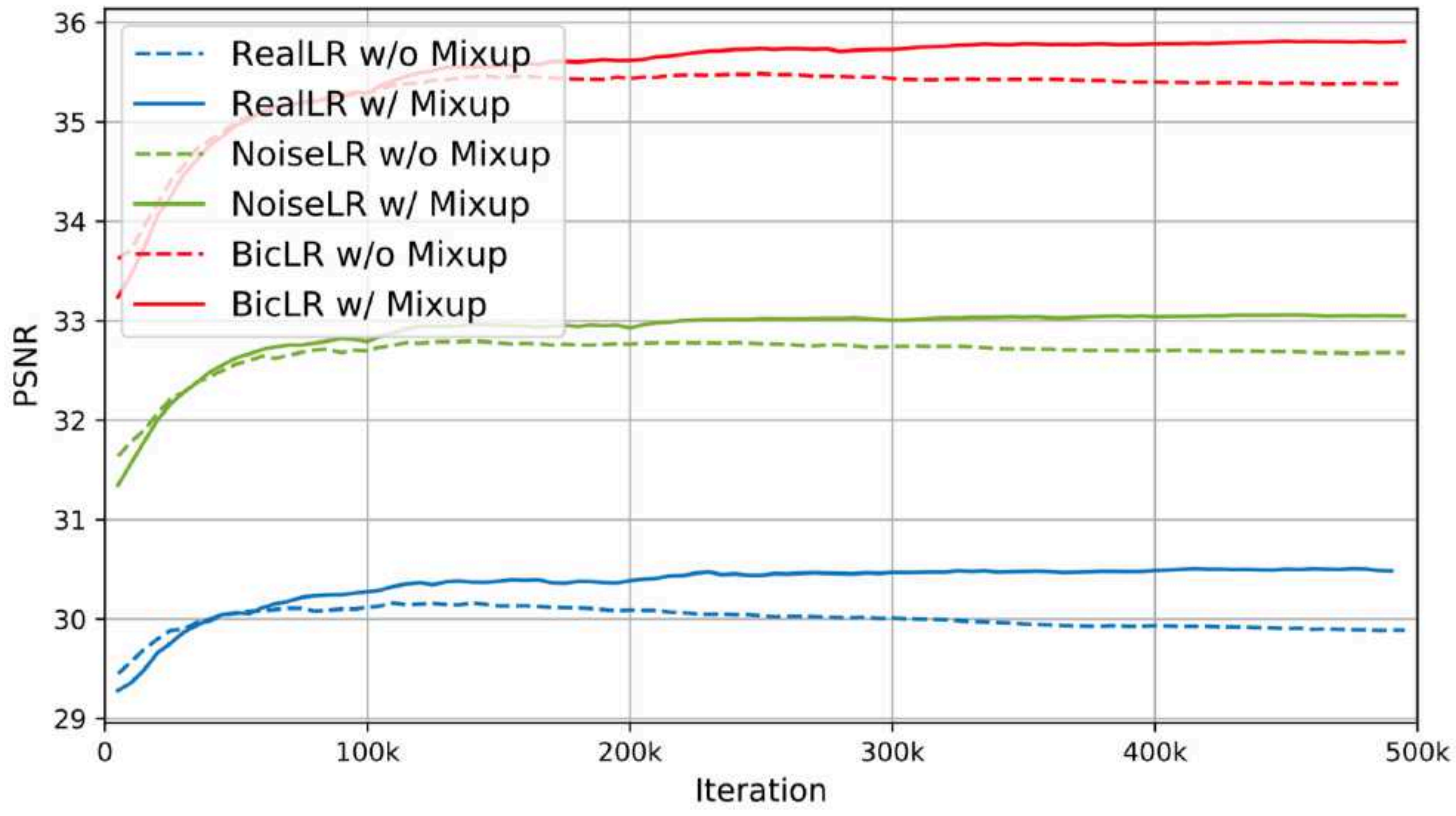
$$\begin{aligned} x' &= \lambda x_i + (1 - \lambda) x_j \\ &= \lambda (D y_i + n_i) + (1 - \lambda)(D y_j + n_j) \\ &= D(\lambda y_i + (1 - \lambda) y_j) + (\lambda n_i + (1 - \lambda) n_j) \\ &= D y' + n' \end{aligned}$$

Mixup Trick for SR

Mixup significantly reduces overfitting and guarantees robust training for 3 degradation models.

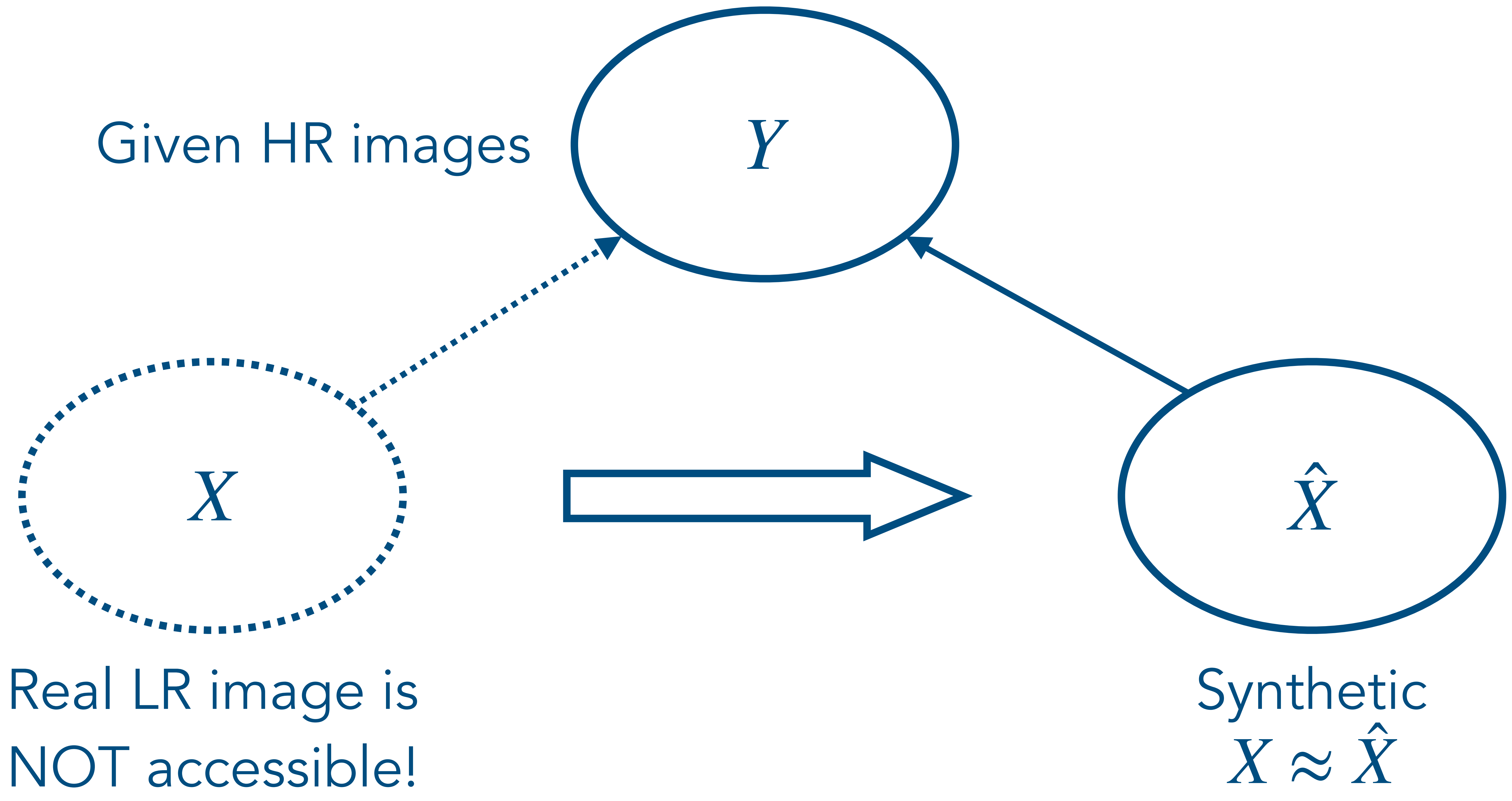
Test on various types of degradation

- RealLR
- Bicubic downsample
- Bicubic downsample + noise



Data Synthesis with Learned Degradation

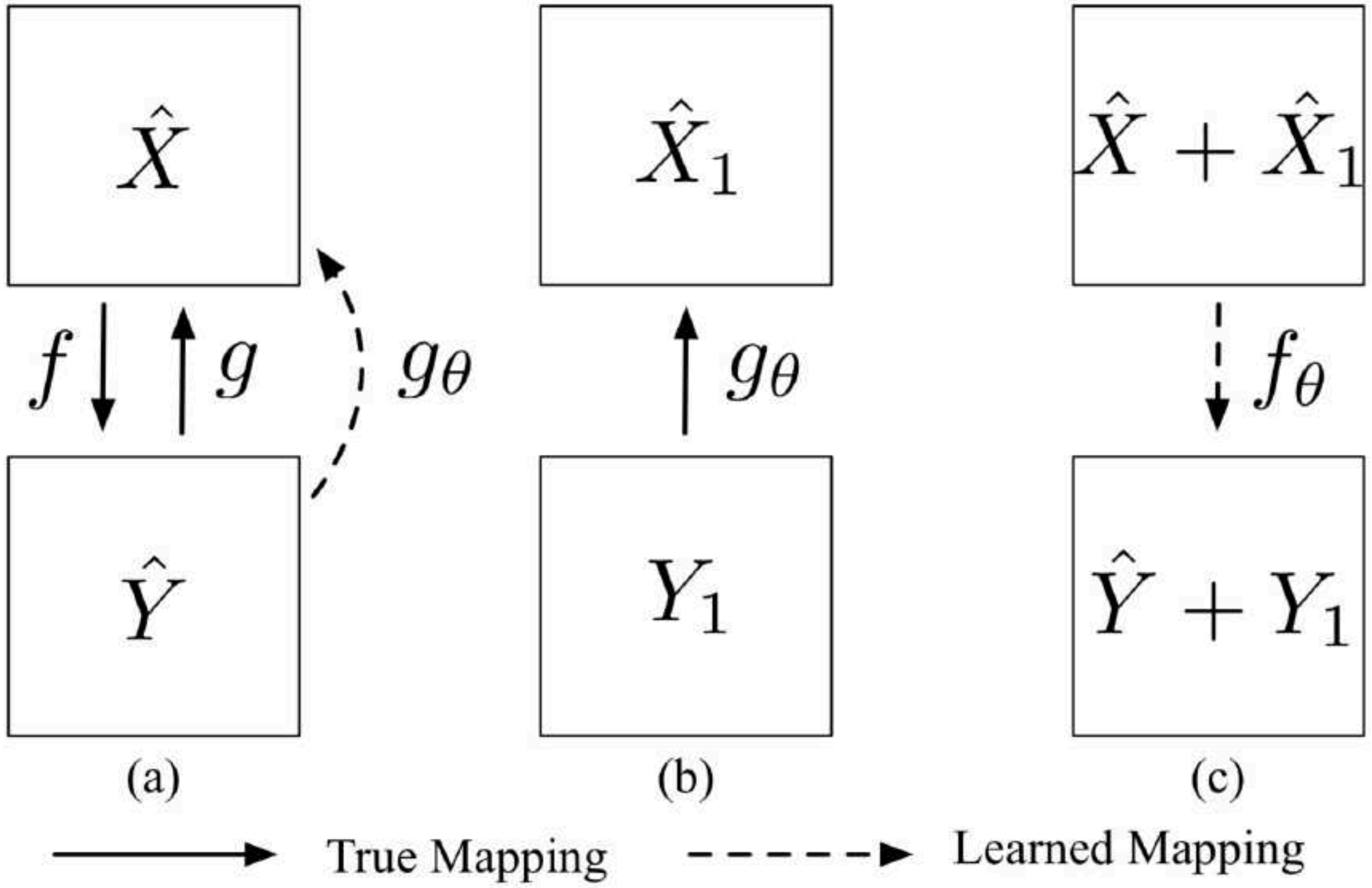
- Provide more image pairs with diverse content.
- Utilize high-quality dataset (DIV2K, Flickr2K, etc.)



Data Synthesis with Learned Degradation

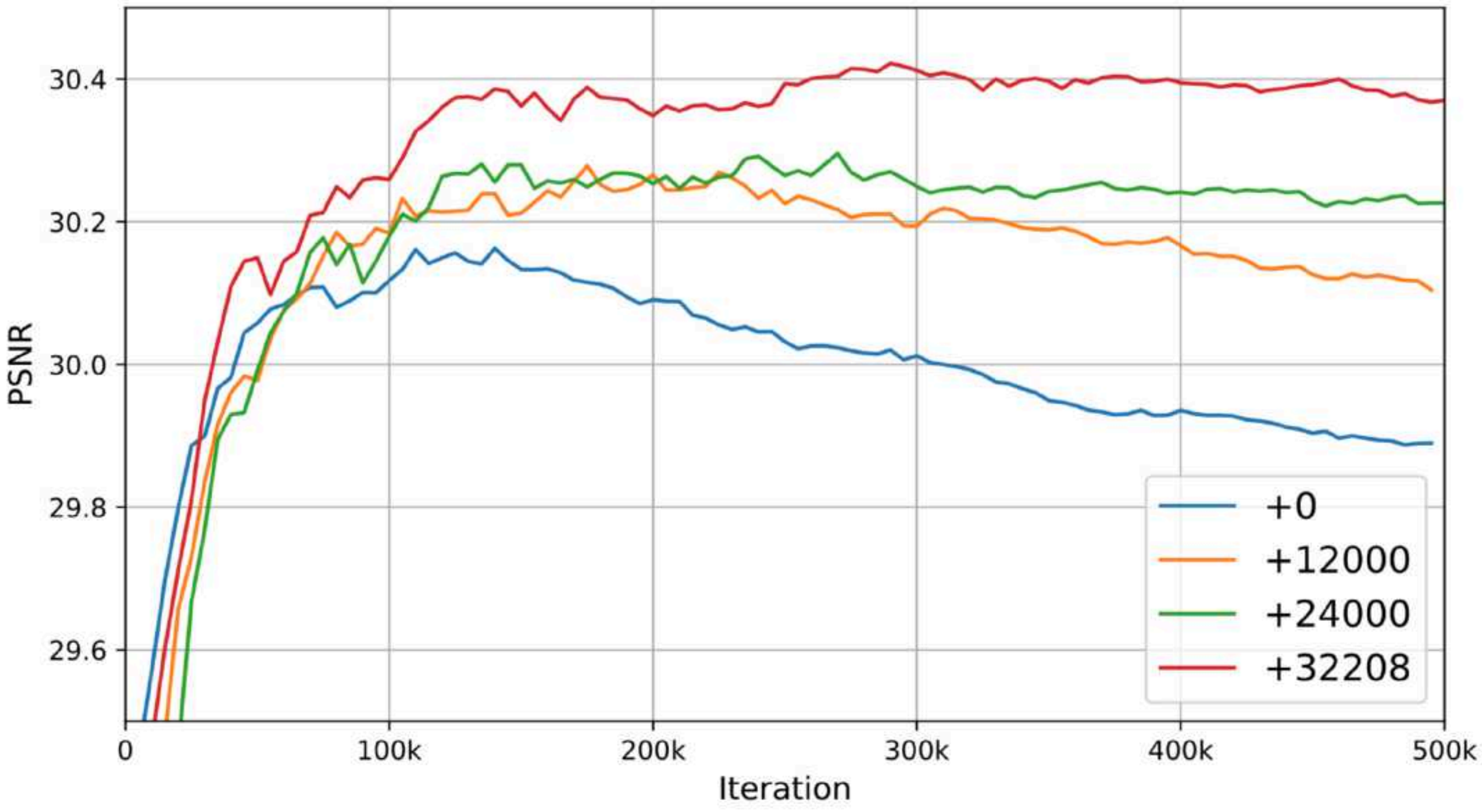
- Provide more image pairs with diverse content.

- Utilize high-quality dataset (DIV2K, Flickr2K, etc.)



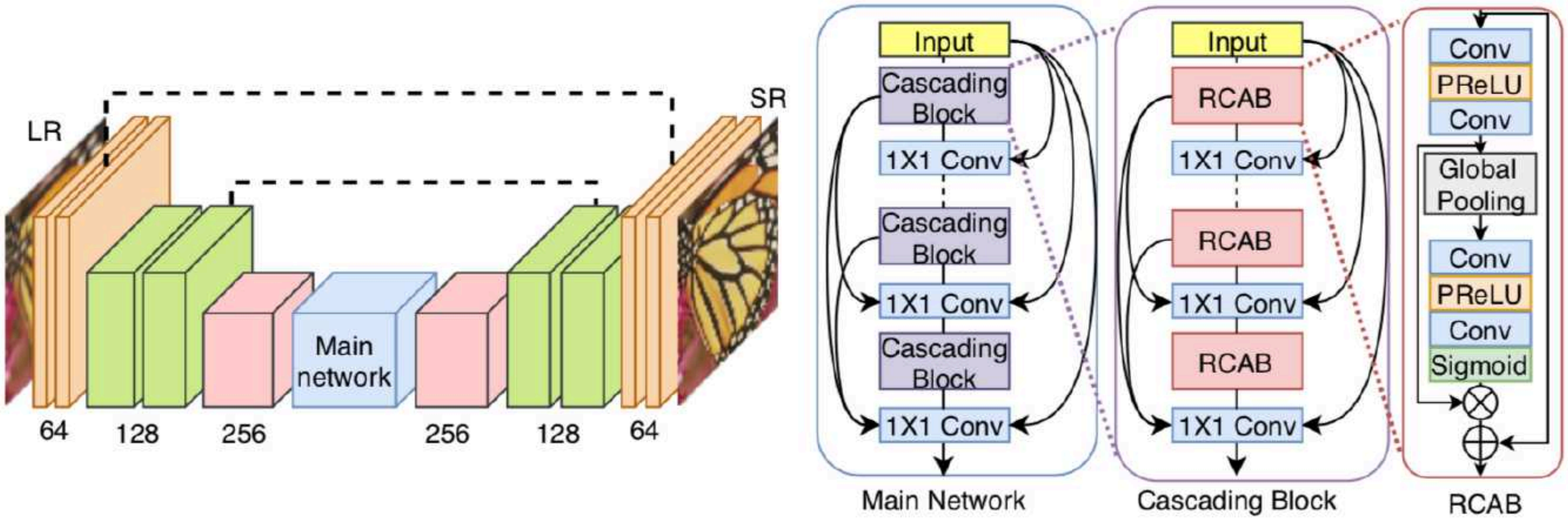
Data Synthesis with Learned Degradation

The observation set is biased sampled and lack diversity. Adding more synthetic data to the training set encourages better generalization.



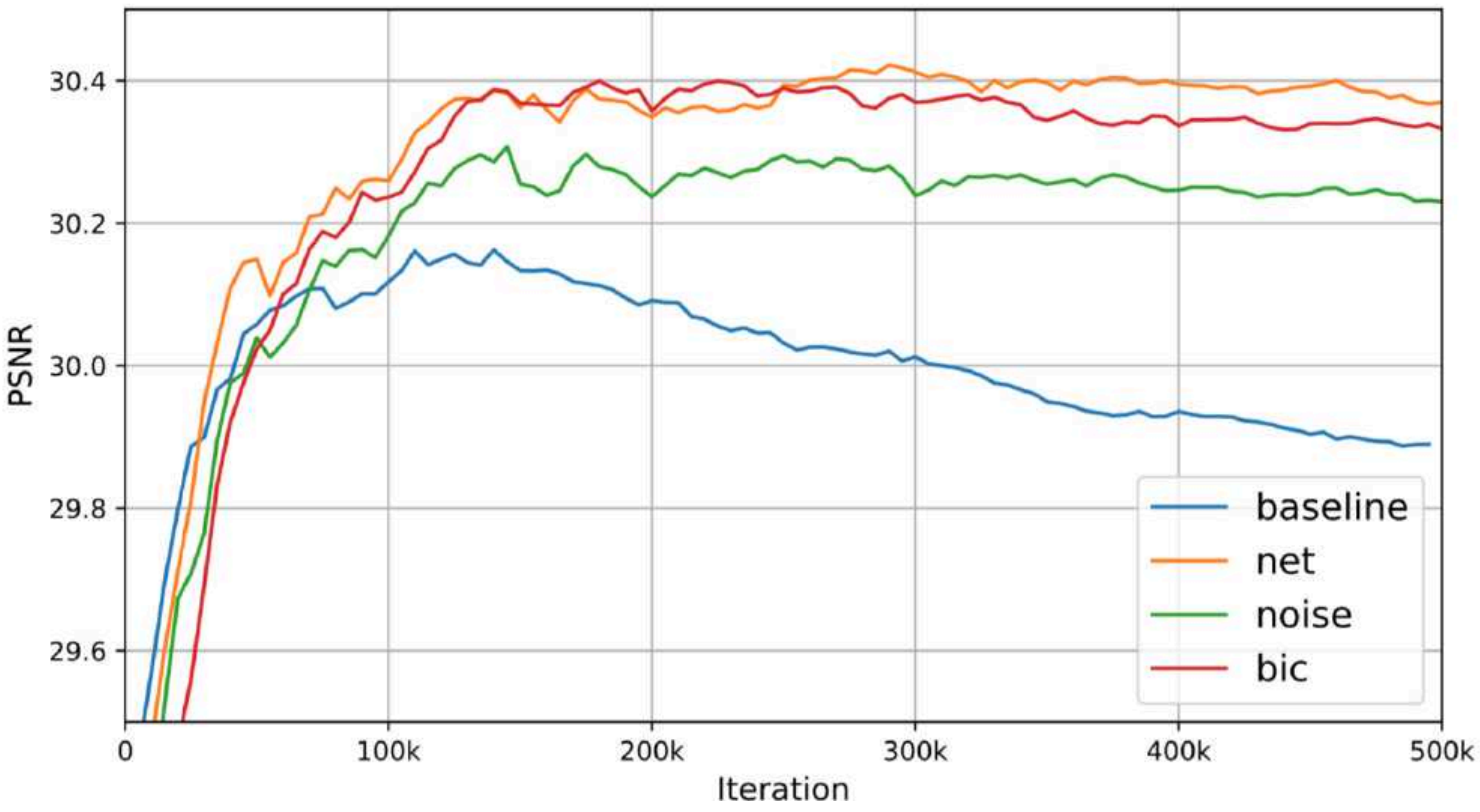
Network Structure

- U-Net shape
- Main network topology motivated by DenseNet.
- Basic block from Residual Channel Attention Blocks.



Combination of Mixup and Data Synthesis

Method	PSNR	SSIM
FSRCNN[13]	28.3394	0.8254
CARN[3] + ES	29.1620	0.8580
RRDB[41] + ES	29.4581	0.8643
RCAN[45] + ES	29.6299	0.8675
U-Net(Ours) + Synthesis	29.8503	0.8731
U-Net(Ours) + MixUp	29.9055	0.8729
U-Net(Ours) + Synthesis + MixUp	30.0278	0.8753
U-Net(Ours)* + Synthesis + MixUp	**30.1624**	**0.8777**

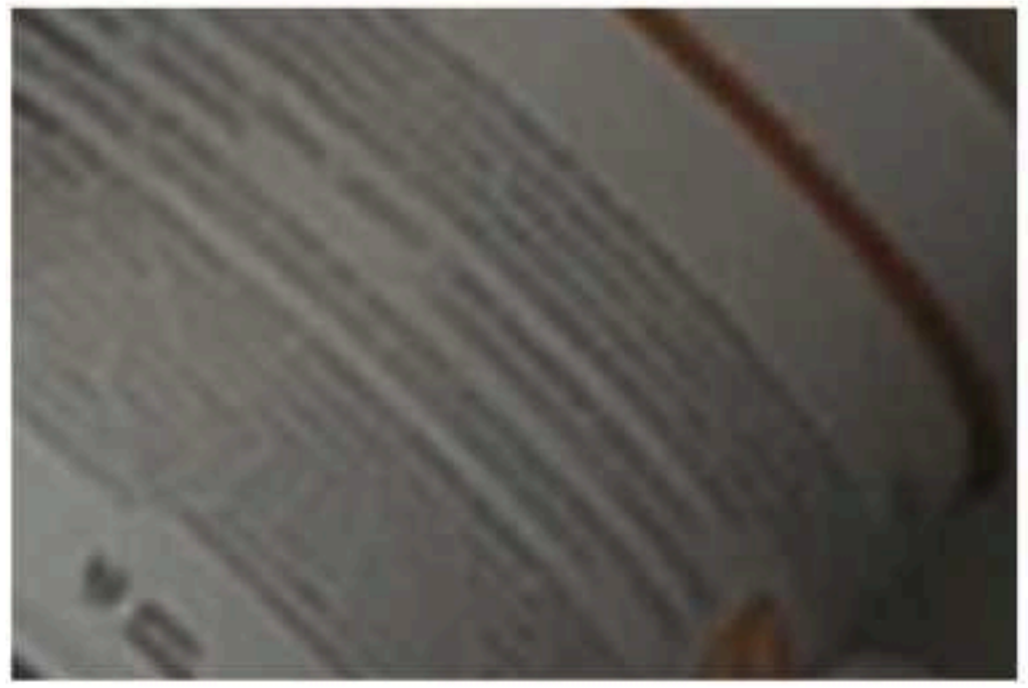


NTIRE 2019 Real Image SR Competition

we achieve competitive performance with only 0.03 difference of the PSNR performance. And we only use almost half of their running time.

Method	PSNR	SSIM	Runtime[s]	Platform	CPU / GPU
1st Method	29.00	0.84	60.00	Pytorch	Tesla V100
Ours	28.97	0.84	36.90	Pytorch	GTX 1080
3rd Method	28.93	0.84	600.00	Pytorch	Titan V
4th Method	28.93	0.83	106.56	Tensorflow	Tesla V100
5th Method	28.88	0.84	38.00	Pytorch	GTX 1080ti
6th Method	28.88	0.83	98.00	Pytorch	Titan Xp
7th Method	28.87	0.83	-	Pytorch	Quadro P6000
8th Method	28.81	0.83	10.00	Pytorch	GTX 1080Ti
9th Method	28.79	0.84	47.08	Pytorch	GTX 1080Ti
10th Method	28.76	0.83	100.28	Pytorch	Titan X
……	…..	…..	…..	…..	…..

Qualitative Comparison



FSCRNN
28.38 / 0.6741

CARN
28.55 / 0.7076

FSRCNN
25.99 / 0.7395

CARN
26.44 / 0.7668

RCAN
28.78 / 0.7341

Ours
29.22 / 0.7840

RCAN
27.17 / 0.7794

Ours
27.62 / 0.7934

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

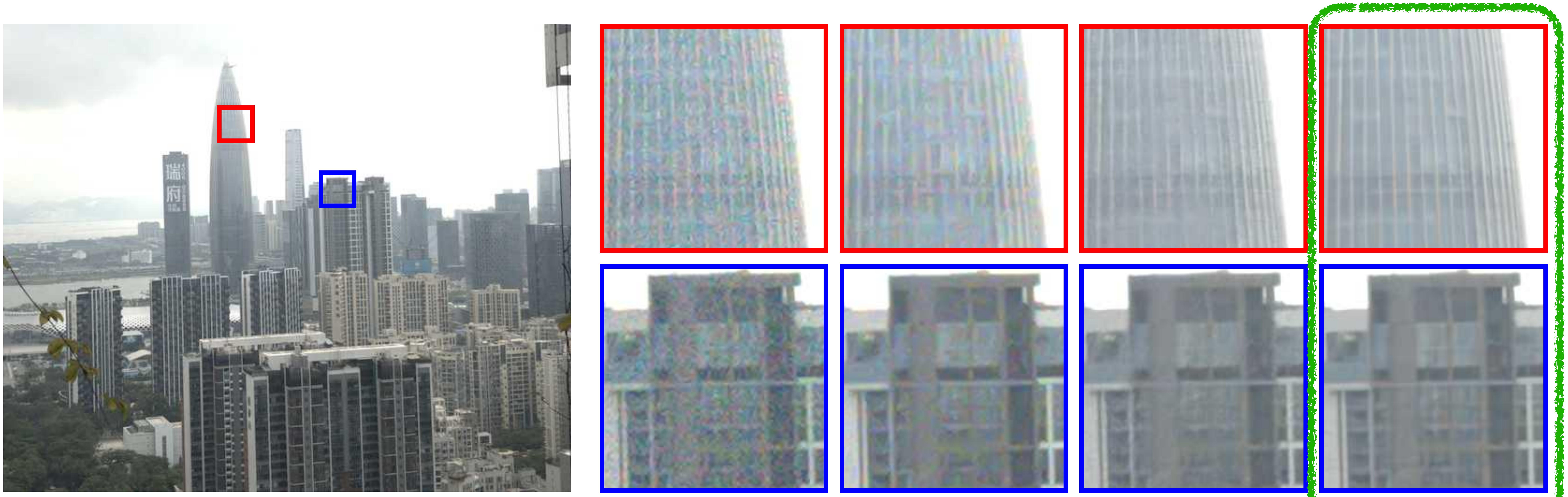
Our product
Vivo X23 Smart Camera Solution

Challenges

- Unknown Downsampling for SR
- Unknown Noise for Denoising
- **Mixture Problem of Denoising and SR**
- Unaligned Multiple Frames
- Time constraint (4K Res Image, Less than 300ms)
- No Powerful GPU

Trinity of Pixel Enhancement: a Joint Solution for Demosaicking, Denoising and Super-Resolution

Guocheng Qian*, Jinjin Gu*, Jimmy S. Ren, Chao Dong, 2019



Real-World Image Captured by iPhone X | dcraw + *CARN | Camera Raw + *CARN | DemosaicNet + *CARN | TENet (ours)

Mixture Problem Brings New Challenges

Low-level tasks interact with each other

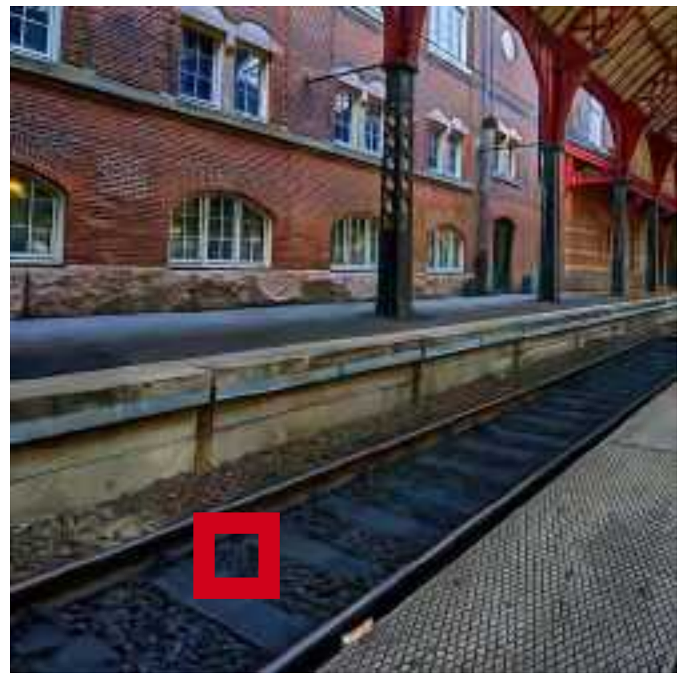
Demoniac
LR Raw
Image

Demoniac
HR Raw
Image

The image demosaiced from an HR image provides better result.

Mixture Problem Brings New Challenges

Low-level tasks interact with each other



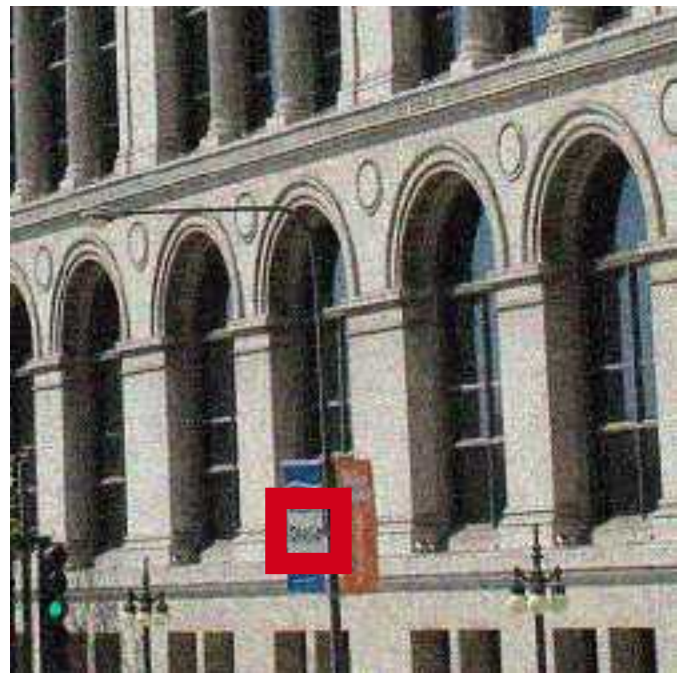
SR result of a denoised LR image

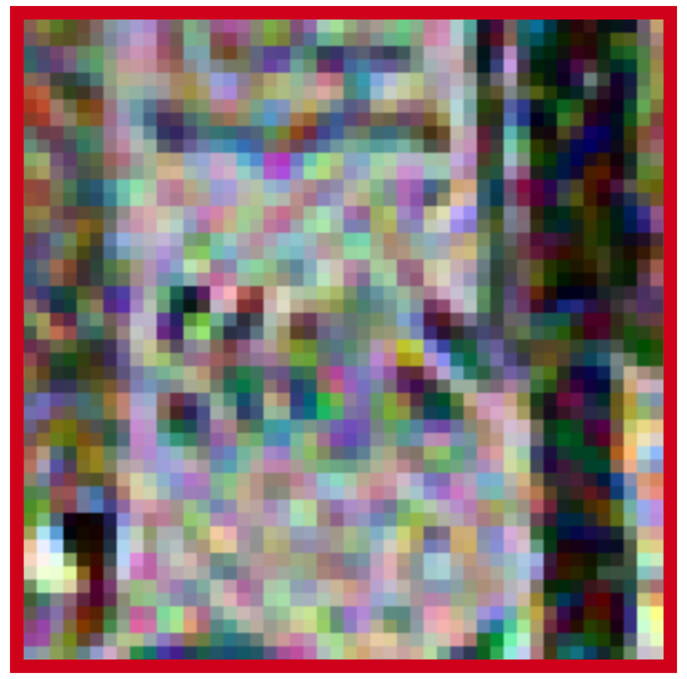
SR result of a clean LR image

The denoising task tends to smooth the high frequency details. This blurry will be magnified by the following SR task.

Mixture Problem Brings New Challenges

Low-level tasks interact with each other





SR result of a noisy LR image

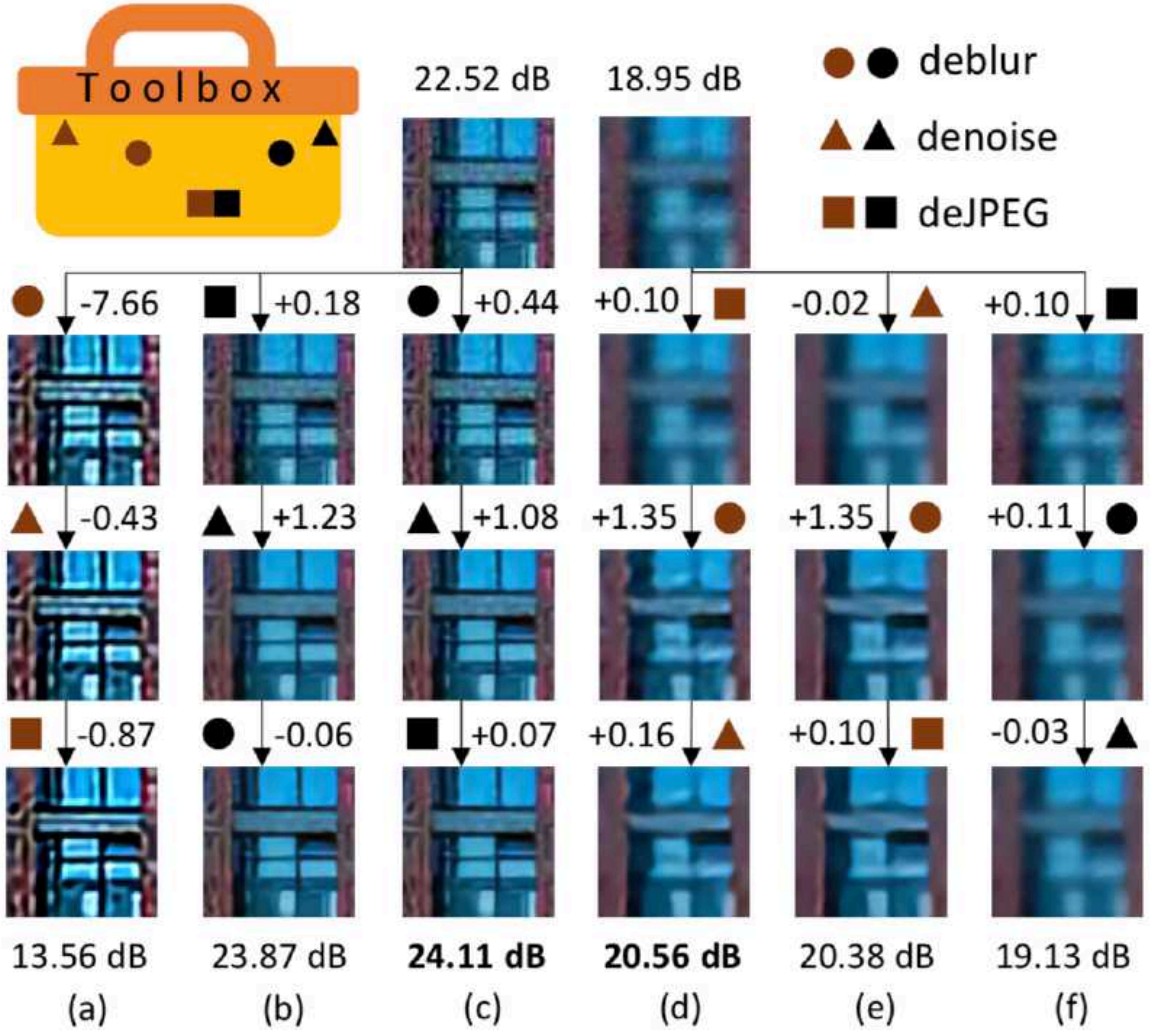
SR result of a clean LR image

Super-resolving a noisy input brings serious artifacts, which can not be easily removed.

Mixture Problem Brings New Challenges

Effect of the "Middle State"

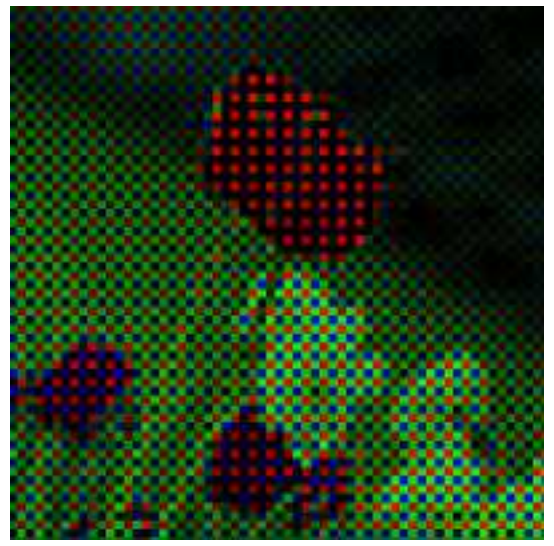
Apply different Pipeline will be different



[Yu, K., Dong, C., Lin, L., & Change Loy, C. Crafting a toolchain for image restoration by deep reinforcement learning. CVPR 2018.]

Pipeline in Computational Photography

Traditional Pipeline:



Input Raw Image

ISP

Color Image

Image Enhancement

Clean High-res Color Image

Pipeline in Computational Photography

Rethink Pipelines:

- We suggest to denoise before other tasks to minimize the effects of noise.

- We propose to super-resolve the raw image to a higher resolution and then perform demosaicing to get the SR color image.

SR Raw Image vs. SR Demosaiced Color Image

There are at least two advantages:

(1) The artifacts caused by super-resolving the defects of the demosaiced images can be avoided.

(2) SR can help demosaicing task to break the limitation of resolution.

Joint Solution

We still cannot totally solve the problem caused by the middle state.

No SR or demosaicing method is designed to compensate the lost high-frequency details.

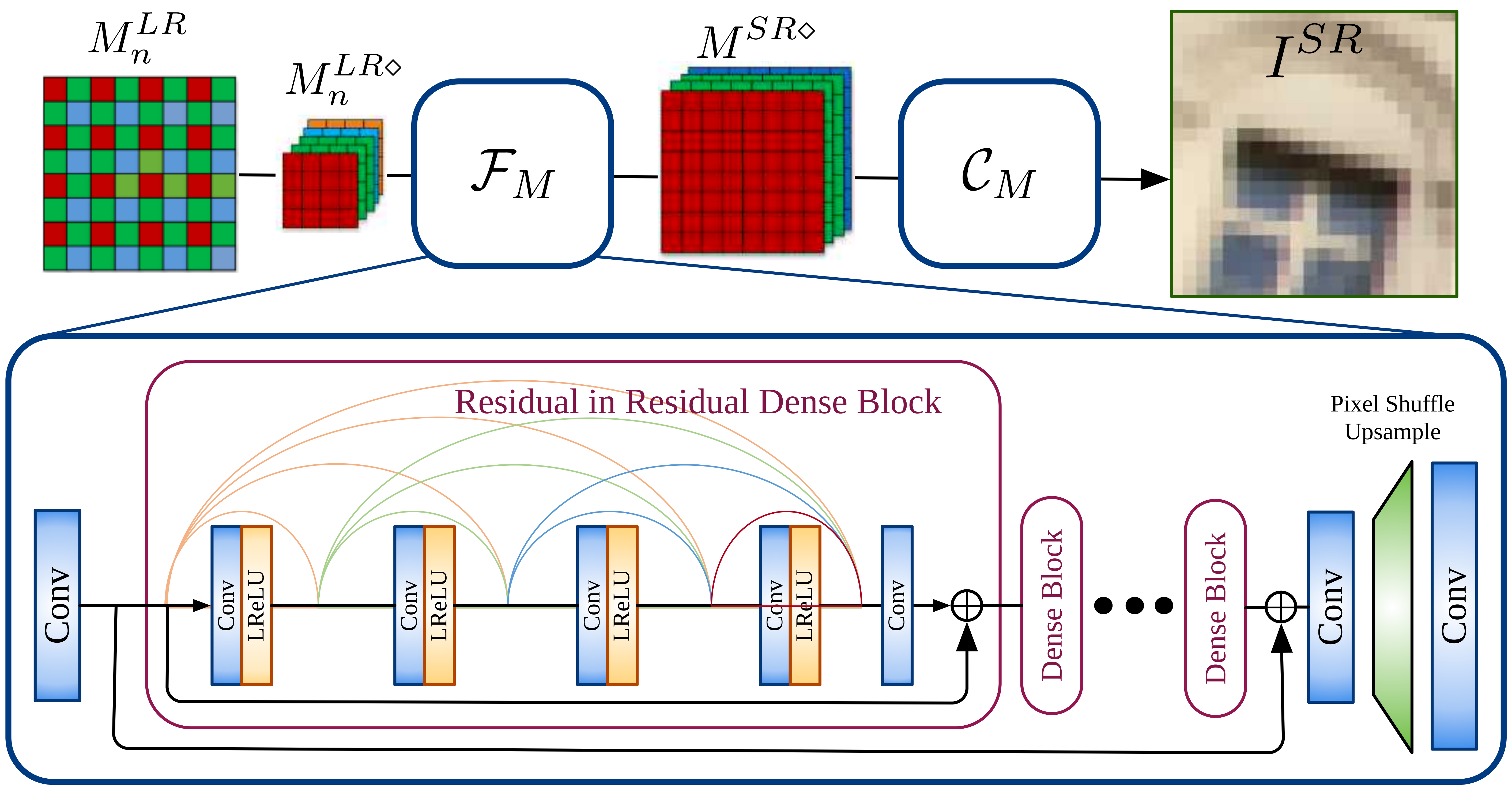
Joint Solution

We still cannot totally solve the problem caused by the middle state.

No SR or demosaicing method is designed to compensate the lost high-frequency details.

Joint perform denoising, SR and demosaicing in an end-to-end manner.

The Proposed Method



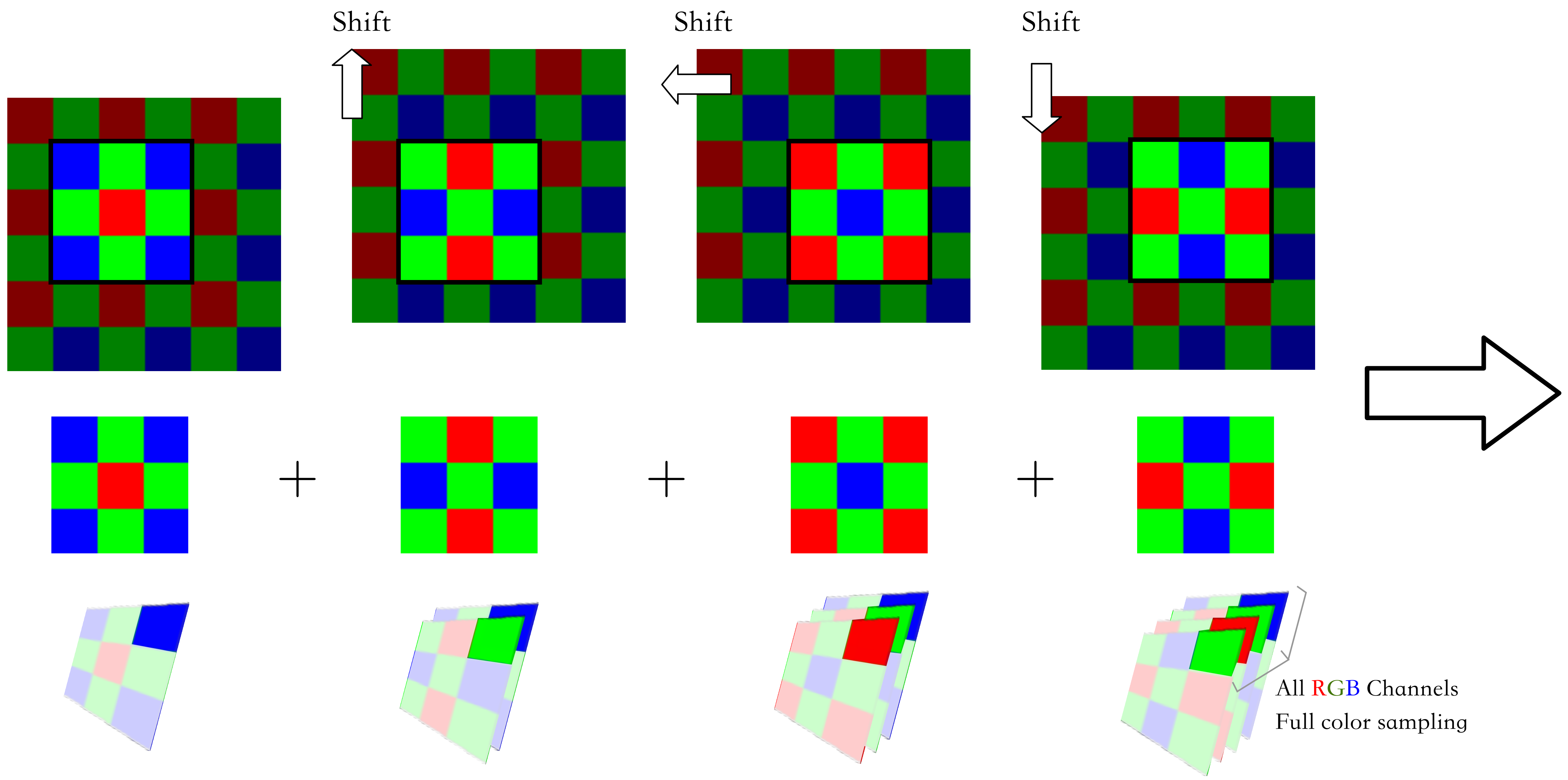
Joint Objective Function

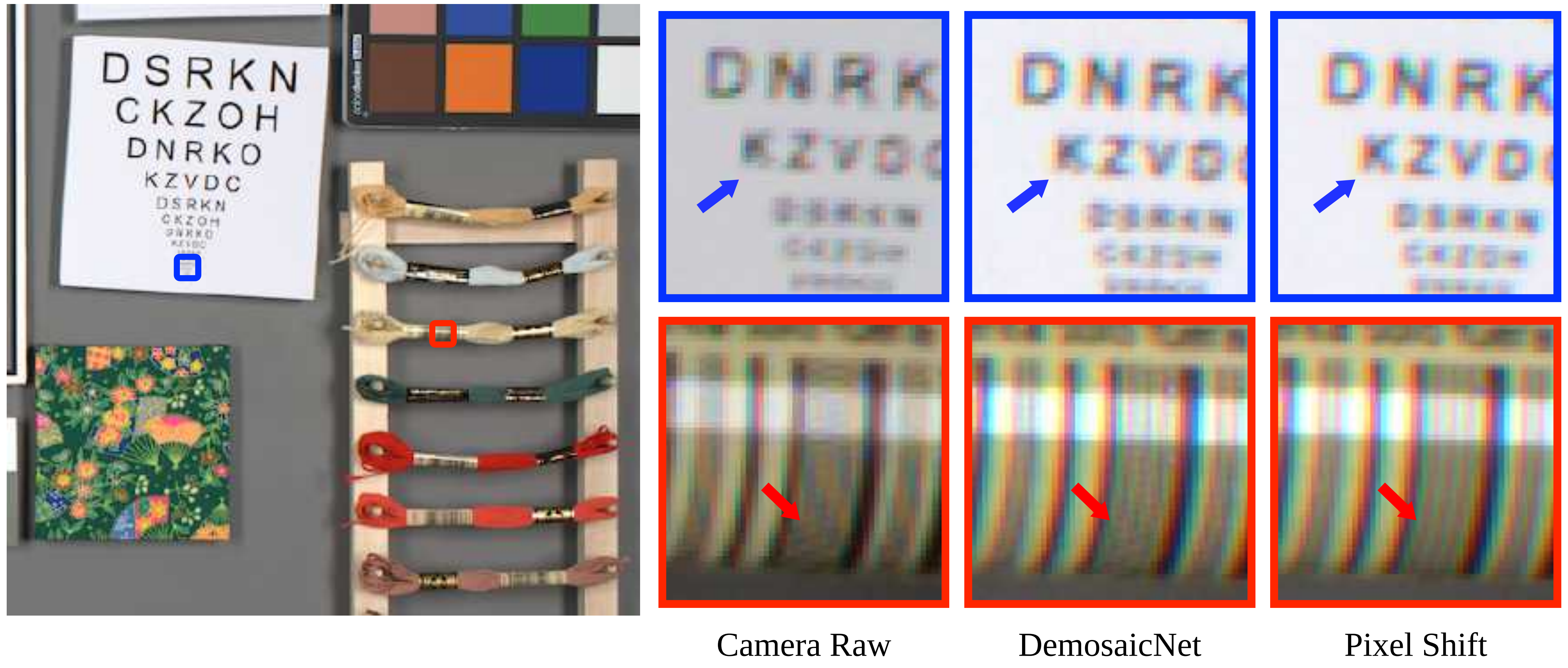
$$L = L_{joint} + \lambda L_{SR}$$

$$L_{joint} = \| C(S_M(D_M(M_n^{LR}))) - I_{gt}^{HR} \|_2^2$$

$$L_{SR} = \| S_M(D_M(M_n^{LR})) - M_{gt}^{HR} \|_2^2$$

PixelShift200: Building Demosaicing Dataset with Real Color Data





Dataset: https://guochengqian.com/pixelshift200/

Qualitative Comparison

Ground Truth

Demosaicing -> Denoising -> SR

Denoising should be perform first

Denoising -> Demosaicing -> SR

Qualitative Comparison

Ground Truth

Denoising -> Demosaicing -> SR

Effectiveness of performing SR before demosaicing

Denoising -> SR -> Demosaicing

Qualitative Comparison

Ground Truth

Denoising -> SR -> Demosaicing

Effectiveness of
joint solution

Our Joint Solution

Quantitative Comparison

Quantitative comparison of different pipelines. In this experiment, the tasks are perform step by step.

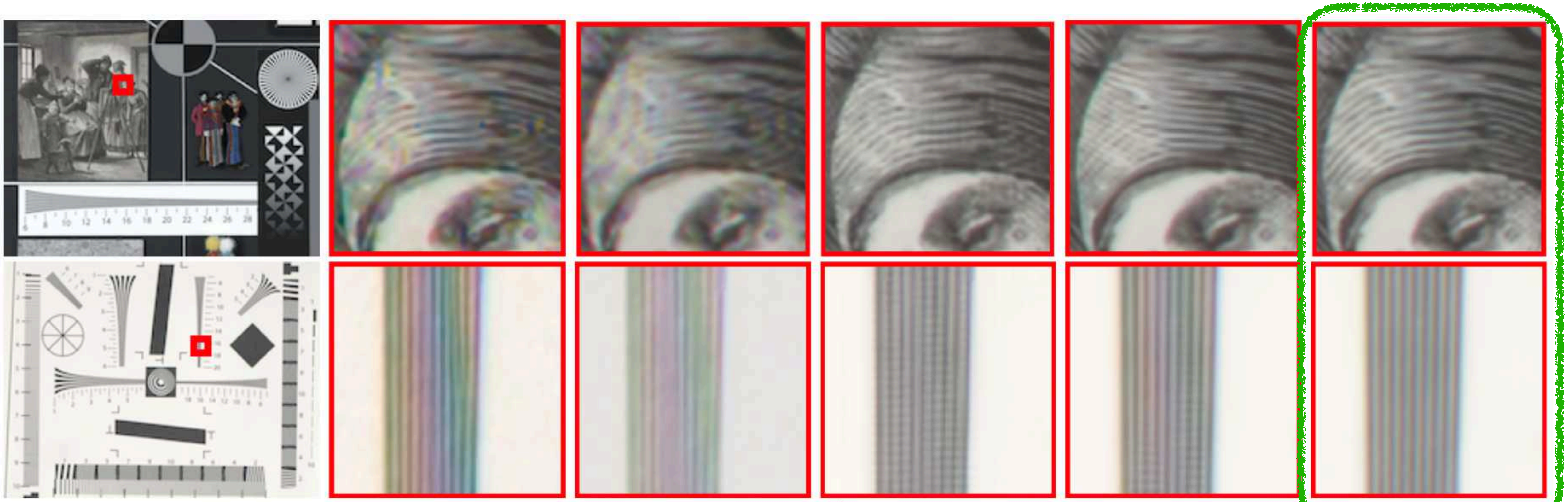
Method	Kodak		Urban100 [19]	
	PSNR	SSIM	PNSR	SSIM
DM $\rightarrow$ SR $\rightarrow$ DN	26.40	0.6495	24.98	0.7029
SR $\rightarrow$ DM $\rightarrow$ DN	26.86	0.6796	25.42	0.7311
SR $\rightarrow$ DN $\rightarrow$ DM	27.28	0.7089	25.86	0.7589
DM $\rightarrow$ DN $\rightarrow$ SR	26.97	0.6991	25.38	0.7491
DN $\rightarrow$ DM $\rightarrow$ SR	28.40	0.8028	26.55	0.8355
DN $\rightarrow$ SR $\rightarrow$ DM	28.45	0.8038	26.75	0.8395

Quantitative comparison of different joint solutions. In this experiment, the tasks are joint or partially joint performed (denoted by +) in different orders.

Method	Kodak		Urban100 [19]	
	PSNR	SSIM	PNSR	SSIM
SR $\rightarrow$ DN + DM	27.27	0.7062	25.89	0.7579
DN $\rightarrow$ DM + SR	28.47	0.8041	26.76	0.8396
DM $\rightarrow$ DN + SR	27.07	0.6869	25.86	0.7468
DM + DN $\rightarrow$ SR	28.43	0.8039	26.59	0.8365
DM + SR $\rightarrow$ DN	26.67	0.6618	25.26	0.7149
DN + SR $\rightarrow$ DM	28.54	0.8048	26.96	0.8437
SR + DN + DM	28.56	0.8050	27.10	0.8451
SR + DN + DM, $w/\mathcal{L}_{SR}$	28.60	0.8051	27.14	0.8458

Qualitative Comparison

Real Raw Image



dcraw +*CARN [3] | Camera Raw +*CARN [3] | †DemosaicNet +*CARN [3] | TENet *w/* 2K training set | TENet *w/ PixelShift200*

Apply Deep Image Restoration in Real World

Smartphone Camera
"AI Smart Camera"

Our product
Vivo X23 Smart Camera Solution

Challenges

- Unknown Downsampling for SR
- Unknown Noise for Denoising
- Mixture Problem of Denoising and SR
- **Unaligned Multiple Frames**
- **Time constraint (4K Res Image, Less than 300ms)**
- **No Powerful GPU**

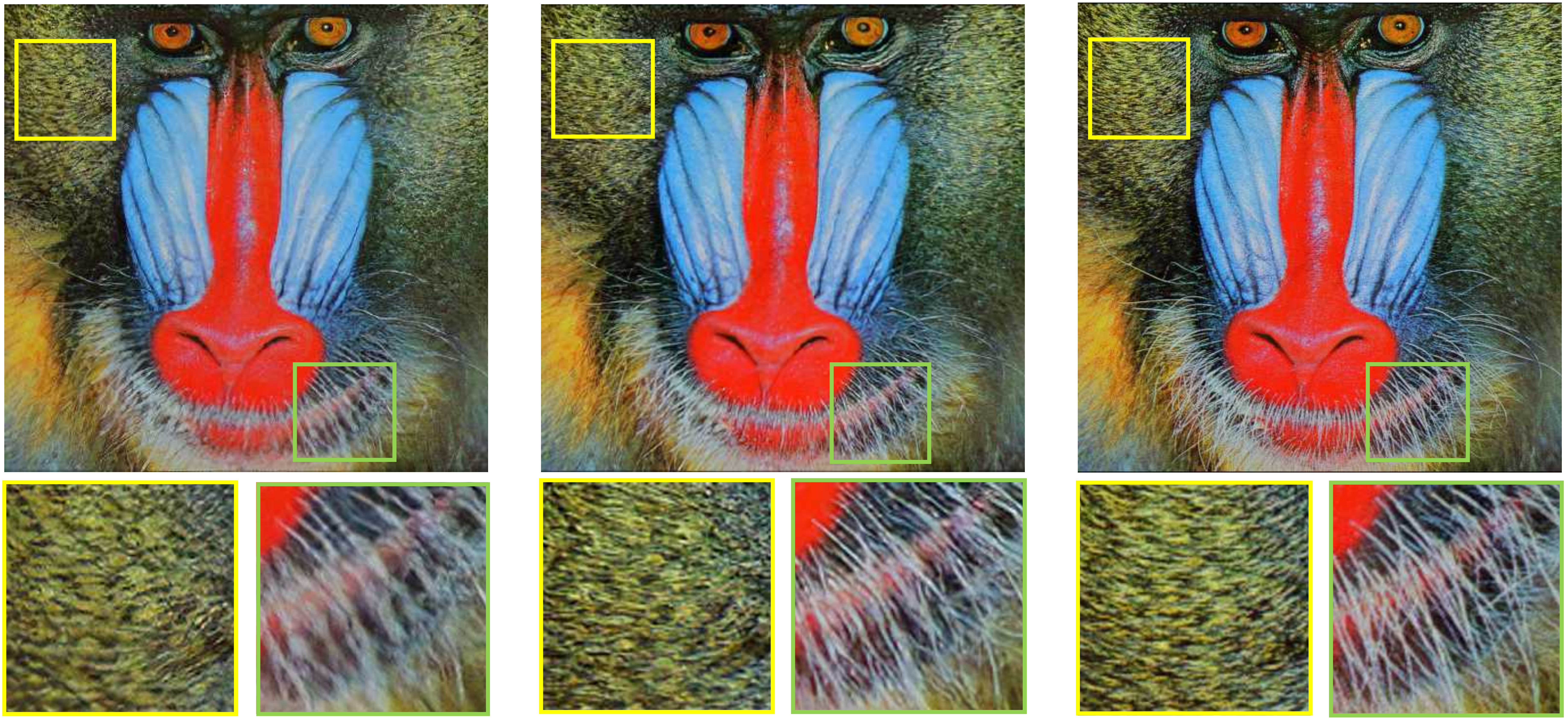
Perceptual Image Restoration

The Perceptual-distortion trade-off

[Blau, Y., & Michaeli, T. The perception-distortion tradeoff. CVPR 2018]

ESRGAN: Enhanced Super-Resolution GAN

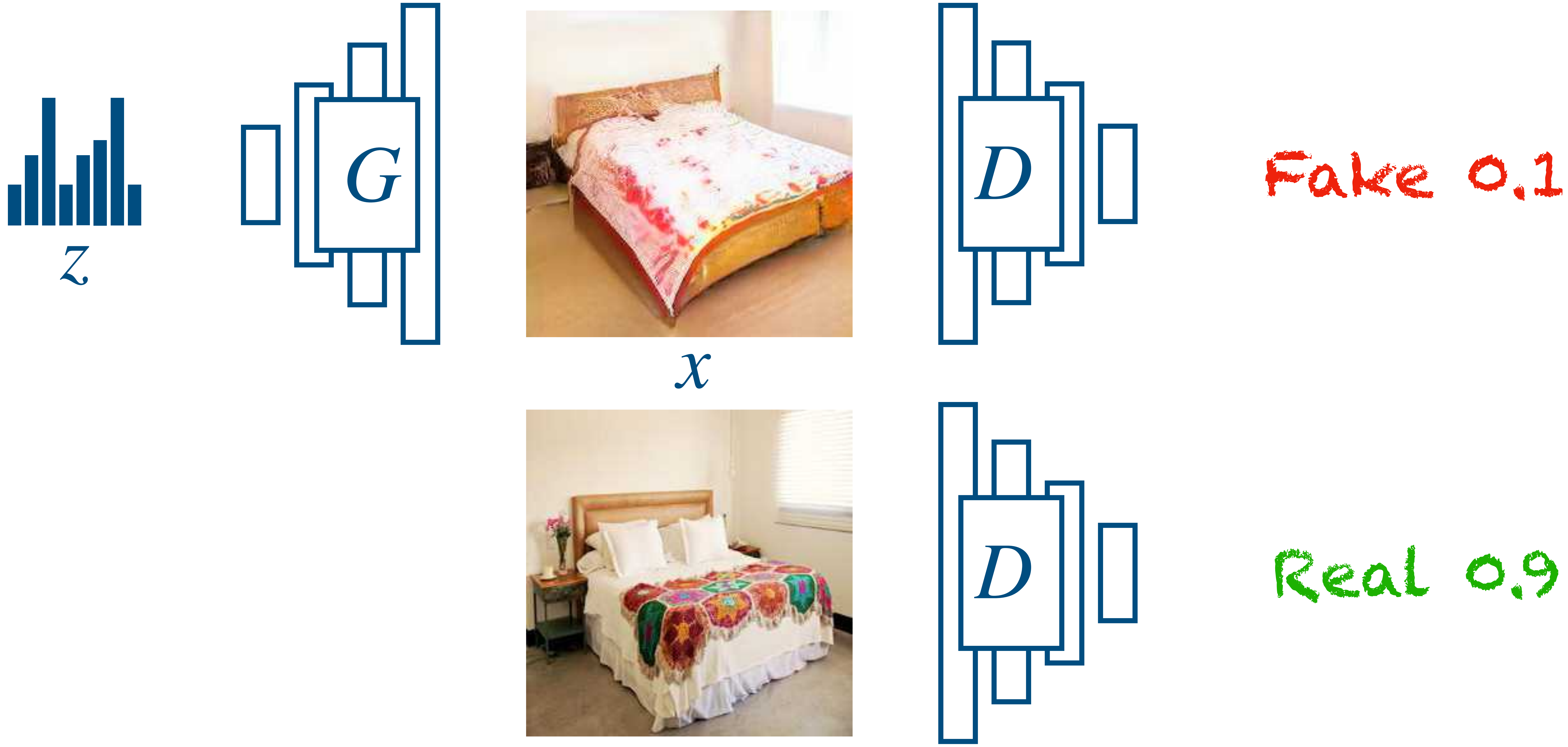
Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang, PIRM 2018 ECCVW, Spotlight Talk



SRGAN ESRGAN Ground Truth

[Ledig C, Theis L, Huszár F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, CVPR. 2017.]

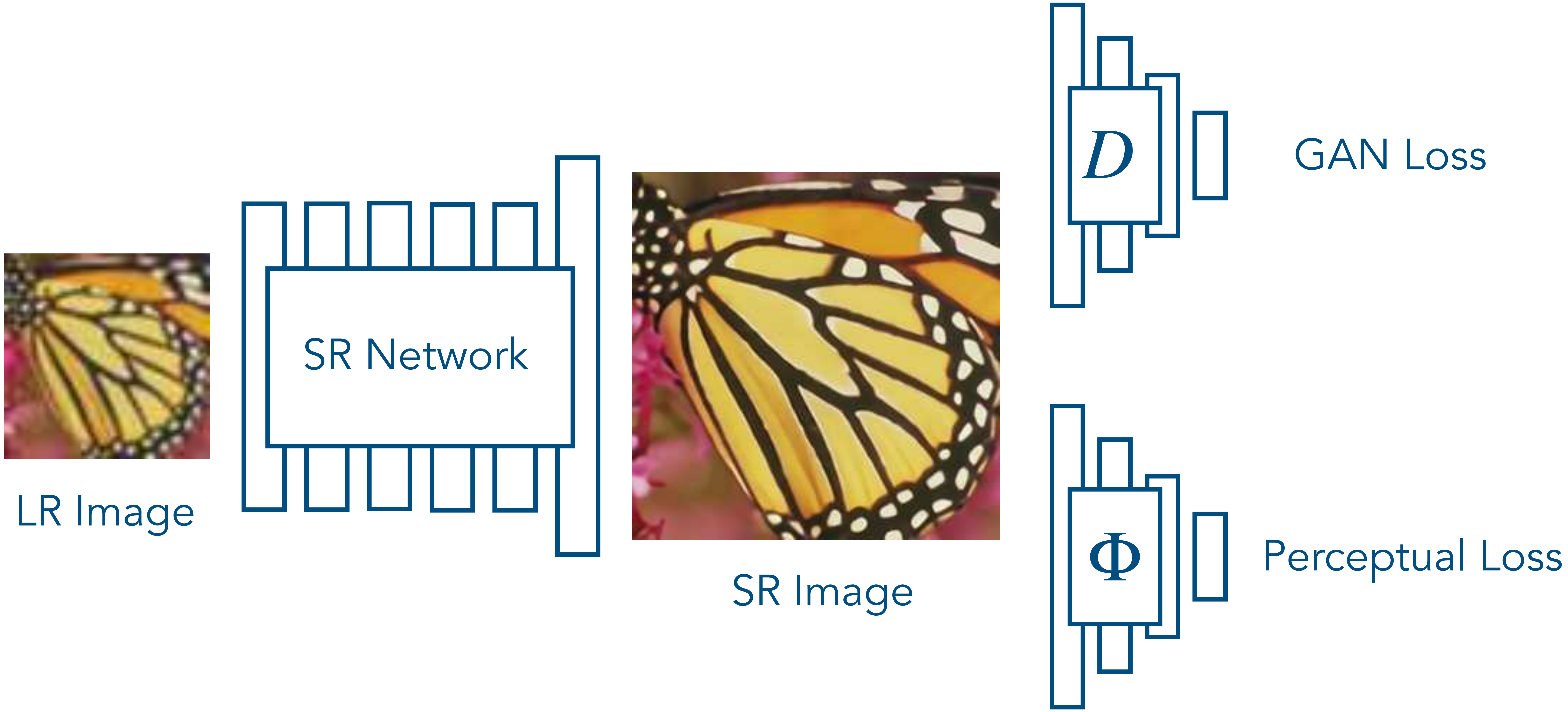
Generative Adversarial Networks



$$\min_G \max_D \mathbb{E}_{z,x}[\boxed{\log D(G(z))} + \log(1 - \boxed{D(x)})]$$

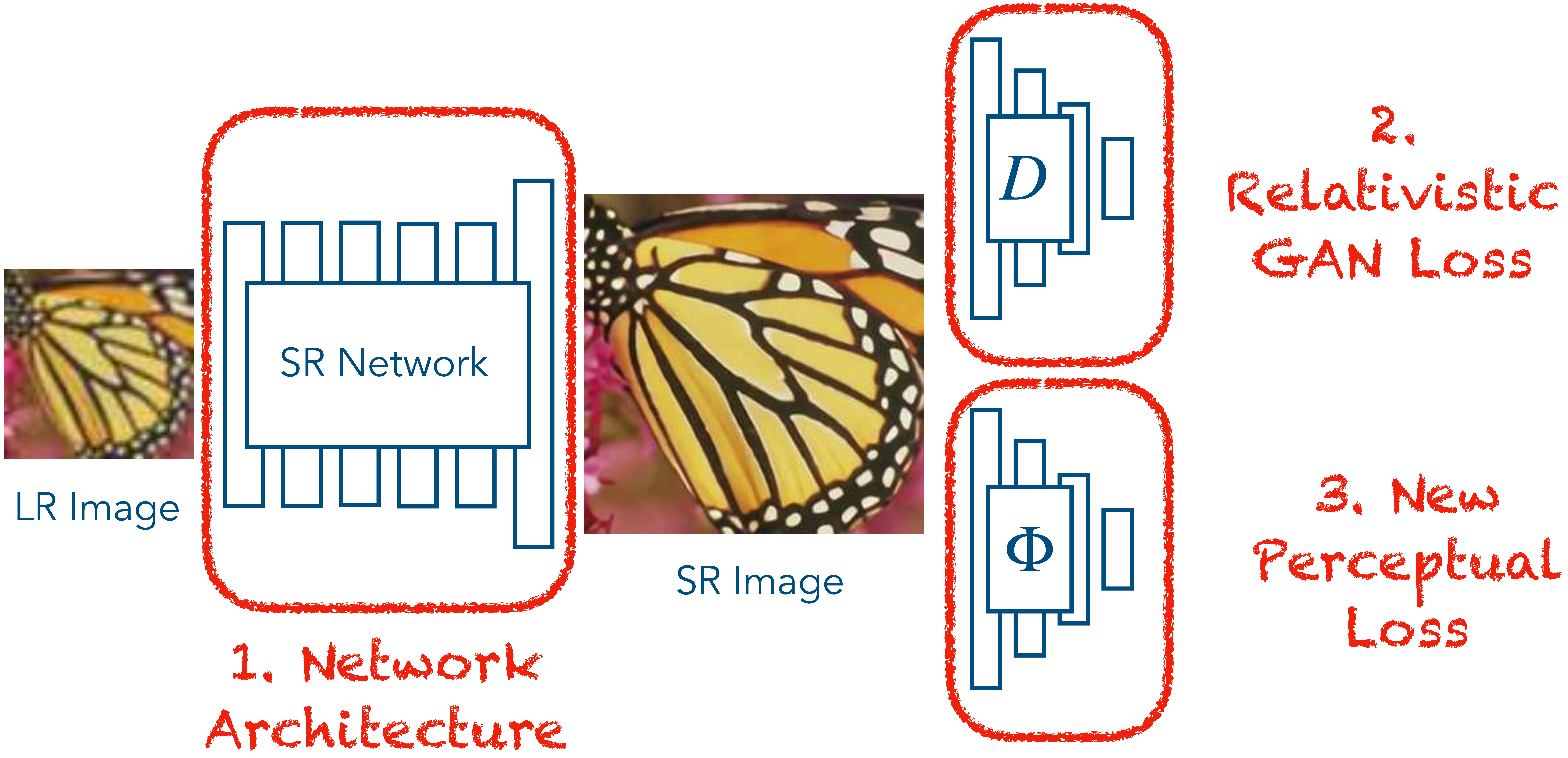
[Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. Generative adversarial nets. NIPS]

Super-Resolution Generative Adversarial Networks



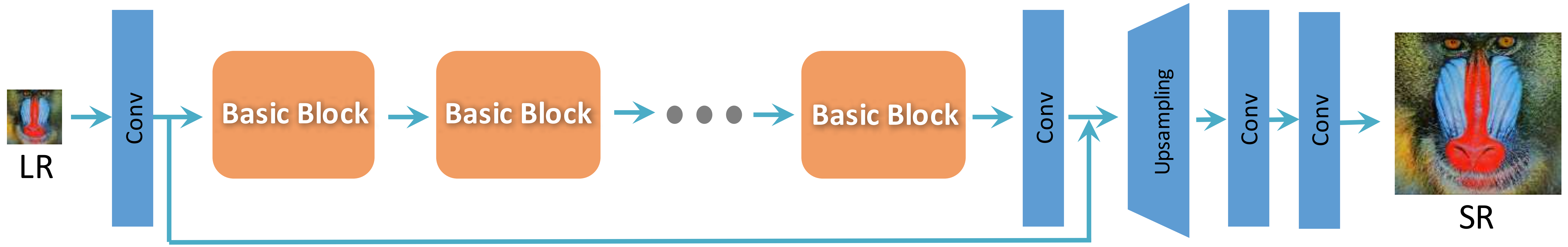
[Ledig C, Theis L, Huszár F, et al. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, CVPR. 2017.]

Enhance SRGAN from three aspects



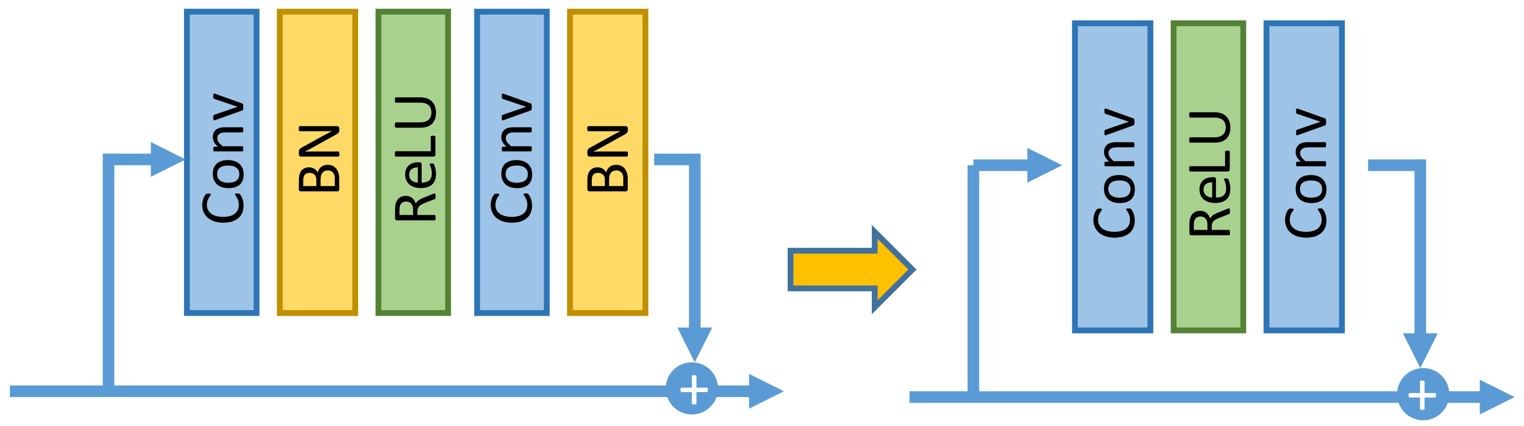
1. New Network Architecture

We employ the high-level architecture of SRGAN.



Replacing the basic block (Residual Block) with the Residual-in-Residual Dense Block (RRDB).

Residual Block (RB)

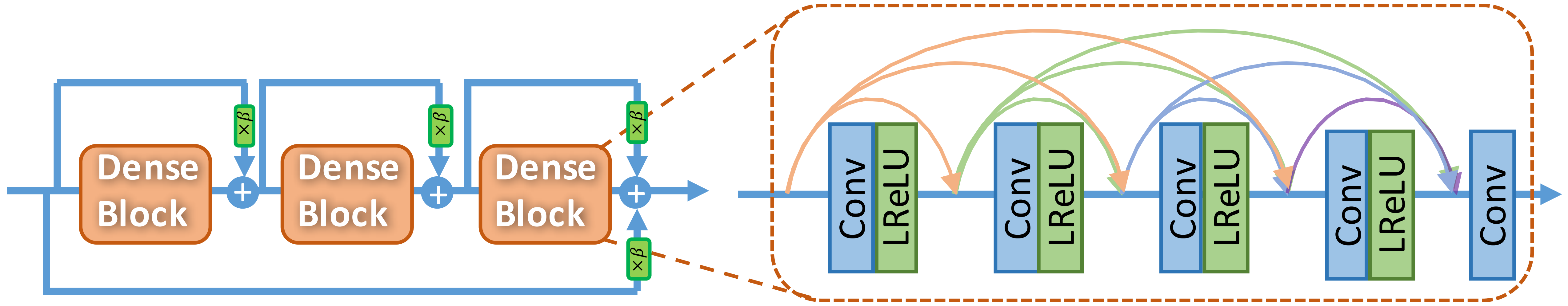


SRGAN

RB w/o BN

~~BN~~

Residual in Residual Dense Block (RRDB)



2. Relativistic GAN Loss

We enhance the discriminator based on the Relativistic GAN

Real?

$$D(x_r) = \sigma(C(\text{Real})) \to 1$$

$$D(x_f) = \sigma(C(\text{Fake})) \to 0$$

Fake?

a) Standard GAN

More realistic than fake data?

$$D_{Ra}(x_r, x_f) = \sigma(C(\text{Real}) - \mathbb{E}[C(\text{Fake})]) \to 1$$

$$D_{Ra}(x_f, x_r) = \sigma(C(\text{Fake}) - \mathbb{E}[C(\text{Real})]) \to 0$$

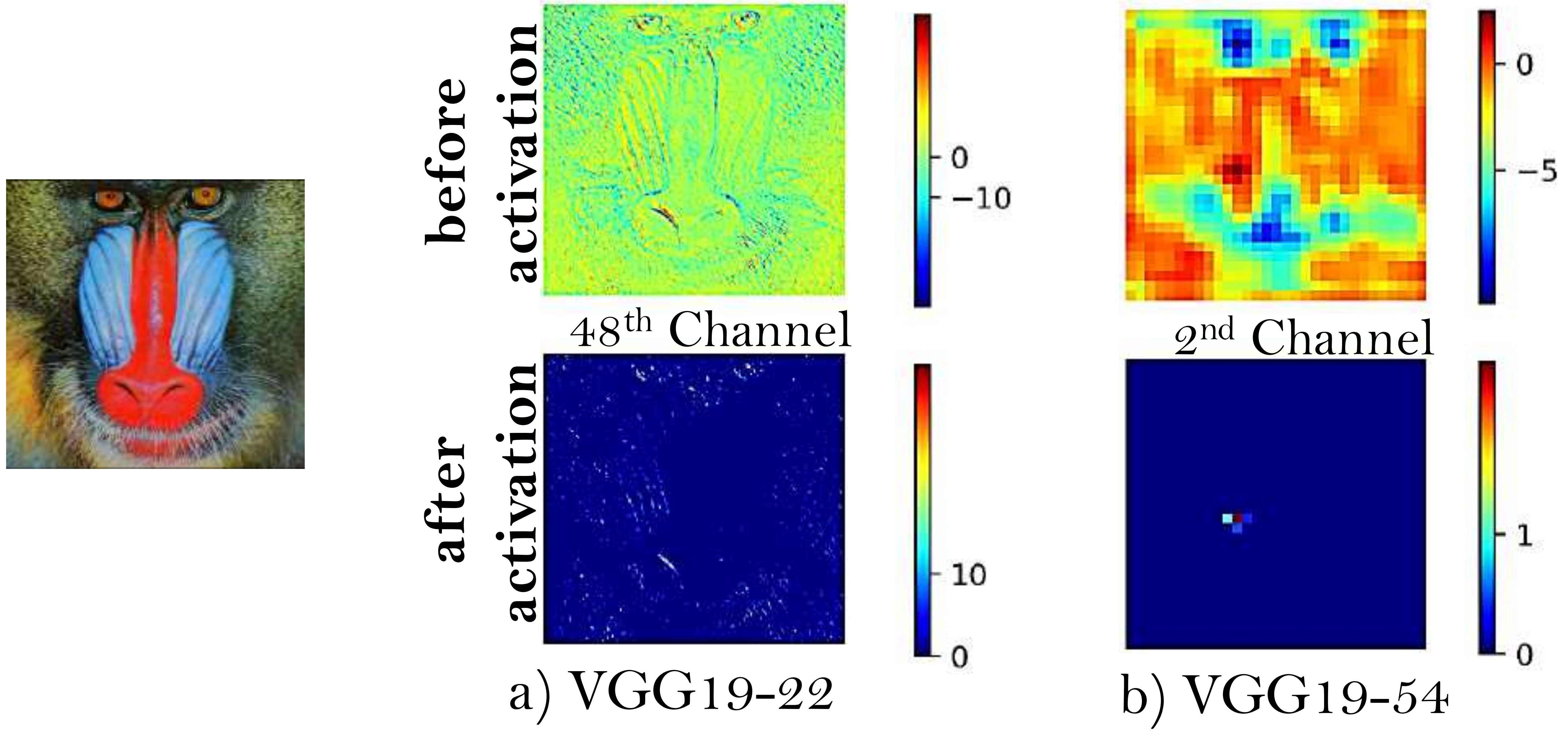
Less realistic than real data?

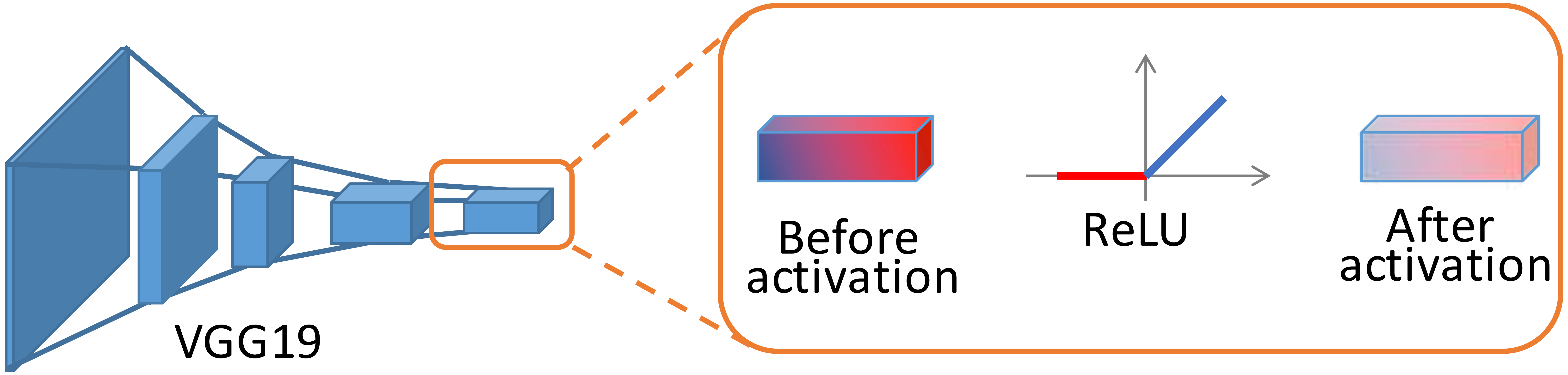
b) Relativistic GAN

[Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734, 2018.]

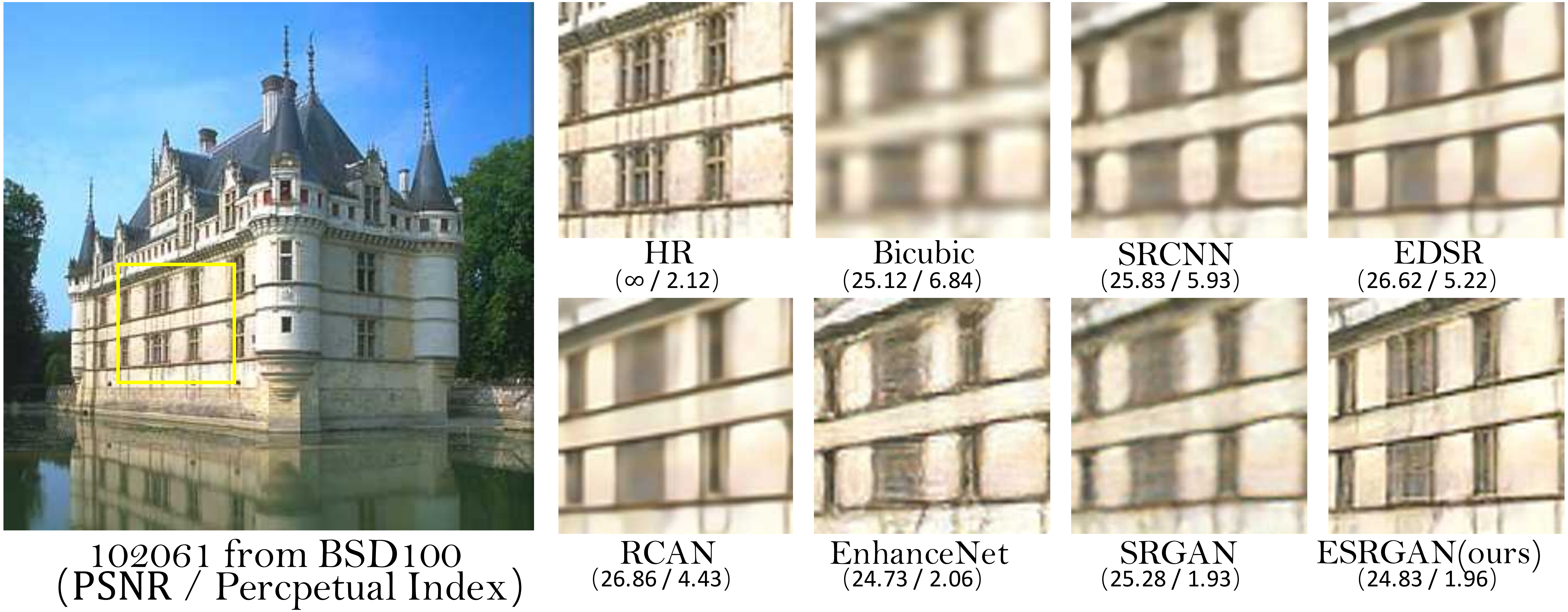
3. Enhanced Perceptual Loss

We develop a more effective perceptual loss by constraining on features before activation

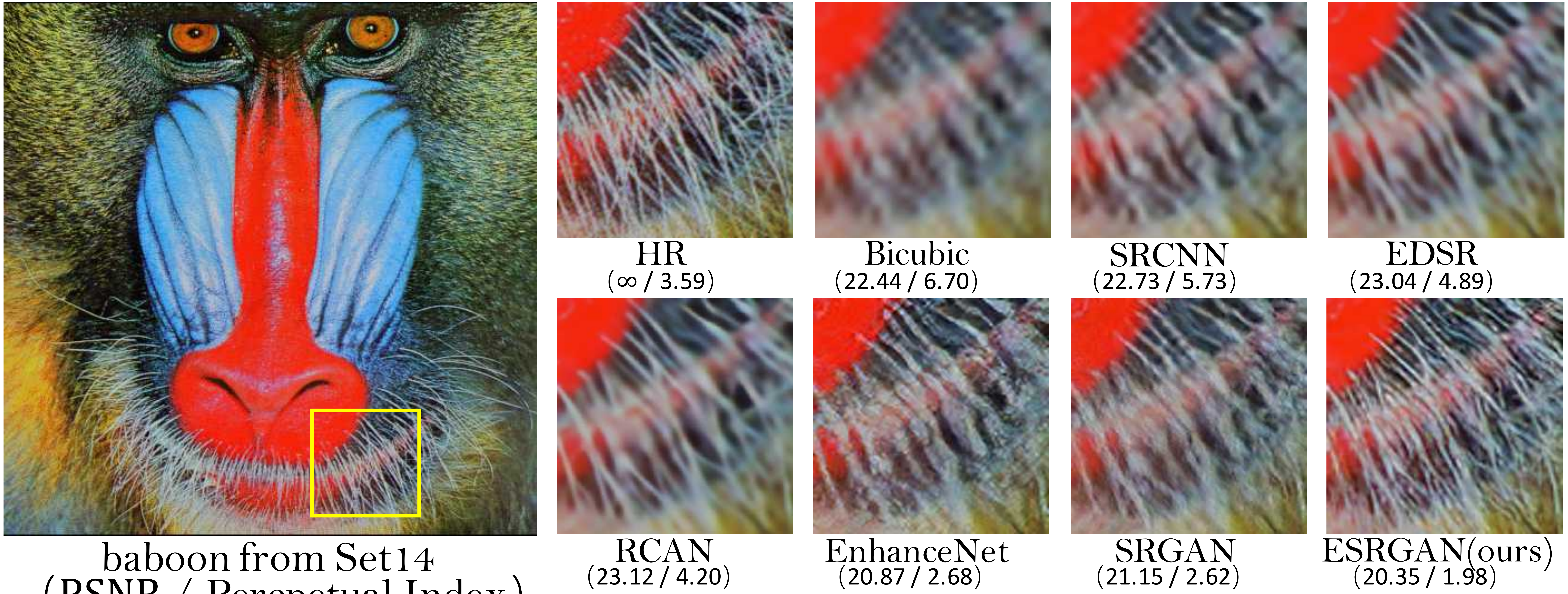




Perceptual SR Comparison

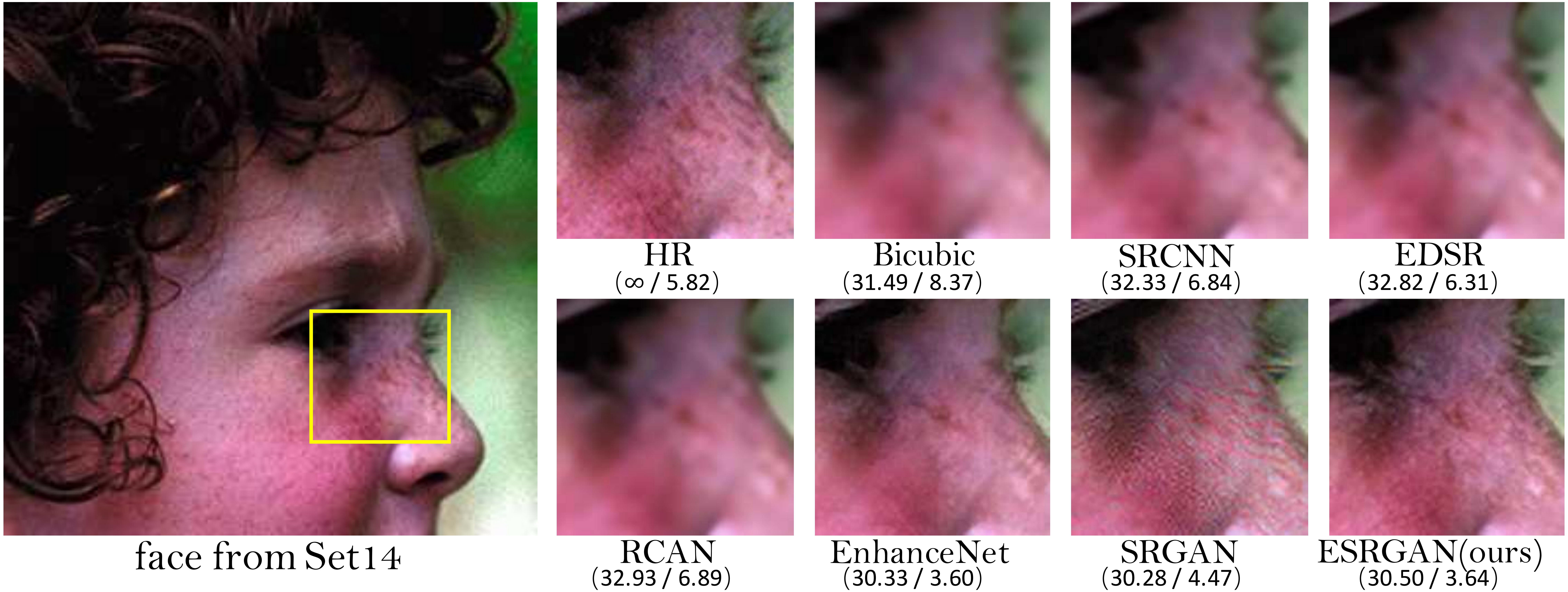


Perceptual SR Comparison



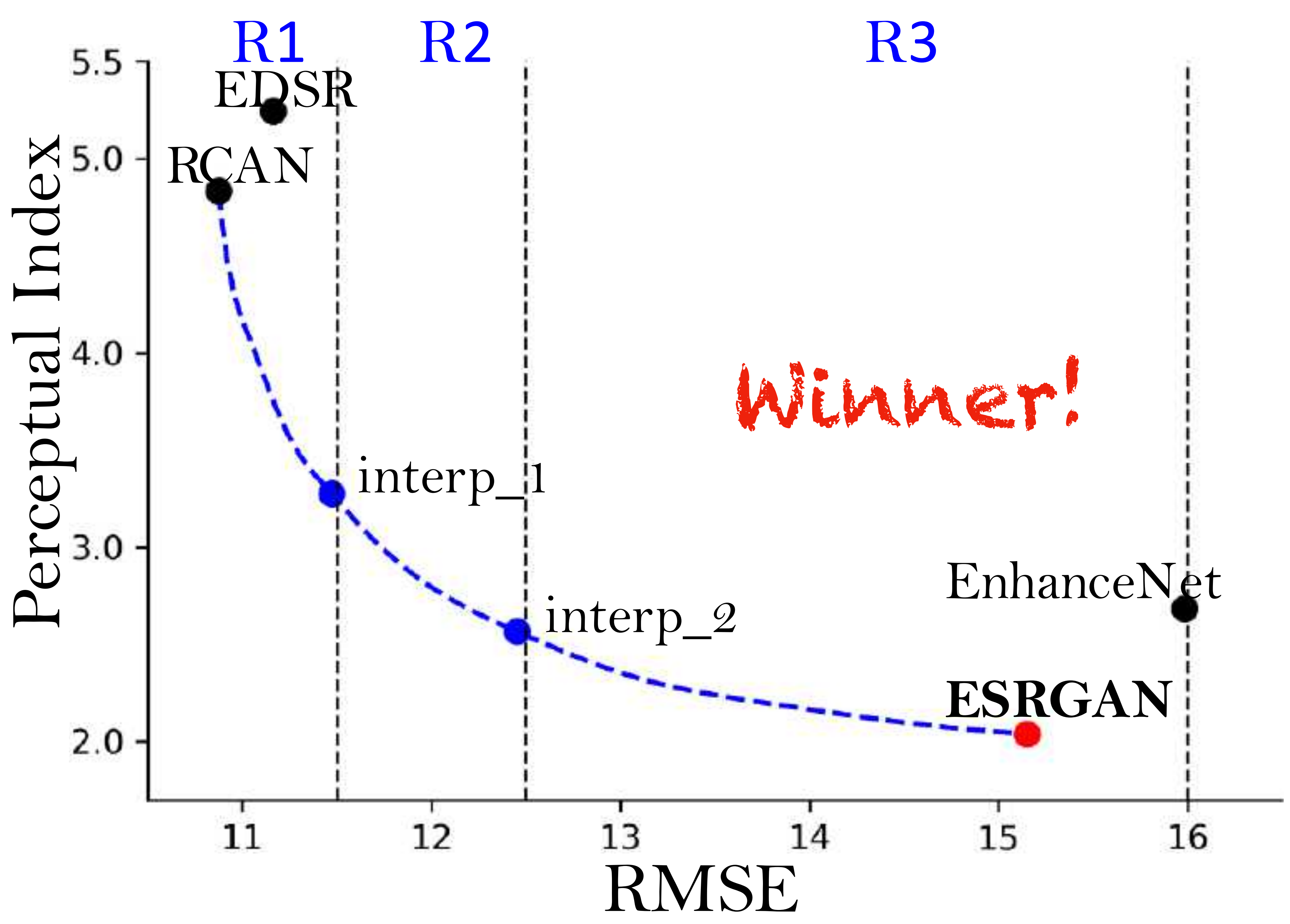
baboon from Set14
(PSNR / Percpetual Index)

Perceptual SR Comparison



PIRM Perceptual SR Challenge (ECCV 2018)

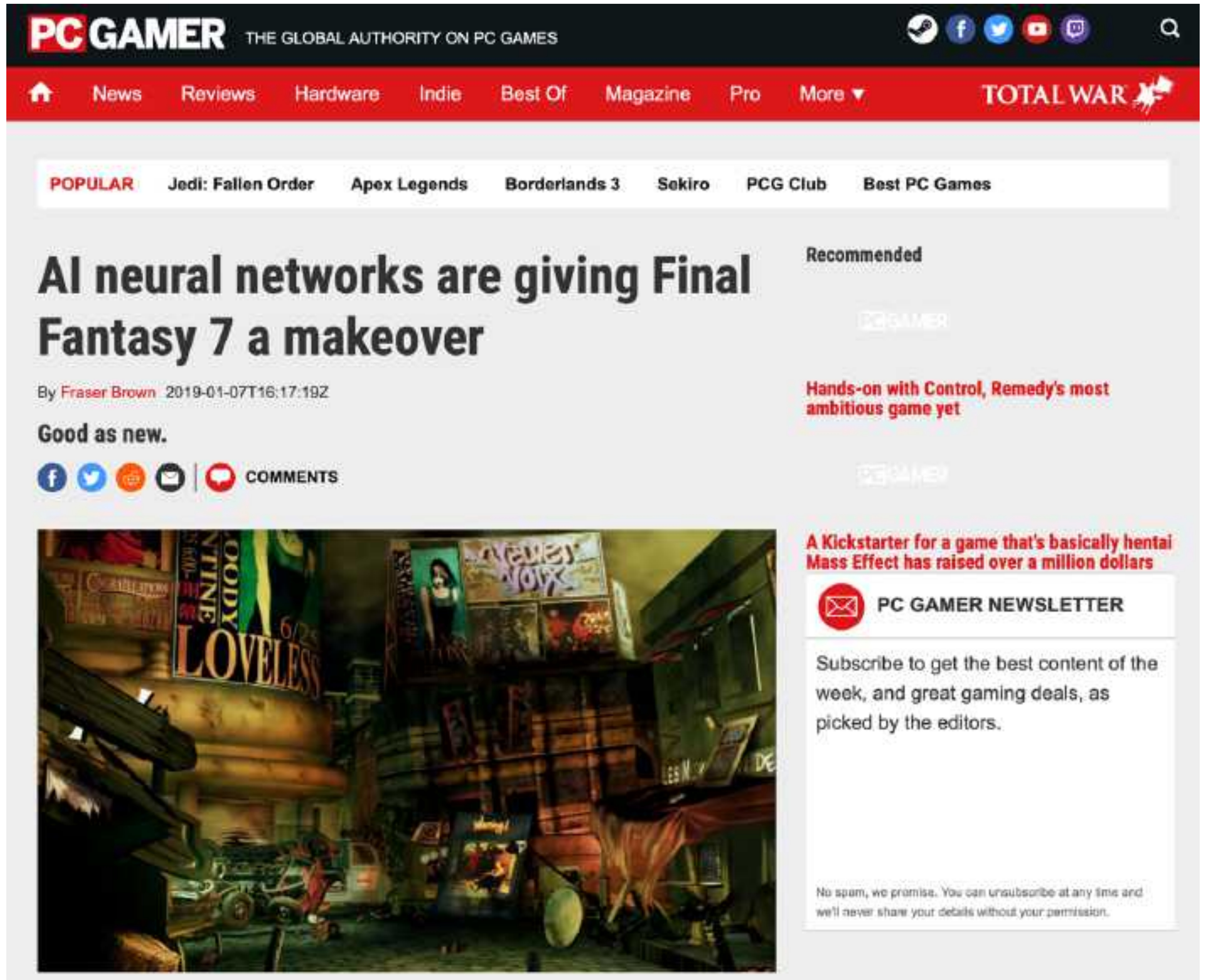
- ESRGAN with 16 residual blocks
- MINC loss for material recognition as a variant of perceptual loss
- Pristine dataset, which is used for learning the perceptual index
- Back projection as post-processing
- Use image interpolation for a lower perceptual index



Perception-distortion plane on PIRM self validation dataset

Apply ESRGAN in Game Remaster.

ESRGAN has been used to remaster various classic games. See DSOGaming, PC Gamer, Extreme Tech, and Synced for media coverage.



PC GAMER THE GLOBAL AUTHORITY ON PC GAMES

News Reviews Hardware Indie Best Of Magazine Pro More ▾ TOTAL WAR

POPULAR Jedi: Fallen Order Apex Legends Borderlands 3 Sekiro PCG Club Best PC Games

AI neural networks are giving Final Fantasy 7 a makeover

By Fraser Brown 2019-01-07T16:17:19Z

Good as new.

COMMENTS

Recommended

Hands-on with Control, Remedy's most ambitious game yet

A Kickstarter for a game that's basically hentai Mass Effect has raised over a million dollars

PC GAMER NEWSLETTER

Subscribe to get the best content of the week, and great gaming deals, as picked by the editors.

No spam, we promise. You can unsubscribe at any time and we'll never share your details without your permission.

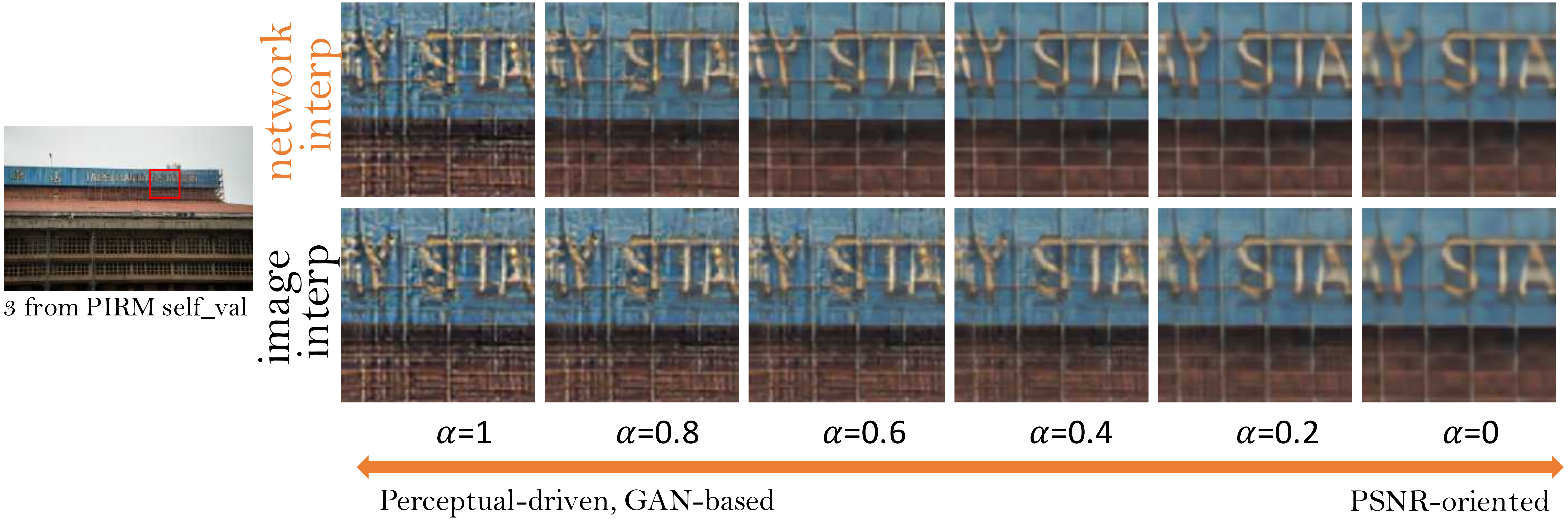
Adjust GAN Effects? Network Interpolation

Interpolates all the corresponding parameters of the PSNR-oriented network and the fine-tuned GAN-based network.

interpolation parameter GAN-based network by fine-tuning

$$\theta_G^{\mathrm{INTERP}} = \alpha\, \boxed{\theta_G^{\mathrm{PSNR}}} + (1-\alpha)\, \boxed{\theta_G^{\mathrm{GAN}}}$$

PSNR-oriented network

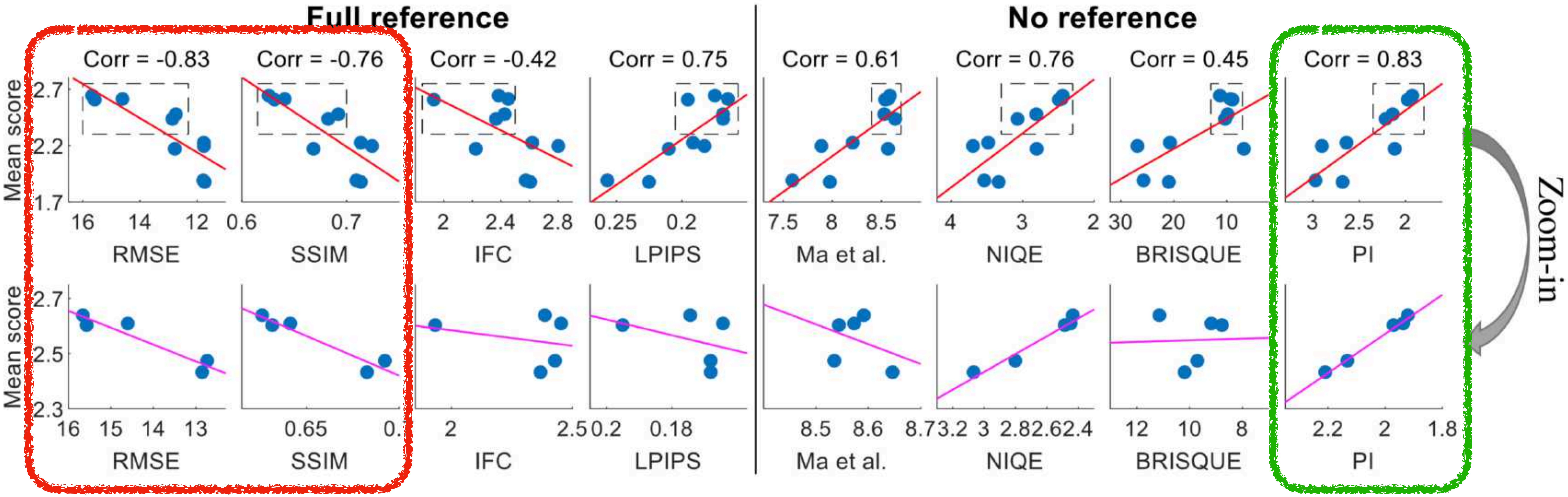


Network Interpolation for Multiple Applications

[Wang, X., Yu, K., Dong, C., Tang, X., & Loy, C. C. Deep network interpolation for continuous imagery effect transition. CVPR 2019]

New Quality Assessment Method In ESRGAN

IQA Methods are Important for Image Restoration



[Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., & Zelnik-Manor, L. (2018). The 2018 PIRM challenge on perceptual image super-resolution. ECCVW 2018.]

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).

Rethink the Gap Between Metric and Human Judgment

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

Before 2018, Evaluation Using PSNR/SSIM

Ground Truth

PSNR / SSIM

23.53 / 0.7056

Good in PSNR and SSIM

19.86 / 0.5530

Preferred by Human

Rethink the Gap Between Metric and Human Judgment

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

After 2018, Evaluation Using PI/NIQE

Ground Truth

Perceptual Index / NIQE

3.80 / **6.47**

Good in PI and NIQE

4.03 / 6.90

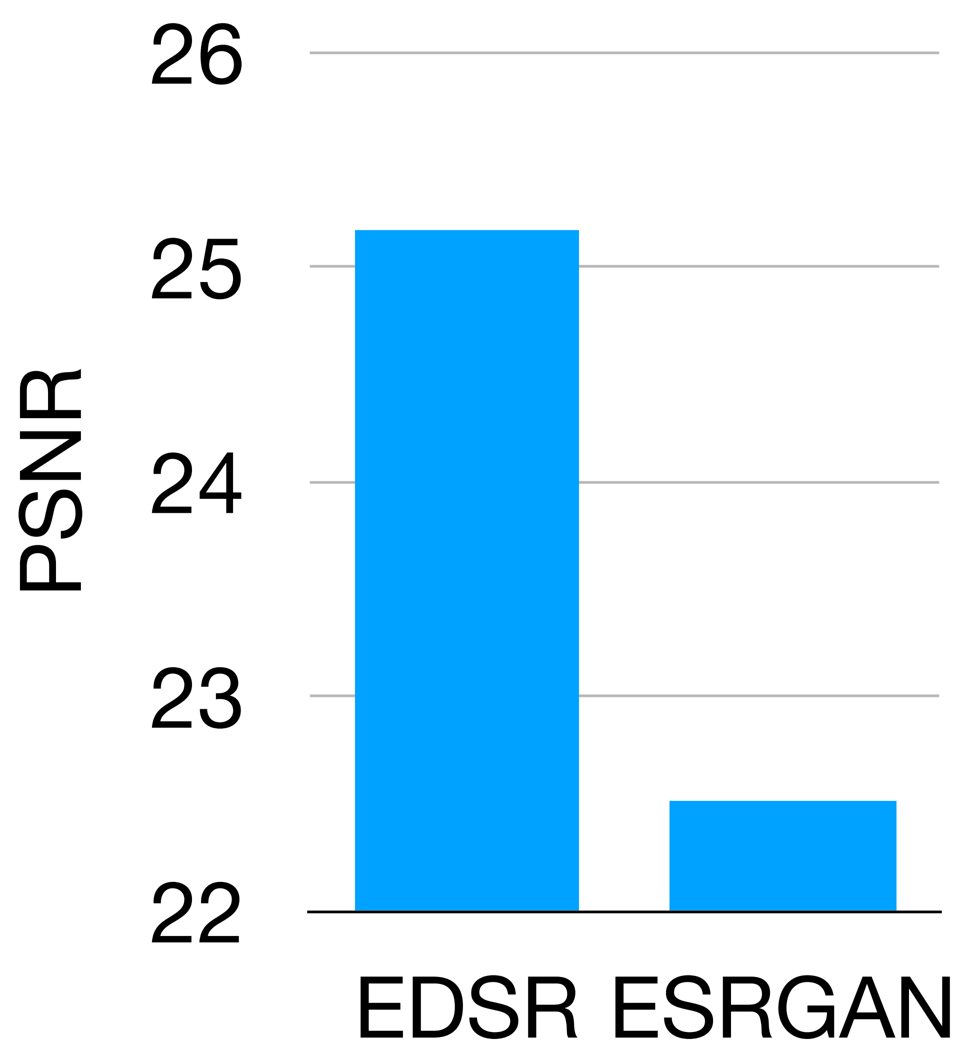
Preferred by Human

Rethink the Gap Between Metric and Human Judgment

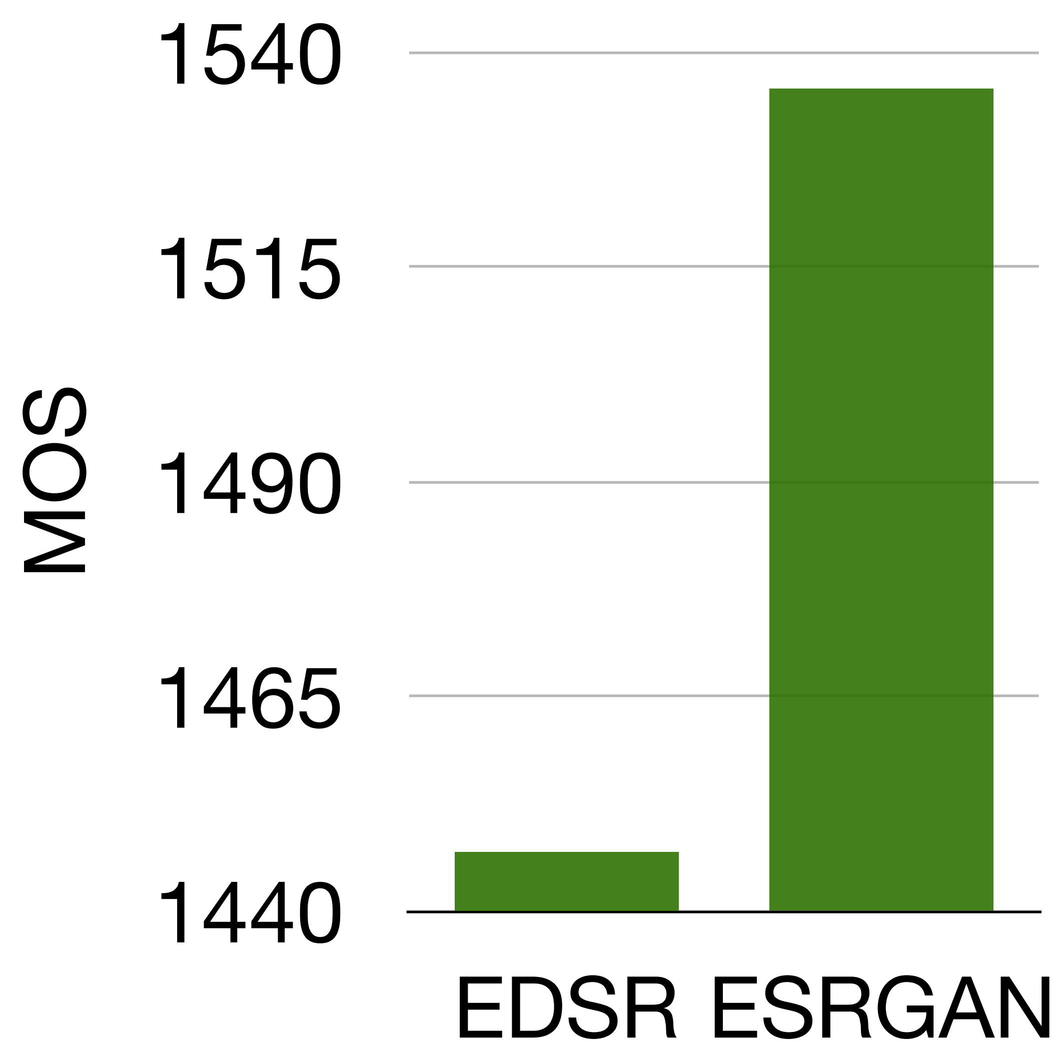
Increasing inconsistency between high numerical performances (PSNR, SSIM, PI, etc.) and perceptual performance.

Method	Year	PSNR	PI	NIQE	MOS
YY [43]	2013	22.78	7.94	10.27	1367.23
TSG [33]	2013	23.01	7.77	9.67	1343.02
A+ [34]	2014	23.29	7.84	9.94	1371.14
SRCNN [9]	2014	23.39	7.41	9.16	1360.13
FSRCNN [10]	2016	23.52	7.76	9.98	1373.59
VDSR [15]	2016	23.59	8.37	11.32	1382.74
EDSR [19]	2017	**24.63**	6.82	9.11	1453.90
SRGAN [18]	2017	22.02	3.92	**5.85**	1501.76
RCAN [51]	2018	**24.66**	6.82	9.26	1453.35
ESRGAN [36]	2018	21.98	4.53	7.11	**1535.67**
BOE [23]	2018	22.29	**3.88**	5.87	1484.29
RankSRGAN [50]	2019	21.61	**3.66**	**5.64**	**1523.97**

Compare EDSR and ESRGAN



Higher is better



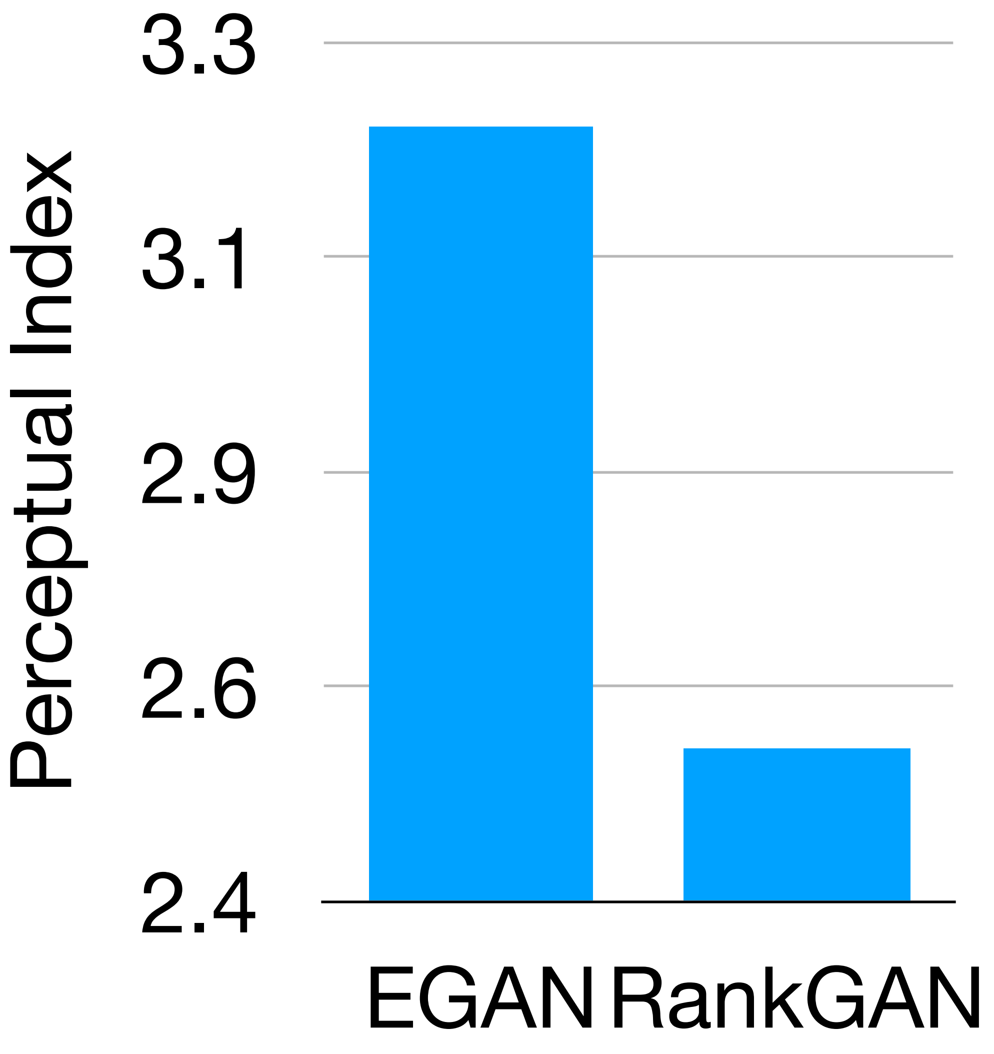
Higher is better

Rethink the Gap Between Metric and Human Judgment

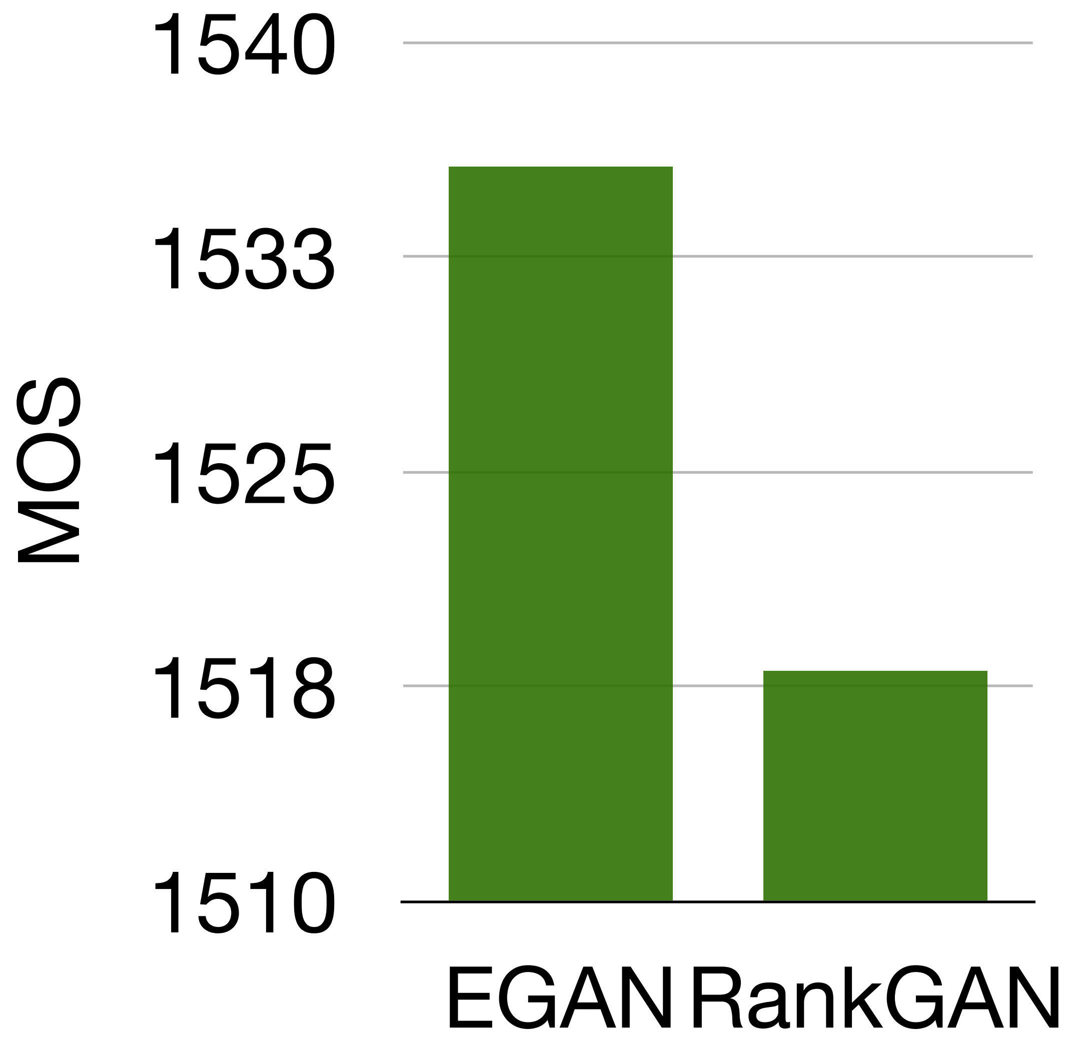
Increasing inconsistency between high numerical performances (PSNR, SSIM, PI, etc.) and perceptual performance.

Method	Year	PSNR	PI	NIQE	MOS
YY [43]	2013	22.78	7.94	10.27	1367.23
TSG [33]	2013	23.01	7.77	9.67	1343.02
A+ [34]	2014	23.29	7.84	9.94	1371.14
SRCNN [9]	2014	23.39	7.41	9.16	1360.13
FSRCNN [10]	2016	23.52	7.76	9.98	1373.59
VDSR [15]	2016	23.59	8.37	11.32	1382.74
EDSR [19]	2017	**24.63**	6.82	9.11	1453.90
SRGAN [18]	2017	22.02	3.92	**5.85**	1501.76
RCAN [51]	2018	**24.66**	6.82	9.26	1453.35
ESRGAN [36]	2018	21.98	4.53	7.11	**1535.67**
BOE [23]	2018	22.29	**3.88**	5.87	1484.29
RankSRGAN [50]	2019	21.61	**3.66**	**5.64**	**1523.97**

Compare ESRGAN and RankSRGAN

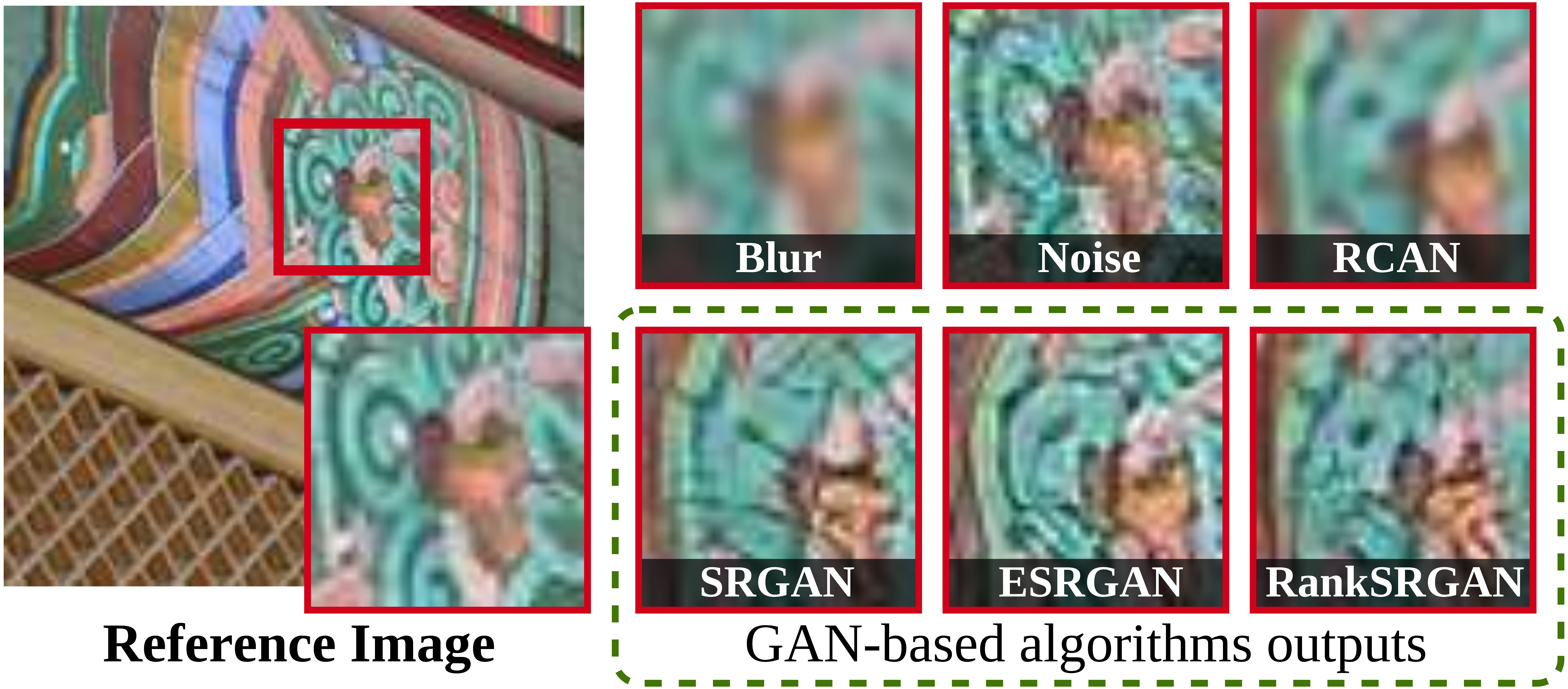


Lower is better



Higher is better

GAN-based Methods Bring Challenges to IQA



Why We Need a New Large-Scale IQA Dataset

- A good dataset will be provided for the IQA community to develop and test new metric.
- It provides new insights for the image restoration community on how to better evaluate new algorithms and what kind of restored image are visual friendly for human.
- With our new algorithm, the real collected MOS score can be used directly to train a better image restoration algorithm.

PIPAL: Large-Scale IQA Dataset for Perceptual Image Restoration

Jinjin Gu, Haoming Cai, Haoyu Chen, Xiaoxing Ye, Jimmy S. Ren, Chao Dong, In Progress

PIPAL: Perceptual Image Processing ALgorithms

250
Reference Images

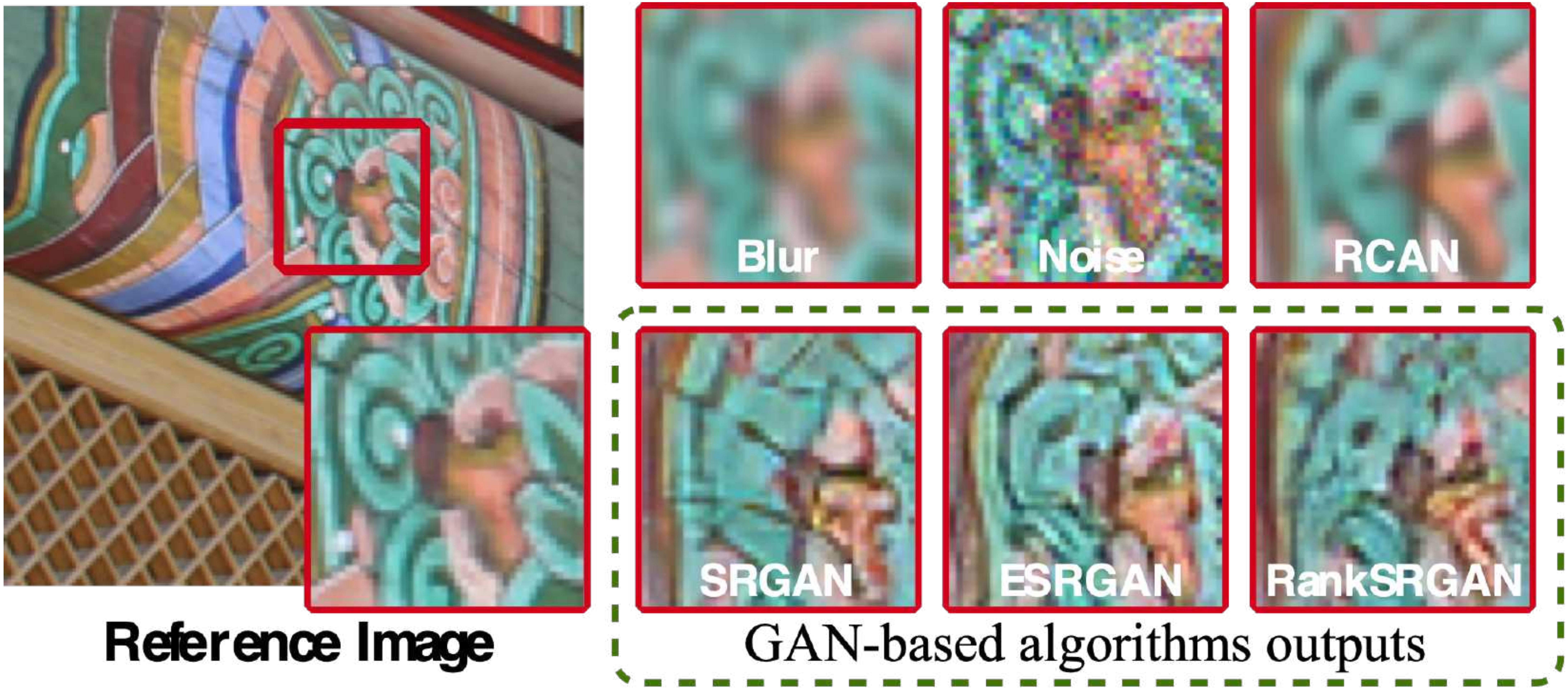
40
Distortion Types

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

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

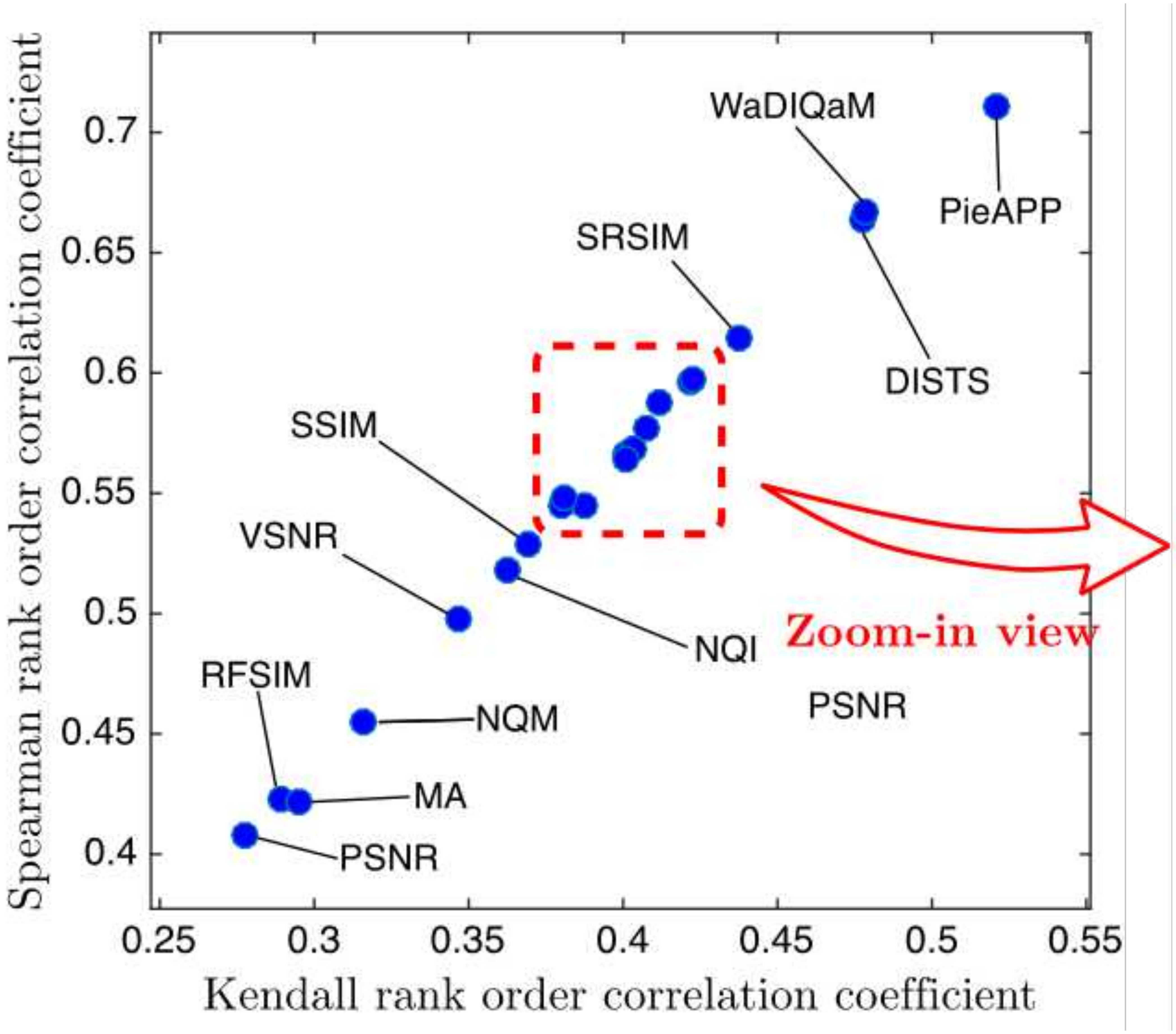
$$R_B^{new} = R_B + K(S_B - P_{B>A})$$

- 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

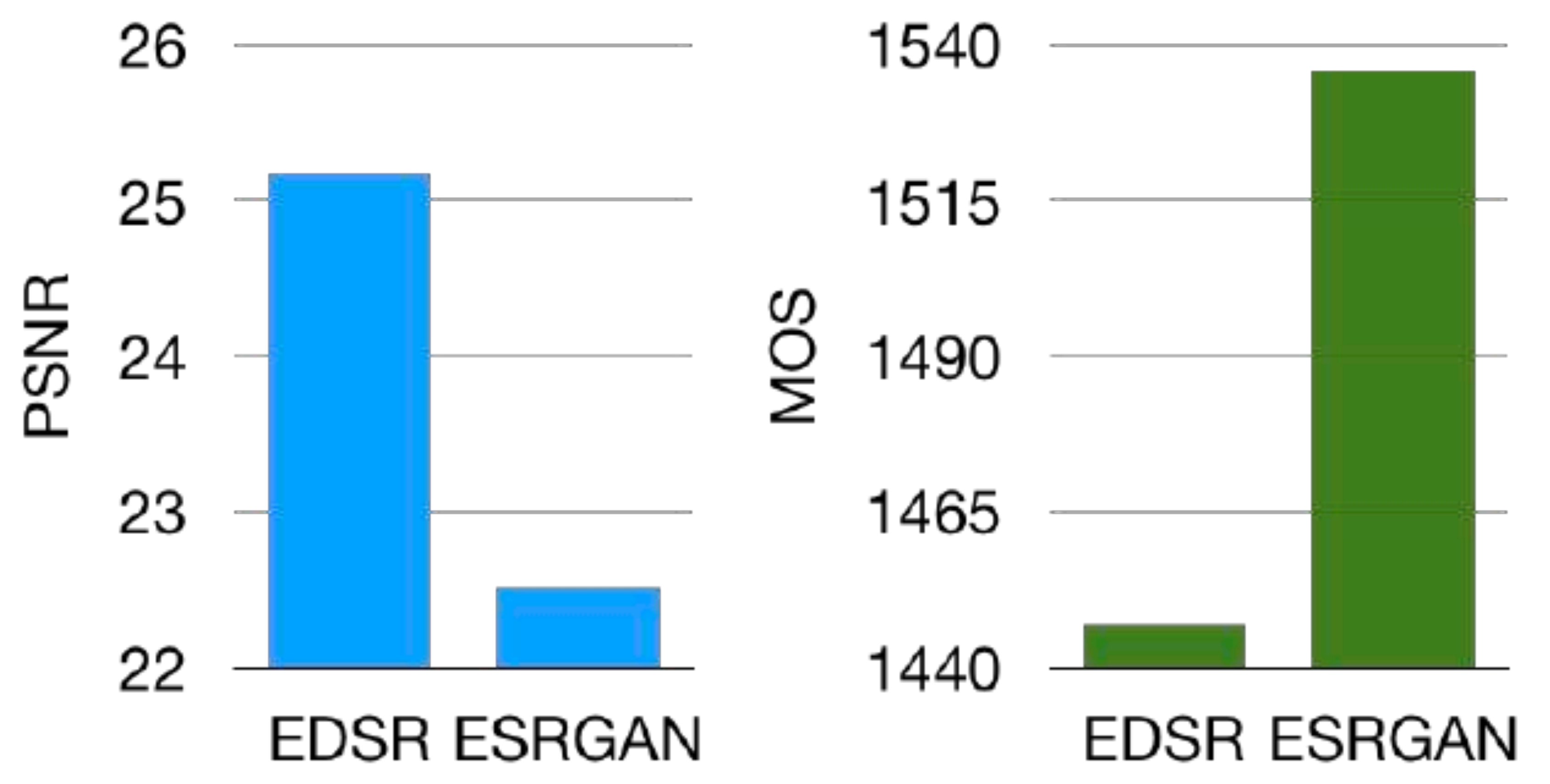
Method	Year	PSNR ↑	NIQE ↓	PI ↓	MOS ↑
YY [52]	2013	23.35^{8}	6.4174^{8}	5.9344^{7}	1367.71^{8}
TSG [45]	2013	23.55^{7}	6.4163^{7}	6.1433^{10}	1387.24^{7}
A+ [46]	2014	23.82^{6}	6.3645^{5}	5.9897^{9}	1354.52^{12}
SRCNN [12]	2014	23.93^{5}	6.5657^{10}	5.9781^{8}	1363.68^{11}
FSRCNN [13]	2016	24.07^{4}	6.9985^{11}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

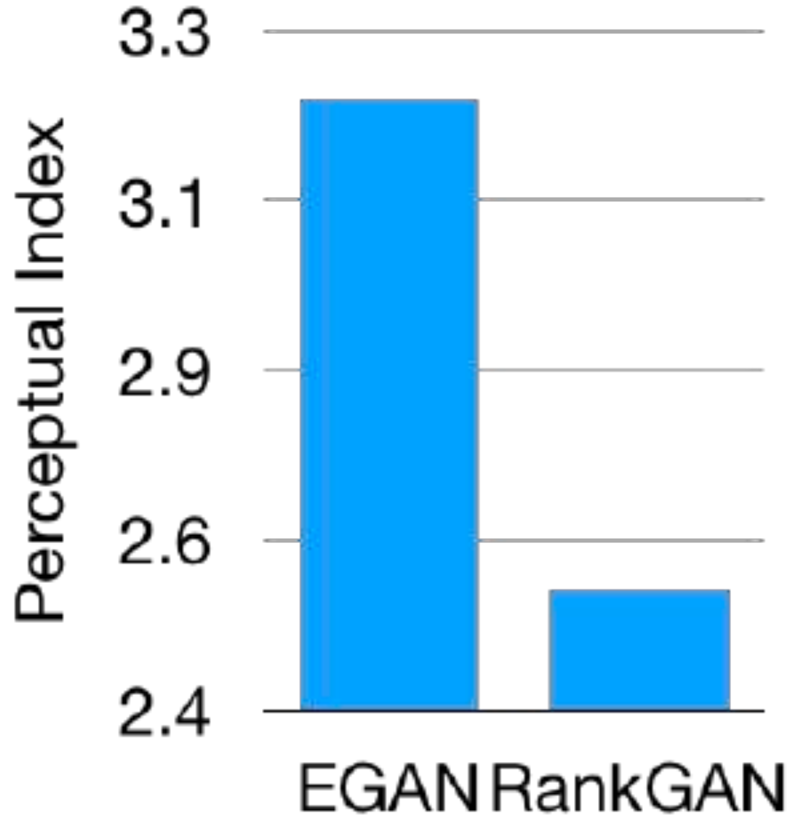


Higher is better **Higher is better**

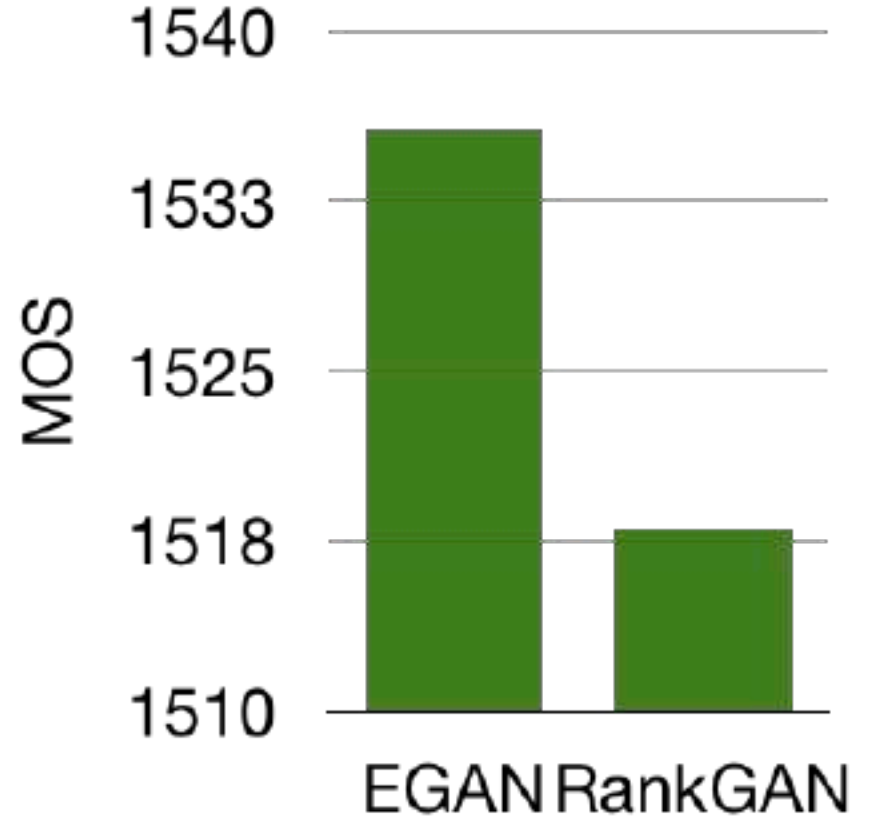
Benchmarking Image Restoration Algorithms

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	**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	**25.21**^{1}	6.4121^{6}	5.2430^{5}	1455.31^{5}
BOE [32]	2018	22.68^{9}	**3.7945**^{1}	**2.6368**^{2}	1481.51^{4}
ESRGAN [47]	2018	22.51^{11}	4.7821^{4}	3.2198^{4}	**1534.25**^{1}
RankSRGAN [61]	2019	22.11^{12}	**3.8155**^{2}	**2.5636**^{1}	**1518.29**^{2}

Compare ESRGAN and RankSRGAN



Lower is better



Higher is better

Benchmarking Super-Resolution Algorithms

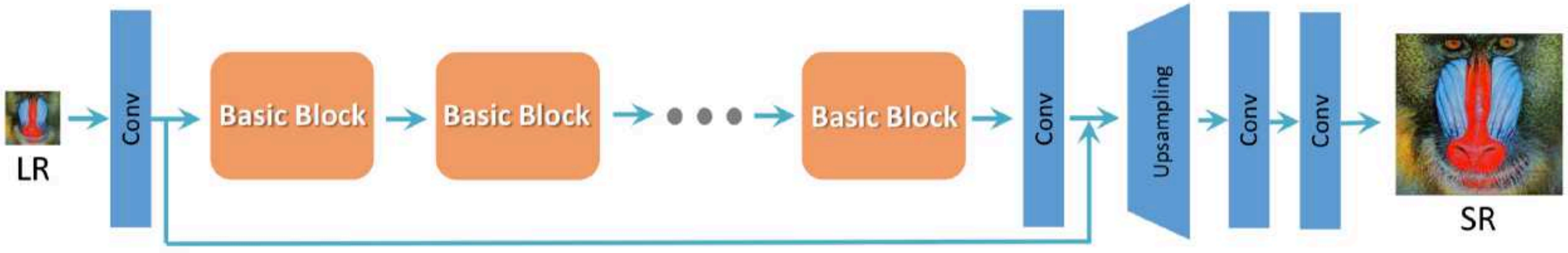
- None of existing IQA methods is always effective in evaluation.
- Excessively optimizing performance on a specific IQA may cause a decrease in perceptual quality

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.

Future Work On Image and Video Restoration

Analyzing deep neural networks for image processing



What does a net **do**?

How does it **do** it?

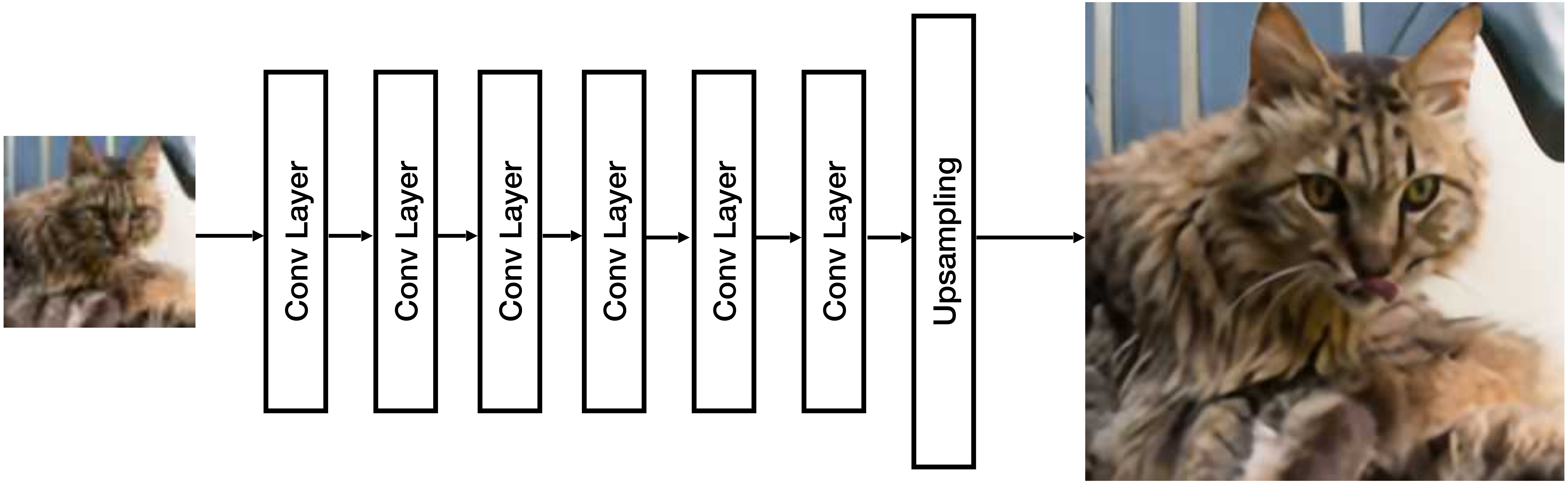
How does it **learn** it?

Interpreting Super-Resolution Networks with Local Attribution Maps

Jinjin Gu, Chao Dong

The University of Sydney,
Shenzhen Institutes of Advanced Technology

Super-Resolution Networks



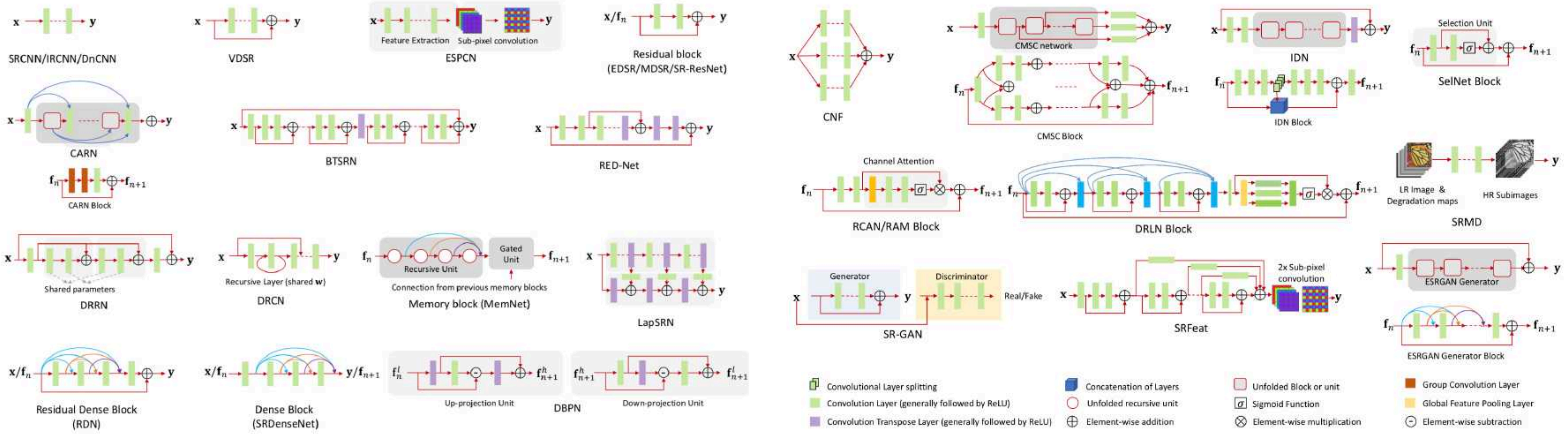
SR networks build up of convolutional layers and upsampling blocks, with parameter θ.

SR networks are trained using thousands of image pairs.

Super-Resolution Networks

Many SR network architectures have been proposed.

What makes their different performance?



[Anwar, S., Khan, S., & Barnes, N. (2019). A Deep Journey into Super-resolution: A survey. arXiv preprint arXiv:1904.07523.]

Information usage in SR networks

In the past, we only have one metric to study SR networks:

The Performance

Add module A,
seems good

Add module B,
seems good

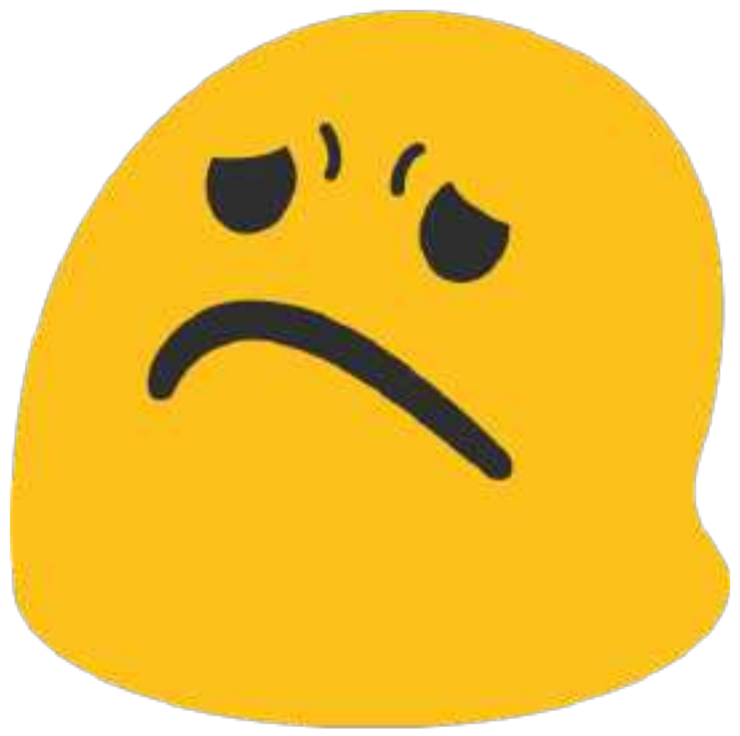
Add module C,
not good

SR networks are still mysterious

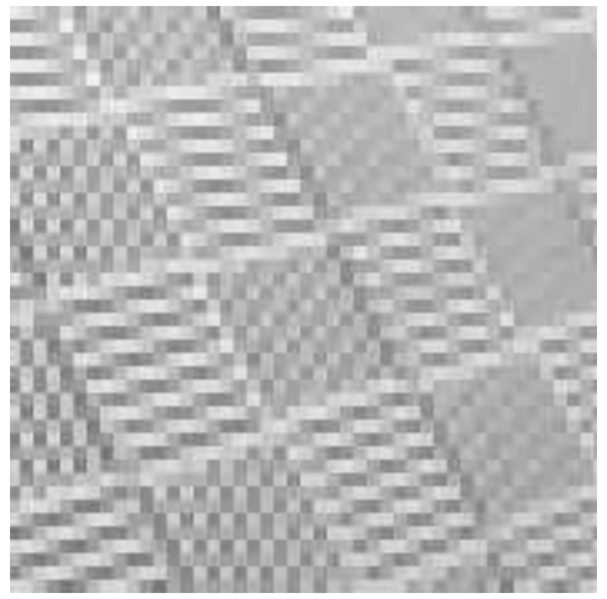
Have you met these scenarios?

- Do you need multi-scale architecture or a larger receptive field?
- Does non-local attention module work as you want?
- Why different SR networks perform differently?

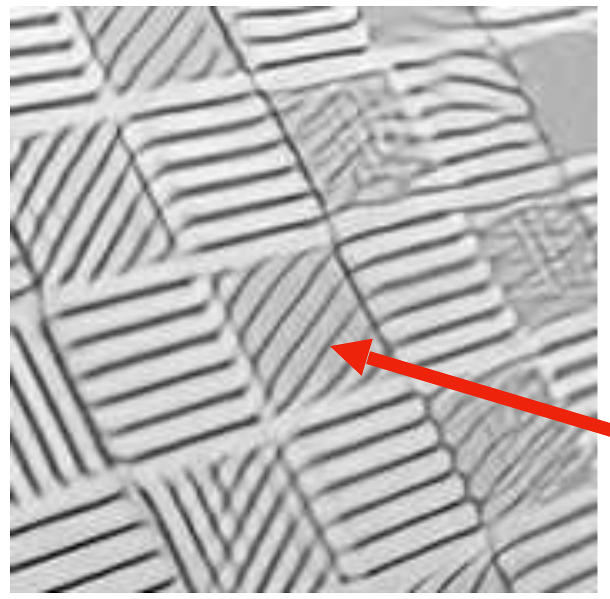
We lack understanding toward these questions
And also research tools



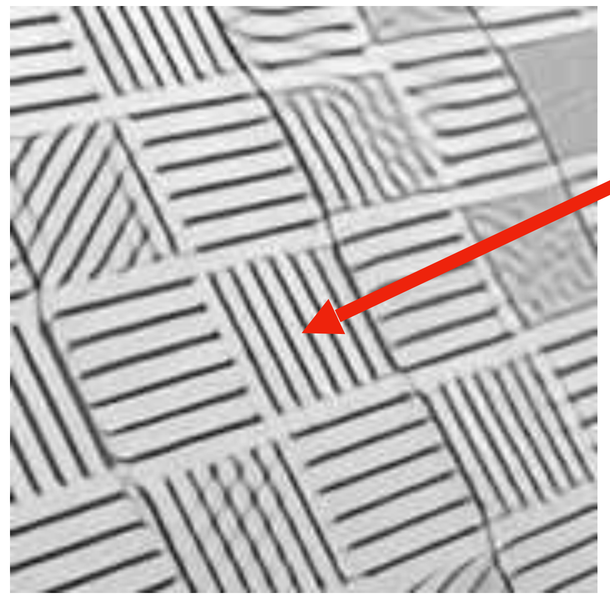
Attribution Analysis



Input image



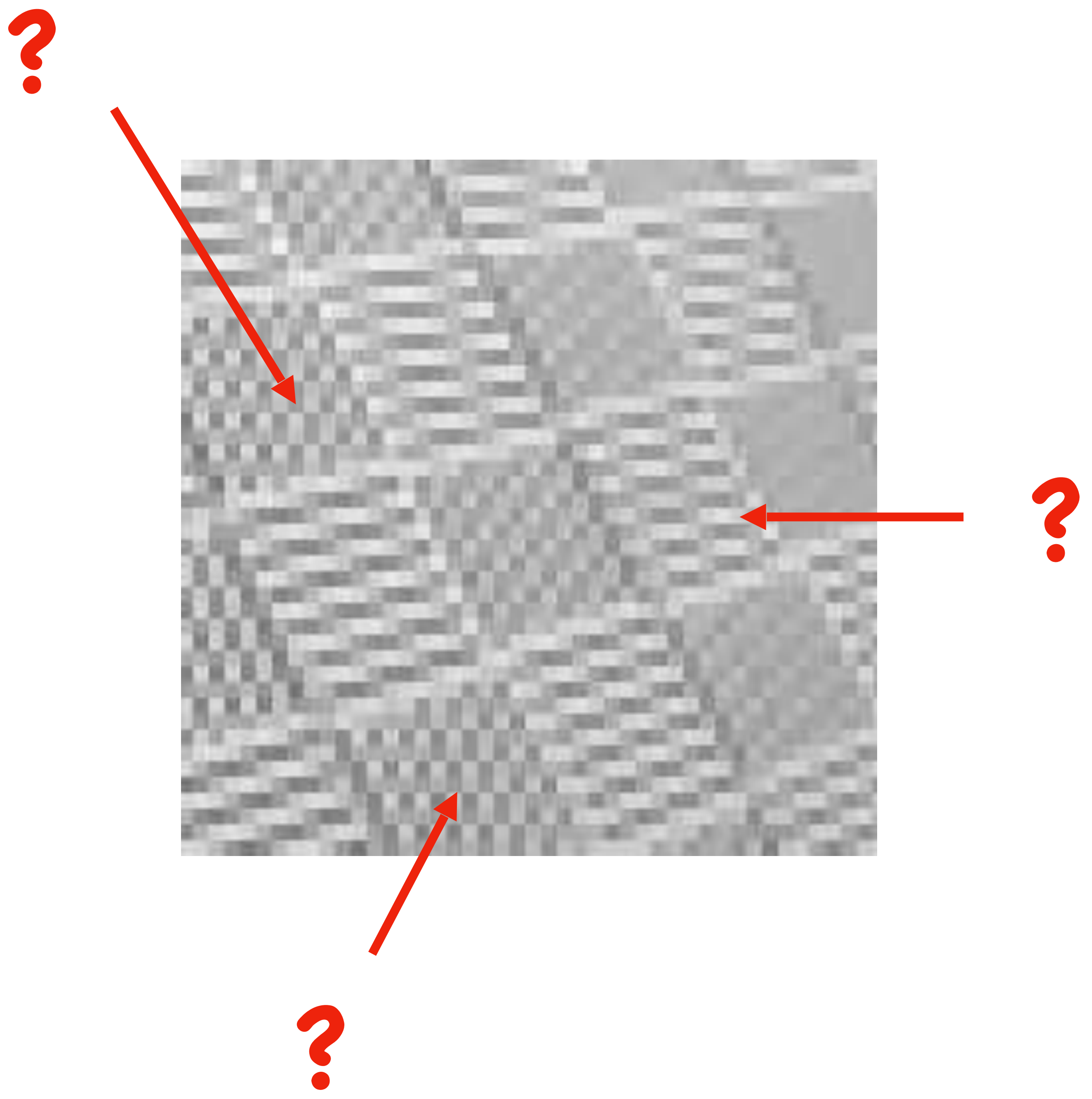
EDSR



RNAN

Why RNAN gives correct results in the center?

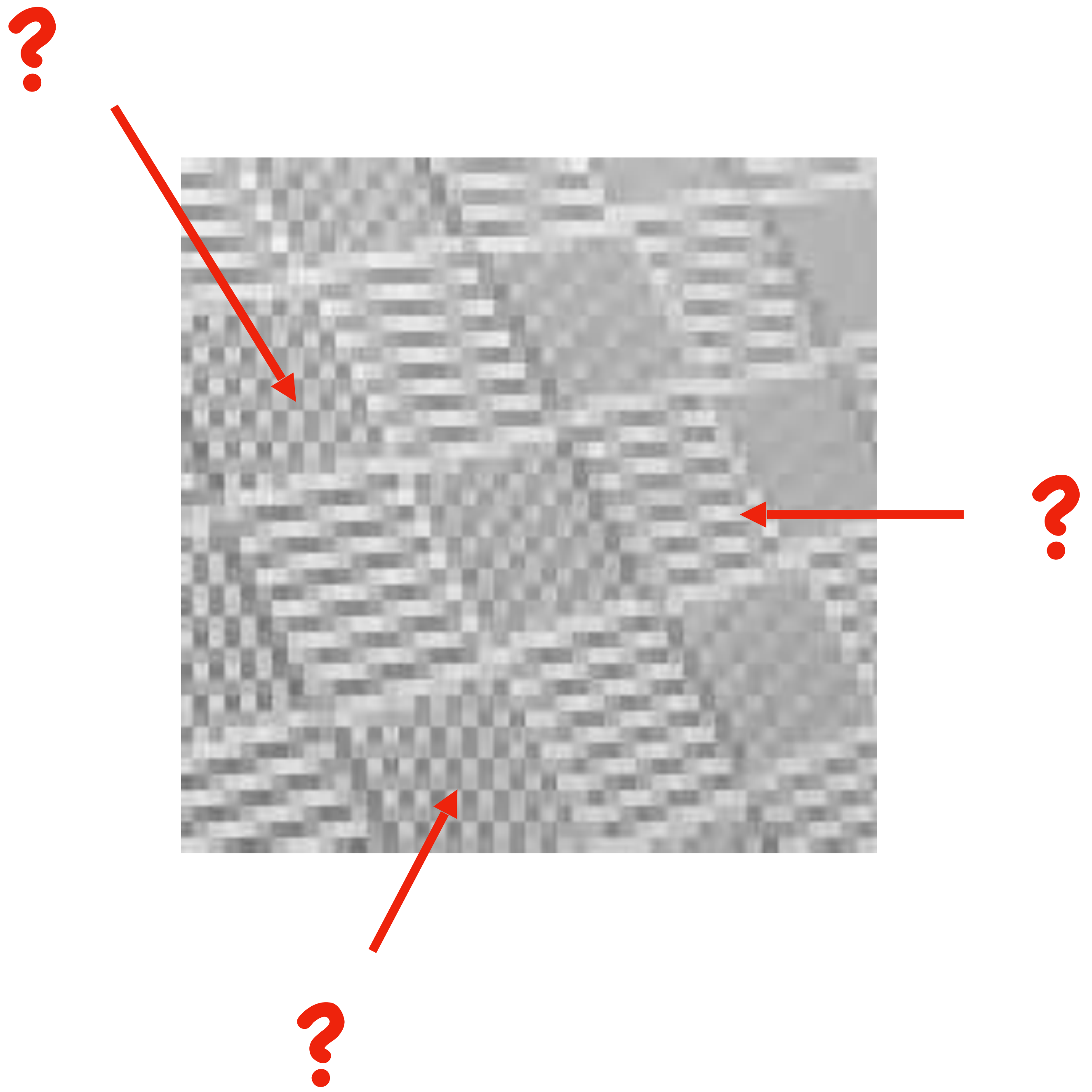
Attribution Analysis



What did RNAN notice from the input that allowed it to make the correct prediction?

Does EDSR notice this information?

Attribution Analysis



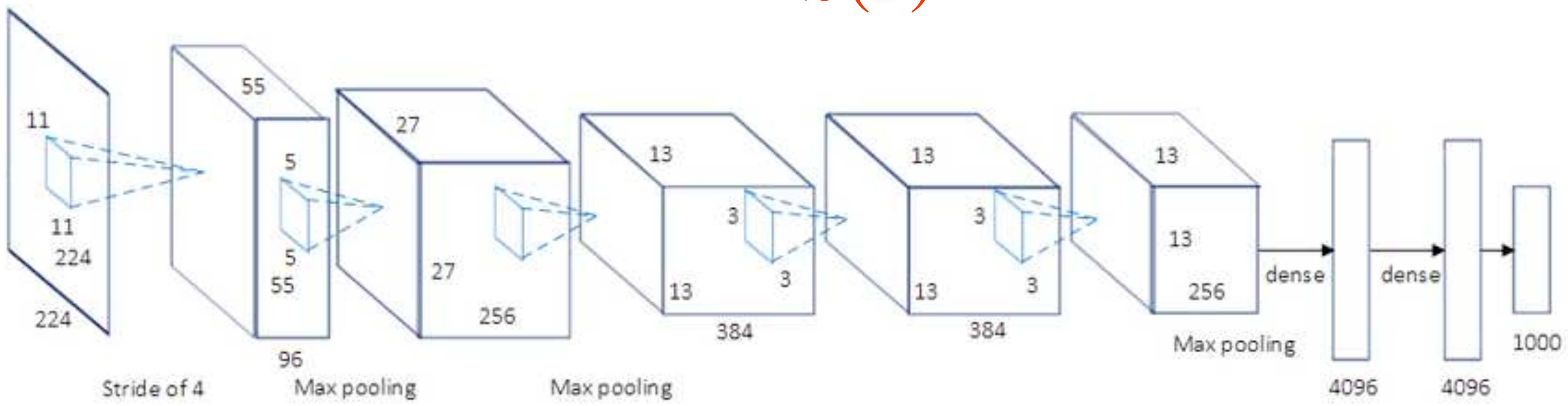
Identify input features responsible for SR results.

Attribution Analysis for High-level Networks

What is S(I) looking at?

$S(I)$

I



98% house finch

10% bird

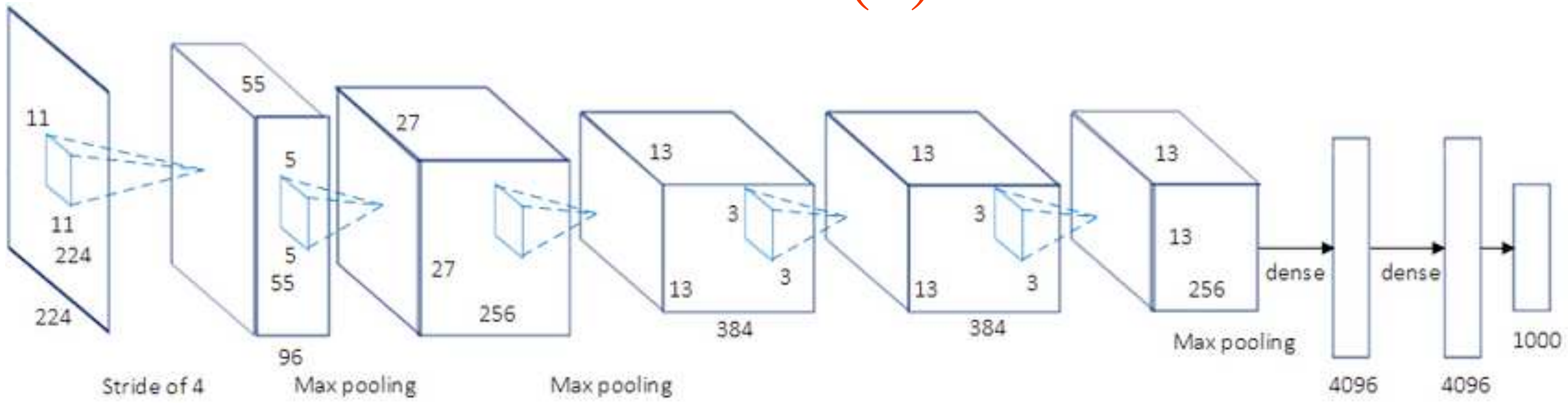
1% People

Attribution Analysis for High-level Networks

What is S(I) looking at?

I

$S(I)$



98% house finch

10% bird

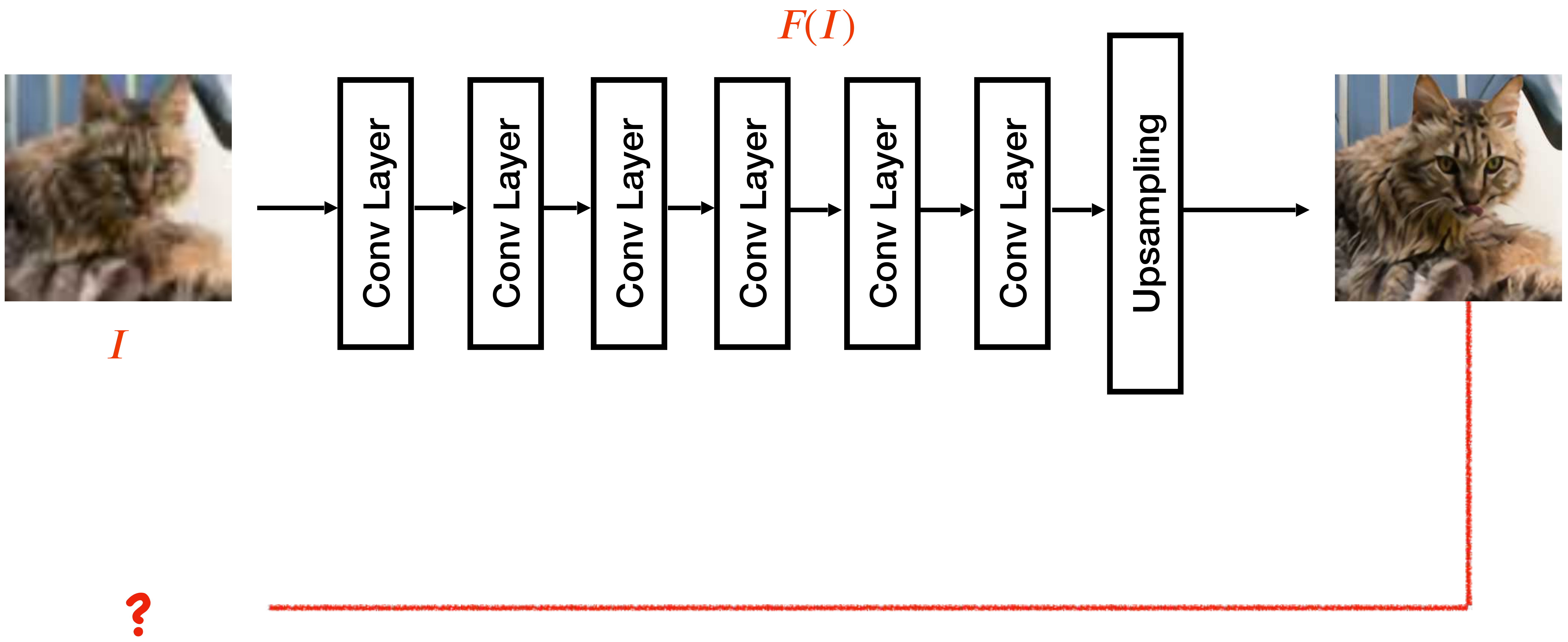
1% People

Backprop methods: gradient

$$\text{Grad}_S(I) = \frac{\partial S(I)}{\partial I}$$

The visualized attribution map

Attribution Analysis for High-level Networks



How to calculate gradient for low-level networks?

Auxiliary Principles

We introduce auxiliary principles for interpreting low-level networks:

- Interpreting local not global

SR networks can not be interpreted globally

Auxiliary Principles

We introduce auxiliary principles for interpreting low-level networks:

- Interpreting local not global
- Interpreting hard not simple

Interpreting simple cases can provide limited help

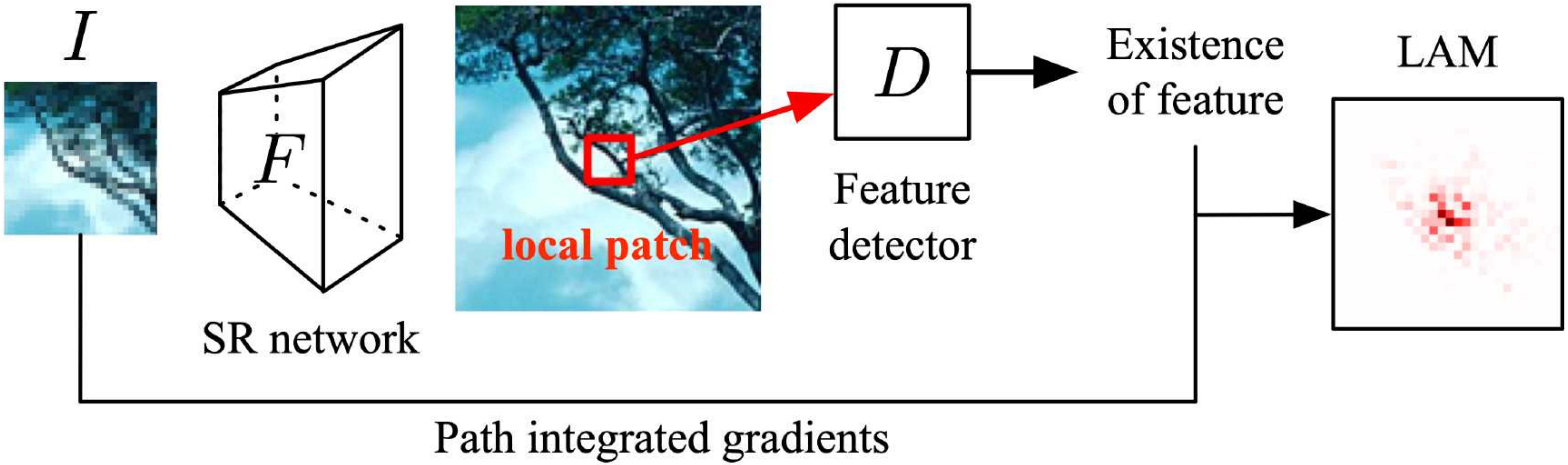
Auxiliary Principles

We introduce auxiliary principles for interpreting low-level networks:

- Interpreting local not global
- Interpreting hard not simple
- Interpreting features not pixels

We convert the problem into
whether there exists edges/textures or not,
instead of why these pixels have such intensities.

Local Attribution Maps (LAM)



Local Attribution Maps (LAM)

We employ Path Integral Gradient

$$\text{LAM}_{F,D}(\gamma)_i := \int_0^1 \frac{\partial D(F(\gamma(\alpha)))}{\partial \gamma(\alpha)_i} \times \frac{\partial \gamma(\alpha)_i}{\partial \alpha} d\alpha \,.$$

SR Network $\boldsymbol{F}$

Feature Detector $\boldsymbol{D}$

Path function $\gamma(\alpha), \alpha \in \mathbb{R}$

Baseline Input $\gamma(0) = \boldsymbol{I'}$

Input $\gamma(1) = \boldsymbol{I}$

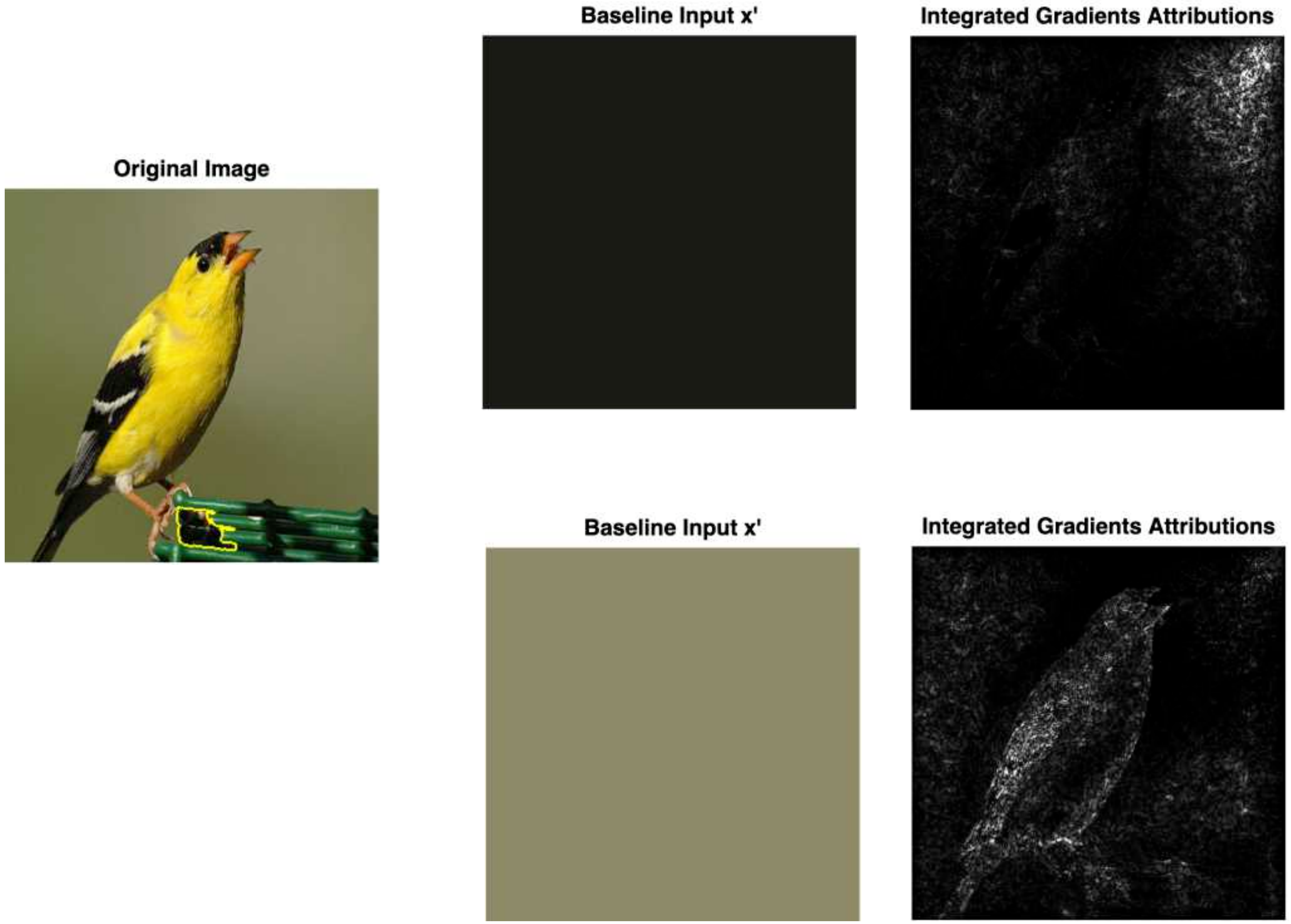
Local Attribution Maps (LAM)

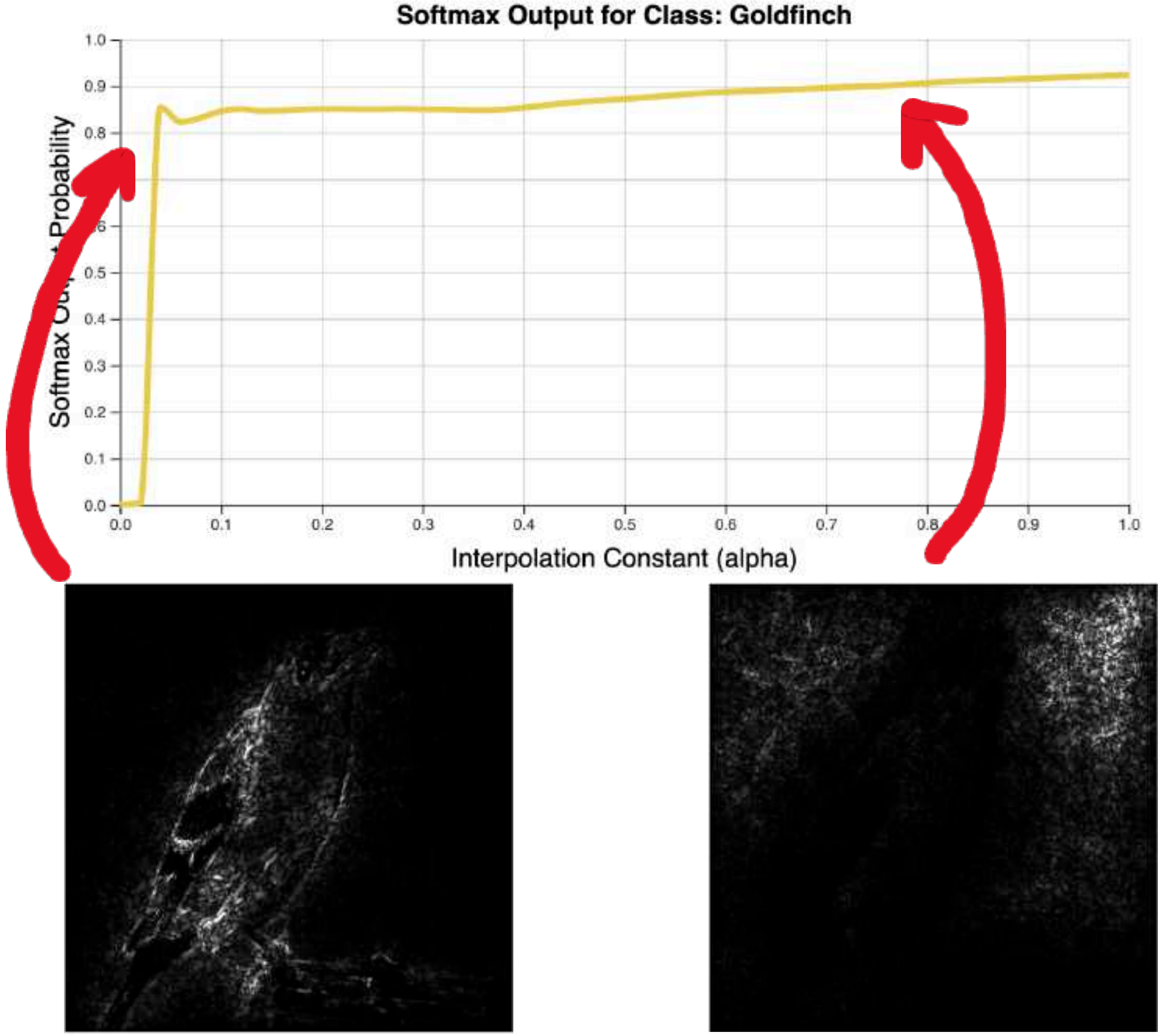
We design the Baseline Input and Path function especially for SR networks.

Blurred image as baseline input: $I' = \omega(\sigma) \otimes I$

Progressive blurring path function: $\gamma_{\mathrm{pb}}(\alpha) = \omega(\sigma - \alpha\sigma) \otimes I$

Local Attribution Maps (LAM)





[Visualizing the Impact of Feature Attribution Baselines]

Local Attribution Maps (LAM)

We employ Path Integral Gradient

$$\mathrm{LAM}_{F,D}(\gamma)_i := \int_0^1 \frac{\partial D(F(\gamma(\alpha)))}{\partial \gamma(\alpha)_i} \times \frac{\partial \gamma(\alpha)_i}{\partial \alpha} d\alpha \,.$$

The Gradient of interpolation

The weight determined by path function

SR Network F

Feature Detector D

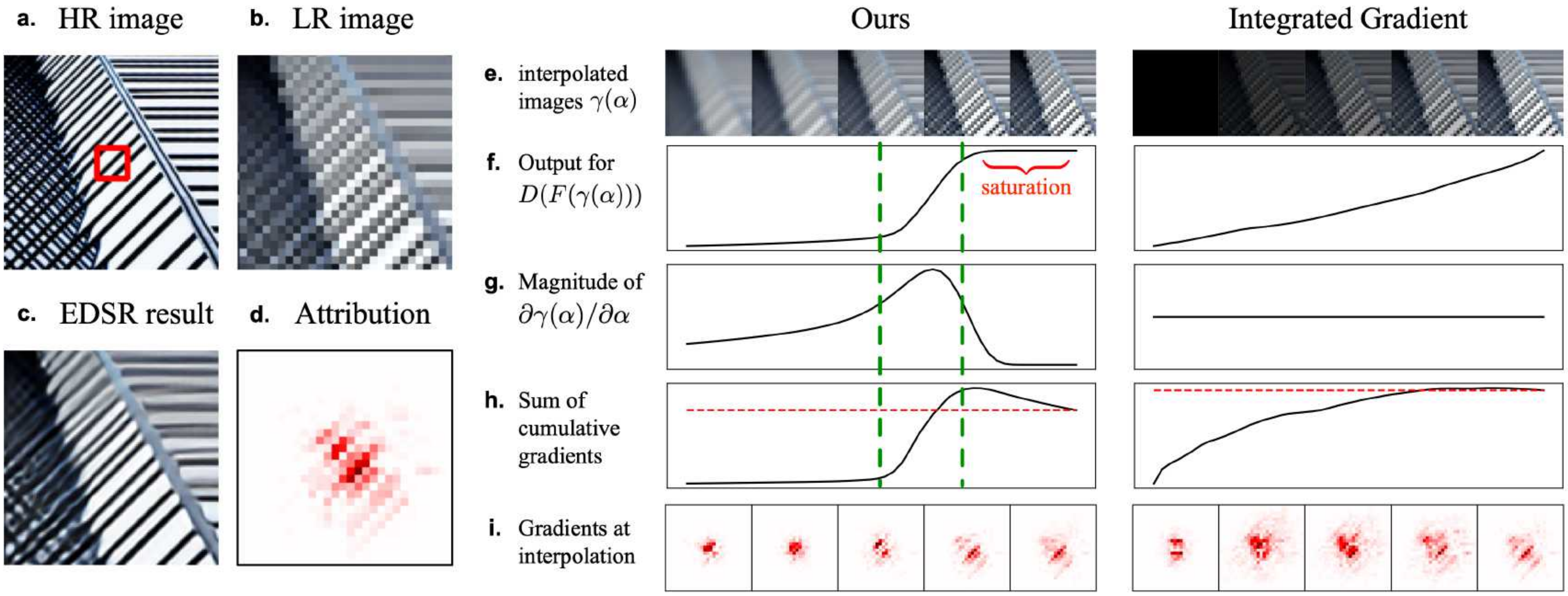
Path function $\gamma(\alpha), \alpha \in \mathbb{R}$

Baseline Input $\gamma(0) = I'$

Input $\gamma(1) = I$

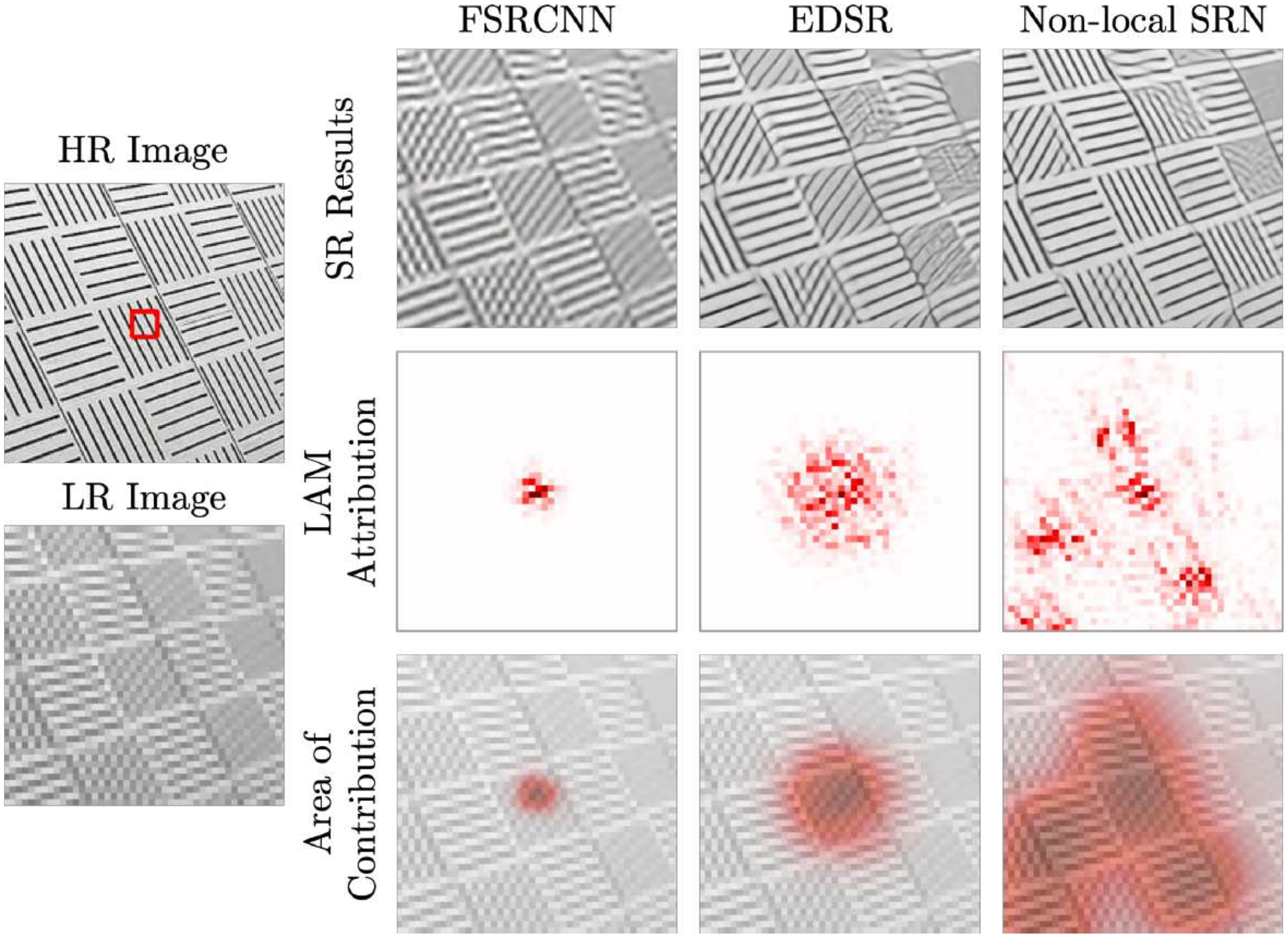
Local Attribution Maps (LAM)

Why using path integral gradient: Gradient Saturation



Local Attribution Maps Results

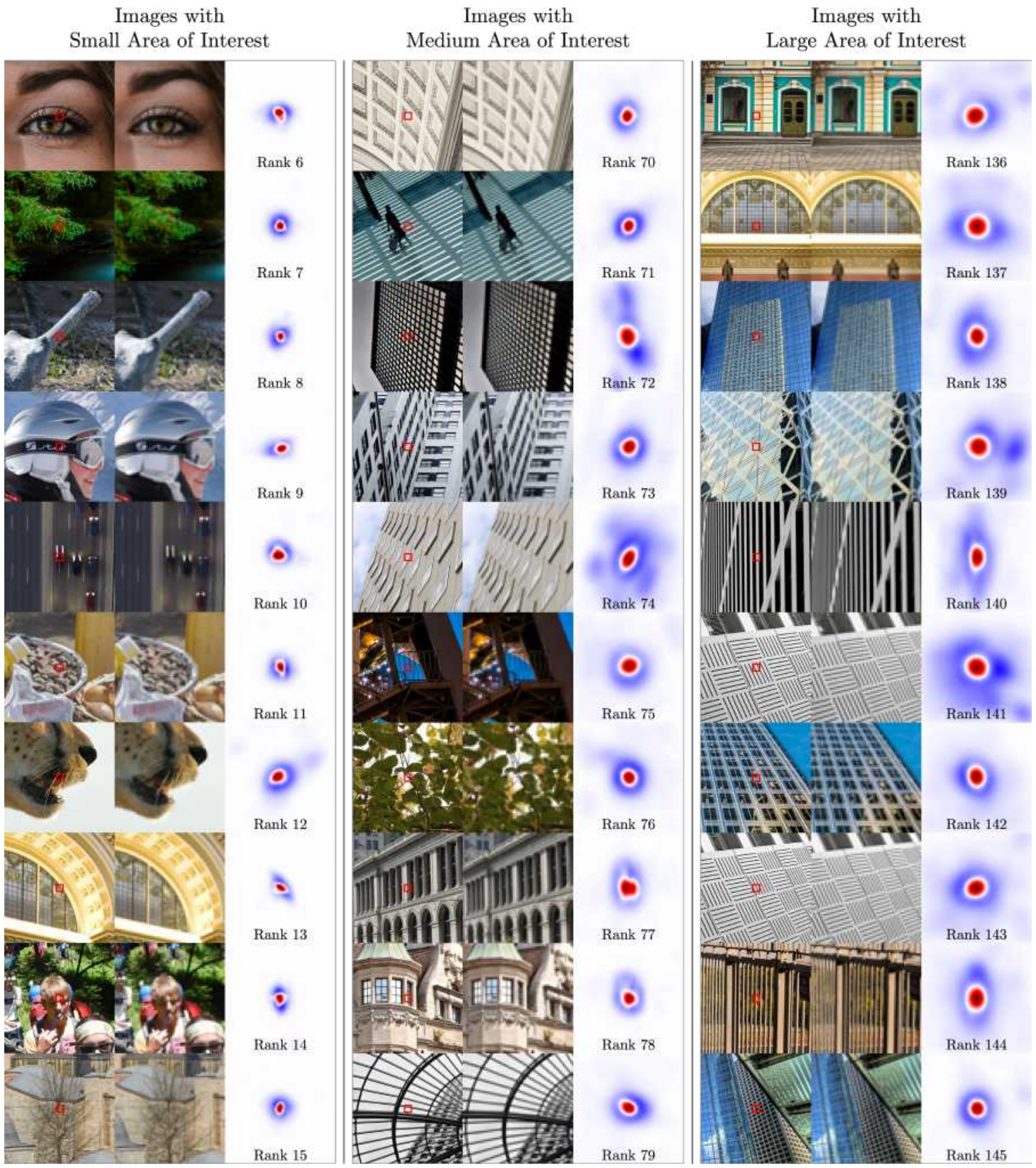
Local Attribution Maps Results



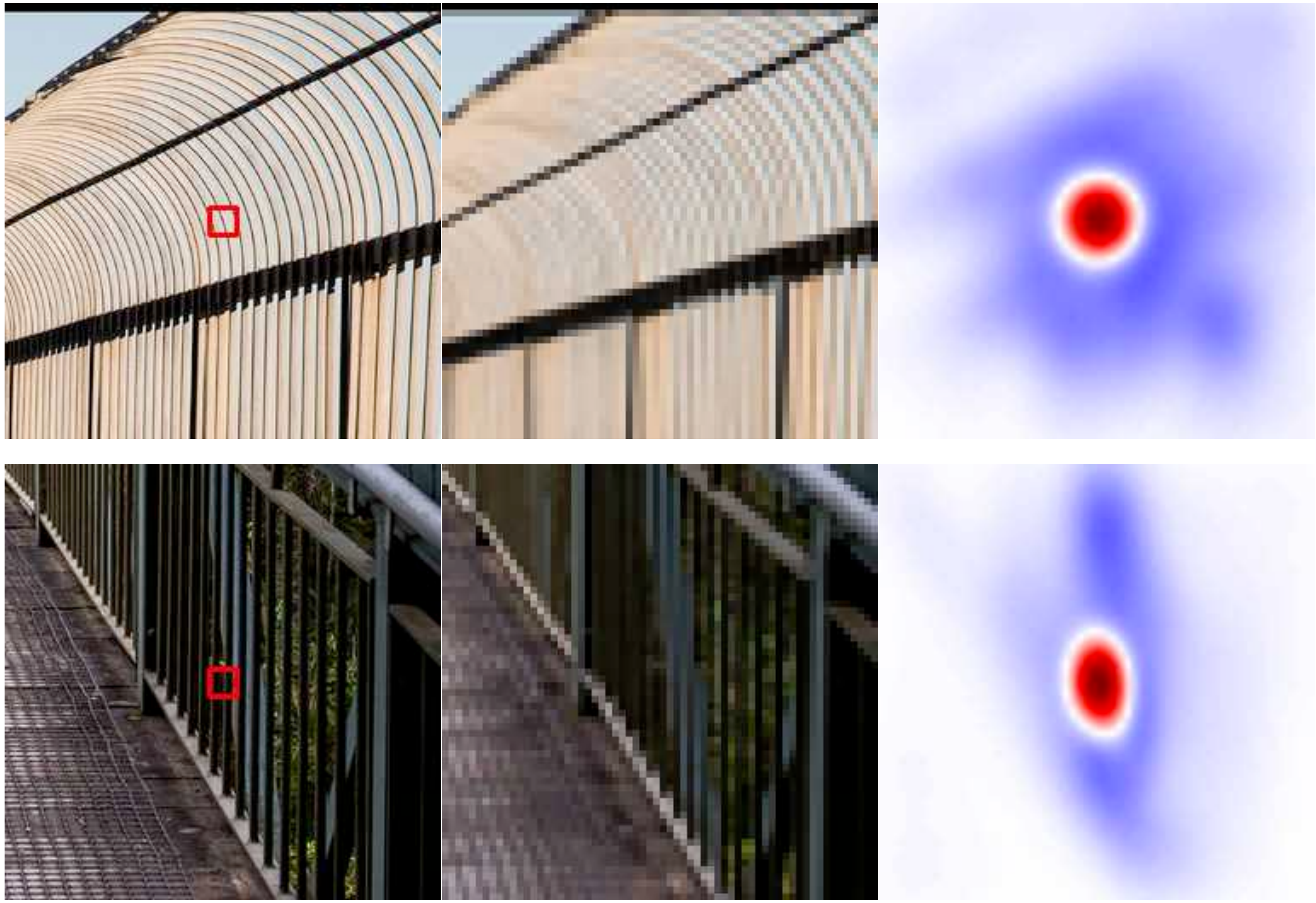
Informative Areas

The similarities and differences of LAM results for different SR networks

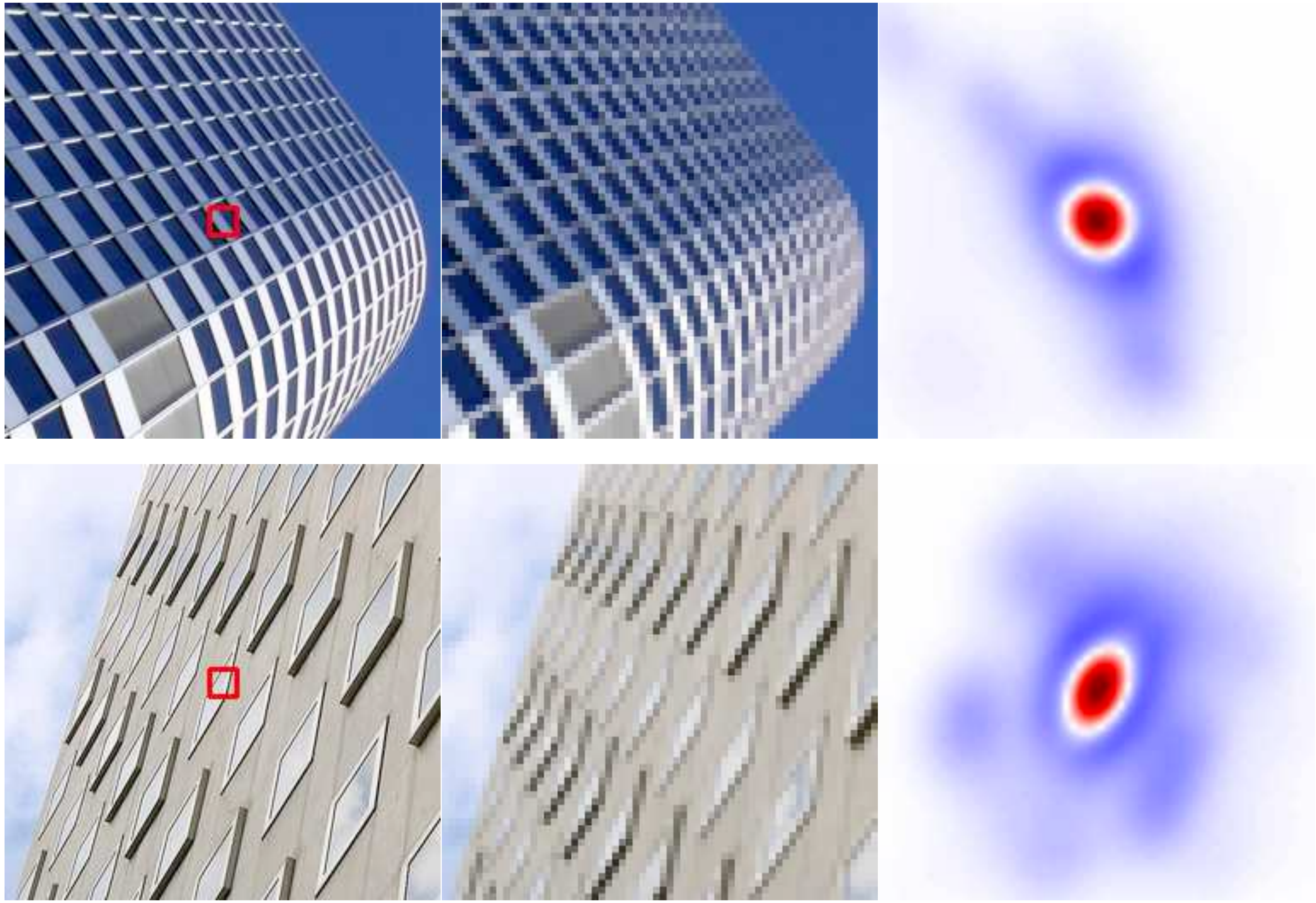
- Red areas can be used for the most preliminary level of SR
- Blue areas show the potential informative areas



Informative Areas



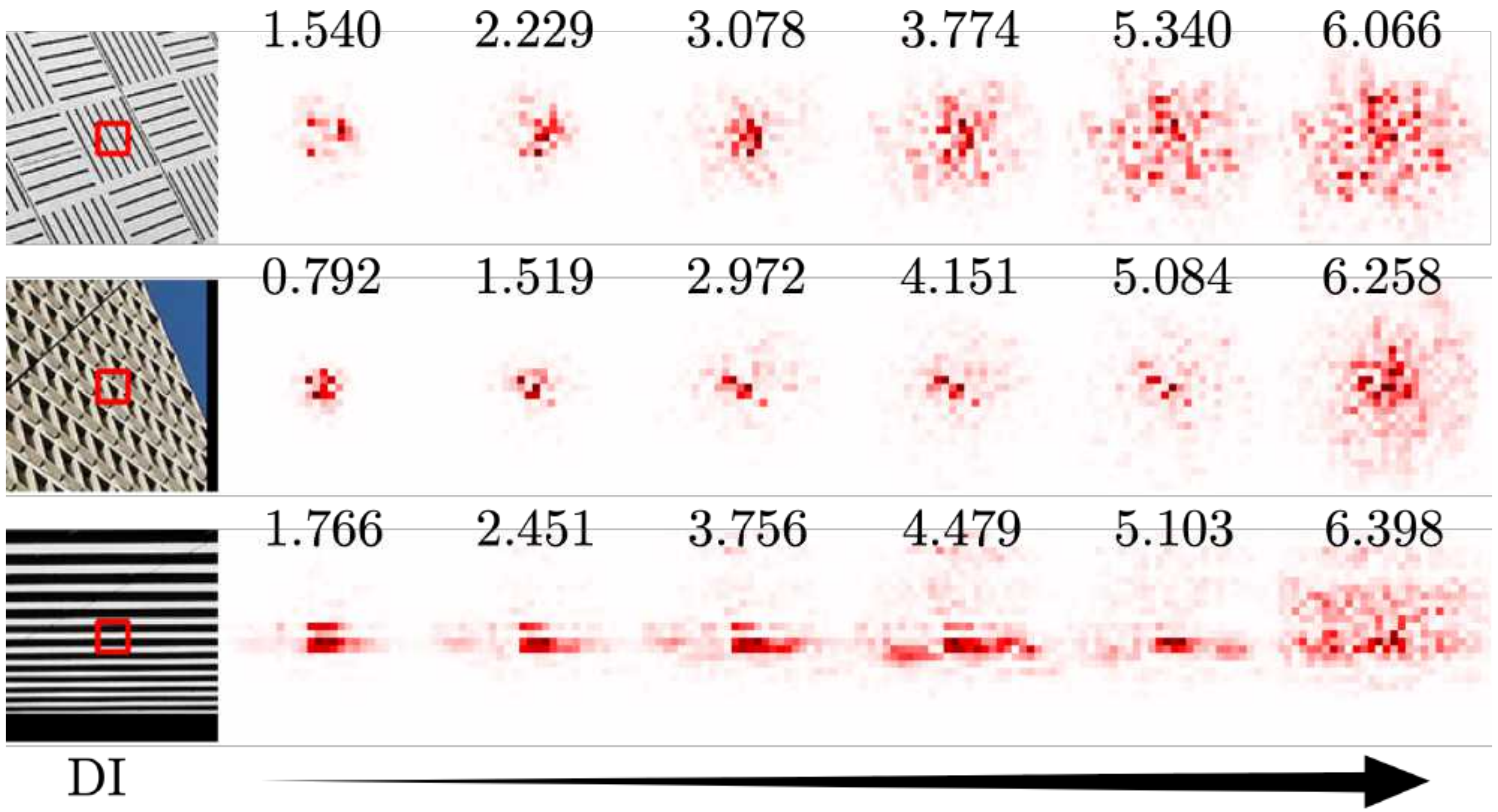
Informative Areas



Exploration with LAM

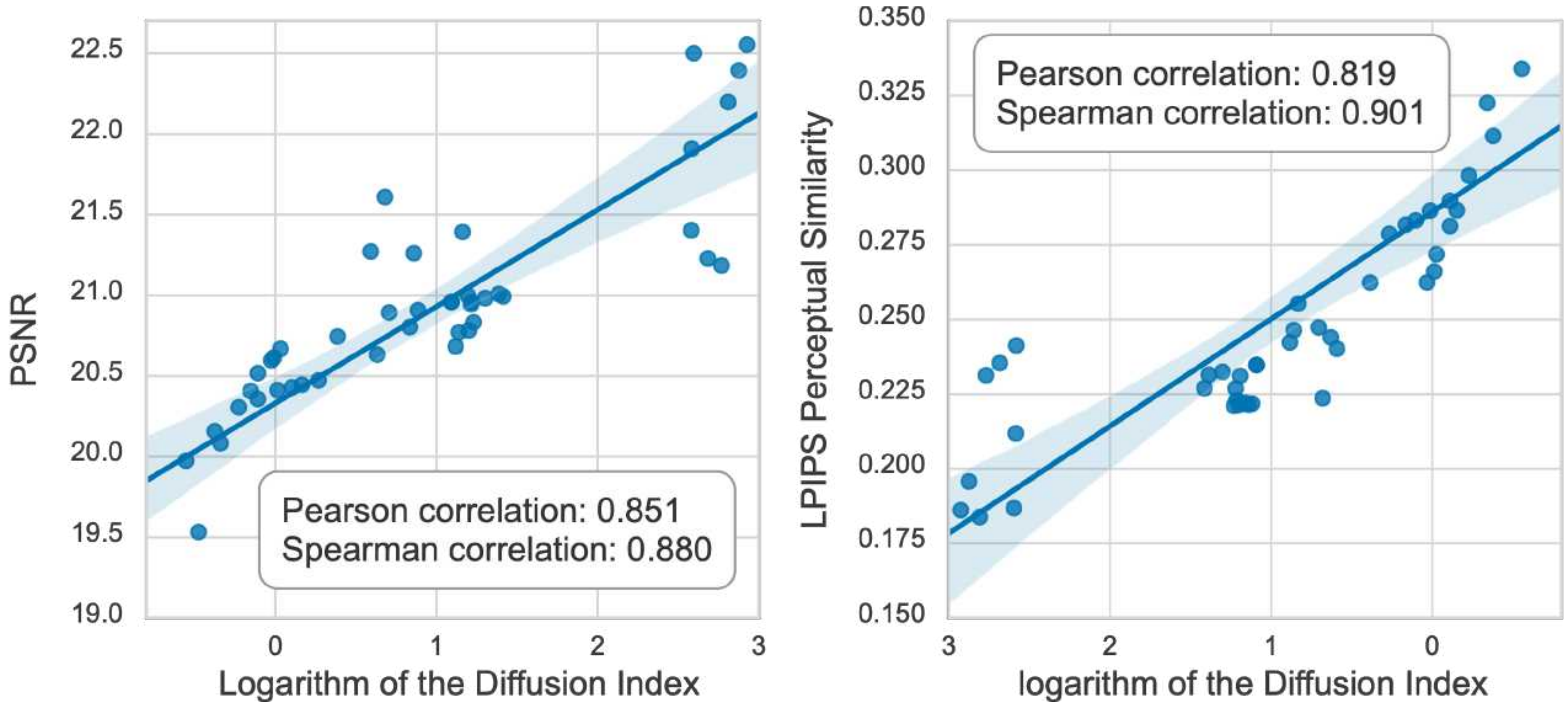
We use Gini Index to indicate the range of involved pixels: $G = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{n}|g_i - g_j|}{2n^2\bar{g}}$

And propose Diffusion Index for quantitative analysis: $\mathrm{DI} = (1 - G) \times 100$



Exploration with LAM

Diffusion Index vs. Network Performances.



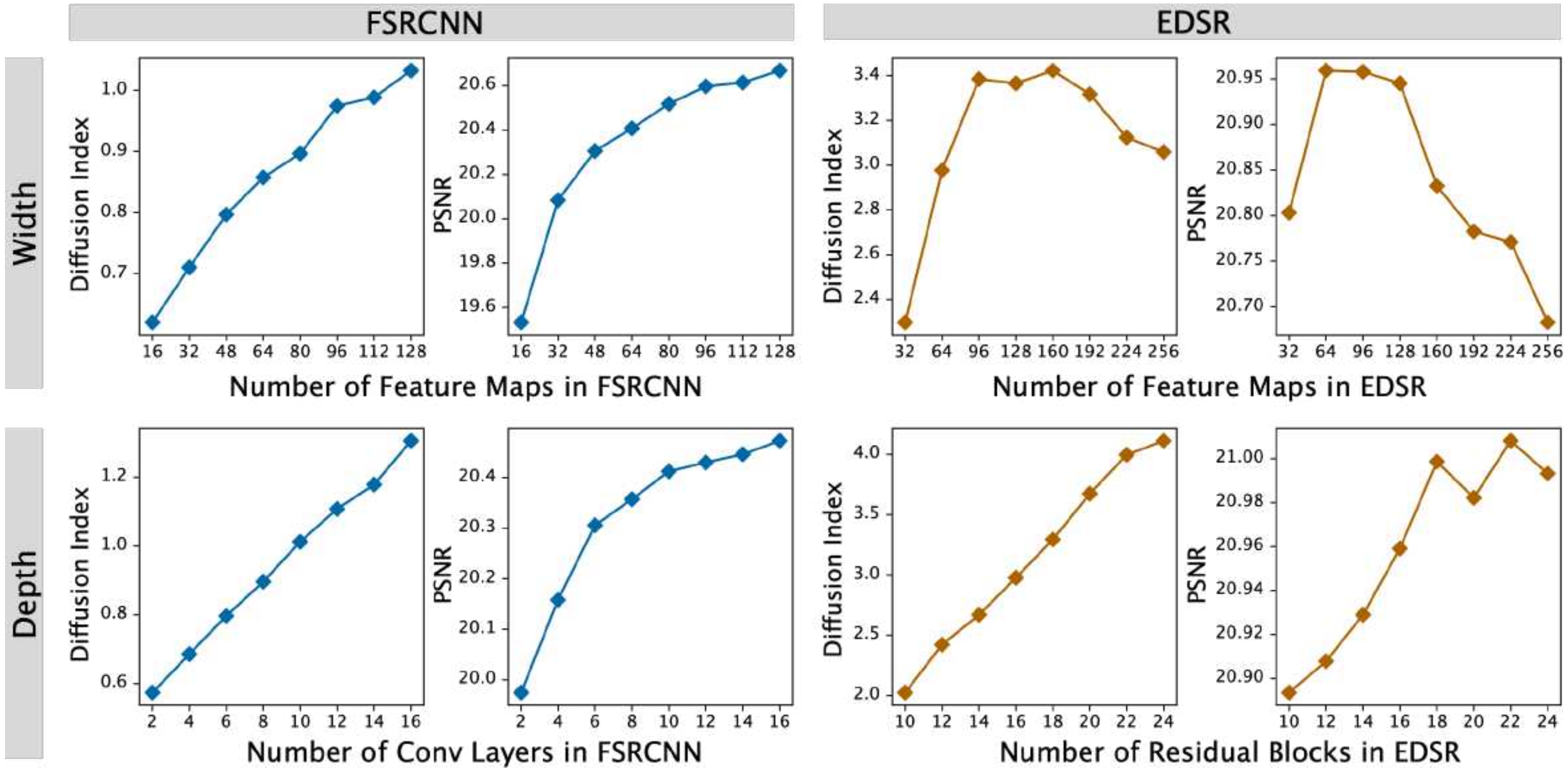
Exploration with LAM

Diffusion Index vs. Receptive Field.

Model	Recpt. Field	PSNR	DI	Remark
FSRCNN	17×17	20.30	0.797	Fully convolution network.
CARN	45×45	21.27	1.807	Residual network.
EDSR	75×75	20.96	2.977	Residual network.
MSRN	107×107	21.39	3.194	Residual network.
RRDBNet	703×703	20.96	13.417	Residual network.
IMDN	global	21.23	14.643	Global pooling.
RFDN	global	21.40	13.208	Global pooling.
RCAN	global	22.20	16.596	Global pooling.
RNAN	global	21.91	13.243	Non-local attention.
SAN	global	22.55	18.642	Non-local attention.

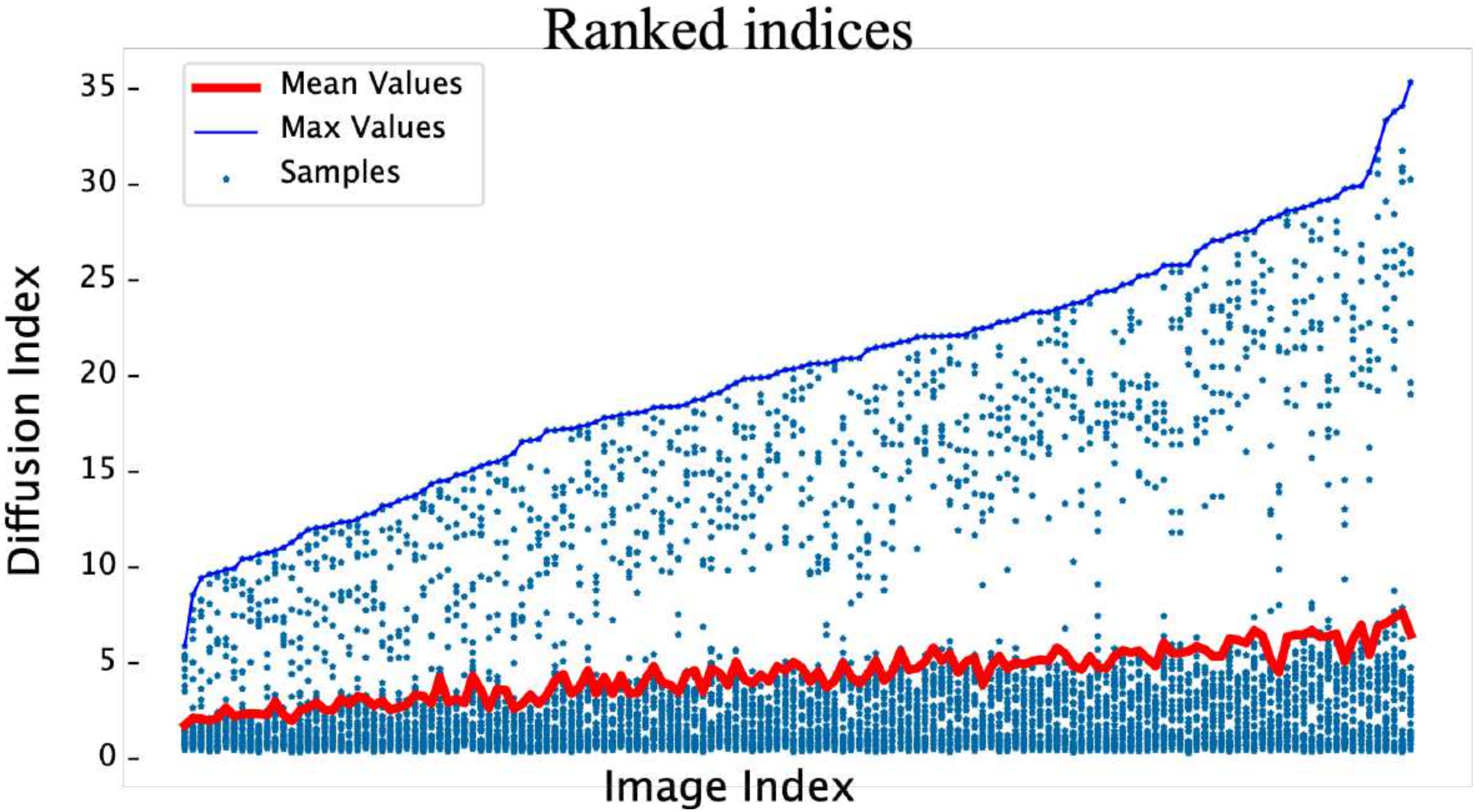
Exploration with LAM

Diffusion Index vs. Network Scale.



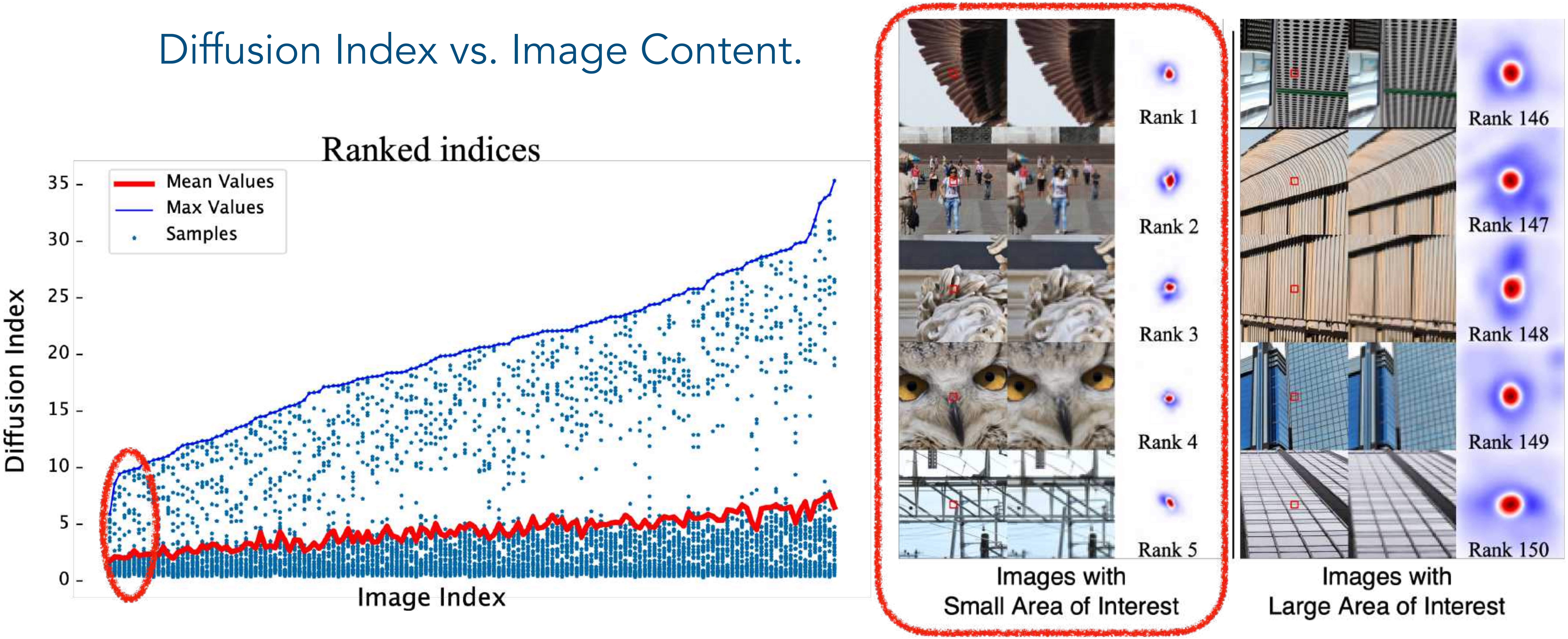
Exploration with LAM

Diffusion Index vs. Image Content.



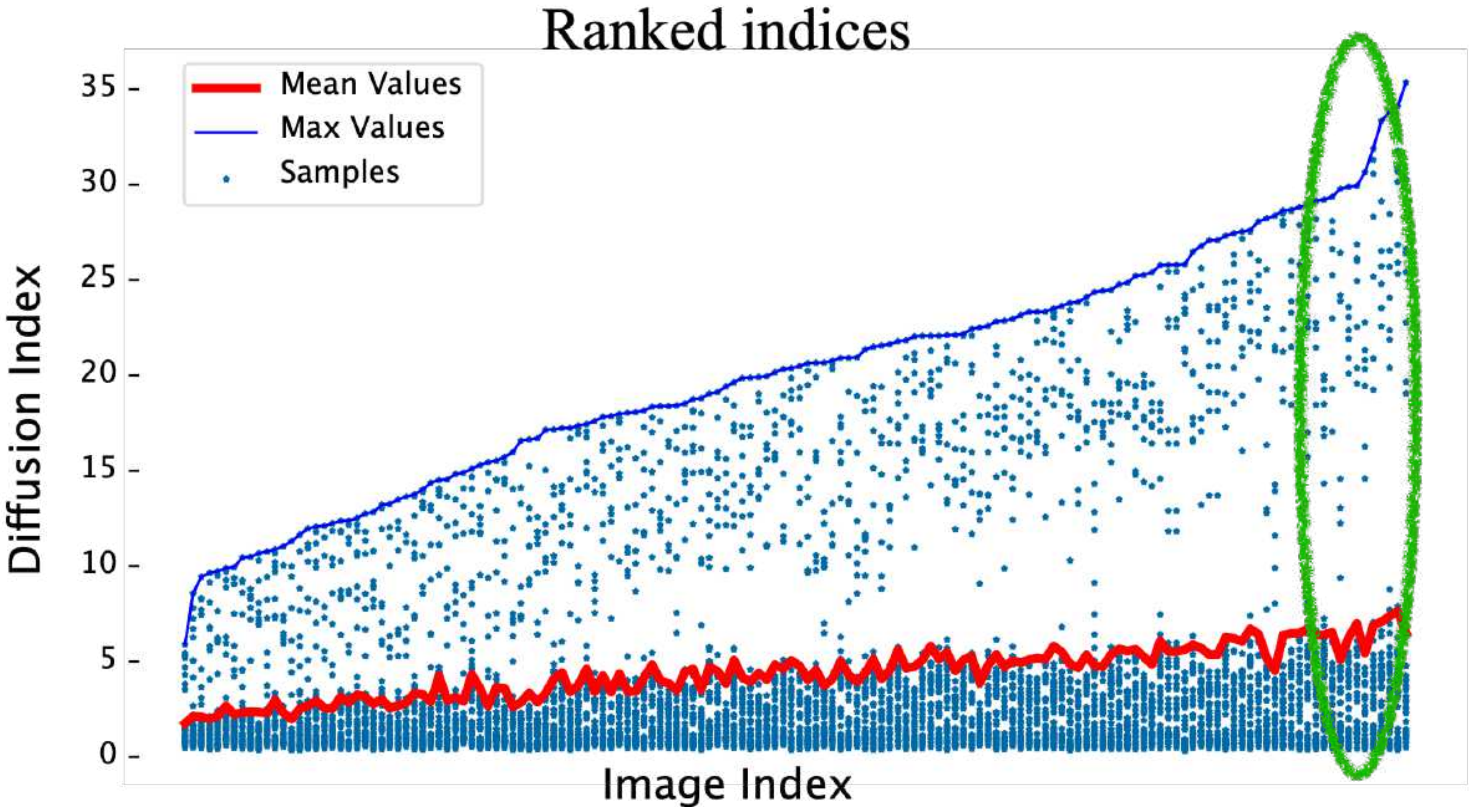
Exploration with LAM

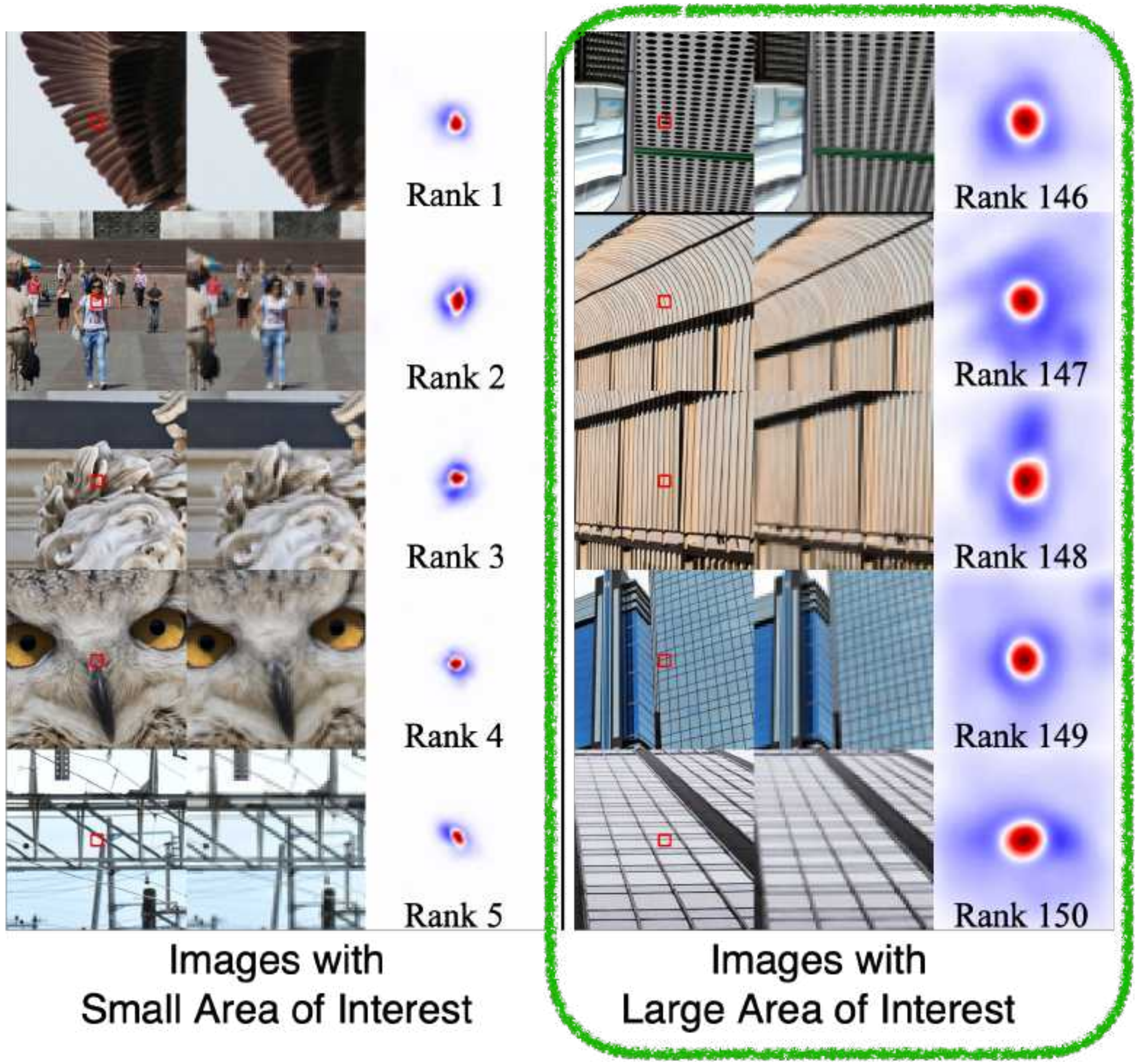
Diffusion Index vs. Image Content.



Exploration with LAM

Diffusion Index vs. Image Content.





LAM Playground

LocalAttributionMapsDemo.ipynb

File Edit View Insert Runtime Tools Help Last edited on November 23

Comment Share

+ Code + Text

Connect Editing

Interpreting Super-Resolution Networks with Local Attribution Maps

Jinjin Gu, Chao Dong

Project Page: https://x-lowlevel-vision.github.io/lam.html

This is an online Demo. Please follow the code and comments, step by step

First, click `file` and then COPY you own notebook file to make sure your changes are recorded. Please turn on the colab GPU switch.

Import packages

[ ]

```
import torch, cv2, os, sys, numpy as np, matplotlib.pyplot as plt
from PIL import Image
```

Load model codes and model files

Research Outline

Real World Image Restoration

Apply advanced image processing technologies in real world applications.

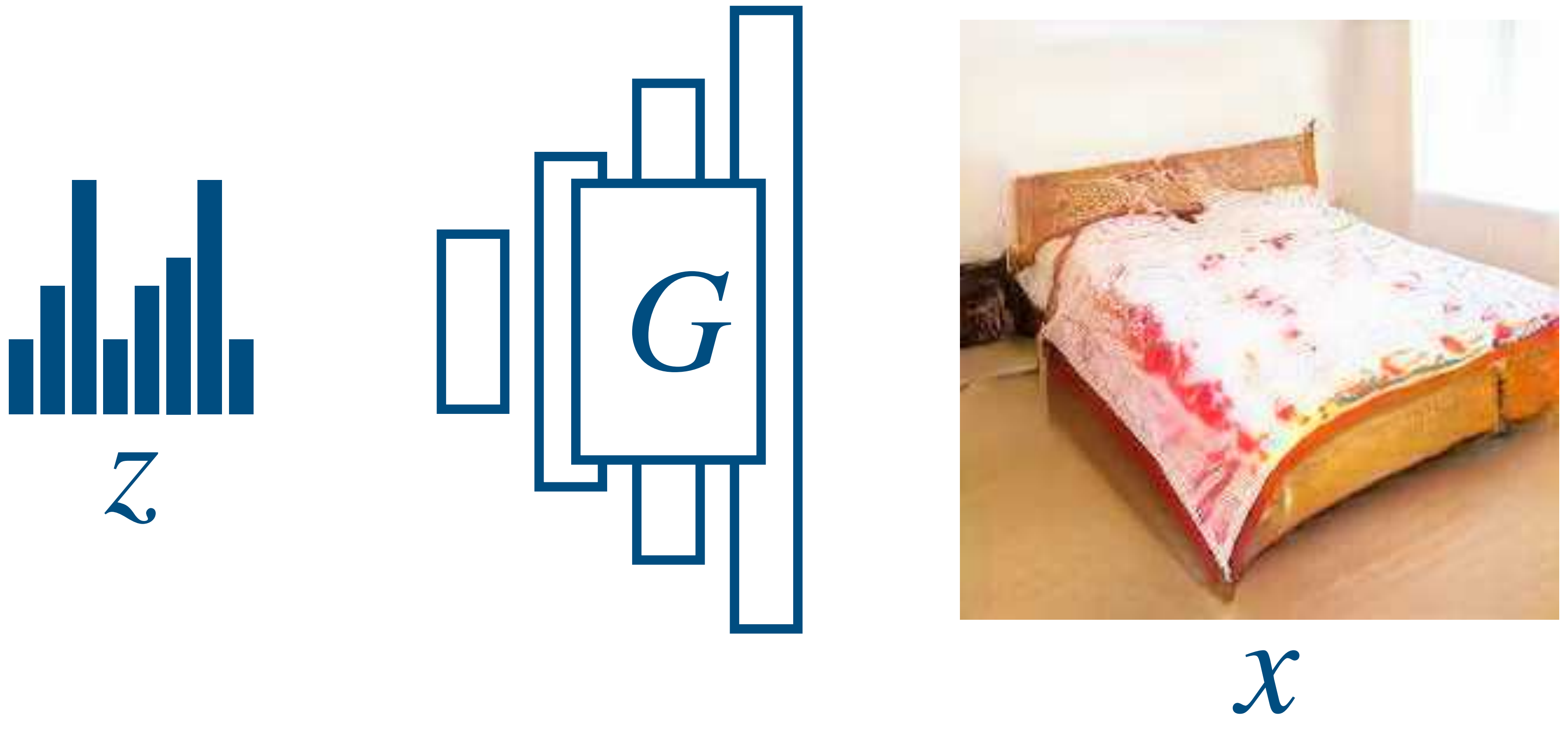
Interpreting Image Generation

Understanding latent space of GANs and make use of GANs' powerful representation.

Applying Deep Signal Processing Technologies in Industrial

Empowering existing industrial facilities with advanced signal processing technologies.

GANs for Synthesizing Images



2014

GAN

[Goodfellow et al.]

2015

DCGAN

[Radford et al.]

2017

PGGAN

[Karras et al.]

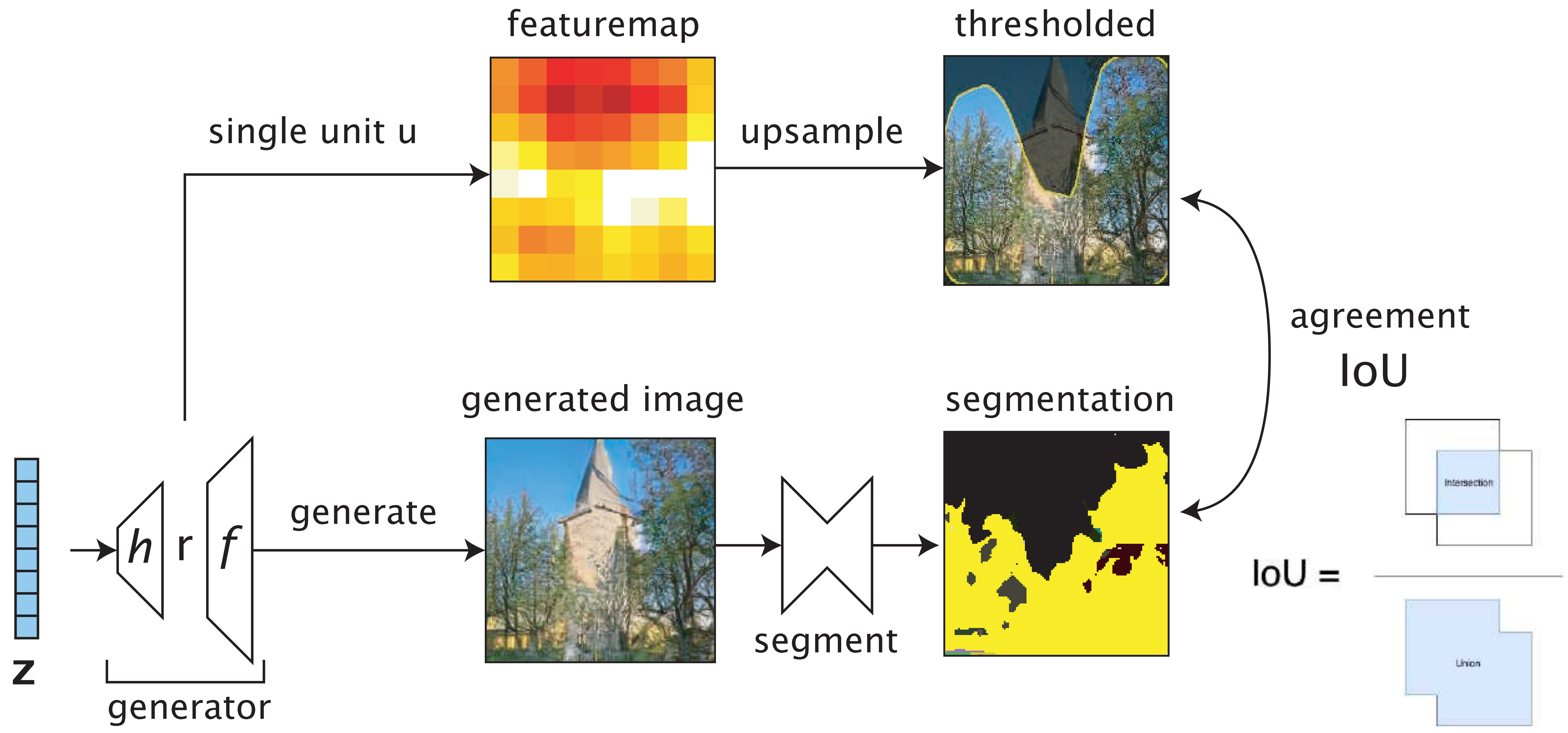
2018

BigGAN

[Brock et al.]

◦ ◦ ◦

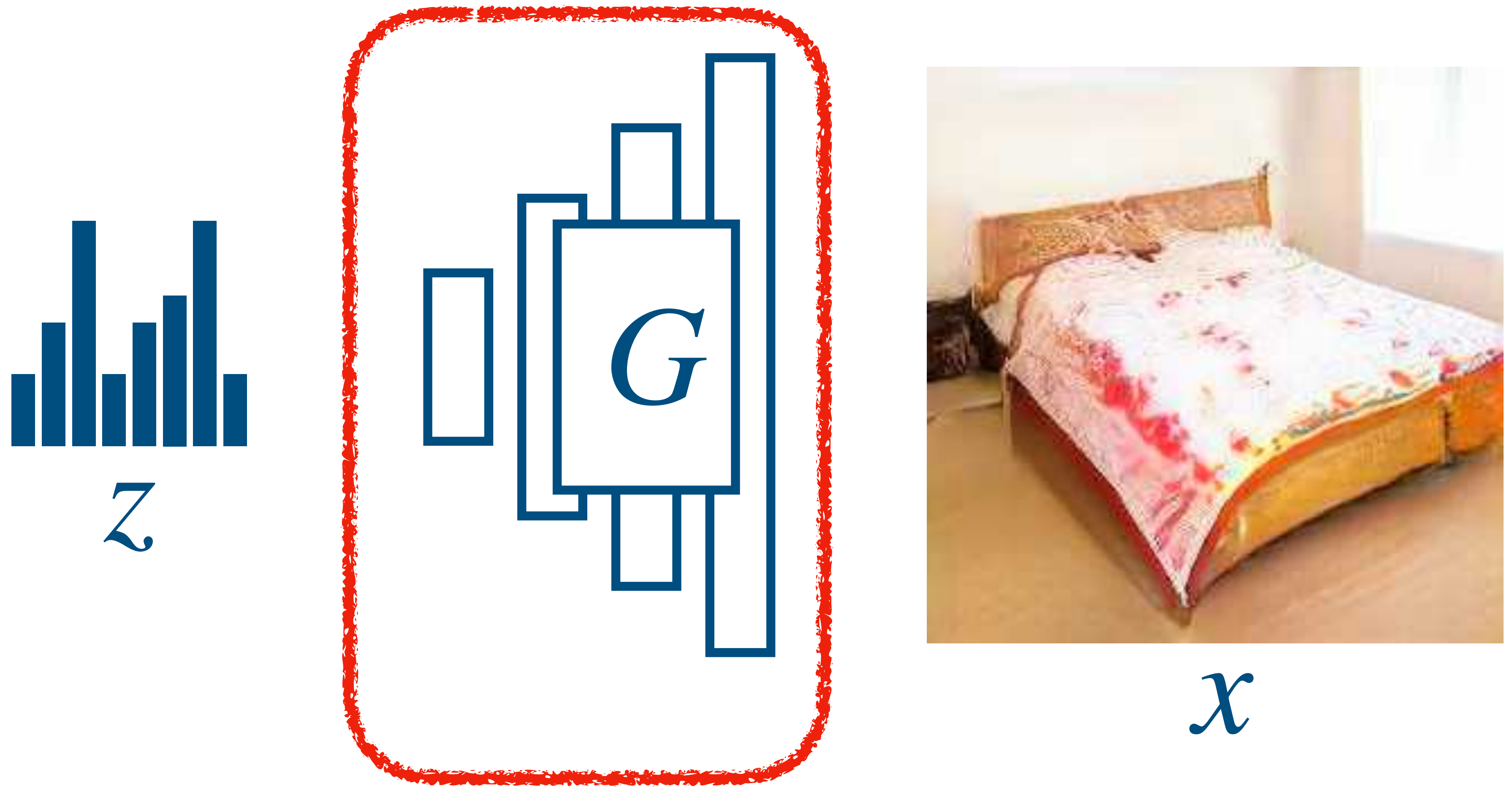
Previous Work on Interpreting Units in GANs



[David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba. GAN Dissection: Visualizing and Understanding Generative Adversarial Networks, ICLR 2019]

Deep Generative Representation

Previous work:

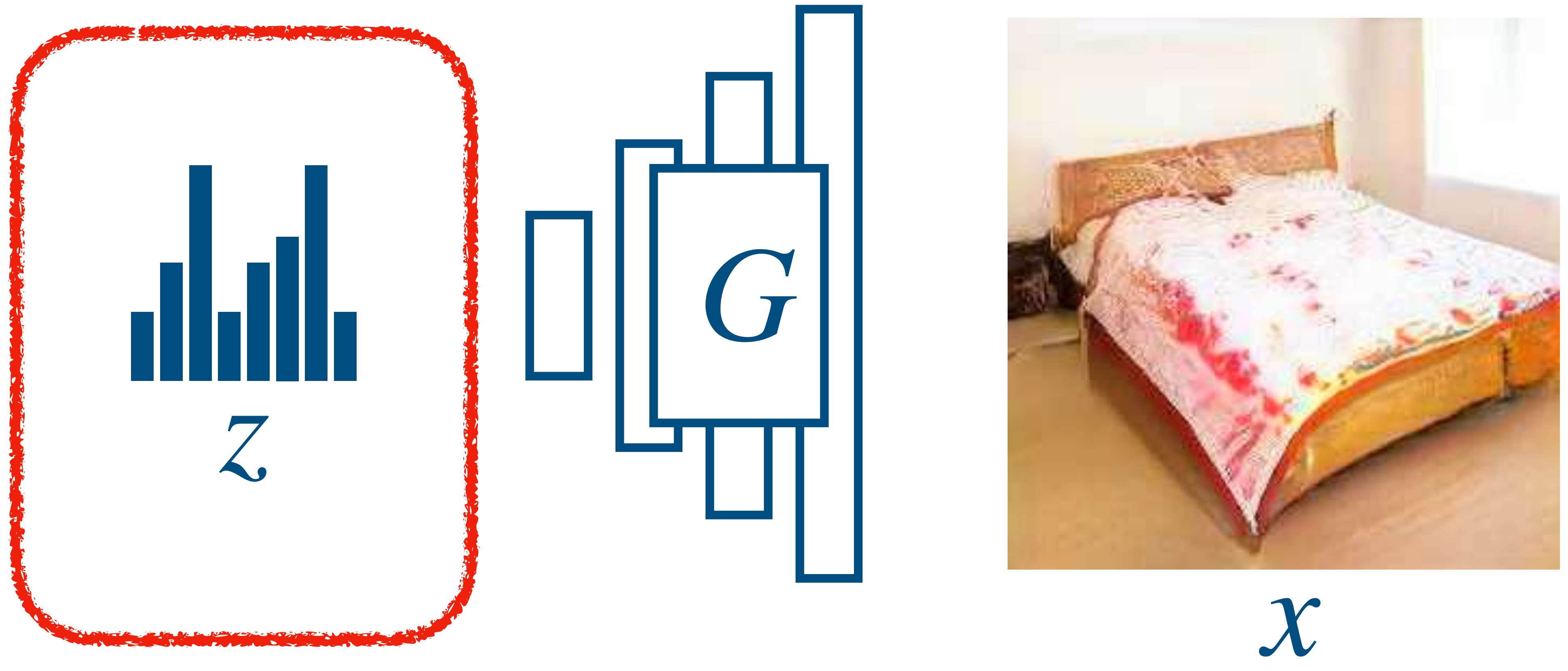


How different units affect the output?

[David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba. GAN Dissection: Visualizing and Understanding Generative Adversarial Networks, ICLR 2019]

Deep Generative Representation

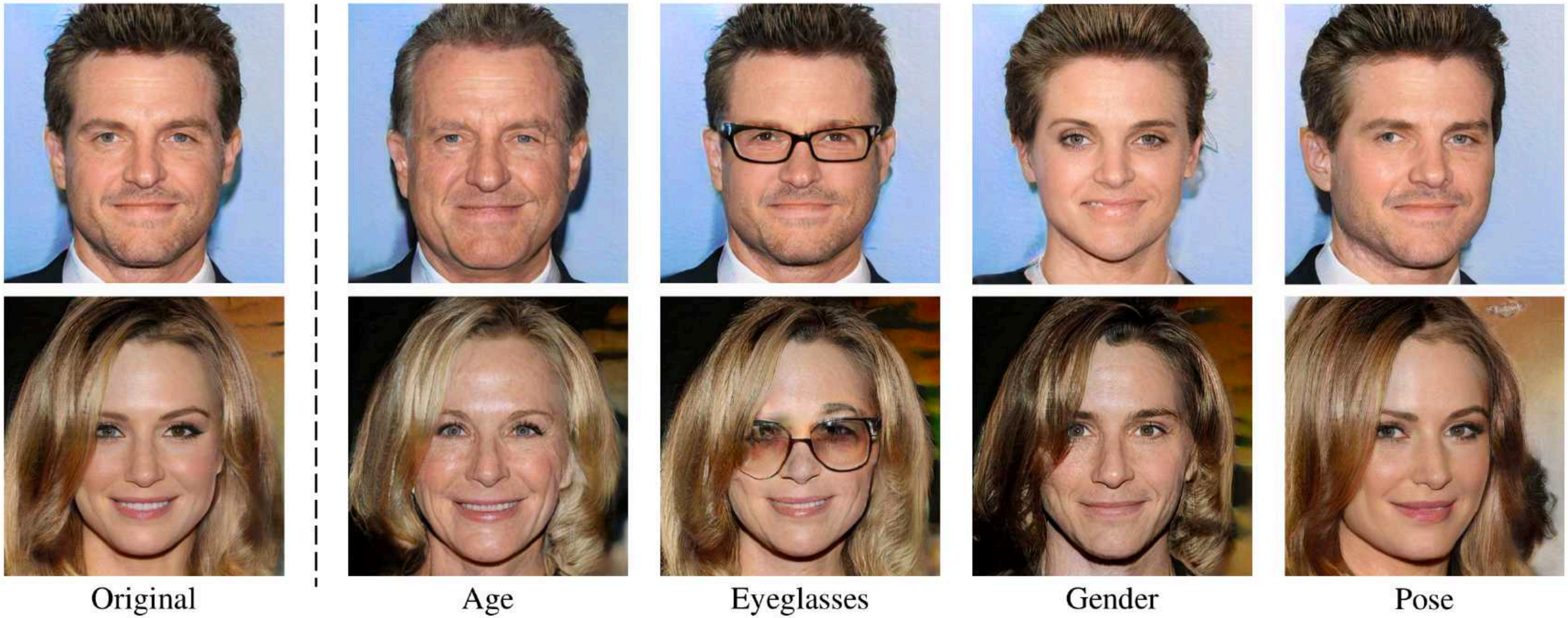
Ours:



How latent code affects the output?

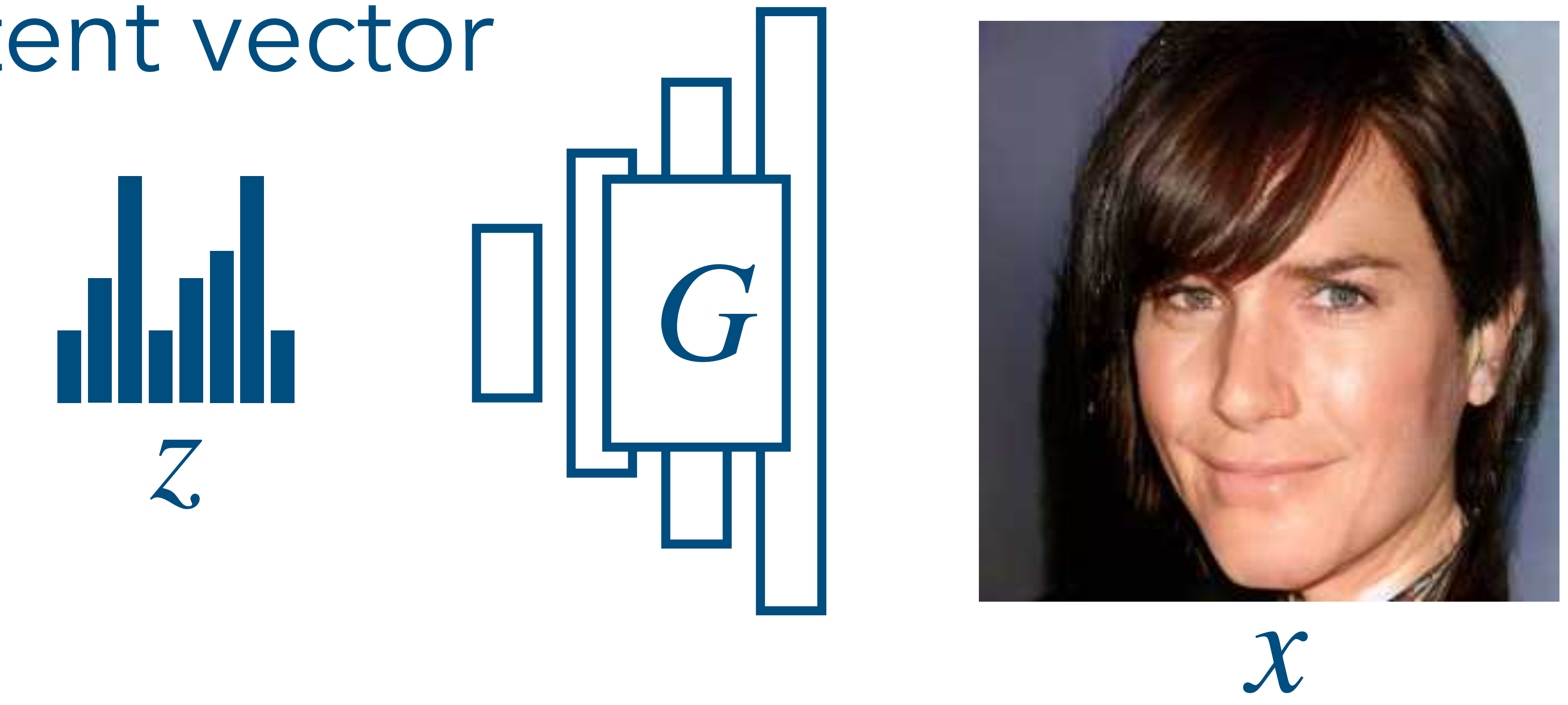
Interpreting the Latent Space of GANs for Semantic Face Editing

Yujun Shen, Jinjin Gu, Xiaoou Tang, Bolei Zhou, CVPR 2020



Understanding GAN Semantic Latent Space

Latent vector

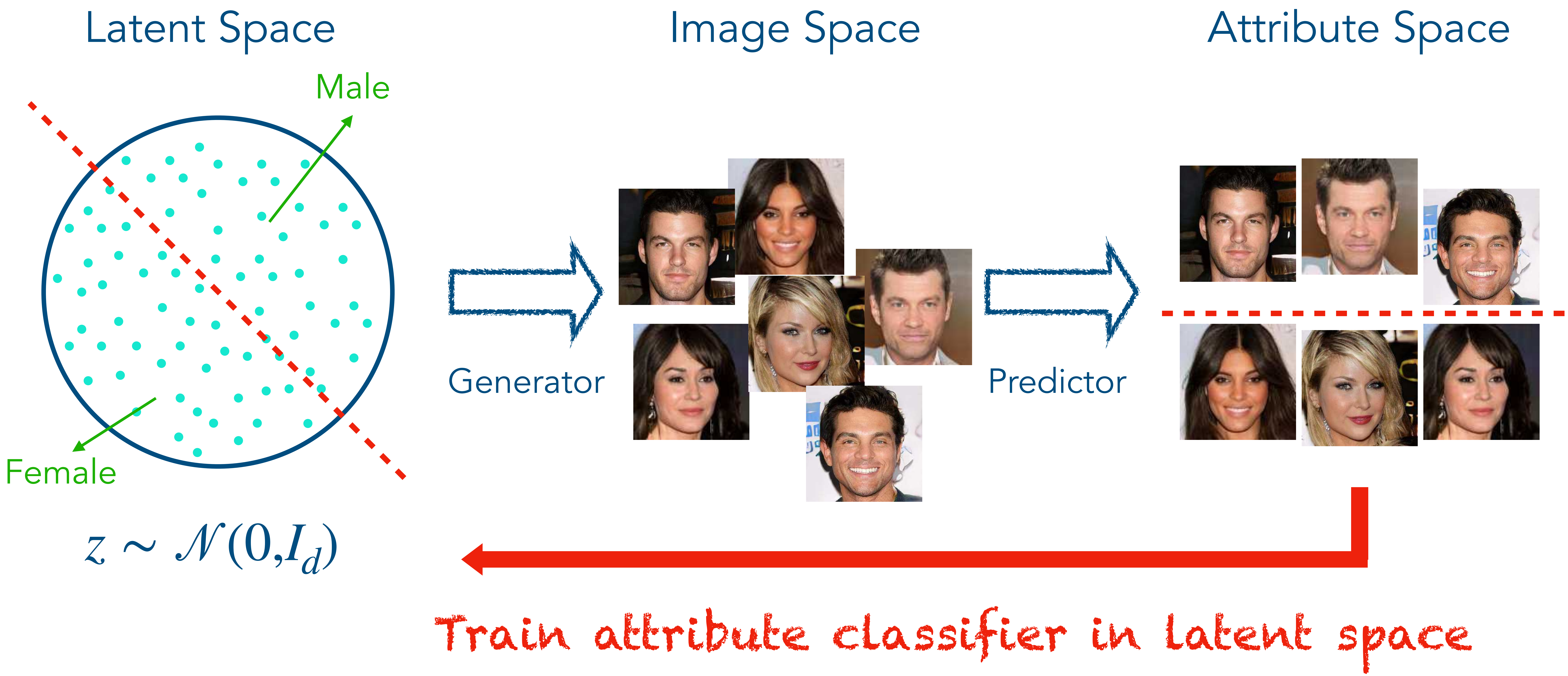


Attributes:

Smile
Pose
Masculine
Old
……

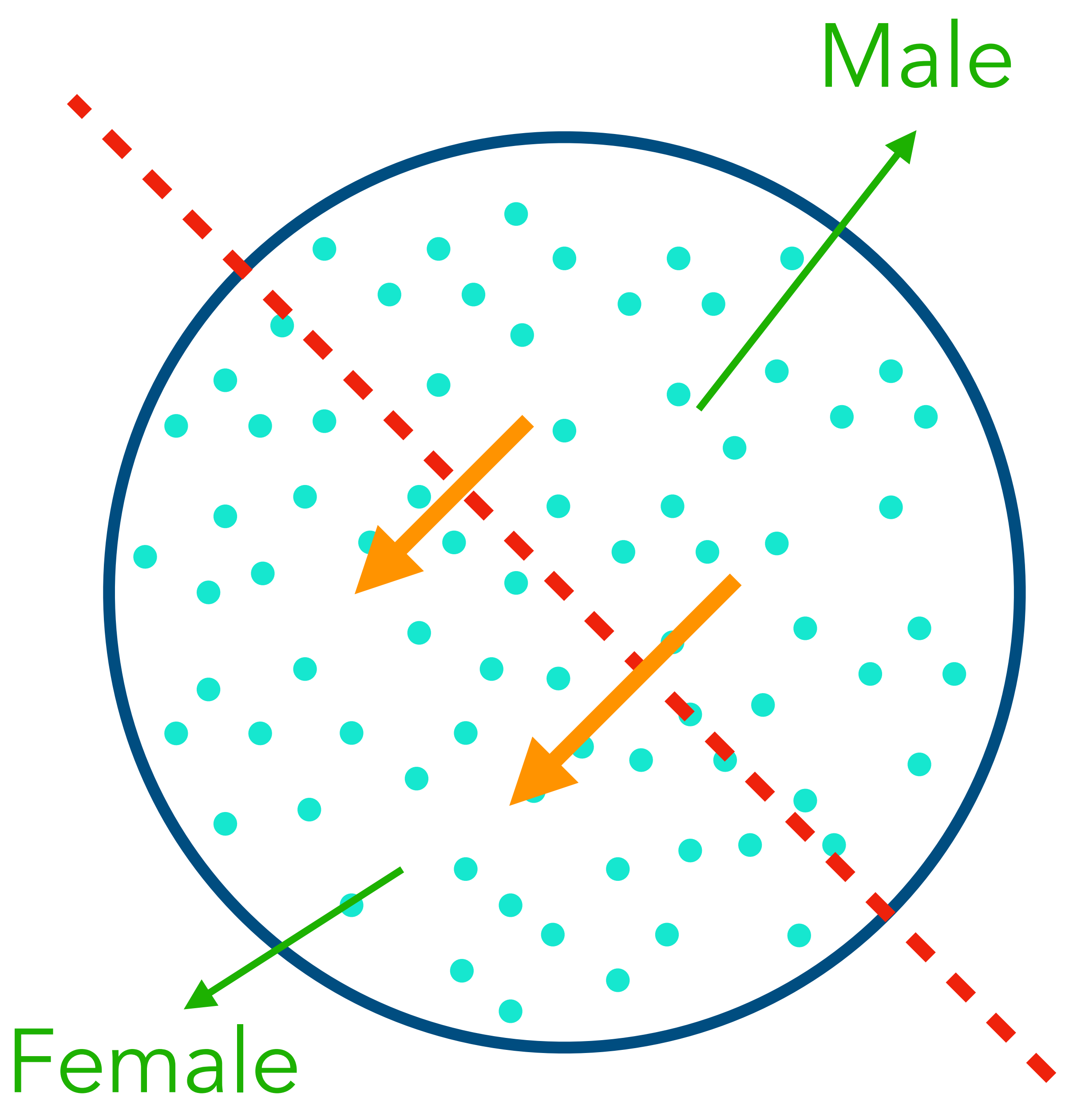
To identify the cause-effect relations

Aligning Latent Space with Attribute Space



Pushing Latent Code through Boundary

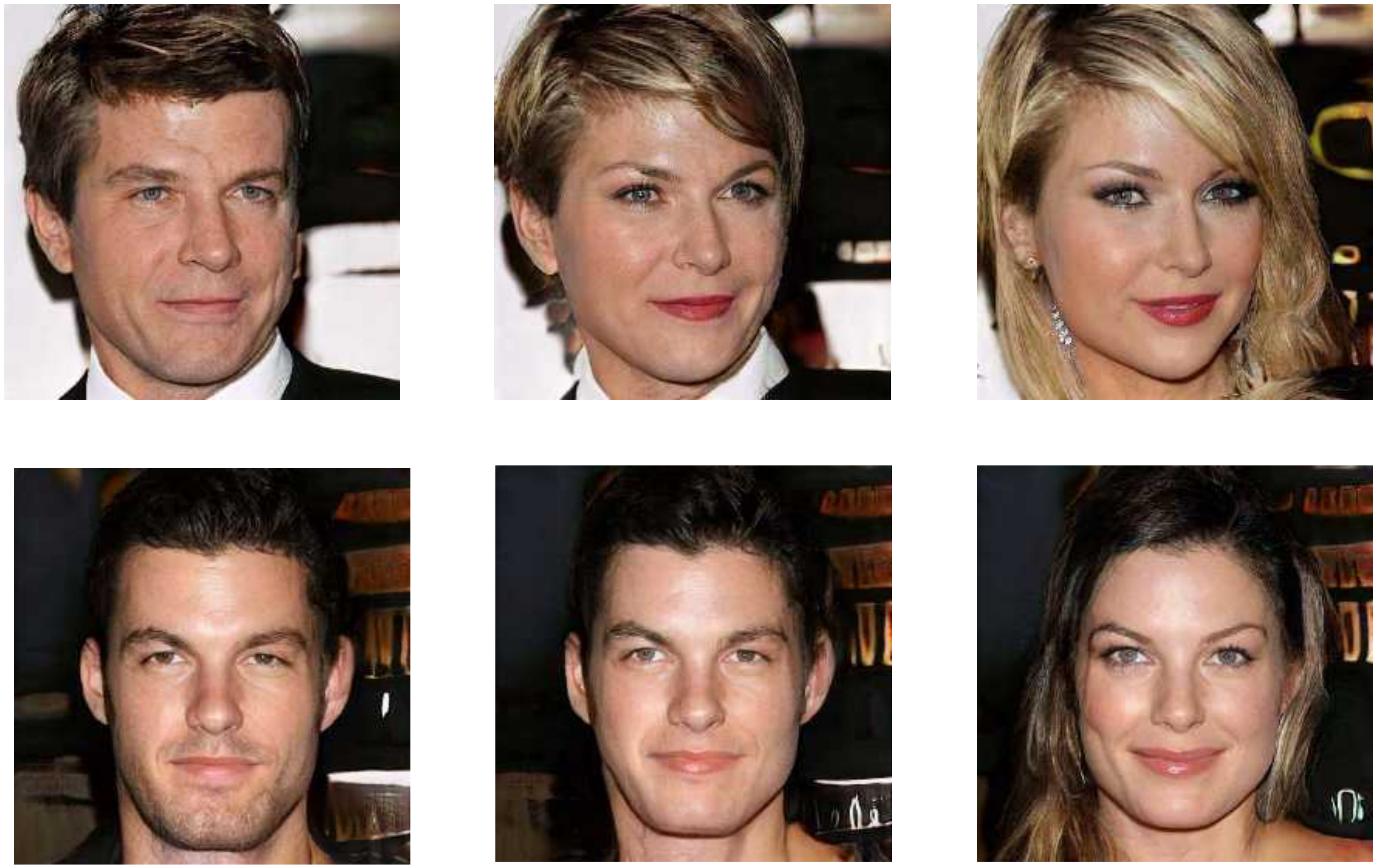
Latent Space



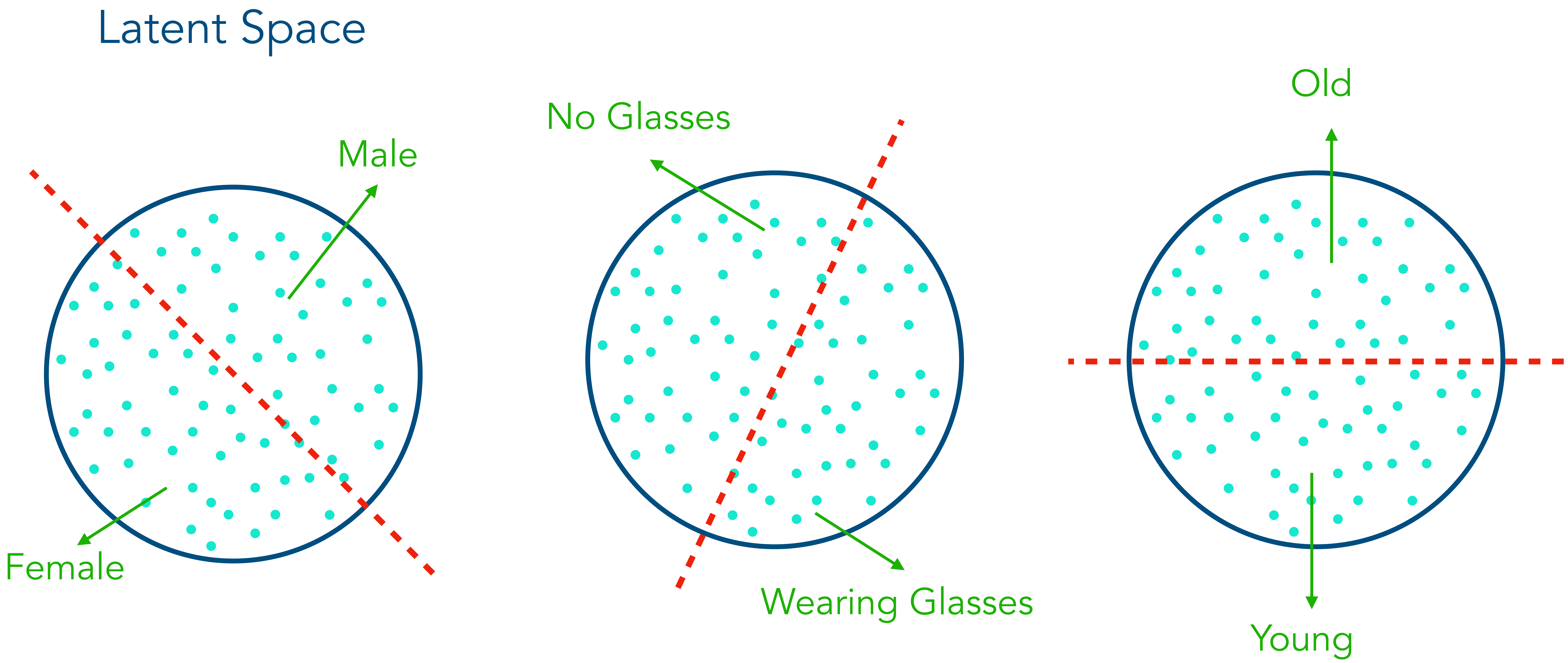
$$z \sim \mathcal{N}(0, I_d)$$

$$G(z) \Longrightarrow G(z + \lambda b)$$

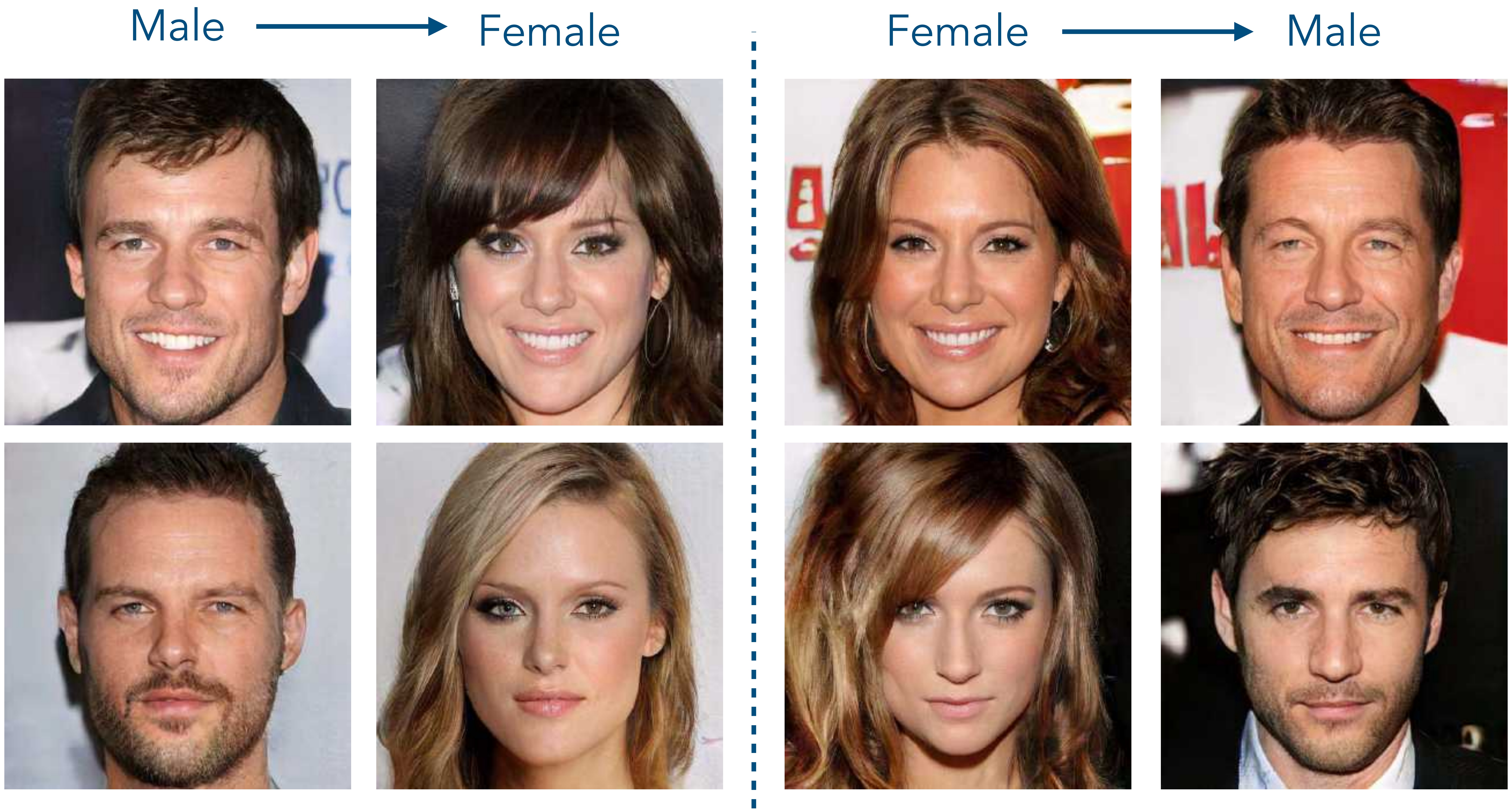
Change Male to Female



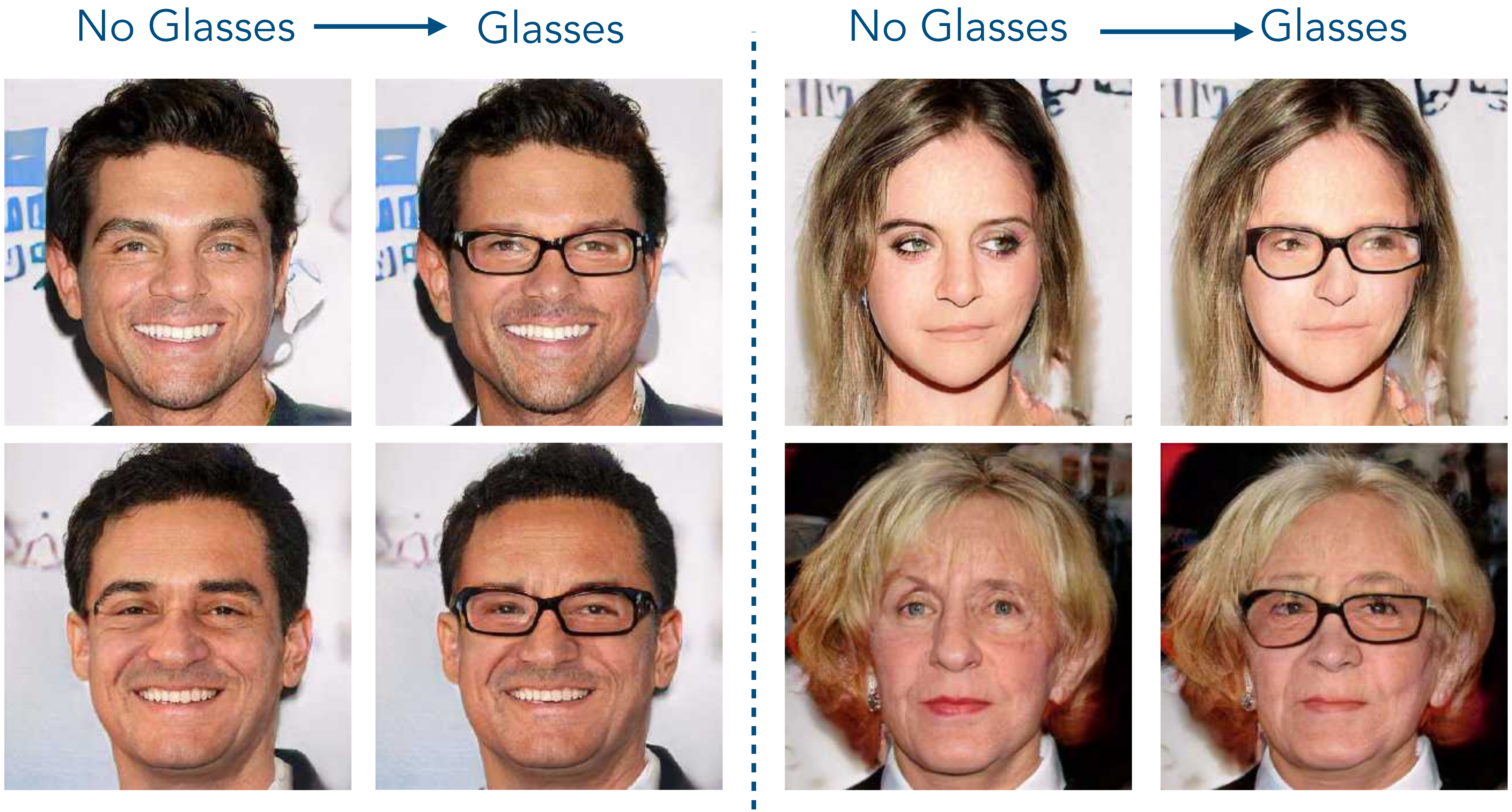
Various Attribute Boundaries in Latent Space



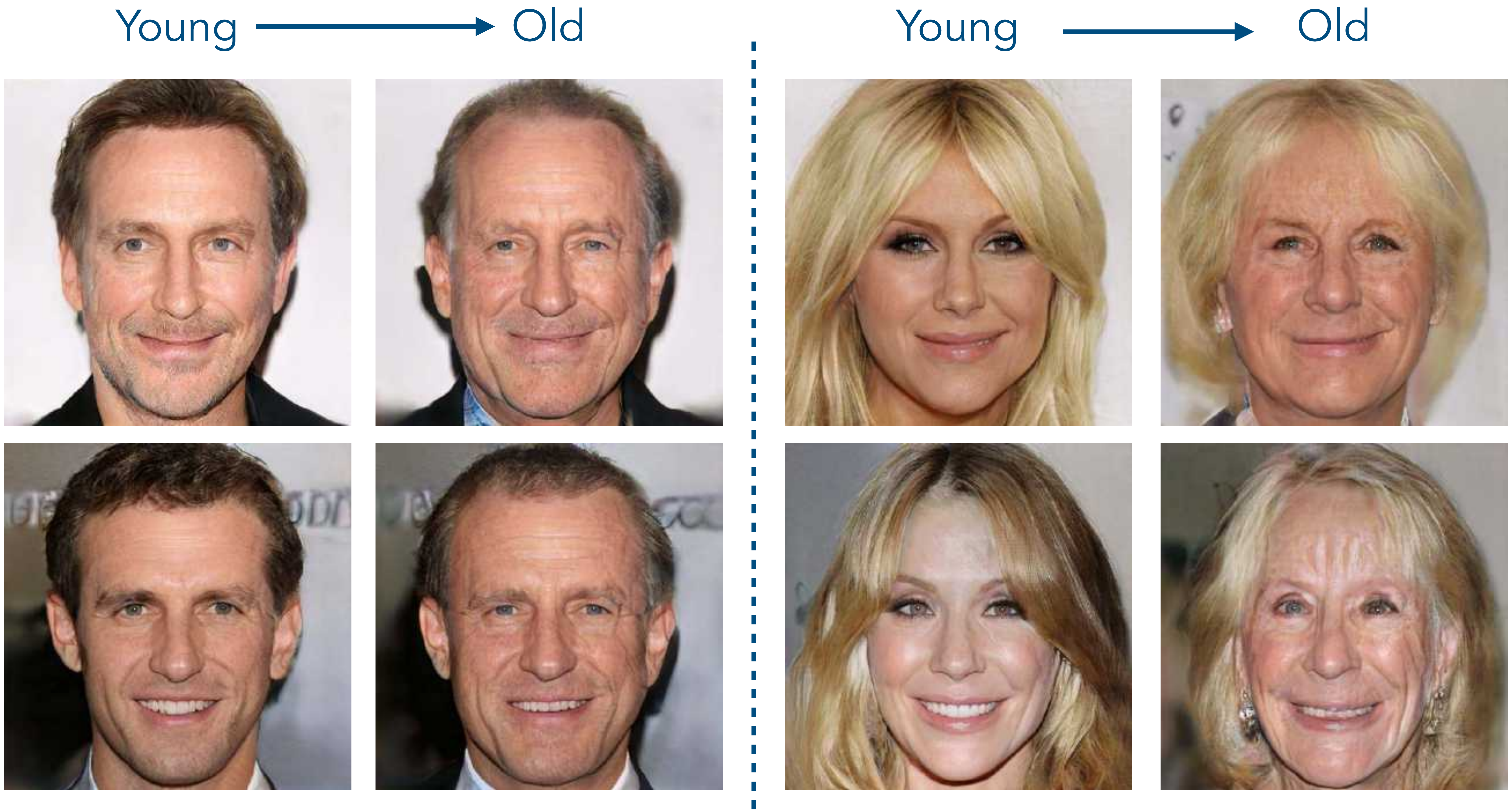
Face Image Editing: Change Gender



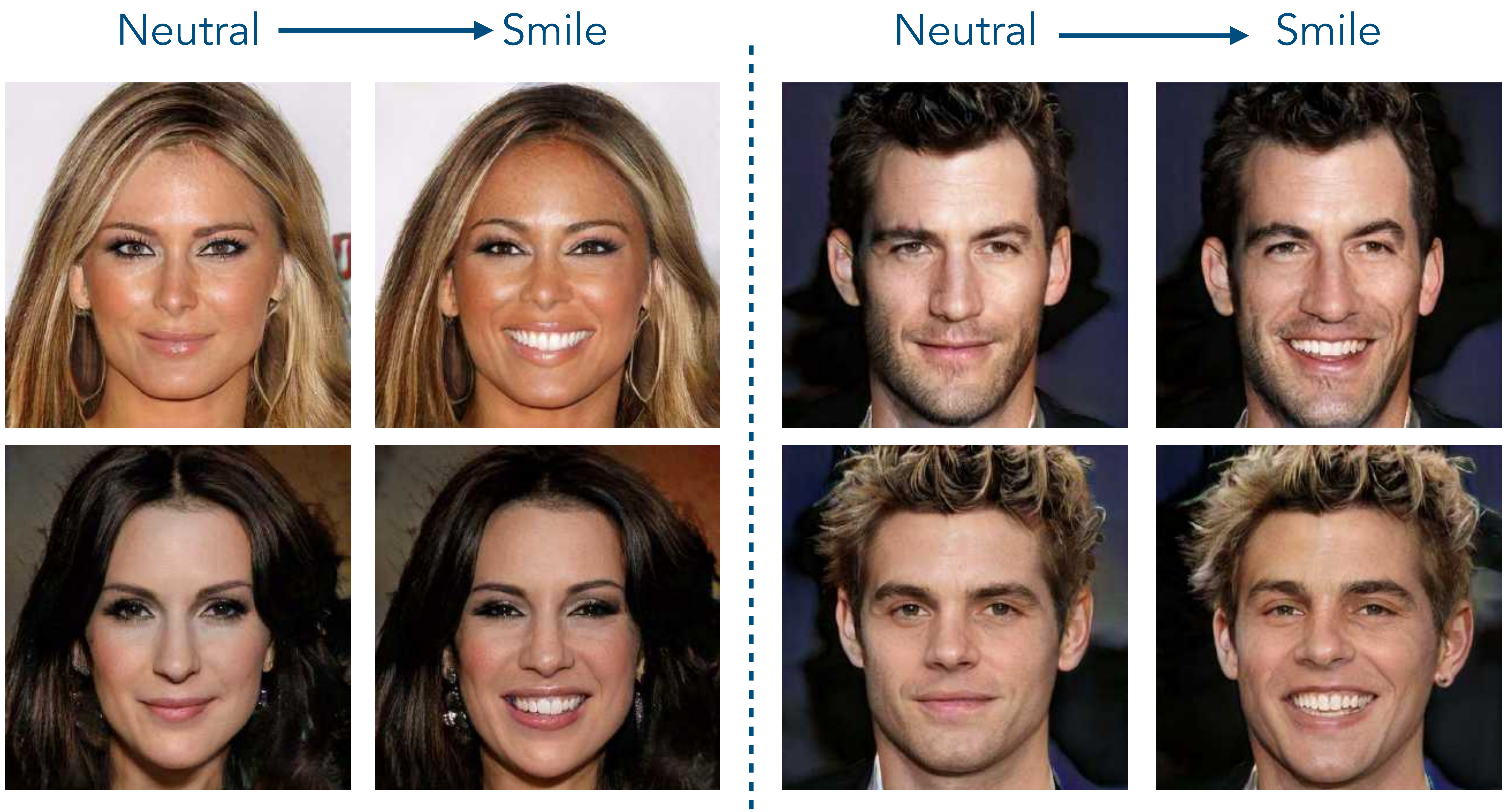
Face Image Editing: Wearing Glasses



Face Image Editing: Change Age

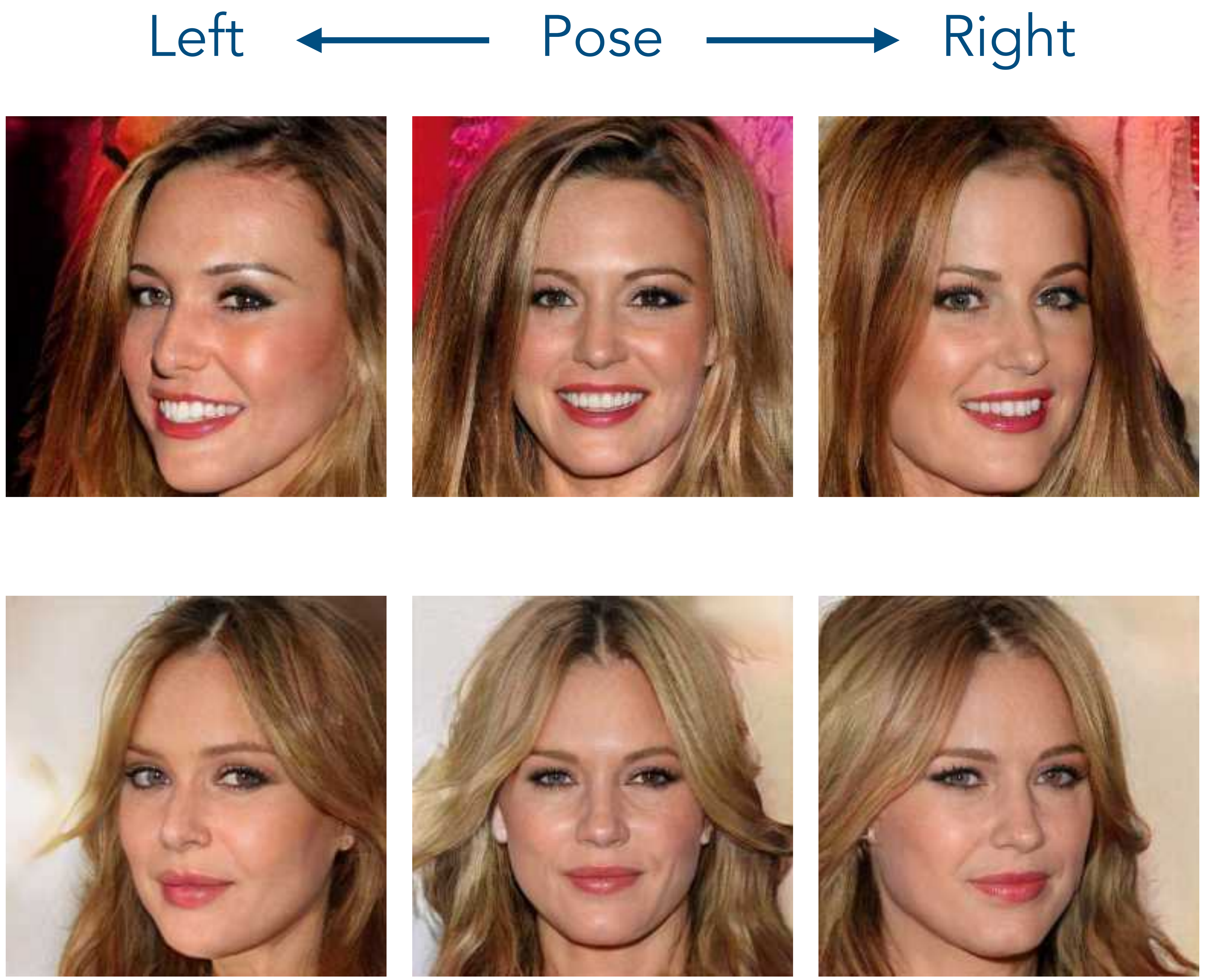


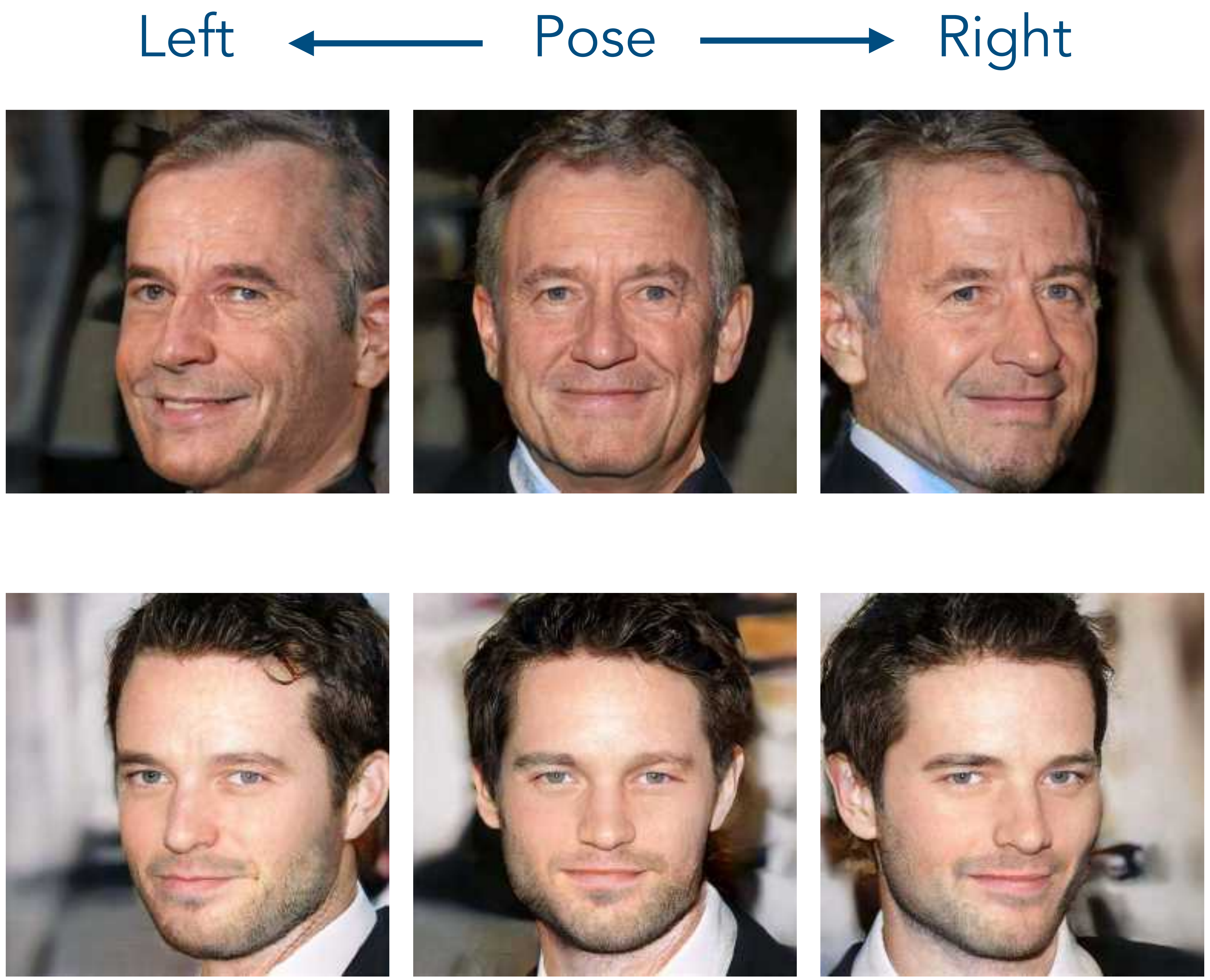
Face Image Editing: Smile!



Face Image Editing: Artifacts Removal

Face Image Editing: Change Pose





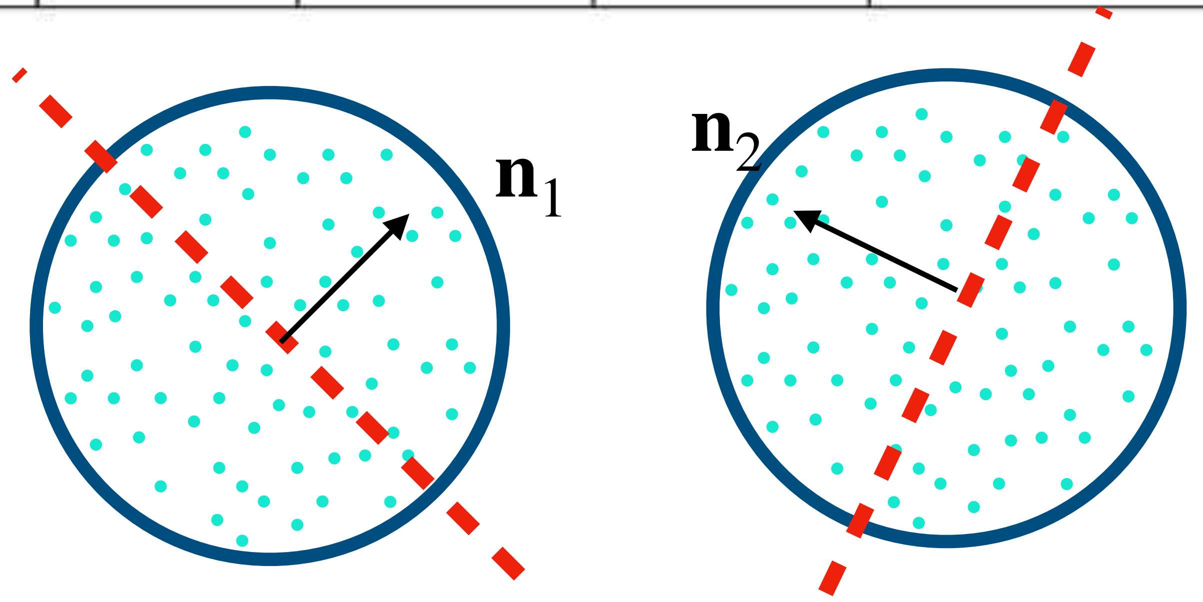
Correlation Between Attributes

Find correlation between attributes with two different metrics.

(1) We compute the cosine similarity between two norm vectors of boundaries.

$$\cos(\mathbf{n_1}, \mathbf{n_2}) = \mathbf{n_1^T n_2}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.04	-0.06	-0.05	-0.04
Smile	-	1.00	0.04	-0.10	-0.05
Age	-	-	1.00	0.49	0.38
Gender	-	-	-	1.00	0.52
Eyeglasses	-	-	-	-	1.00



(2) We treat each attribute score as a random variable, compute the correlation coefficient.

$$\rho_{A_1,A_2} = \frac{Cov(A_1, A_2)}{\sigma_{A_1}\sigma_{A_2}}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.01	-0.01	-0.02	0.00
Smile	-	1.00	0.02	-0.08	-0.01
Age	-	-	1.00	0.42	0.35
Gender	-	-	-	1.00	0.47
Eyeglasses	-	-	-	-	1.00

Correlation Between Attributes

Find correlation between attributes with two different metrics.

(1) We compute the cosine similarity between two norm vectors of boundaries.

$$\cos(\mathbf{n_1}, \mathbf{n_2}) = \mathbf{n_1^T n_2}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.04	-0.06	-0.05	-0.04
Smile	-	1.00	0.04	-0.10	-0.05
Age	-	-	1.00	0.49	0.38
Gender	-	-	-	1.00	0.52
Eyeglasses	-	-	-	-	1.00

(2) We treat each attribute score as a random variable, compute the correlation coefficient.

$$\rho_{A_1,A_2} = \frac{Cov(A_1, A_2)}{\sigma_{A_1}\sigma_{A_2}}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.01	-0.01	-0.02	0.00
Smile	-	1.00	0.02	-0.08	-0.01
Age	-	-	1.00	0.42	0.35
Gender	-	-	-	1.00	0.47
Eyeglasses	-	-	-	-	1.00

Attribute with low correlations
disentangled representation

Correlation Between Attributes

Find correlation between attributes with two different metrics.

(1) We compute the cosine similarity between two norm vectors of boundaries.

$$\cos(\mathbf{n}_1, \mathbf{n}_2) = \mathbf{n}_1^{\mathrm{T}}\mathbf{n}_2$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.04	-0.06	-0.05	-0.04
Smile	-	1.00	0.04	-0.10	-0.05
Age	-	-	1.00	0.49	0.38
Gender	-	-	-	1.00	0.52
Eyeglasses	-	-	-	-	1.00

(2) We treat each attribute score as a random variable, compute the correlation coefficient.

$$\rho_{A_1,A_2} = \frac{Cov(A_1, A_2)}{\sigma_{A_1}\sigma_{A_2}}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.01	-0.01	-0.02	0.00
Smile	-	1.00	0.02	-0.08	-0.01
Age	-	-	1.00	0.42	0.35
Gender	-	-	-	1.00	0.47
Eyeglasses	-	-	-	-	1.00

Attribute with high correlations

Correlation Between Attributes

Find correlation between attributes with two different metrics.

(1) We compute the cosine similarity between two norm vectors of boundaries.

$$\cos(\mathbf{n}_1, \mathbf{n}_2) = \mathbf{n}_1^T \mathbf{n}_2$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.04	-0.06	-0.05	-0.04
Smile	-	1.00	0.04	-0.10	-0.05
Age	-	-	1.00	0.49	0.38
Gender	-	-	-	1.00	0.52
Eyeglasses	-	-	-	-	1.00

(2) We treat each attribute score as a random variable, compute the correlation coefficient.

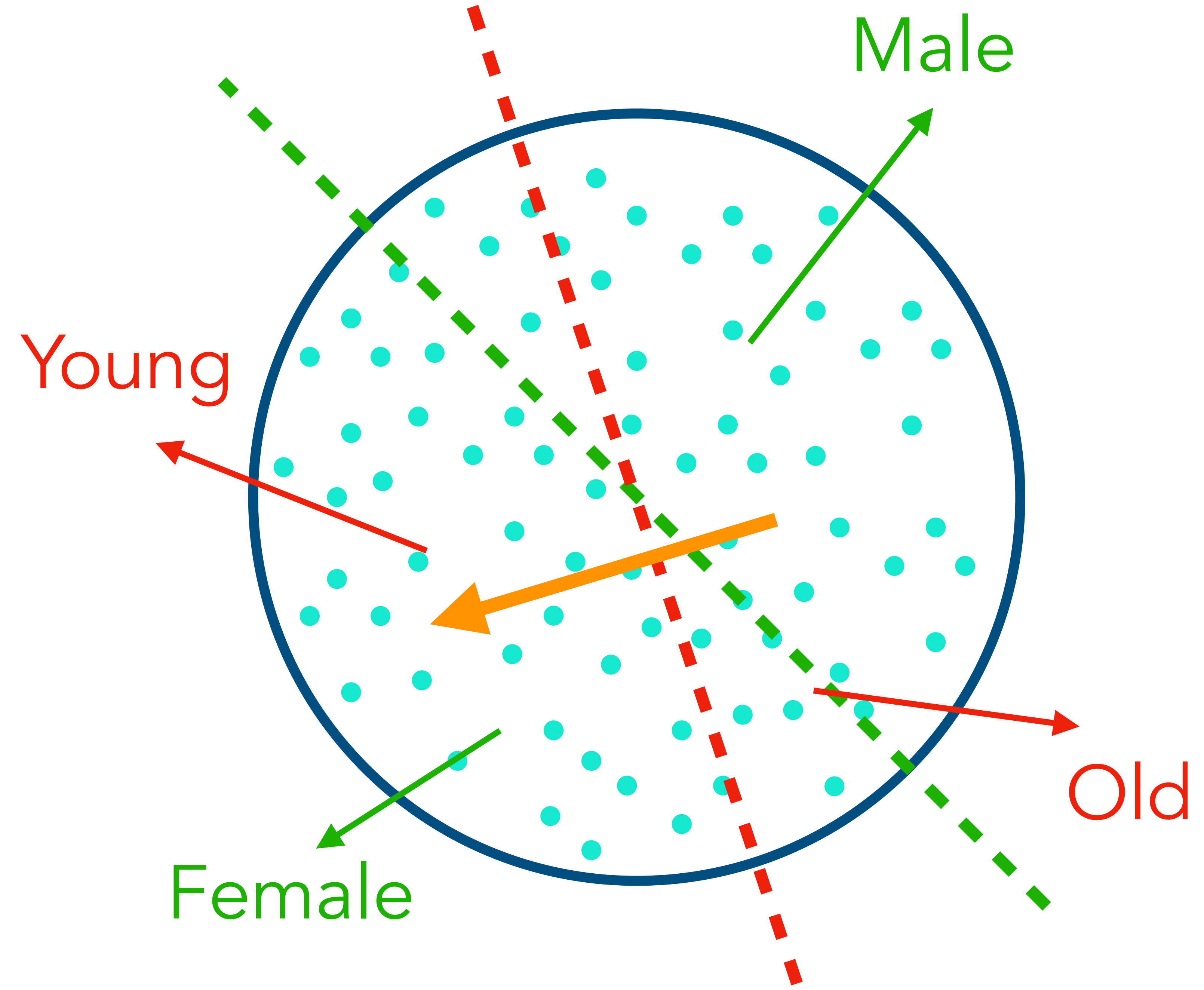
$$\rho_{A_1,A_2} = \frac{Cov(A_1, A_2)}{\sigma_{A_1}\sigma_{A_2}}$$

	Pose	Smile	Age	Gender	Eyeglasses
Pose	1.00	-0.01	-0.01	-0.02	0.00
Smile	-	1.00	0.02	-0.08	-0.01
Age	-	-	1.00	0.42	0.35
Gender	-	-	-	1.00	0.47
Eyeglasses	-	-	-	-	1.00

Does data affect the learned representation?

Conditional Manipulation

High correlation between "Age" and "Gender", which is of 0.49.

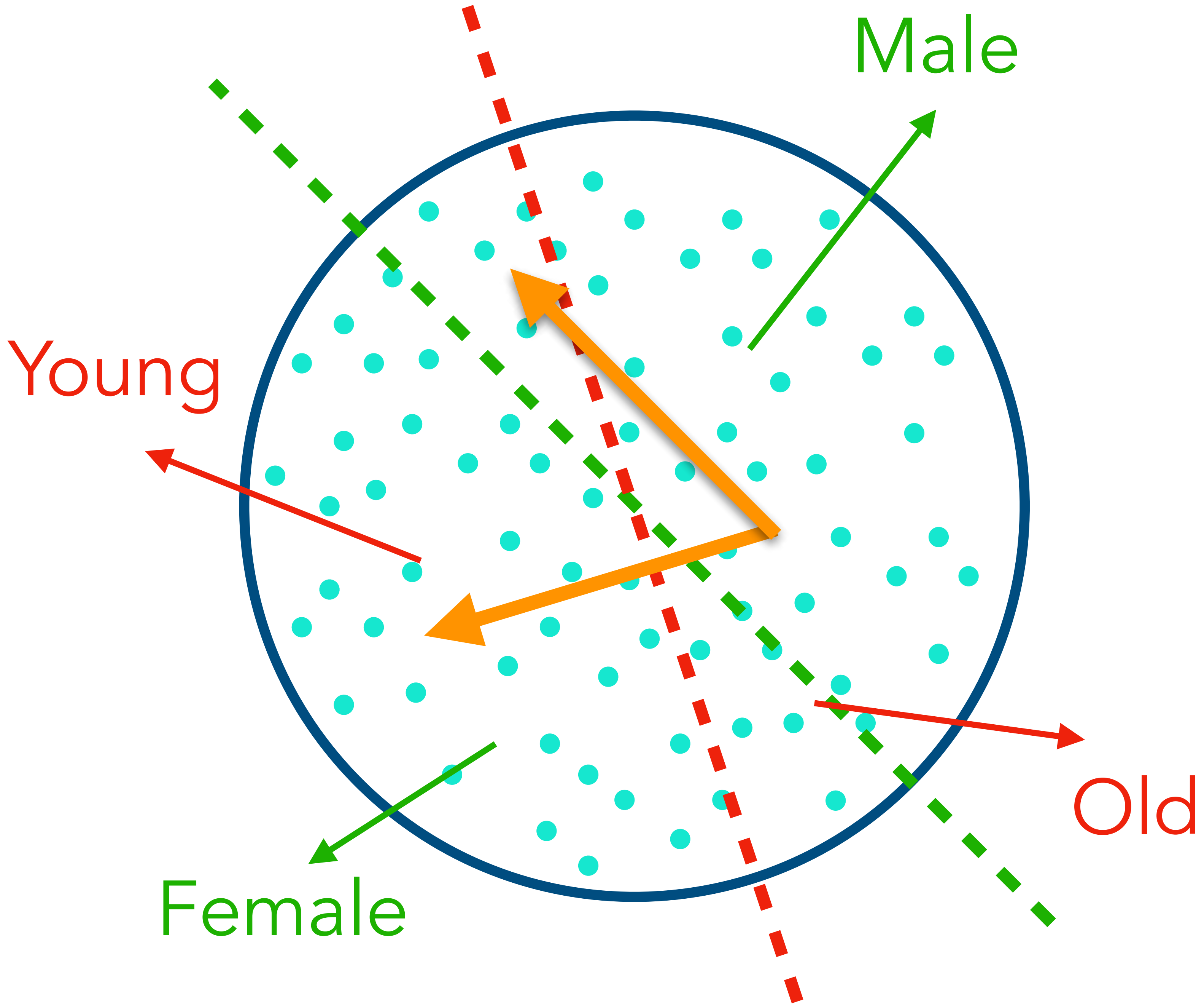


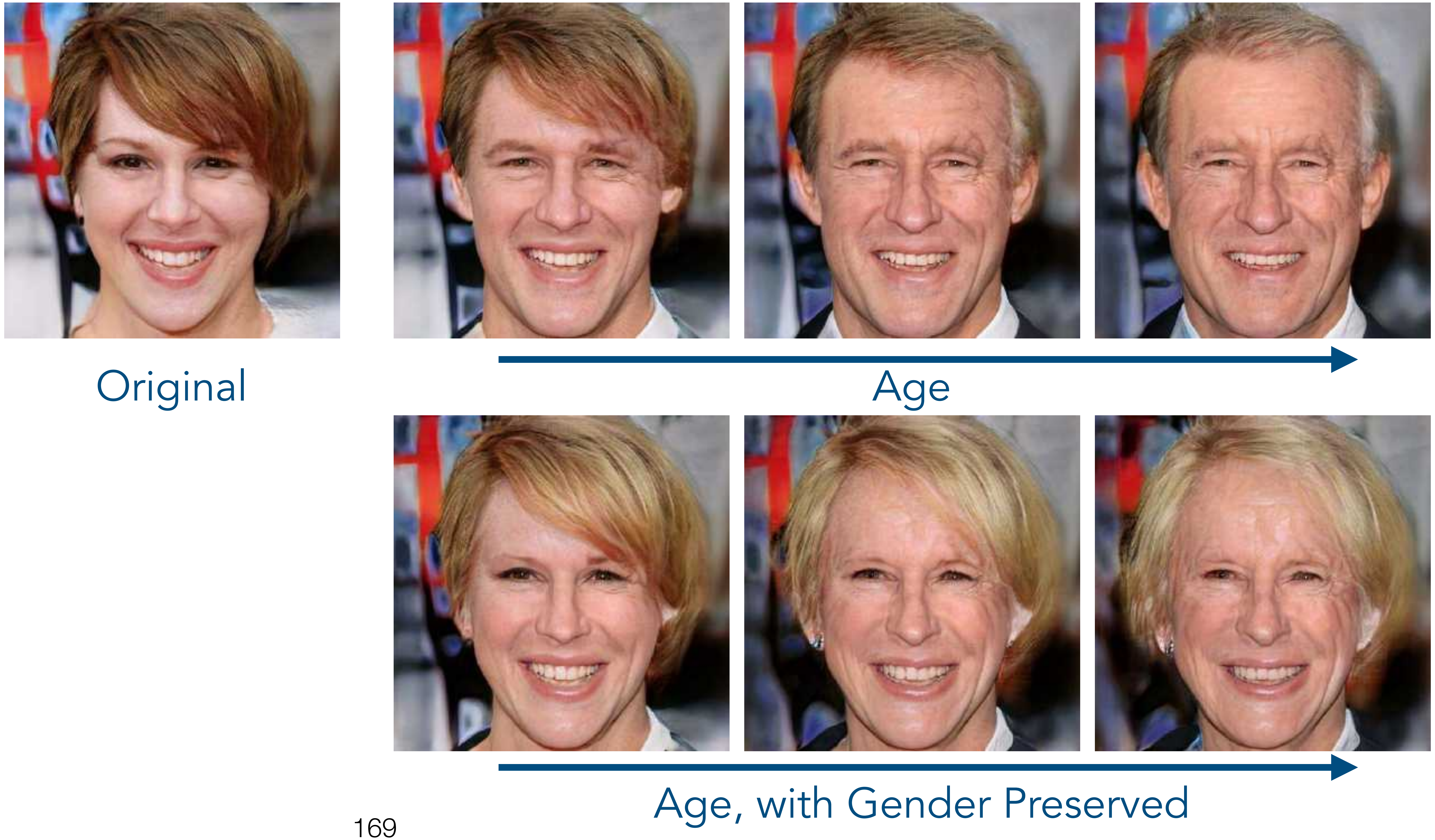
Original

Change Age, the gender changes as well

Conditional Manipulation

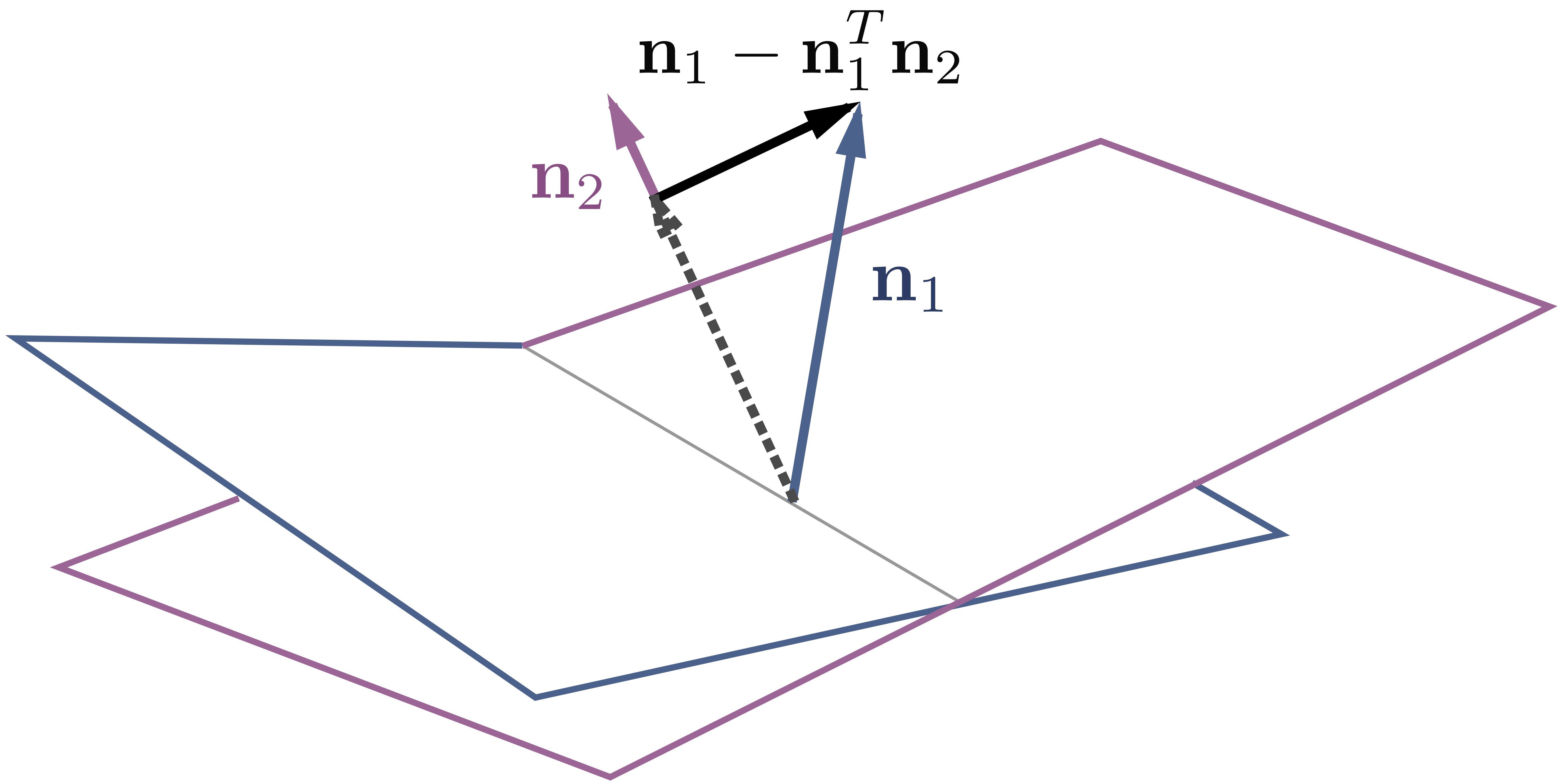
High correlation between "Age" and "Gender", which is of 0.49.



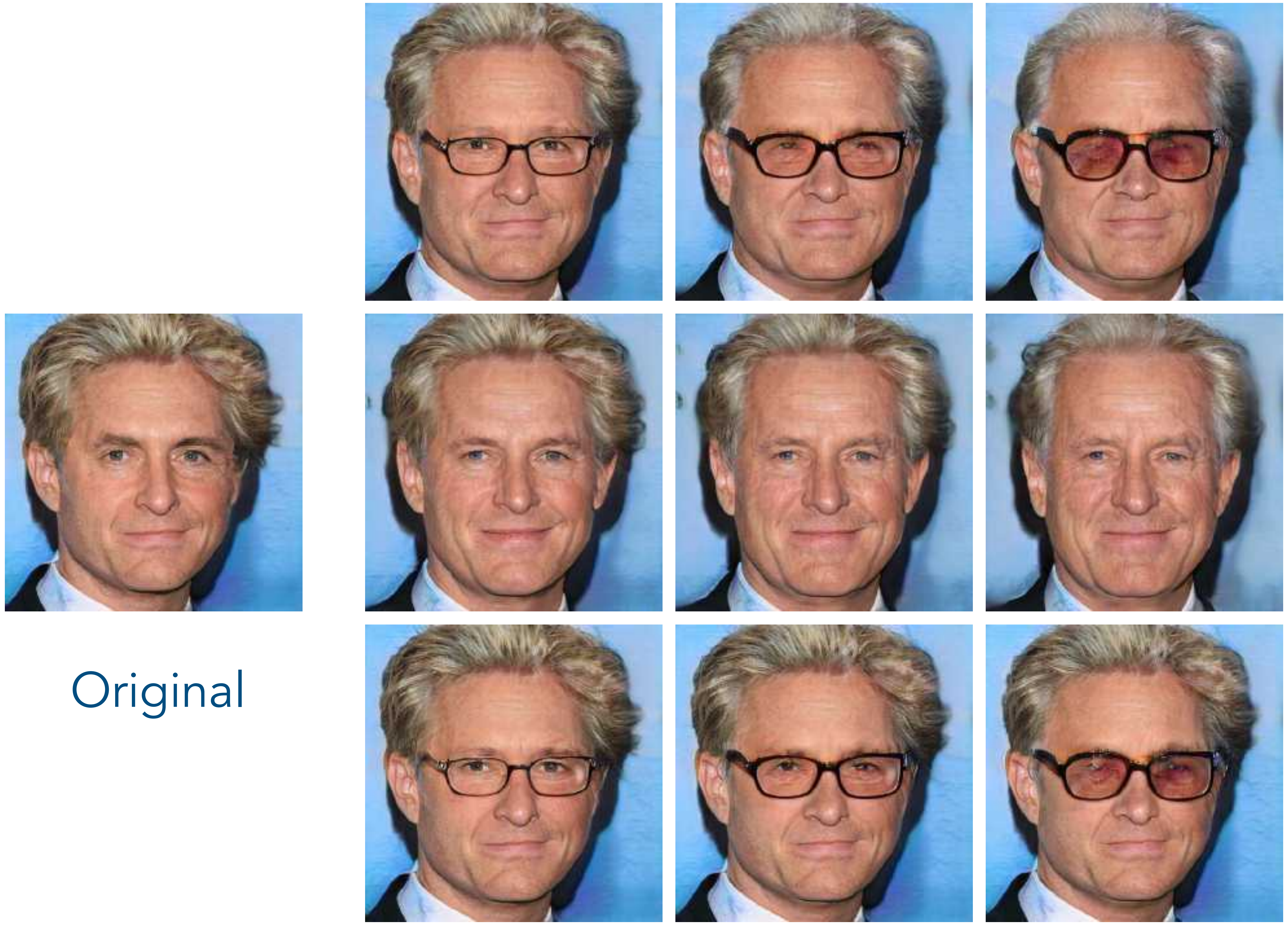


Conditional Manipulation

We can find projected direction, such that moving samples along this new direction can change "attribute 1" without affecting "attribute 2".



Conditional Manipulation



Glasses

Age

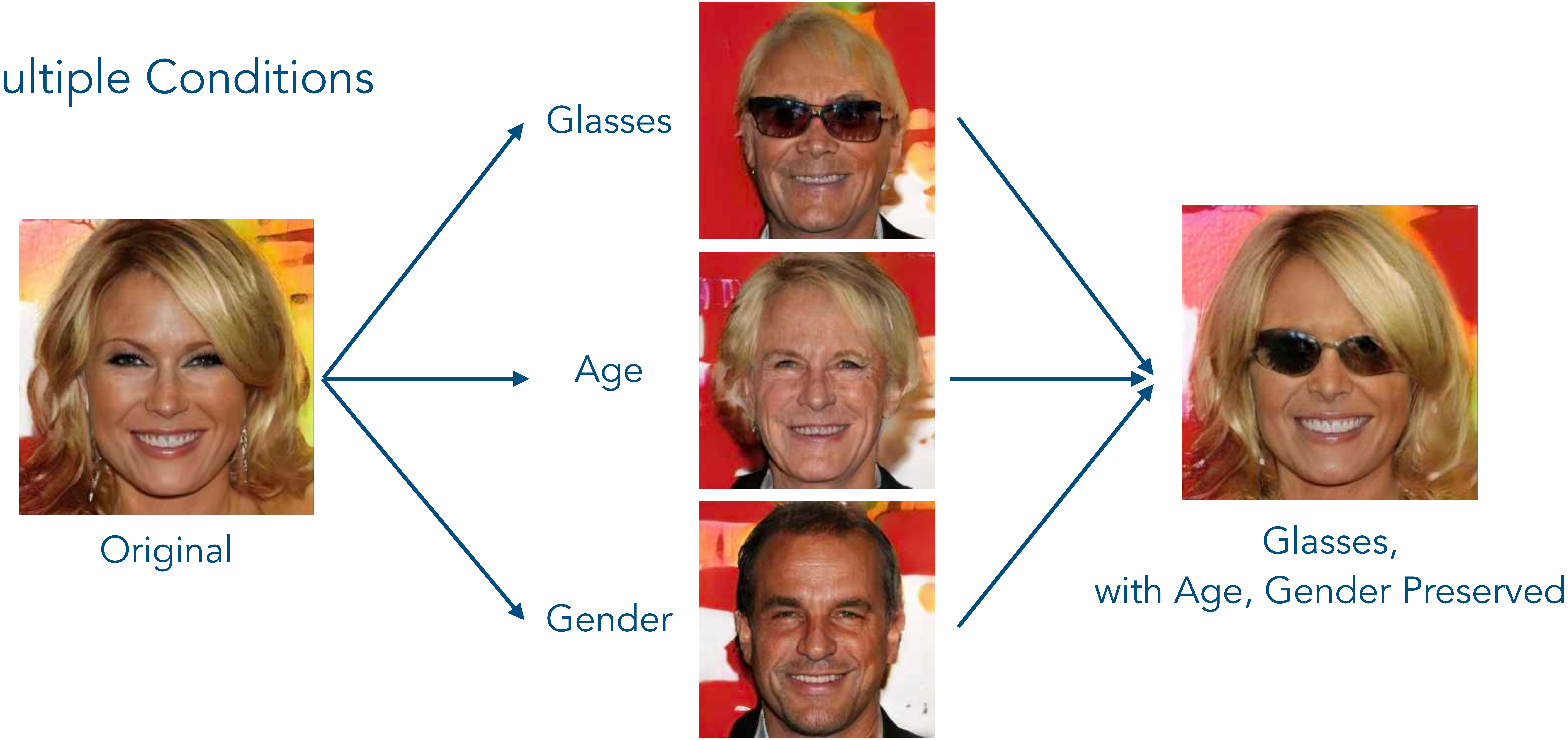
Original

Glasses,

with Age Preserved

Conditional Manipulation

Multiple Conditions



Manipulate Synthesize Scene Images

How to Manipulate Real Image?

So far, we can only manipulate synthesize images.

How to manipulate an arbitrary real image?

Spider-Man

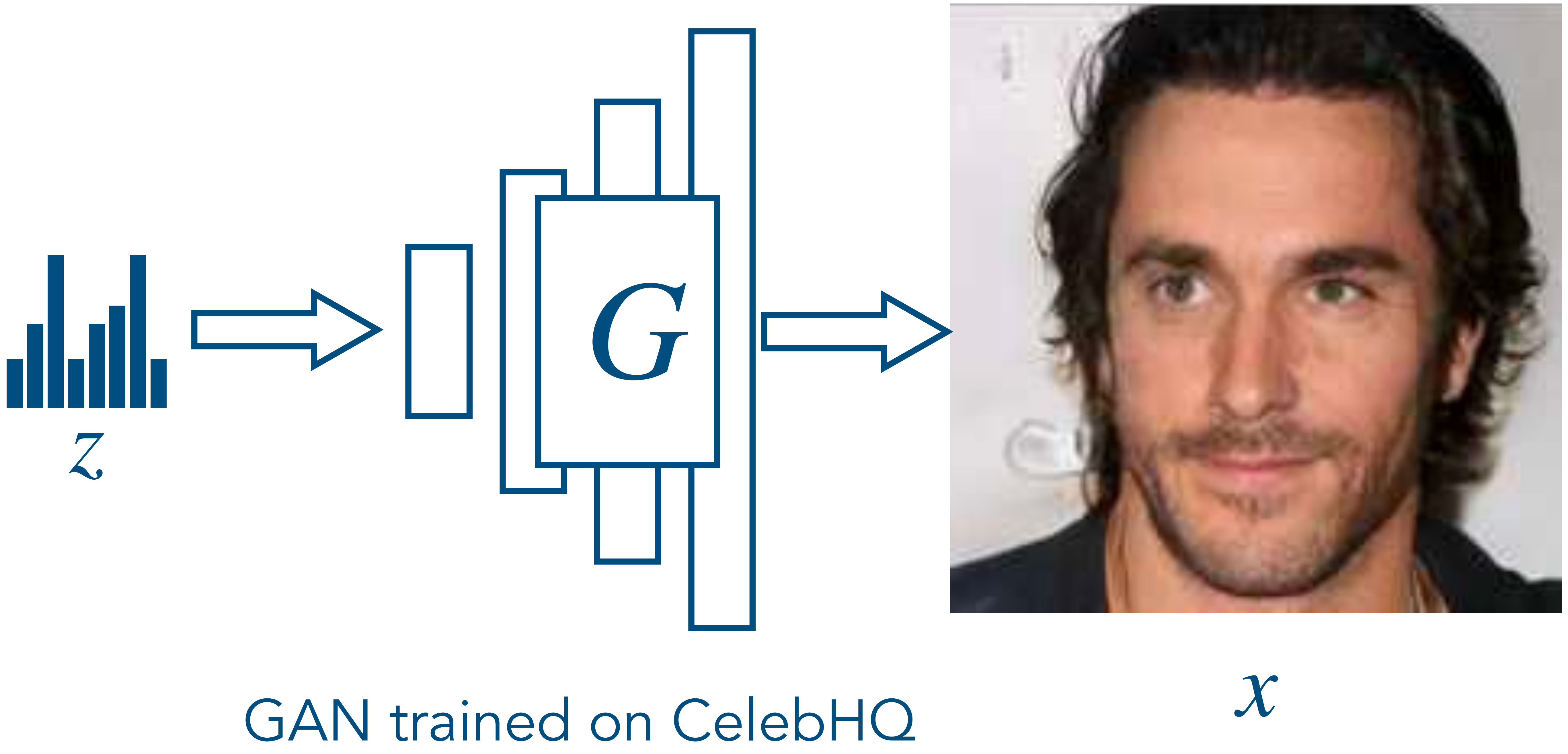
Thor

Loki

How to manipulate attributes for them?

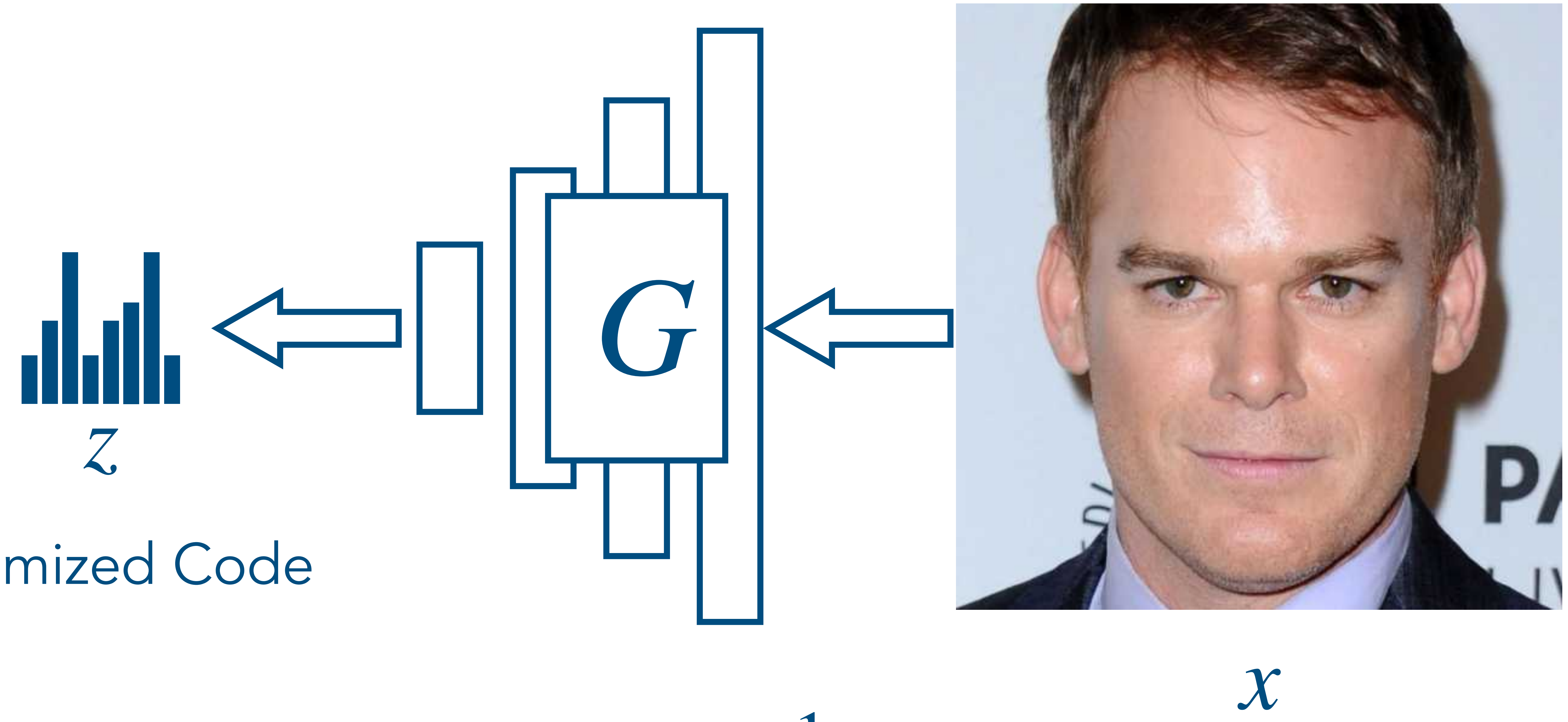
GAN Inversion for Processing Real Images

Find latent code for real image.



GAN Inversion for Processing Real Images

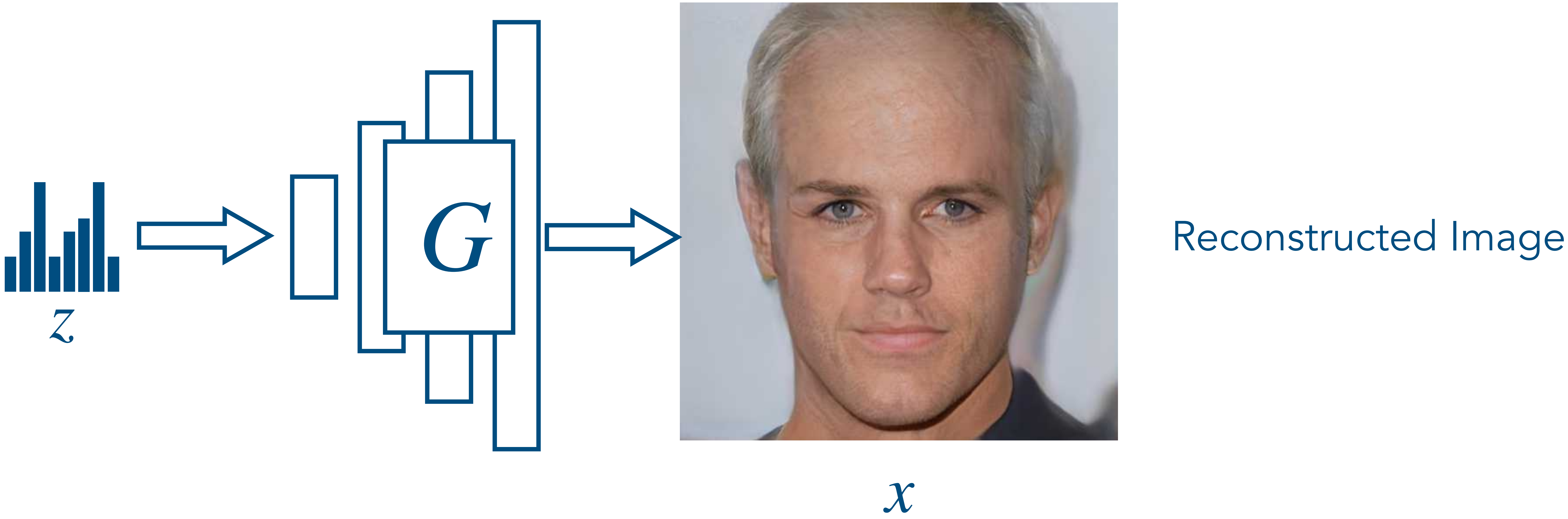
Find latent code by directly optimizing latent code.



$$z^* = \arg\min_{z} \frac{1}{2} \|G(z) - x\|_2^2$$

GAN Inversion for Processing Real Images

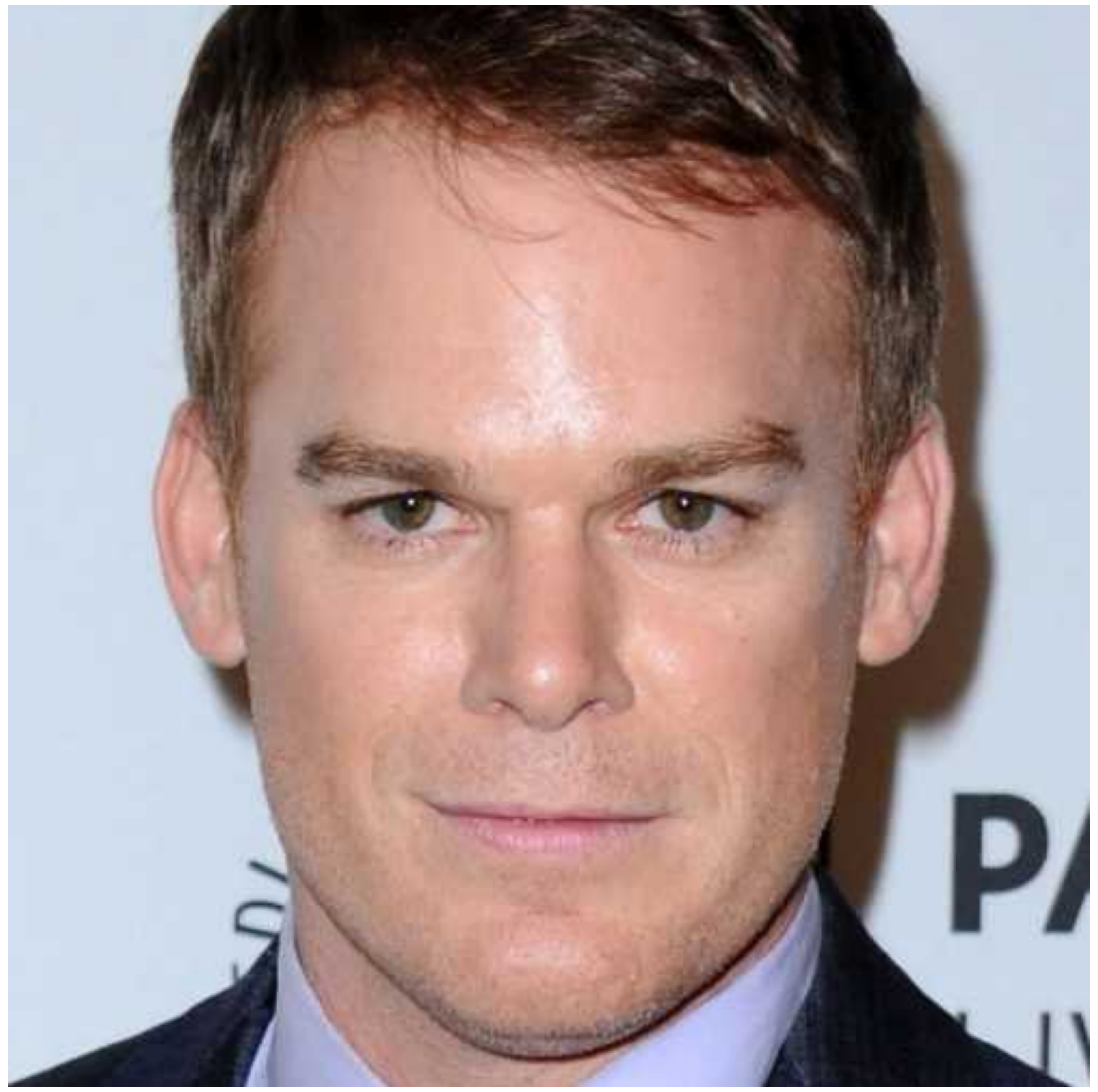
Find latent code by directly optimizing latent code.



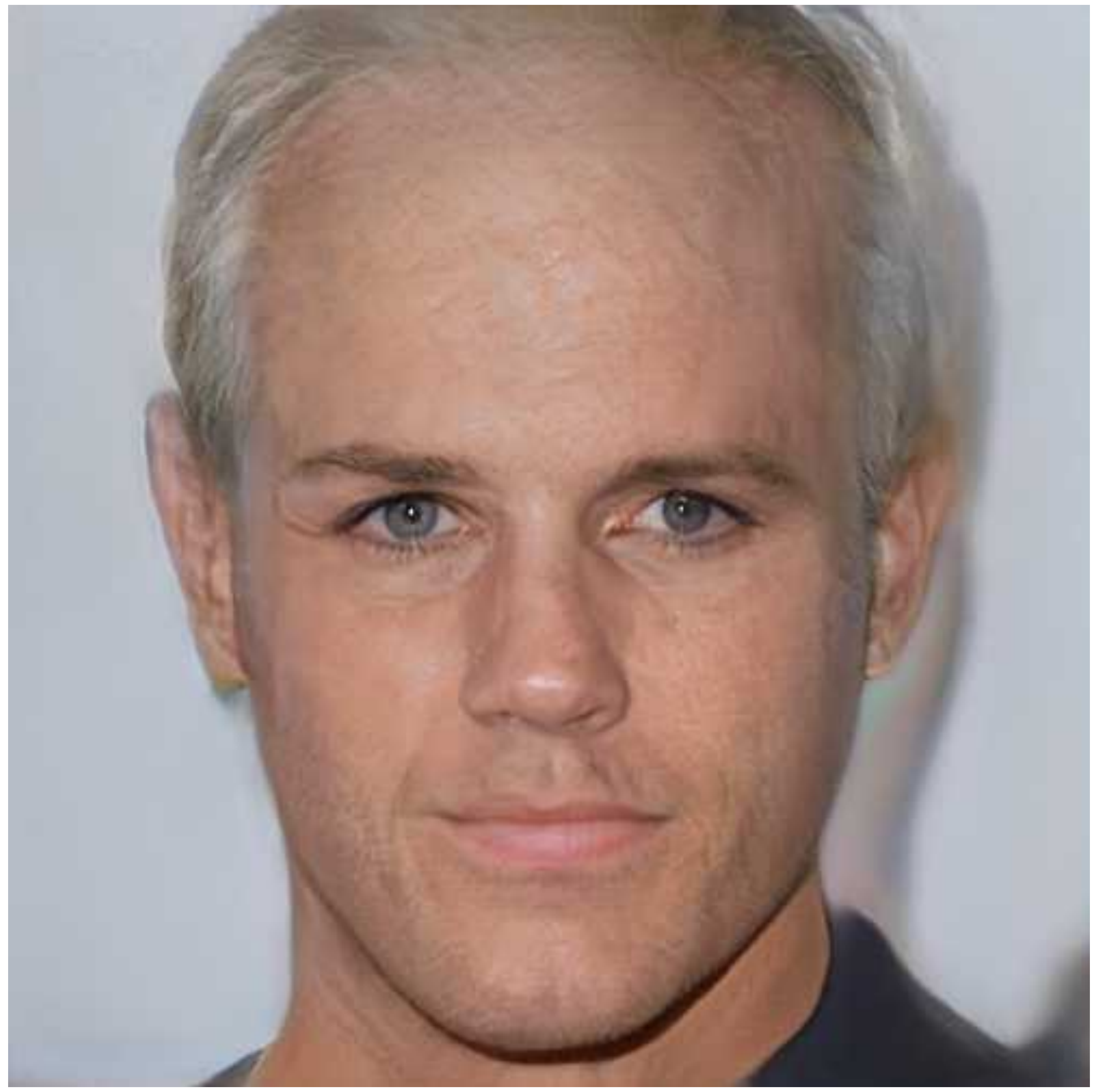
$$z^* = \arg\min_z \frac{1}{2} \|G(z) - x\|_2^2$$

Bias in GANs

Real Face



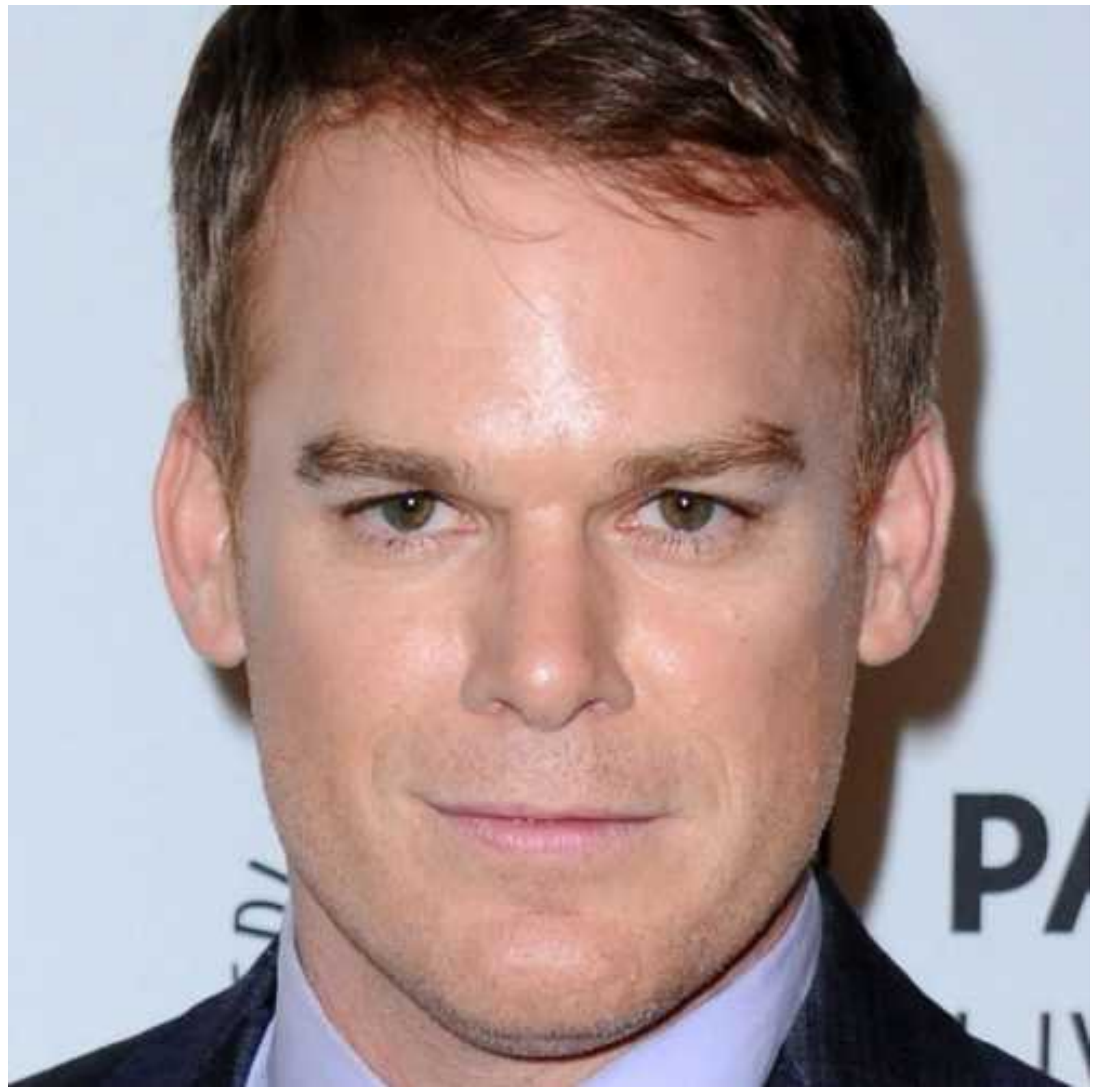
Reconstructed Face



Far from ideal!
Can't support real
image processing

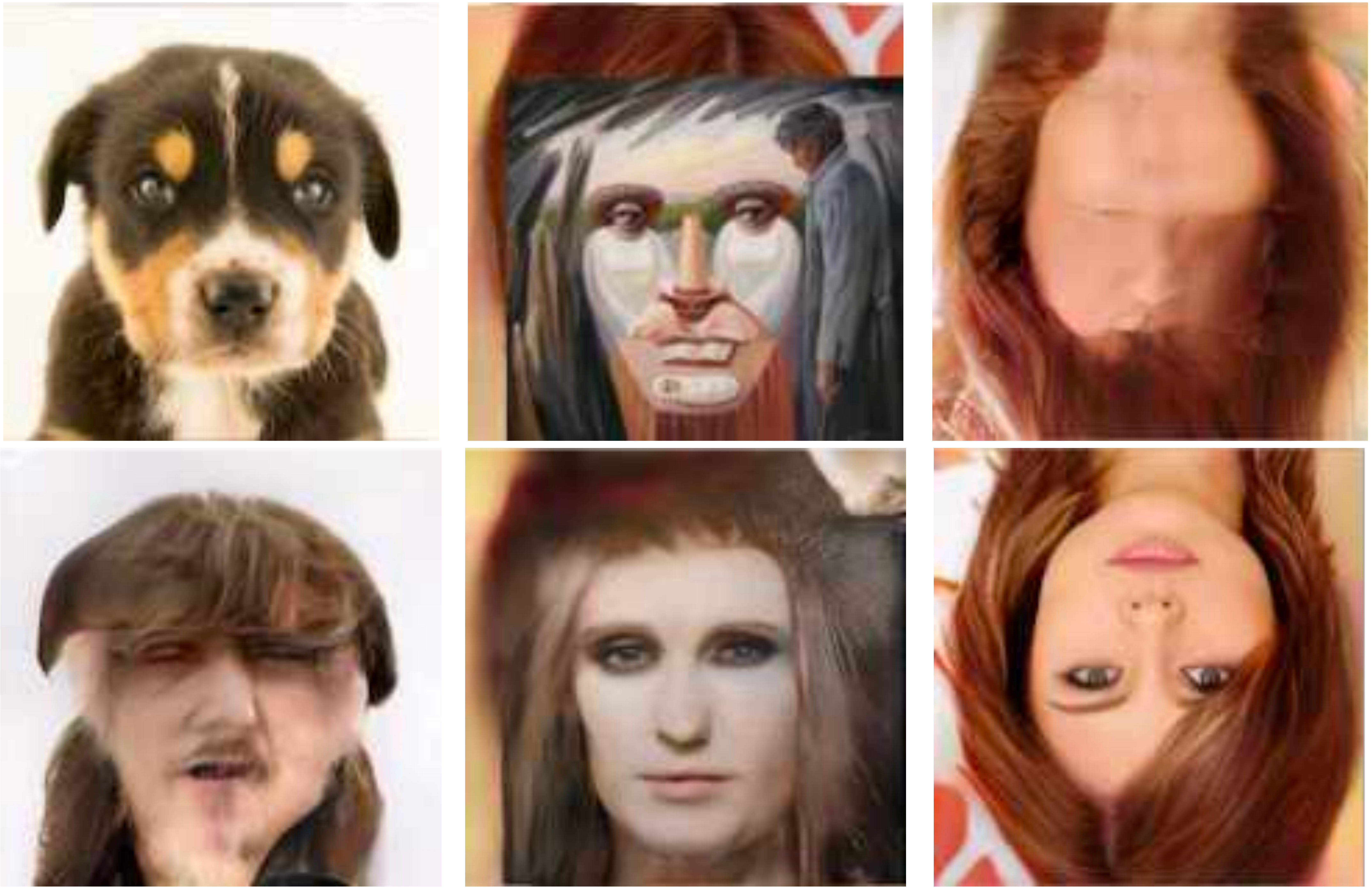
Bias in GANs

Real Face

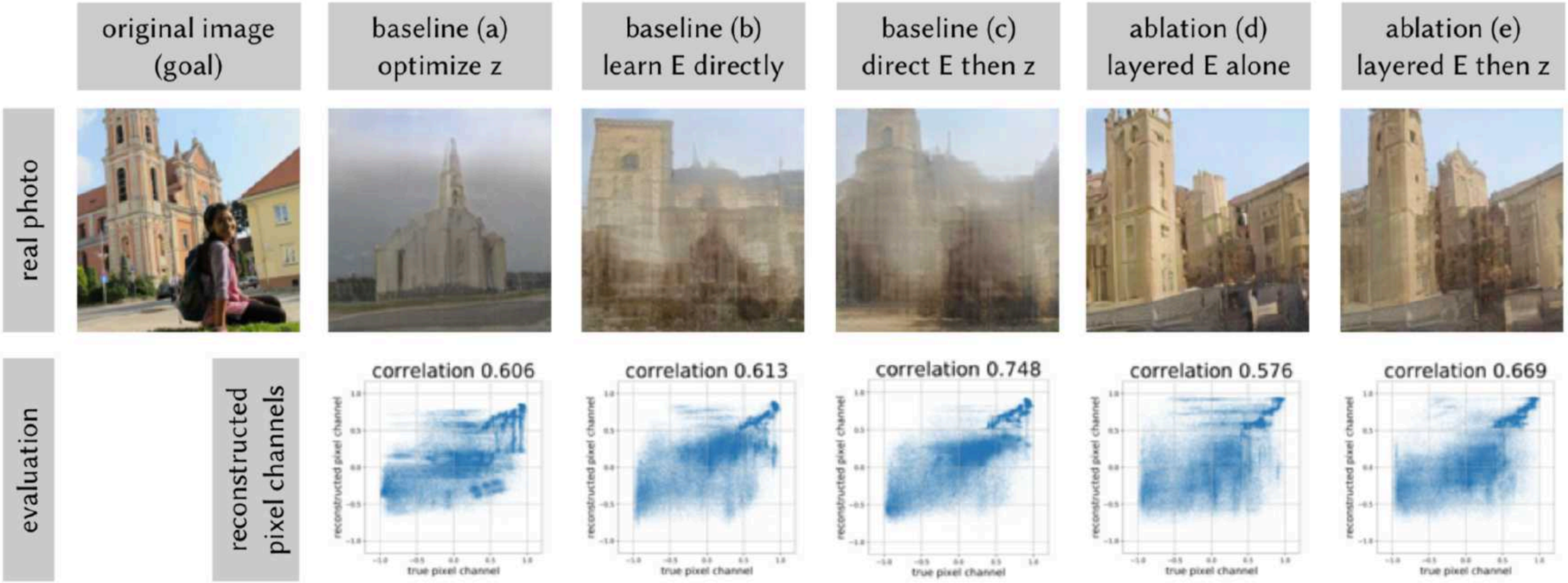


Reconstruct Multiple Times

GAN Inversion of Non-face Images Using Face Model



Previous GAN Inversion Methods



[Bau, D., Zhu, J. Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., & Torralba, A. Seeing what a gan cannot generate. ICCV 2019]

Why GAN Inversion is challenging?

$$z^* = \arg\min_z \frac{1}{2} ||G(z) - x||_2^2$$

1. This optimization Problem is non-convex and non-smooth.

2. The expressiveness of the latent code is limited.

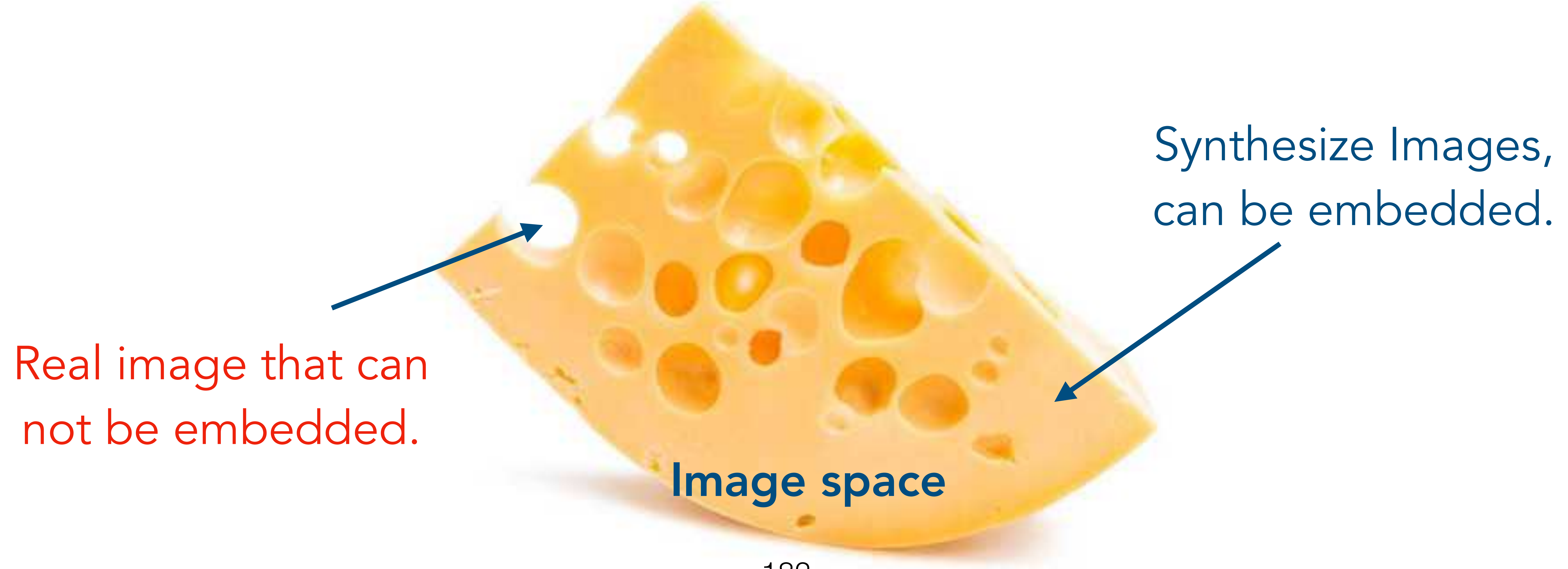
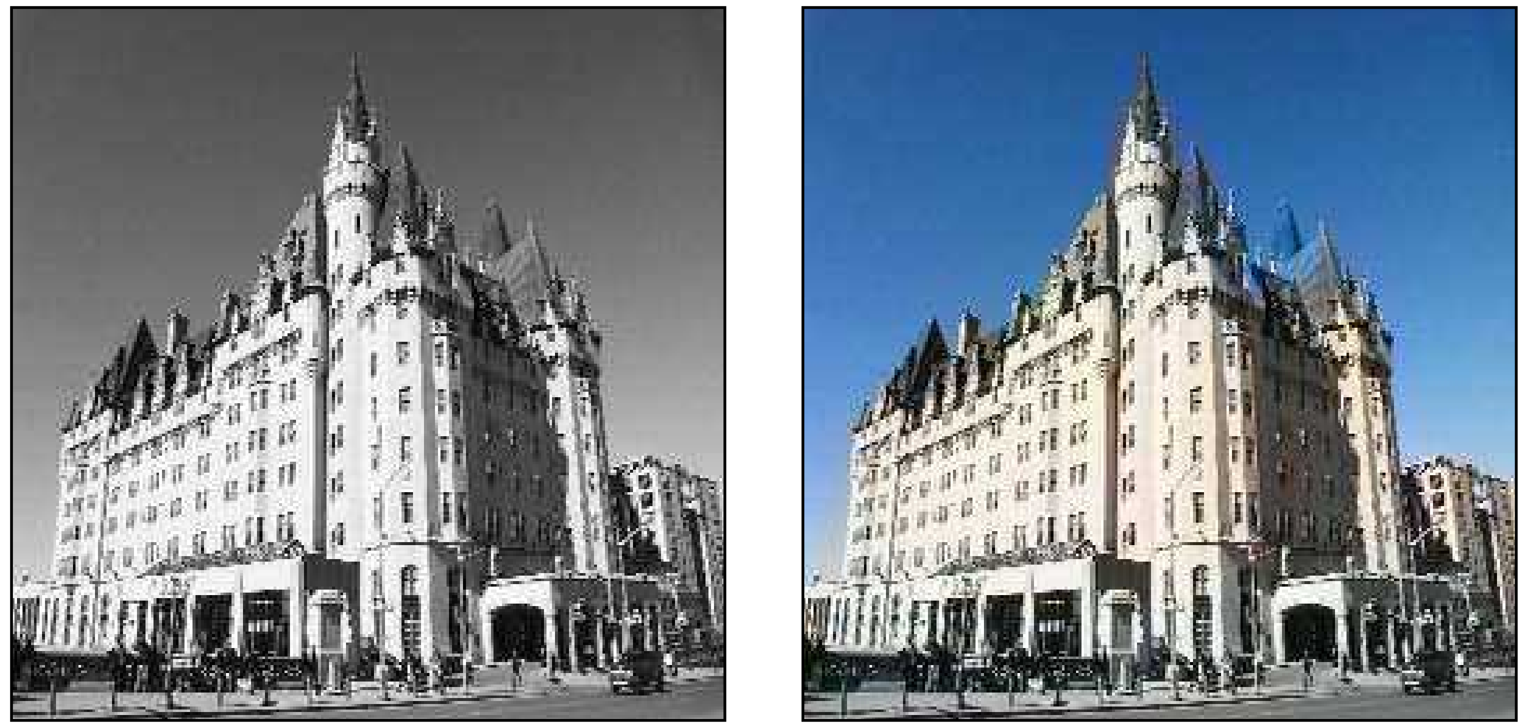


Image Processing Using Multi-Code GAN Prior

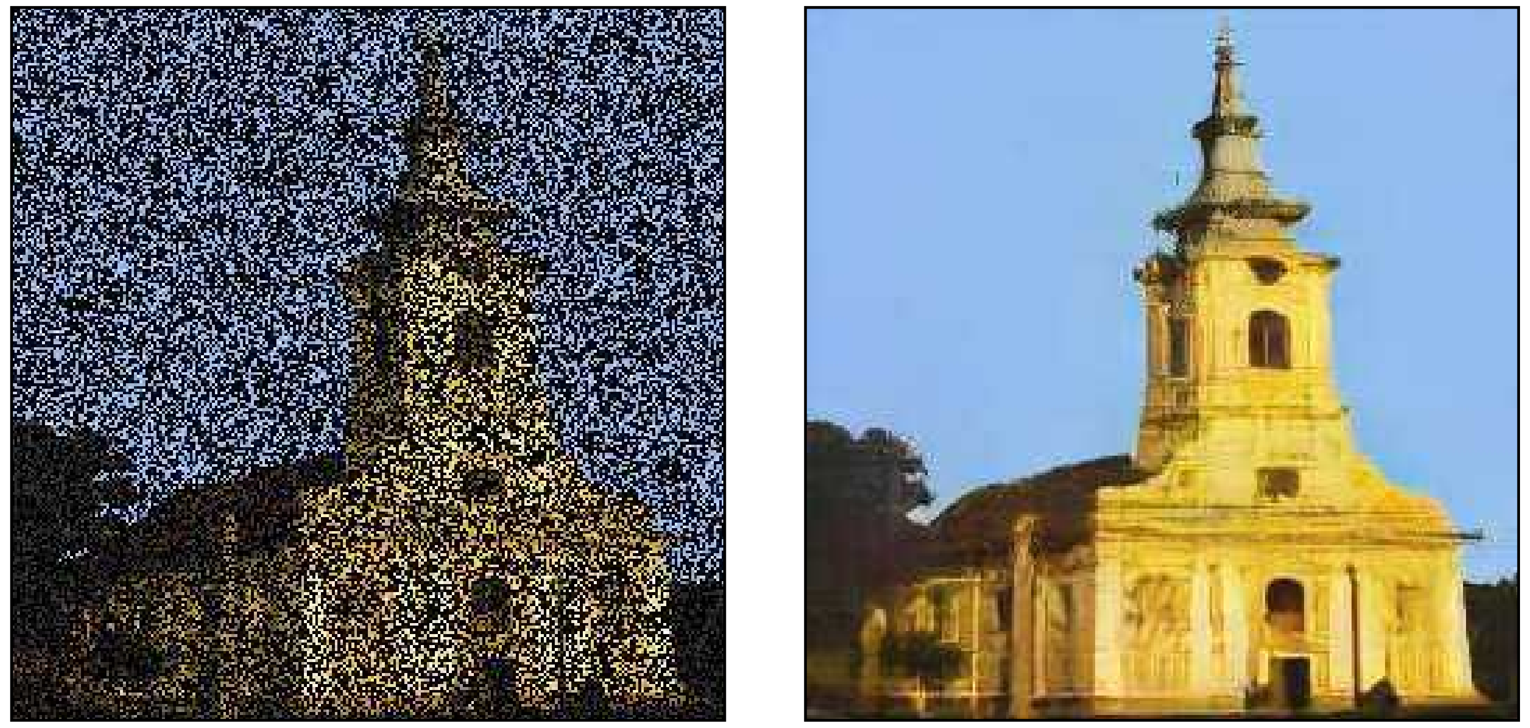
Jinjin Gu, Yujun Shen, Bolei Zhou, CVPR 2020

(a) Image Reconstruction

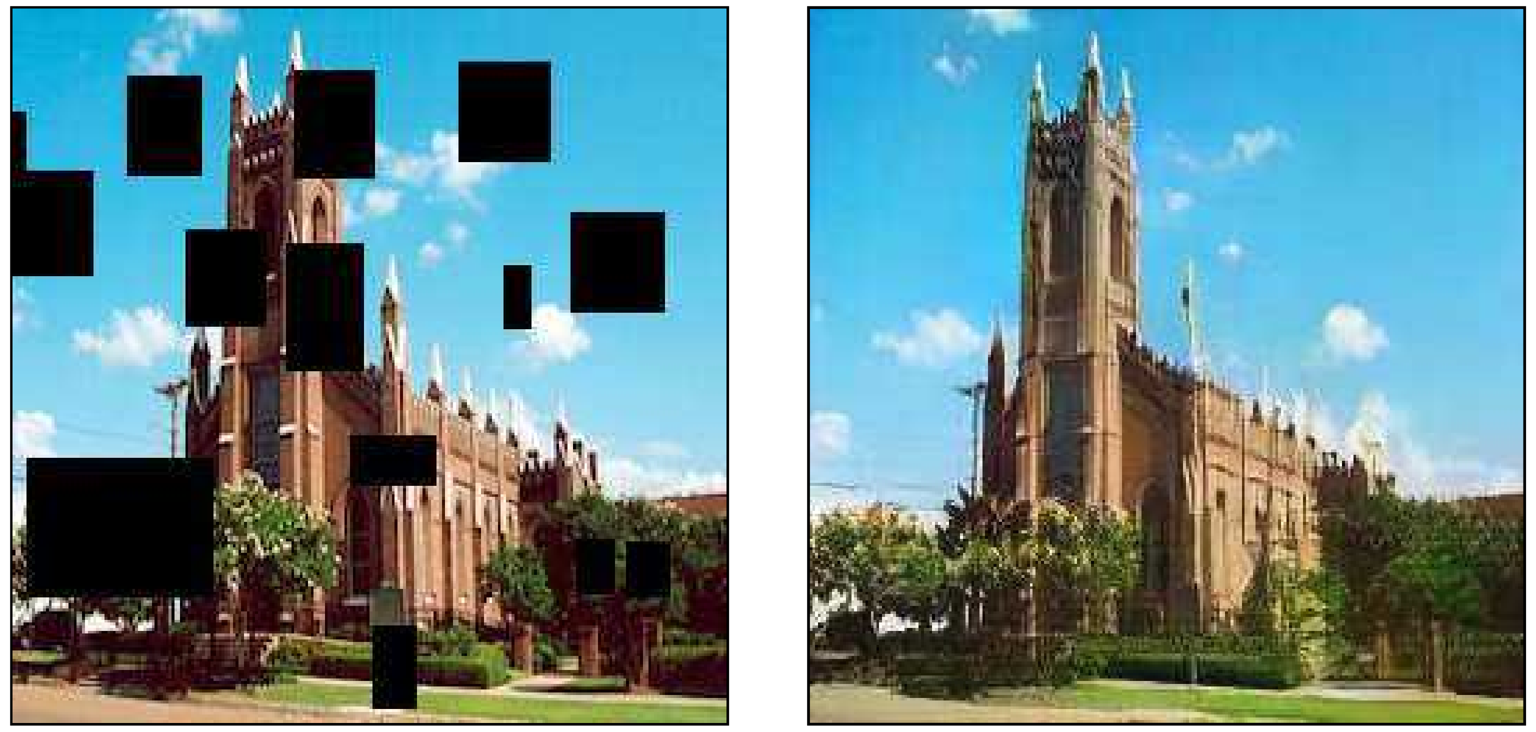


(b) Image Colorization

(c) Image Super-Resolution



(d) Image Denoising



(e) Image Inpainting

(f) Semantic Manipulation

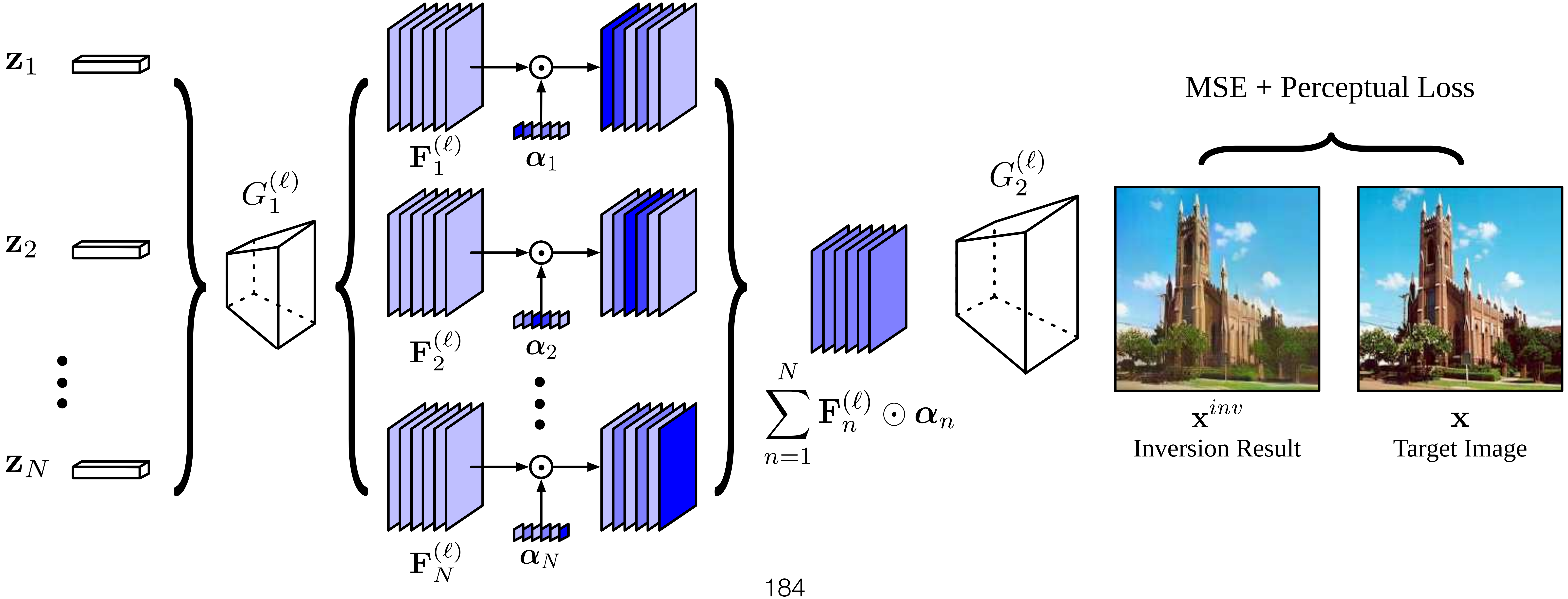
Multi-Code GAN Inversion

$$\{\mathbf{z}_n^*\}_{n=1}^N, \{\alpha_n^*\}_{n=1}^N = \underset{\{\mathbf{z}_n\}_{n=1}^N,\{\alpha_n\}_{n=1}^N}{\arg\min} L(\boxed{x^{inv}}, x)$$

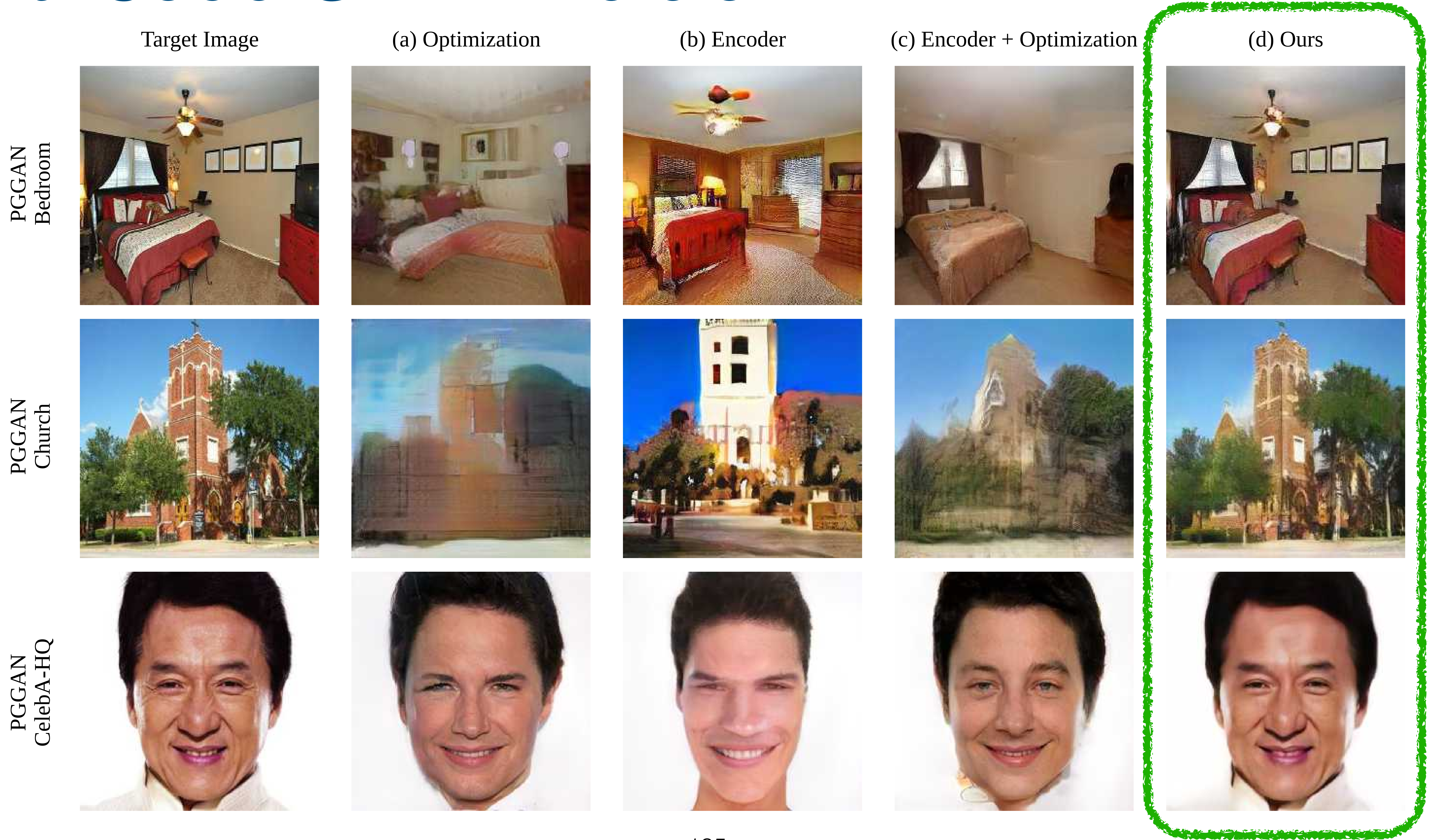
$$x^{inv} = G_2^{(l)}\Big(\sum_{n=1}^N \boxed{\mathbf{F}_n^{(l)} \odot \alpha_n}\Big)$$

$$\{\mathbf{F}_n^{(l)} \odot \alpha_n\}_{i,j,c} = \{\mathbf{F}_n^{(l)}\}_{i,j,c} \times \{\alpha_n\}_c$$

Optimization Objective



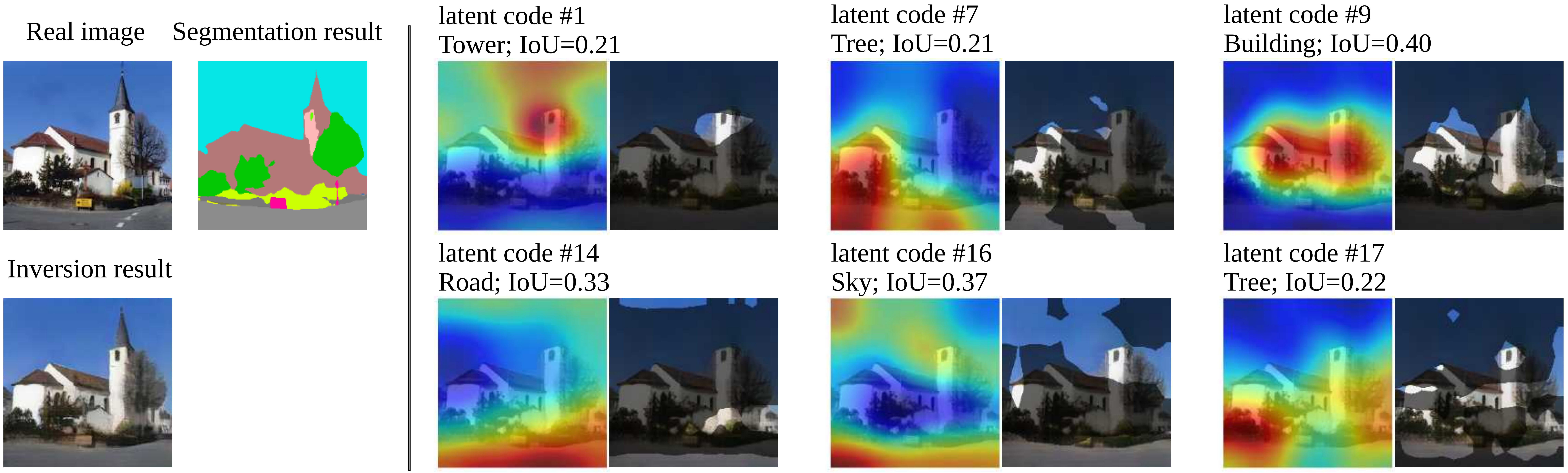
Multi-Code GAN Inversion



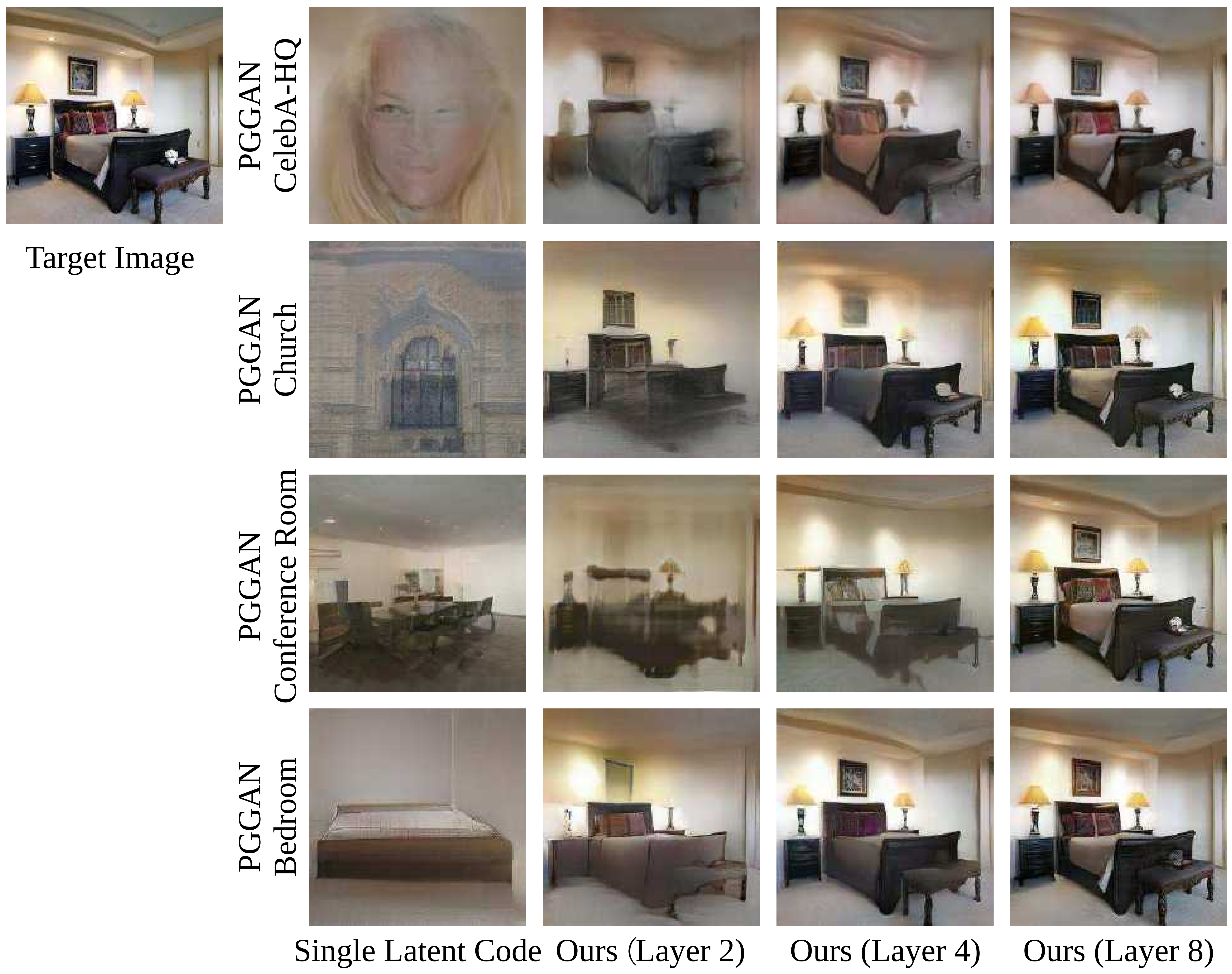
Multi-Code GAN Inversion

$$\mathrm{IoU}_{\mathbf{z}_n,c} = \frac{(\mathbf{r}'_n > t) \wedge \mathbf{s}_c(x^{inv})}{(\mathbf{r}'_n > t) \vee \mathbf{s}_c(x^{inv})}$$

Role of each latent code in Multi-Code Inversion



Knowledge Representation in GANs

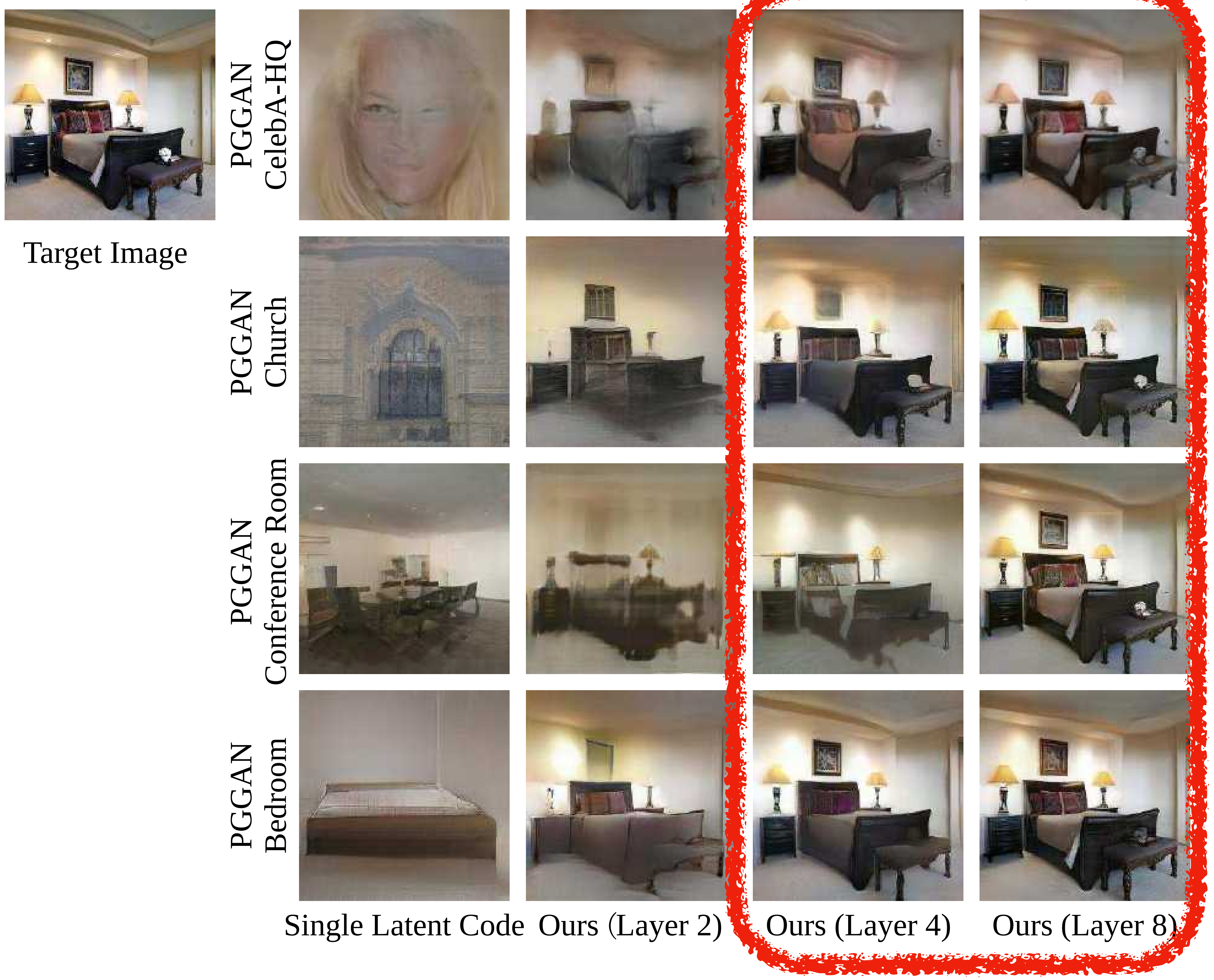


Knowledge Representation in GANs, Different Layers

Reconstruction with early layers.

Using high-level components, such as objects and regions.

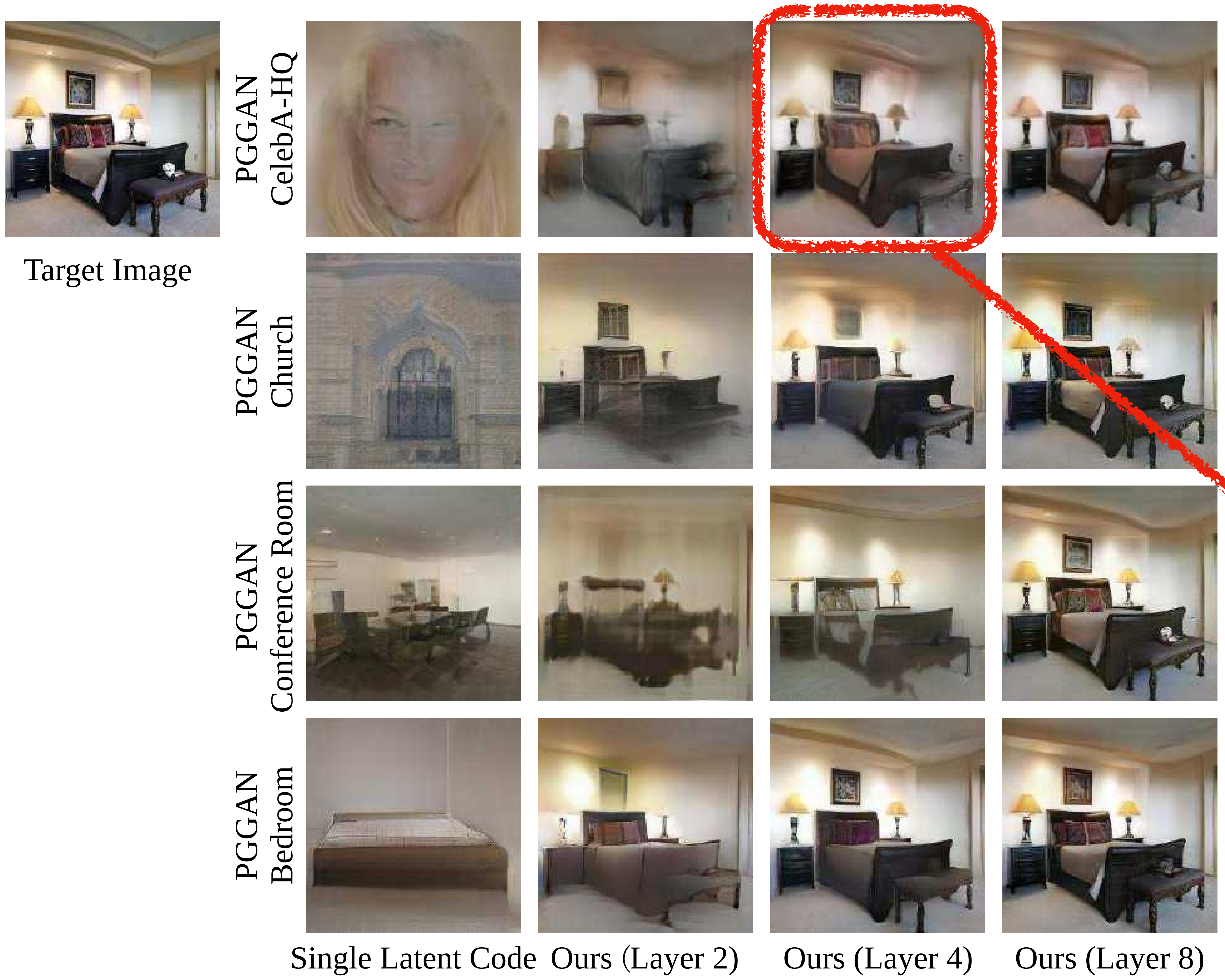
Knowledge Representation in GANs, Different Layers



Reconstruction with later layers.

Using low-level components, such as lines and edges.

Knowledge Representation in GANs, Different Models

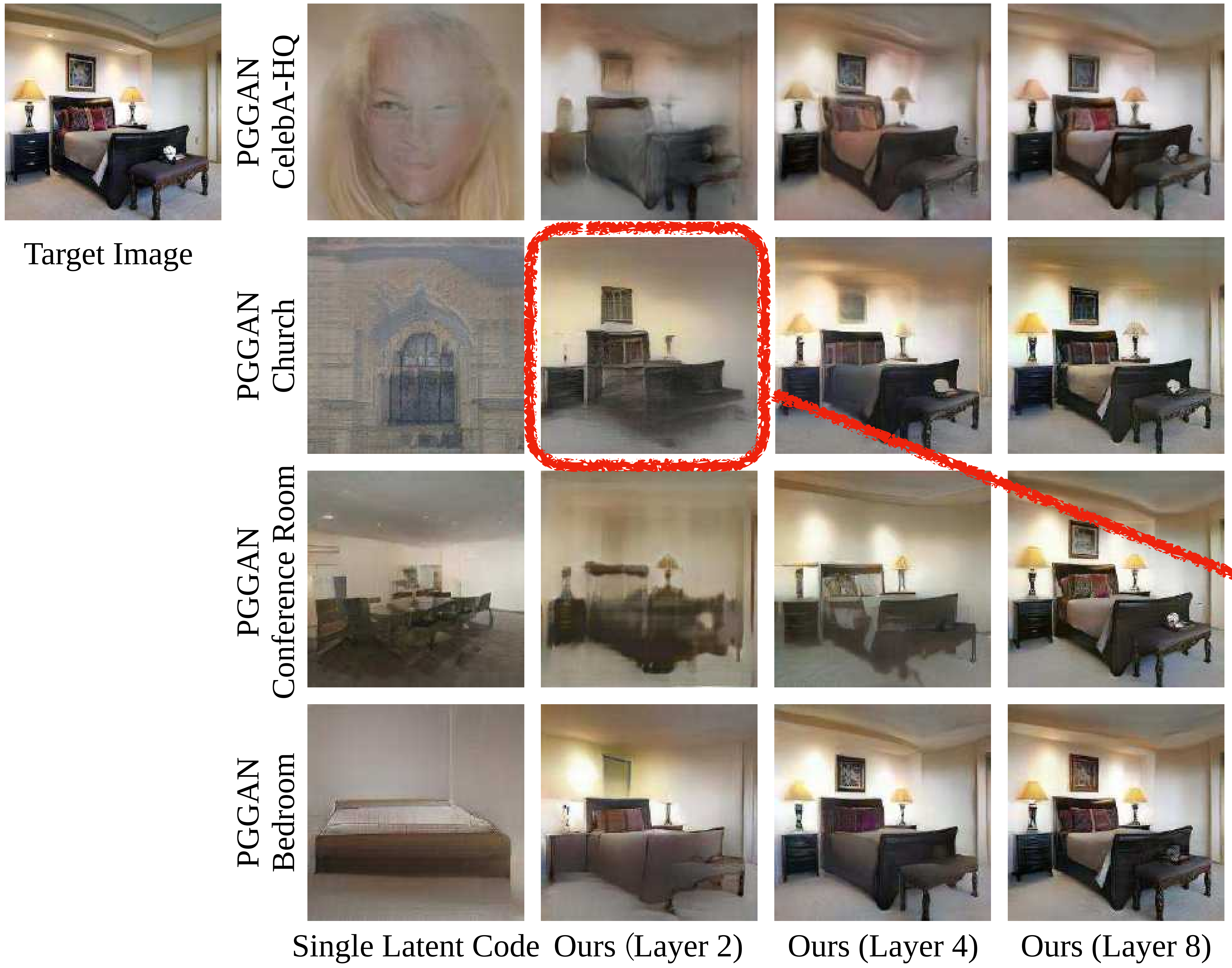


Target Image

Face model

Hardly generate straight lines.

Knowledge Representation in GANs, Different Models



Target Image

Church model

church window as decorative picture.

Knowledge Representation in GANs, Different Models

Target Image

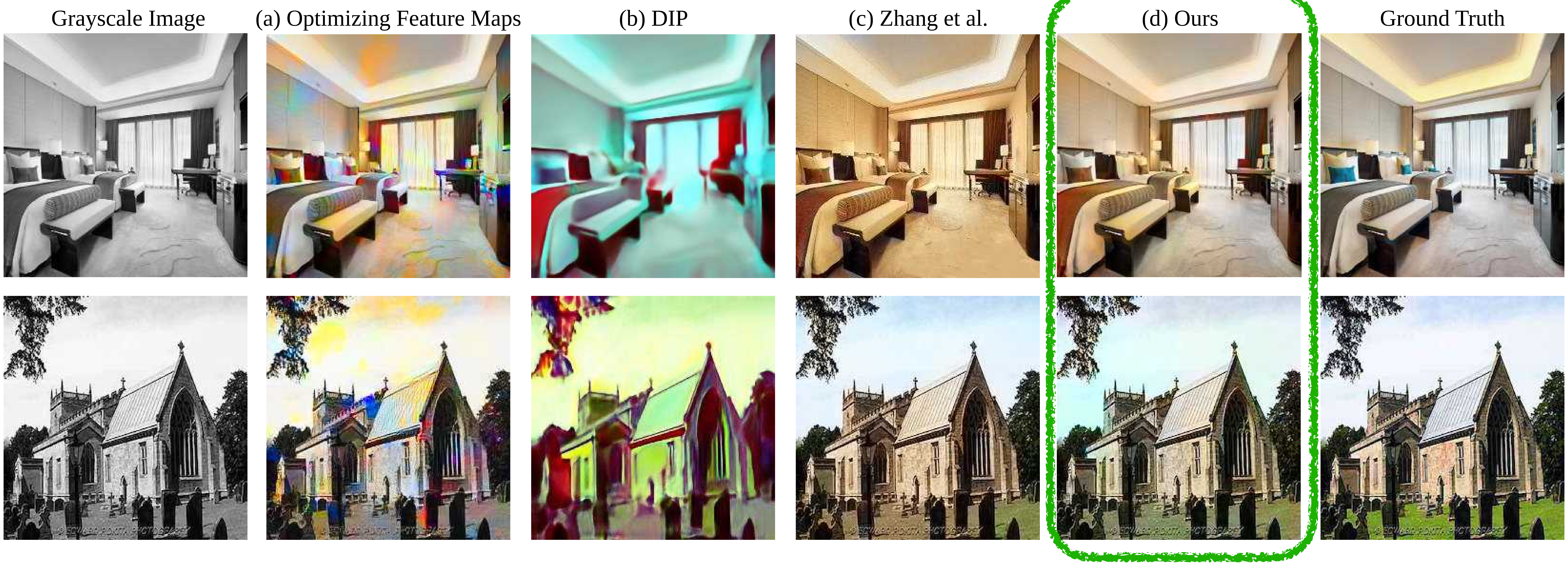
Bedroom model

Correct Prior on Bedroom Images.

Multi-Code GAN Prior for Image Colorization

For image colorization task, we expect the inversion result to have the same gray channel with given grayscale image.

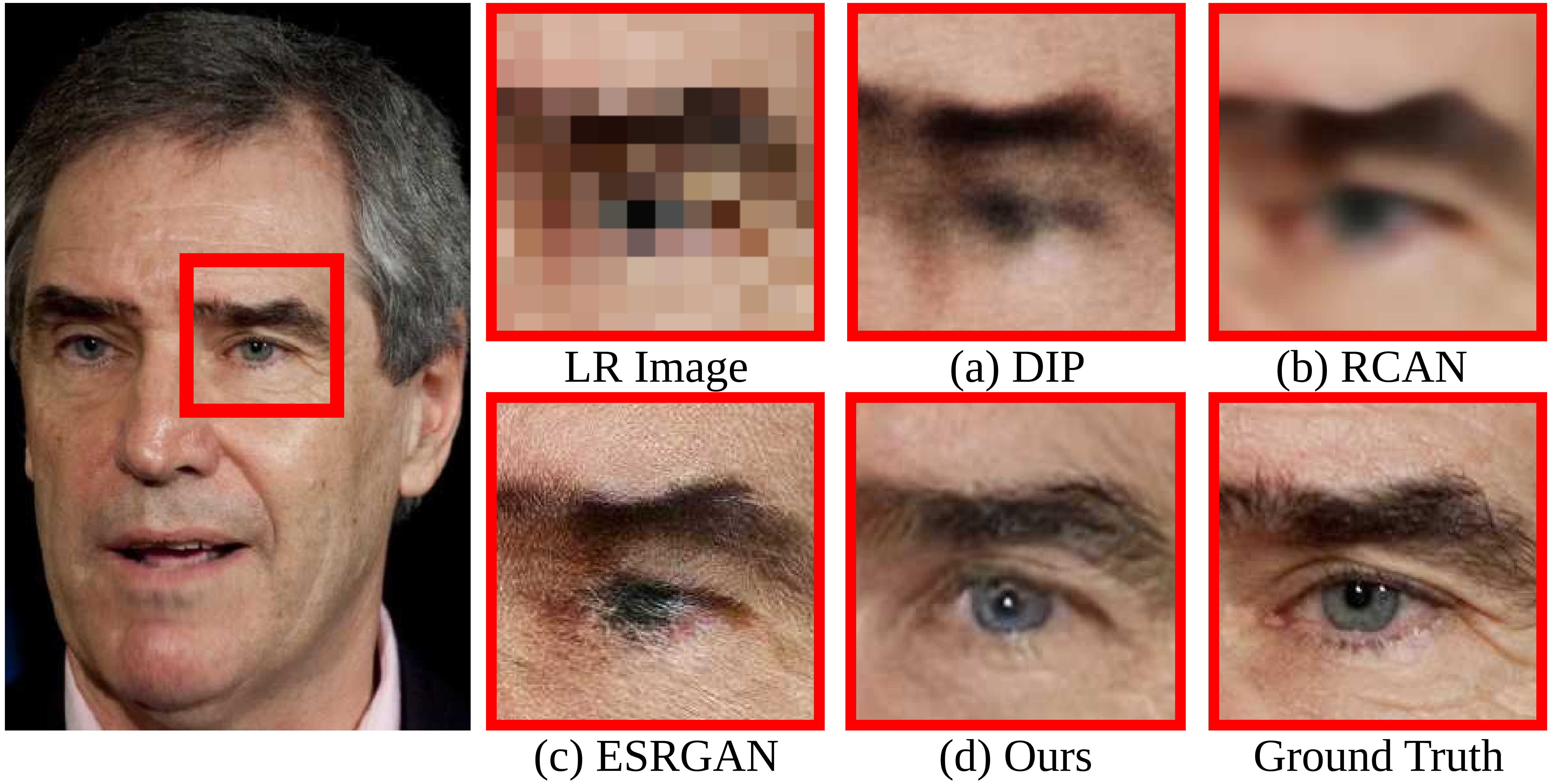
$$L_{color} = L(\mathrm{gray}(\mathbf{x}^{inv}), I_{gray})$$



Multi-Code GAN Prior for Image Super-Resolution

For image super-resolution task, we downsample the inversion result to approximate the given low-res Image.

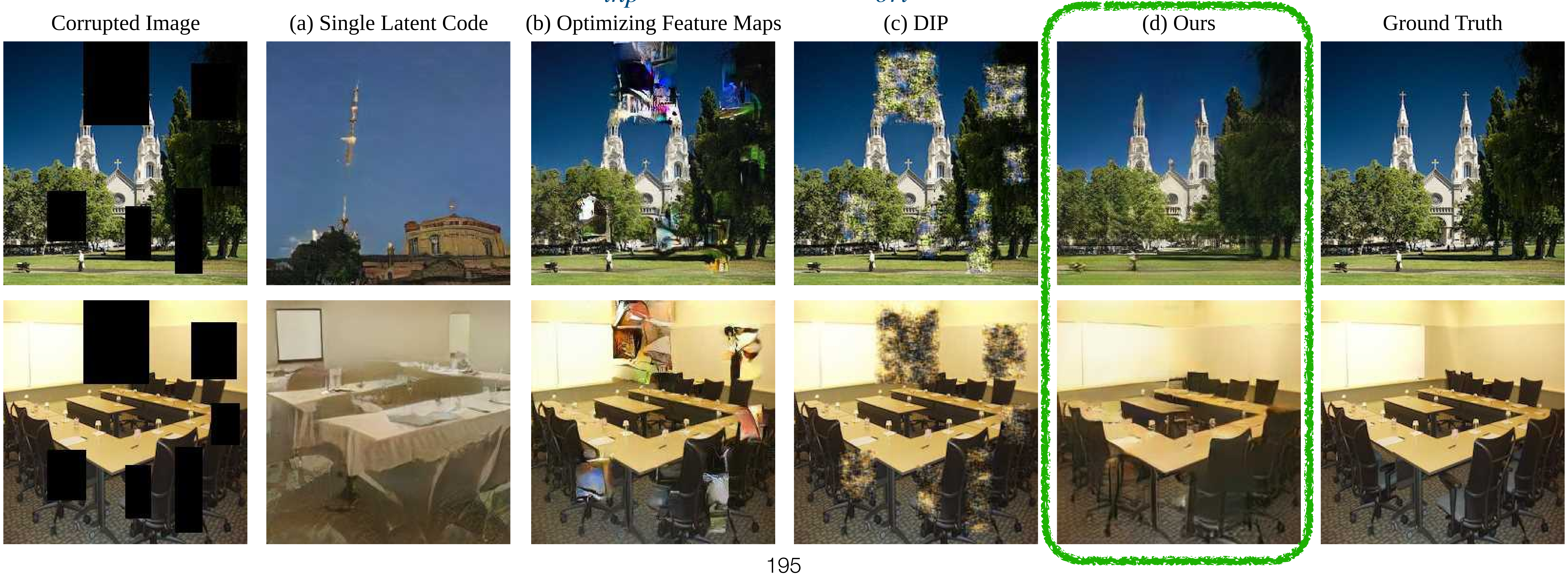
$$L_{SR} = L(\mathrm{down}(\mathbf{x}^{inv}), I_{LR})$$



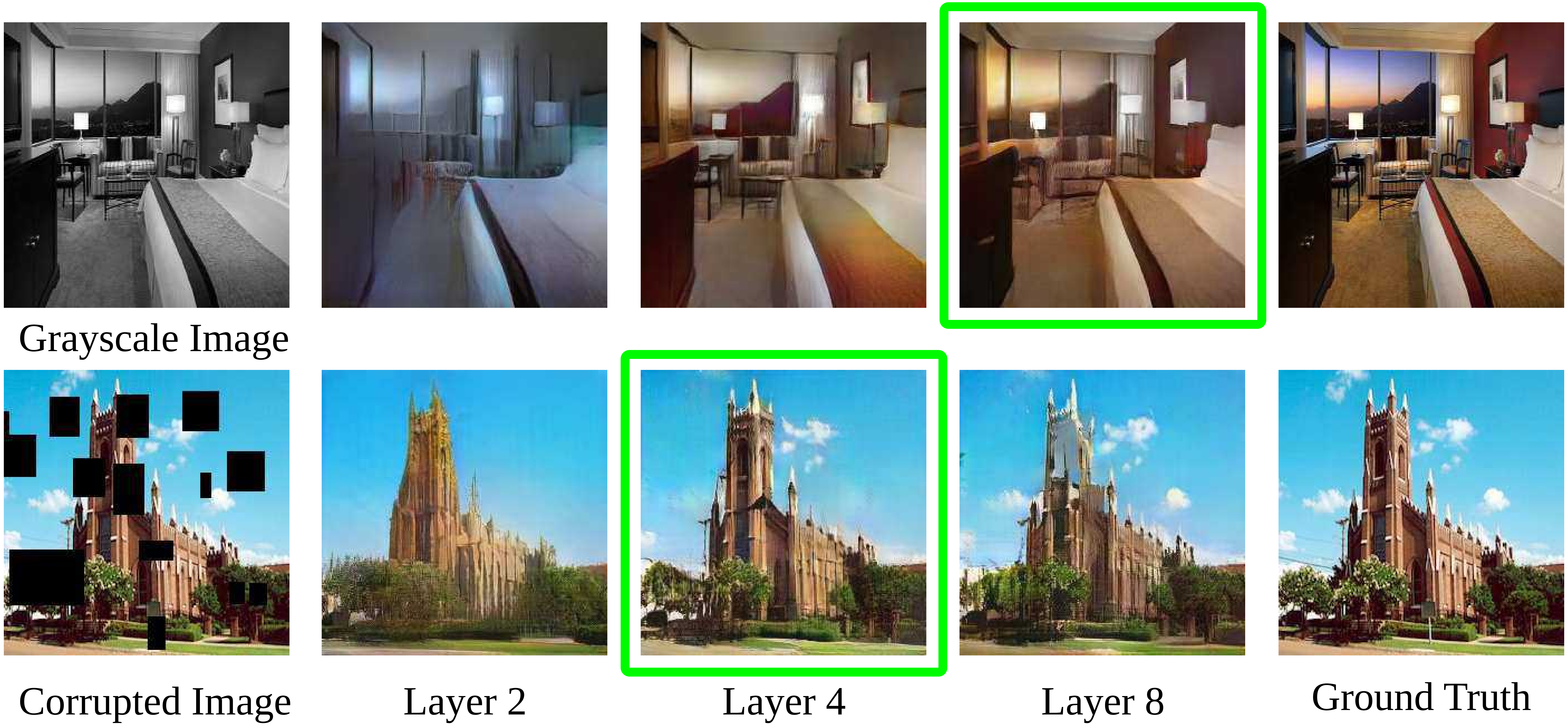
Multi-Code GAN Prior for Image Inpainting

For image inpainting task, we reconstruct the incorrupt parts and let the GAN model fill in the missing pixels automatically.

$$L_{inp} = L(\mathbf{x}^{inv} \circ \mathbf{m}, I_{ori} \circ \mathbf{m})$$



Different Applications With Different Composition Layers



Multi-Code GAN Prior for Semantic Manipulation

Spider-Man

Thor

Loki

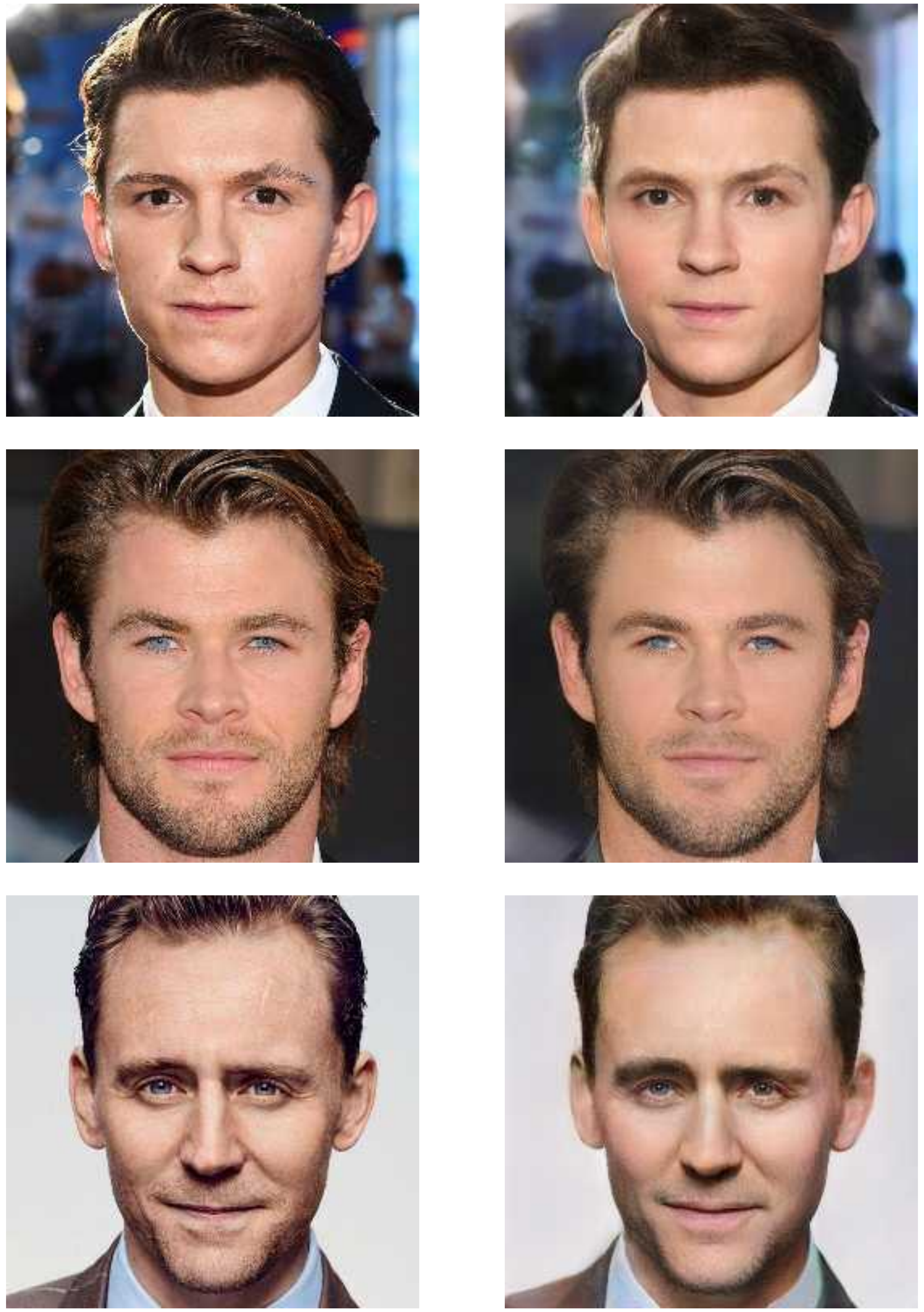
Manipulate attributes for them.

Multi-Code GAN Prior for Semantic Manipulation

Spider-Man

Thor

Loki



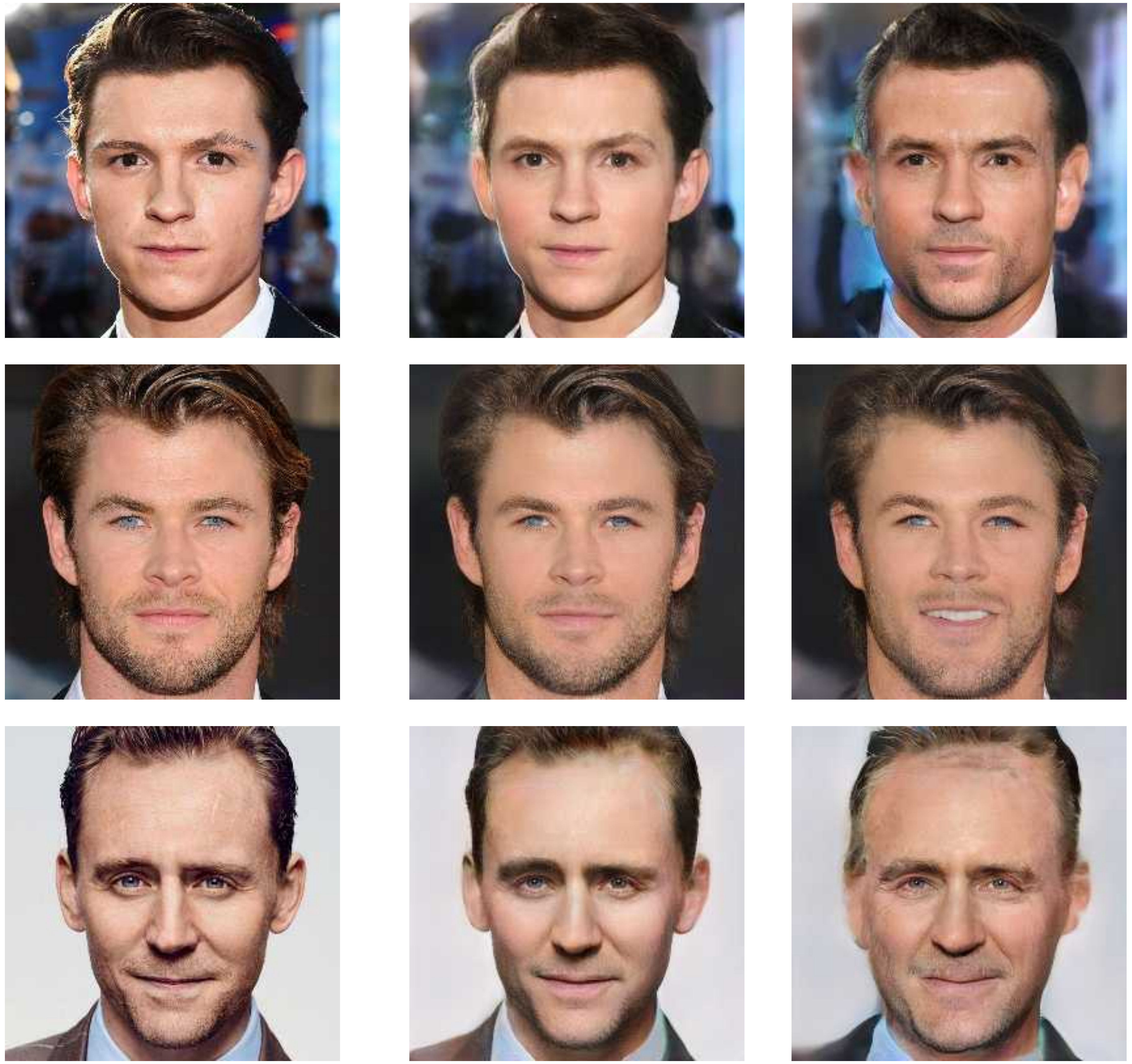
First invert the target face.

Multi-Code GAN Prior for Semantic Manipulation

Spider-Man

Thor

Loki



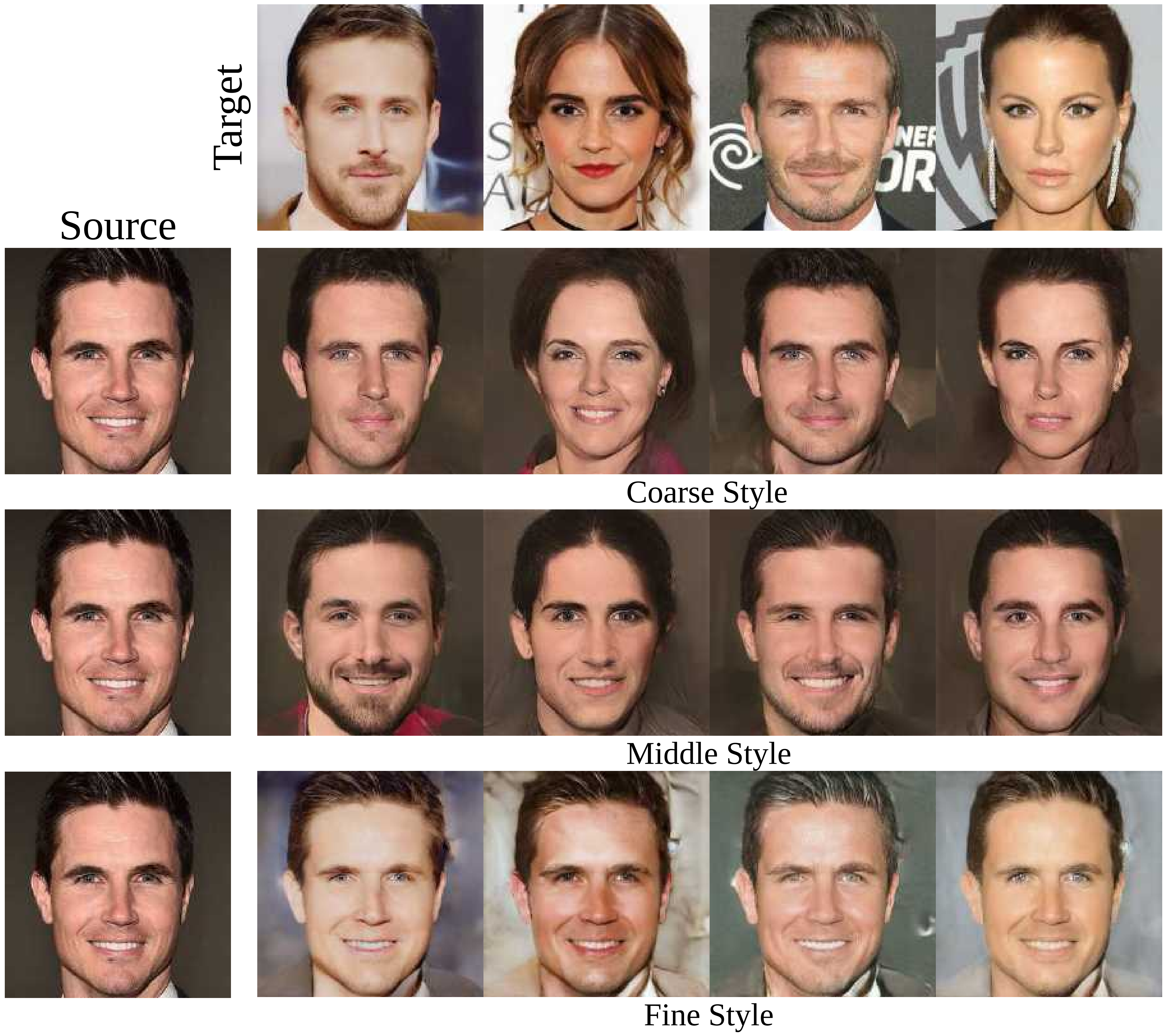
Gender

Smile

Age

Manipulate using previous method

Multi-Code GAN Prior for Sylte Mixing



Research Outline

Real World Image Restoration

Apply advanced image processing technologies in real world applications.

Interpreting Image Generation

Understanding latent space of GANs and make use of GANs' powerful representation.

Applying Deep Signal Processing Technologies in Industrial

Empowering existing industrial facilities with advanced signal processing technologies.

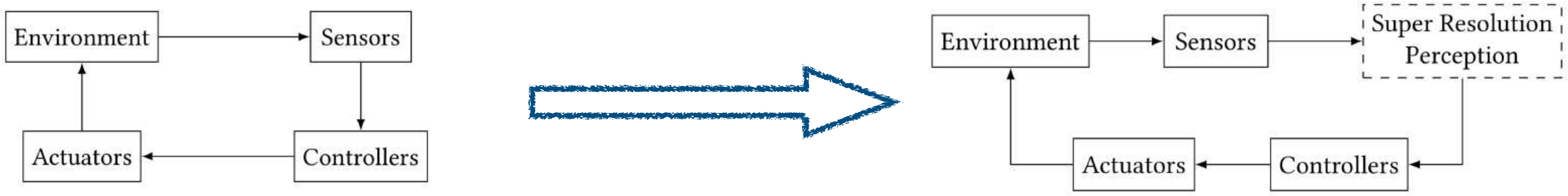
Super-Resolution Perception for Industrial Sensor Data

- Sensors (smart meters, PMUs, etc.) play an important role in industrial systems

- A balance between sensor sampling frequency and meter cost, communication cost and storage cost

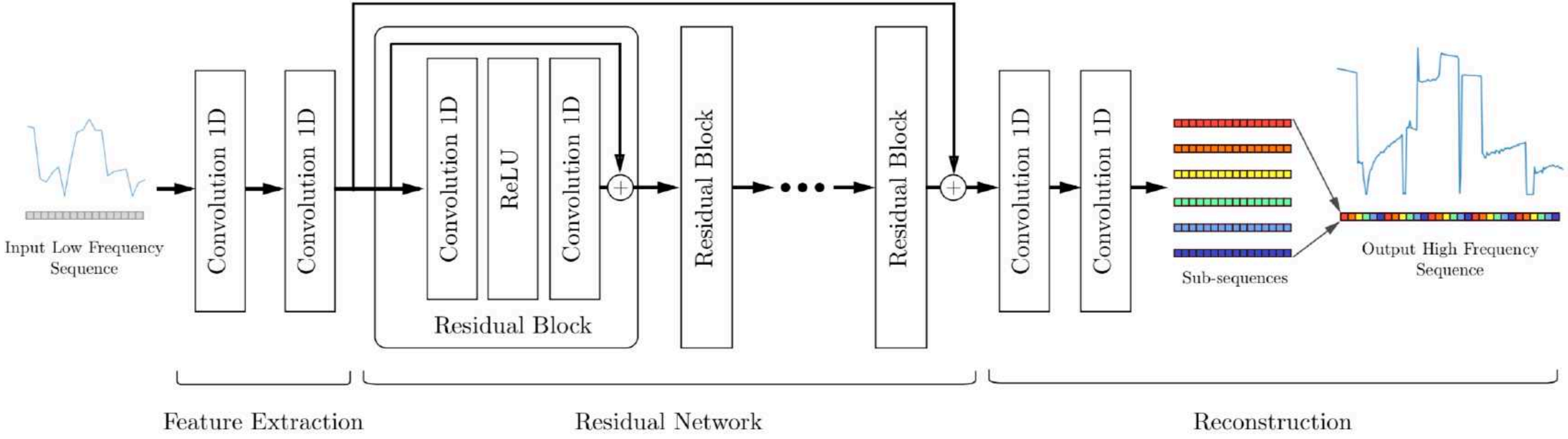
- High-frequency sampling data helps accurate perception of system status

- Sampling data at high frequencies also increases costs



Super-Resolution Perception for Industrial Sensor Data

Jinjin Gu, Haoyu Chen, Guolong Liu, Gaoqi Liang, Junhua Zhao, 2019



Super-Resolution Perception for Industrial Sensor Data

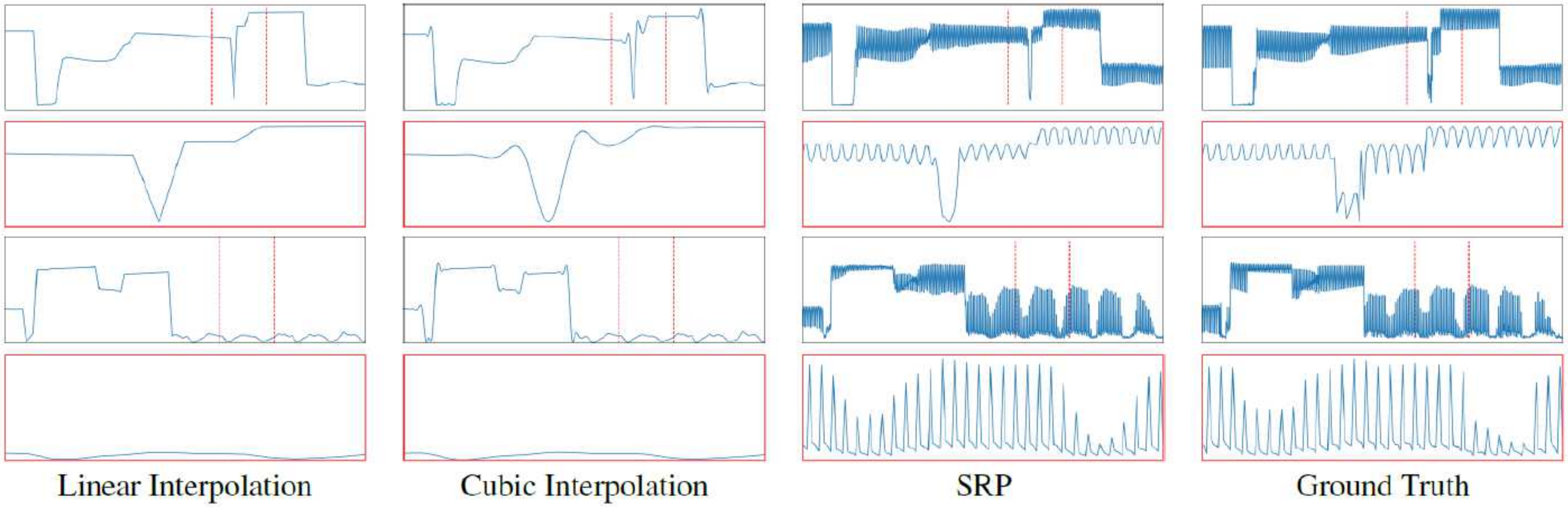
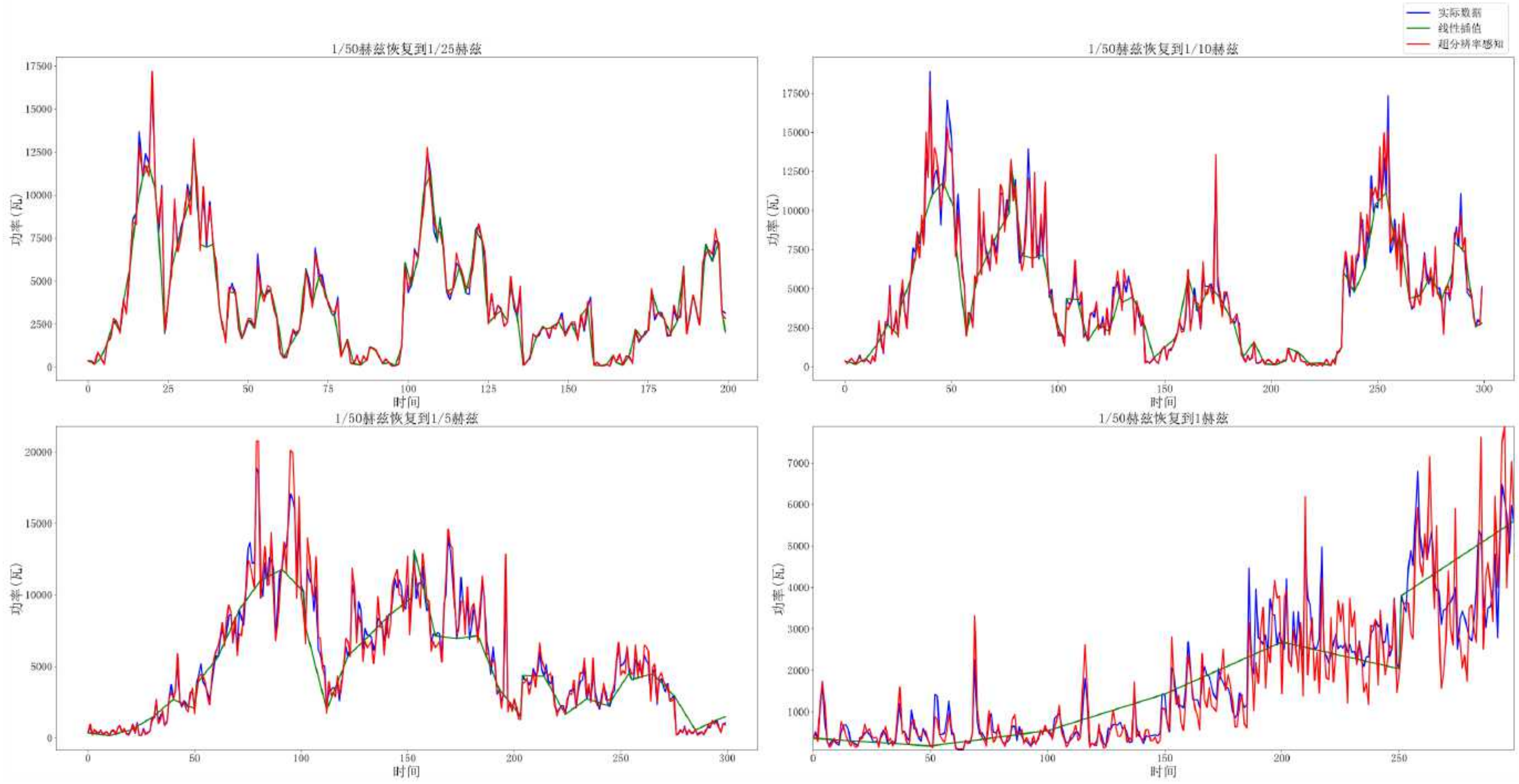


Figure 5: The SRP results of experiment with $f_l = 10\mathbf{Hz}$ and $\alpha = 10$.

Compared with the traditional interpolation method, the super-perception can accurately recover the details of the waveform.

Can significantly reduce the communication and storage burden of data (SR factor of 10, can reduce communication and storage costs by 90%).

Super-Resolution Perception for Industrial Sensor Data



SRP results for real industrial sensor data.

Super-Resolution Perception for Industrial Sensor Data

Super-Resolution Perception helps Nonintrusive Load Monitoring in Smart Gird.

Method	Experiment	trained on LF test on LF	trained on HF test on SRP	trained on HF test on HF	Gain from SRP	Not recovered by SRP
KNN	$f_l = 10$ **Hz**, $\alpha = 10$	0.599	0.593	0.595	−0.006	0.002
	$f_l = 100$ **Hz**, $\alpha = 10$	0.595	0.615	0.685	+0.020	0.070
	$f_l = 10$ **Hz**, $\alpha = 100$	0.599	0.617	0.685	+0.018	0.068
Decision Tree	$f_l = 10$ **Hz**, $\alpha = 10$	0.617	0.639	0.652	+0.022	0.013
	$f_l = 100$ **Hz**, $\alpha = 10$	0.652	0.679	0.714	+0.027	0.035
	$f_l = 10$ **Hz**, $\alpha = 100$	0.617	0.651	0.714	+0.034	0.063
SVM	$f_l = 10$ **Hz**, $\alpha = 10$	0.666	0.670	0.674	+0.004	0.004
	$f_l = 100$ **Hz**, $\alpha = 10$	0.674	0.715	0.749	+0.041	0.034
	$f_l = 10$ **Hz**, $\alpha = 100$	0.666	0.693	0.749	+0.027	0.056
BP-ANN	$f_l = 10$ **Hz**, $\alpha = 10$	0.654	0.681	0.692	+0.027	0.011
	$f_l = 100$ **Hz**, $\alpha = 10$	0.687	0.724	0.737	+0.037	0.013
	$f_l = 10$ **Hz**, $\alpha = 100$	0.654	0.676	0.737	+0.022	0.061
SCN	$f_l = 10$ **Hz**, $\alpha = 10$	0.725	0.796	0.801	+0.068	0.003
	$f_l = 100$ **Hz**, $\alpha = 10$	0.743	0.814	0.822	+0.066	0.011
	$f_l = 10$ **Hz**, $\alpha = 100$	0.725	0.796	0.822	+0.064	0.036
CNN	$f_l = 10$ **Hz**, $\alpha = 10$	0.861	0.877	0.882	+0.016	0.005
	$f_l = 100$ **Hz**, $\alpha = 10$	0.882	0.908	0.912	+0.026	0.004
	$f_l = 10$ **Hz**, $\alpha = 100$	0.861	0.883	0.912	+0.022	0.029

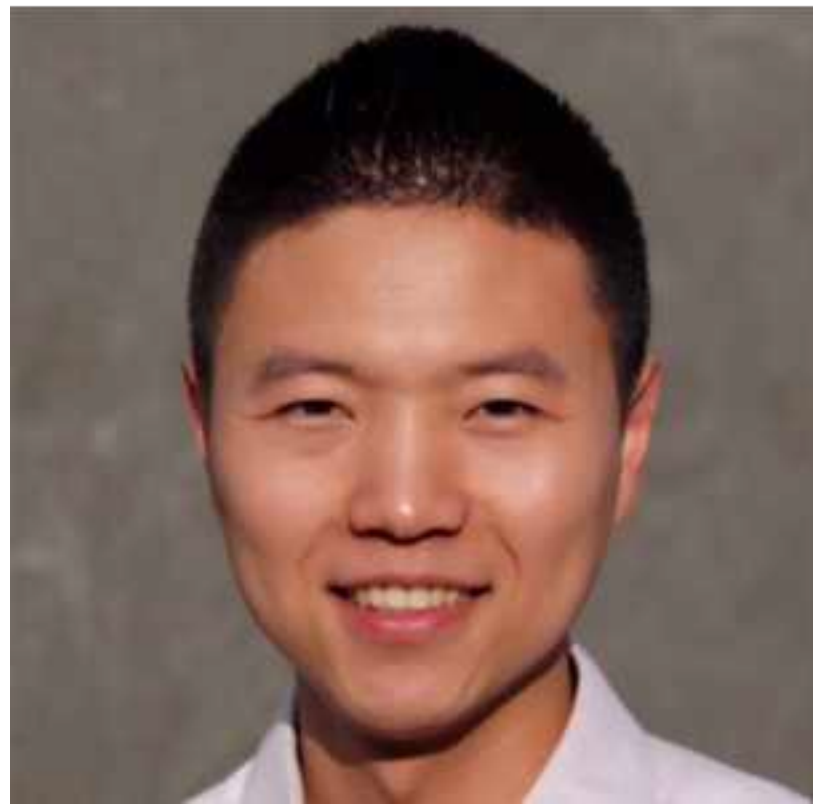
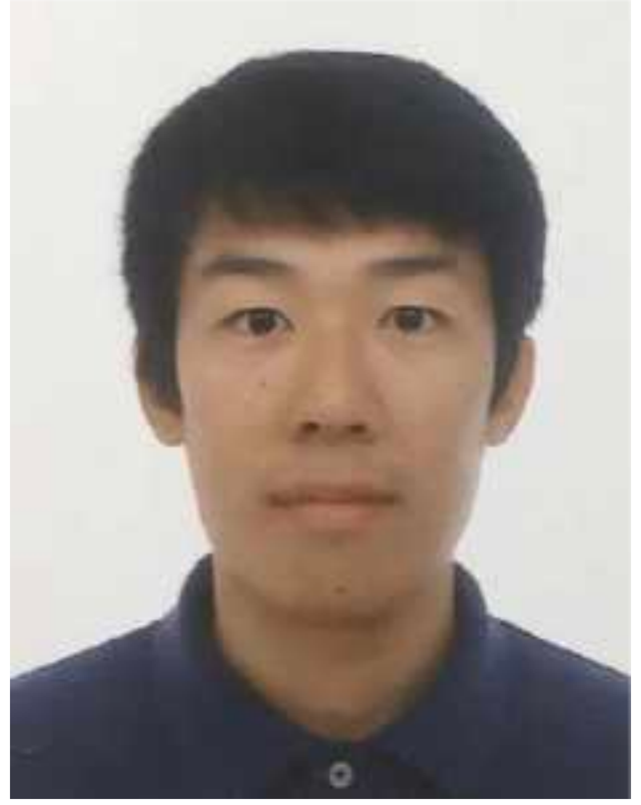
Thanks to my cooperators

AI Imaging @ SenseTime

Xpixel.Group



mmlab @ CUHK

Thanks

JINJIN GU
School of Electrical and Information Engineering,
The University of Sydney.
May. 2021