Interpreting Super-Resolution Networks

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

Interpreting Super-Resolution Networks

Interpretability in Low-Level Vision:

- **Pixel**: What pixels contribute most to restoration?
- **Feature**: Where can we find semantics in SR networks?

Research works about interpreting and explaining low-level vision networks.

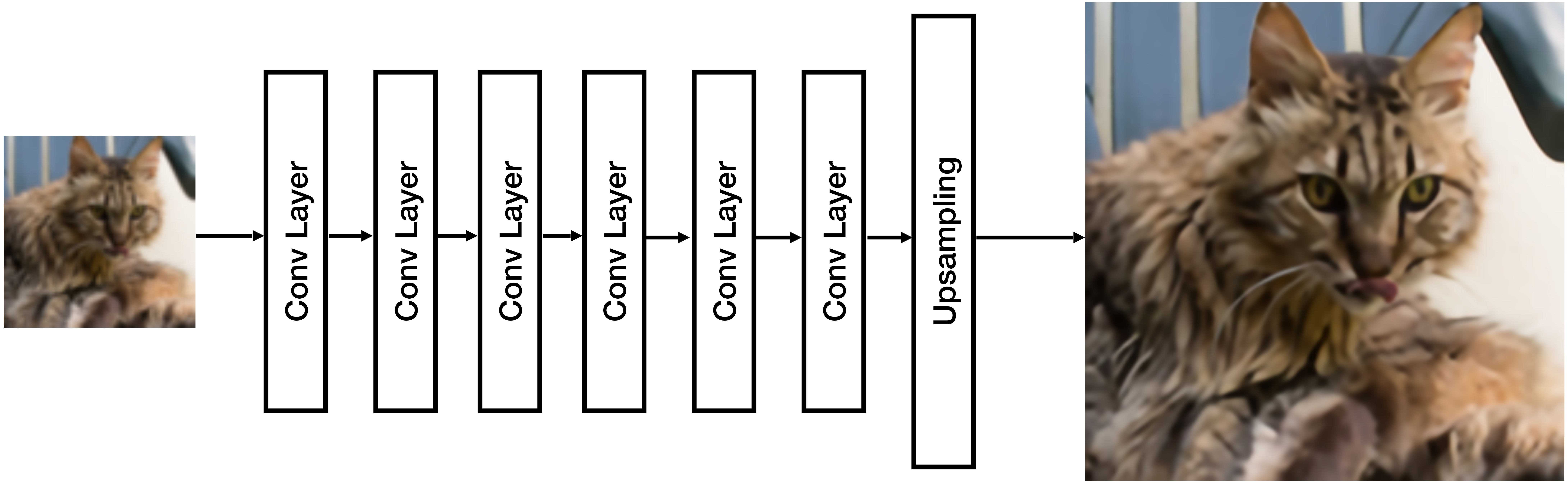
CVPR VIRTUAL JUNE 19-25

Interpreting Super-Resolution Networks with Local Attribution Maps

Jinjin Gu, Chao Dong

The University of Sydney,
Shenzhen Institutes of Advanced Technology

Super-Resolution Networks



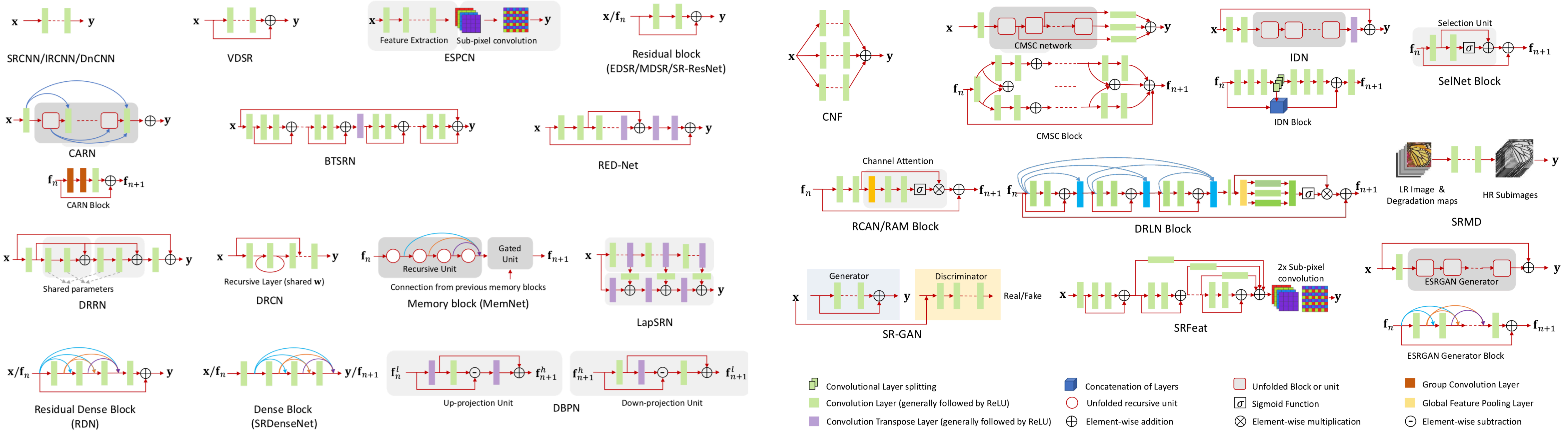
SR networks build up of convolutional layers and upsampling blocks, with parameter θ.

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.]

SR networks are still mysterious

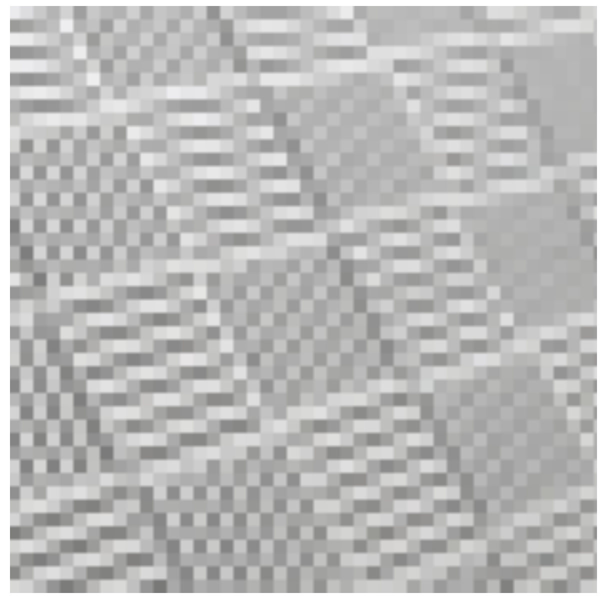
Have you met these scenarios?

- 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?

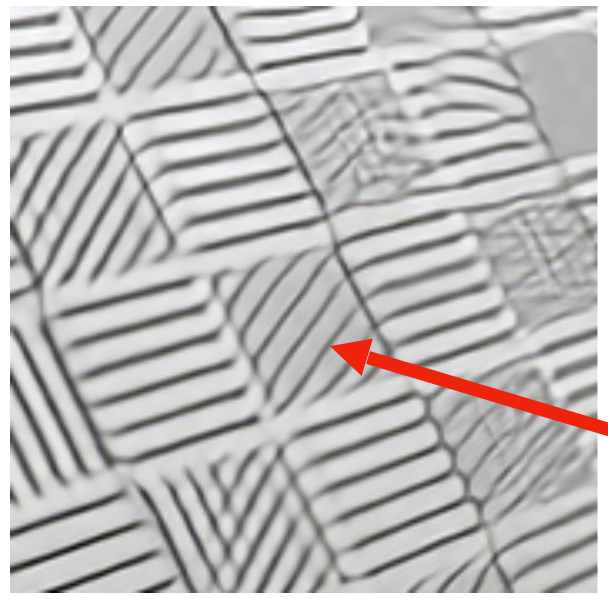
We lack understanding toward these questions

And also research tools

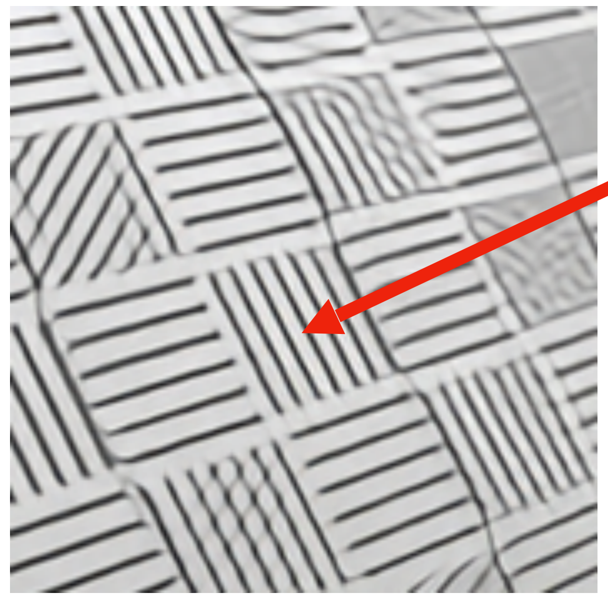
Attribution Analysis



Input image



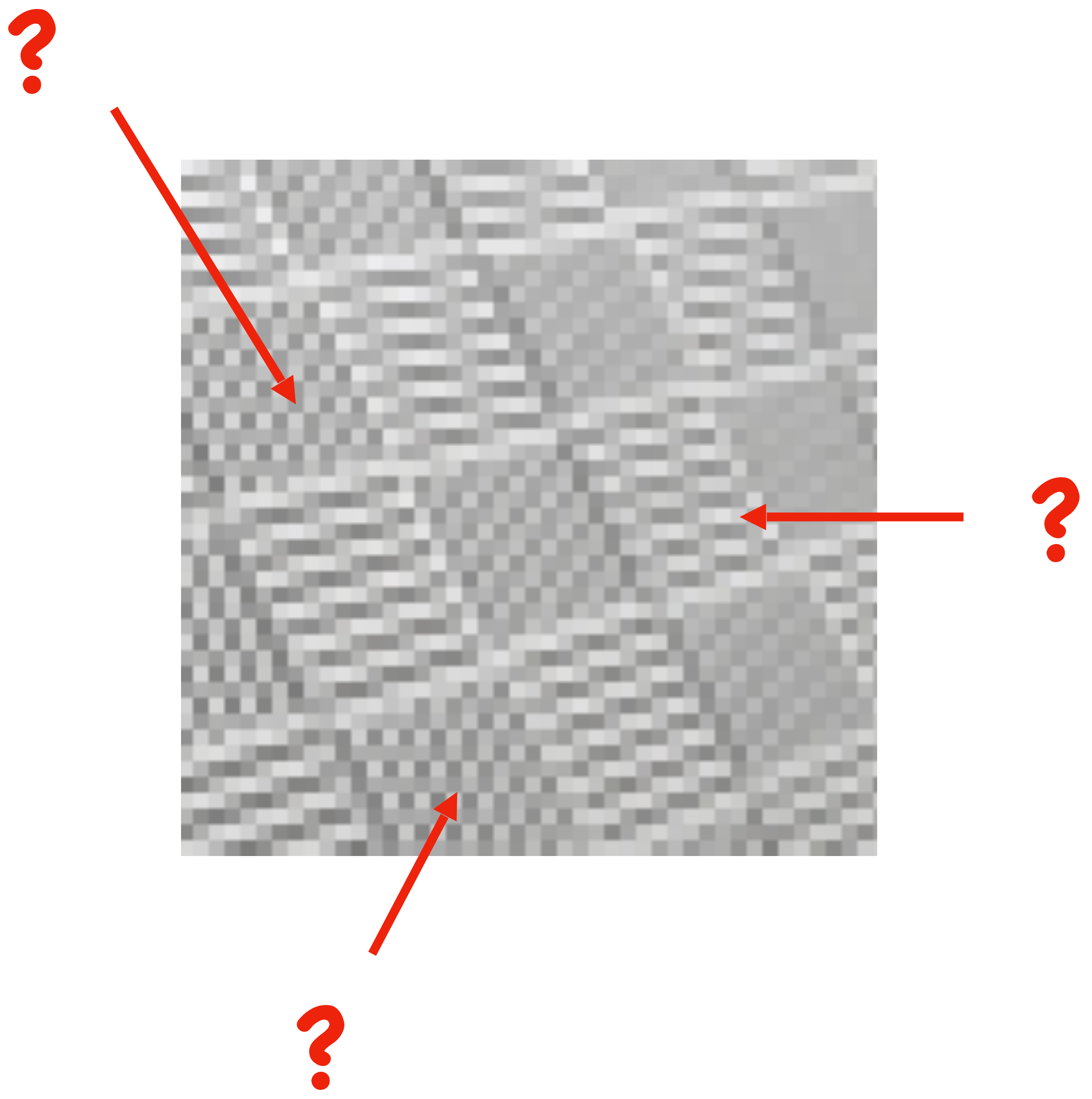
EDSR



RNAN

Why RNAN gives correct results in the center?

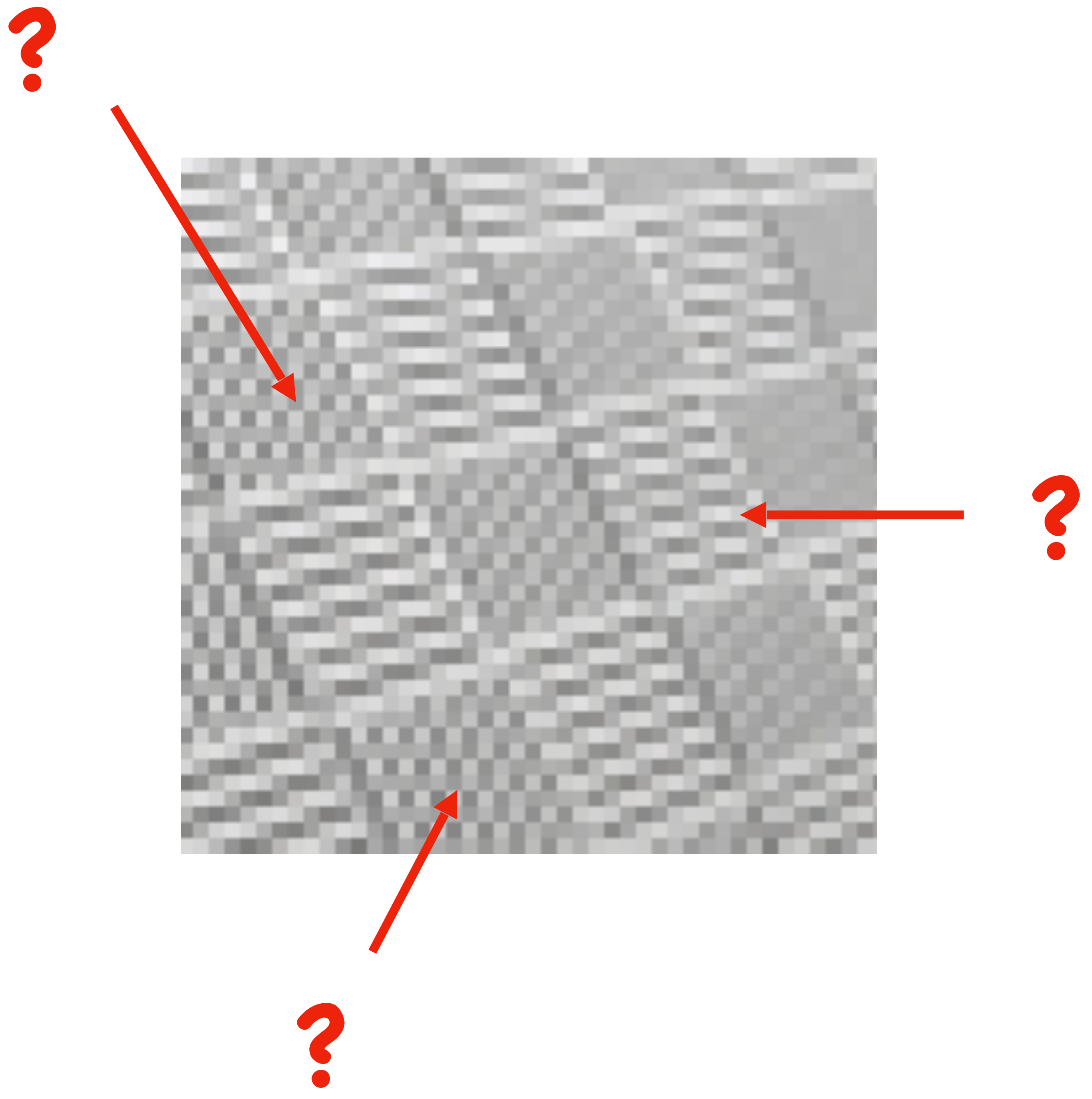
Attribution Analysis



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

Does EDSR notice this information?

Attribution Analysis



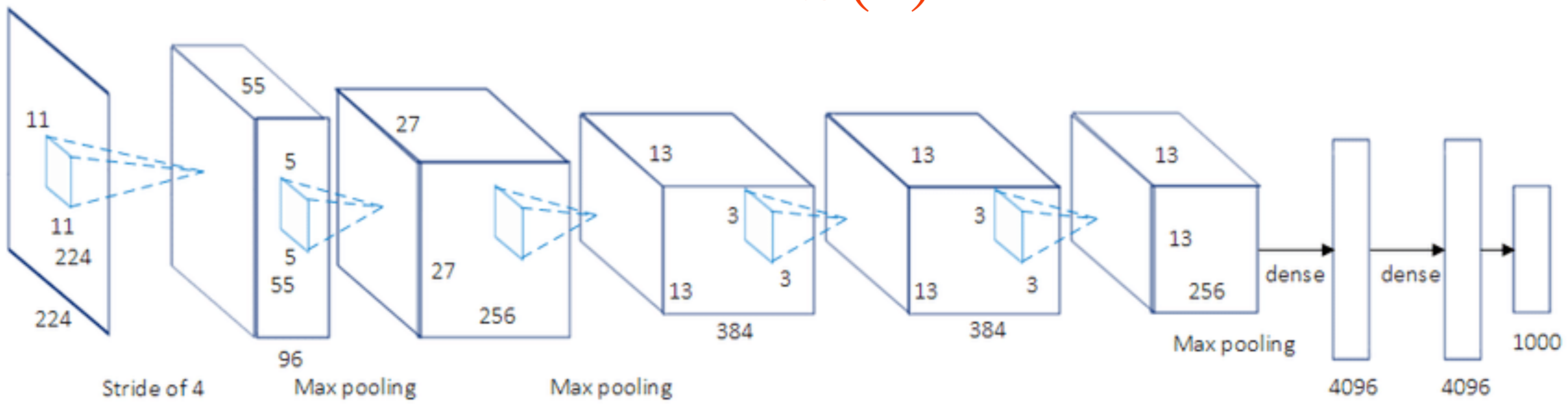
Identify input features responsible for SR results.

Attribution Analysis for High-level Networks

What is S(I) looking at?

$S(I)$

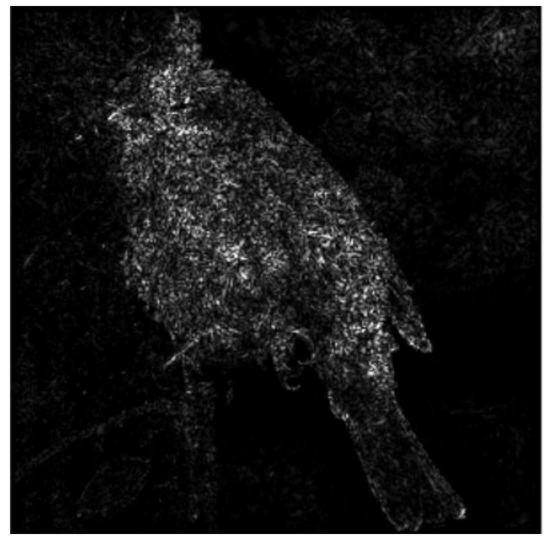
I



98% house finch

10% bird

1% People

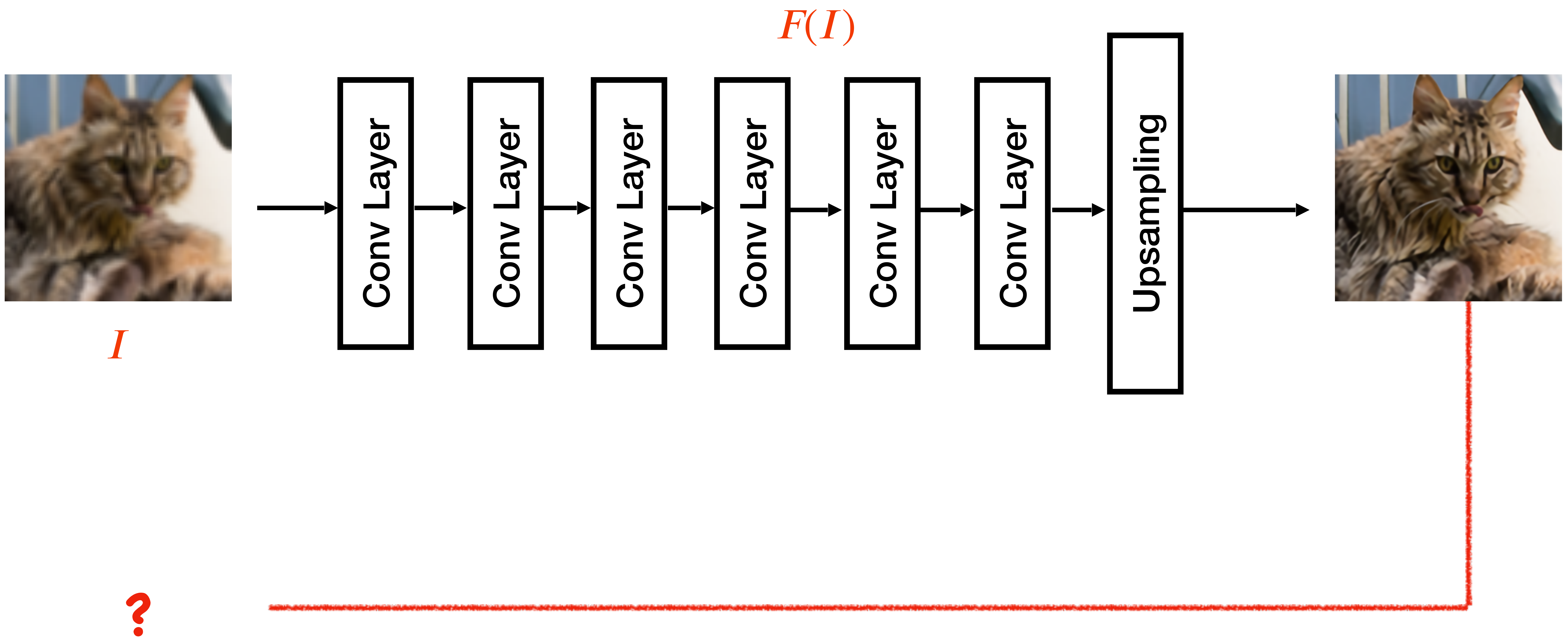


The visualized attribution map

Backprop methods: gradient

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

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

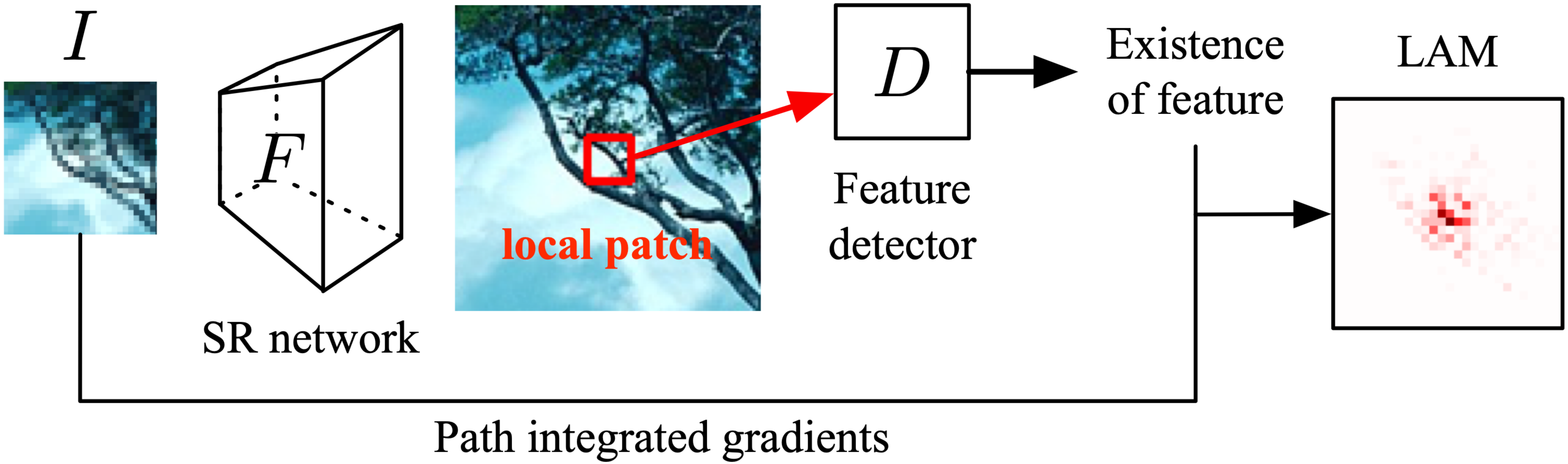
Auxiliary Principles

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 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)

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

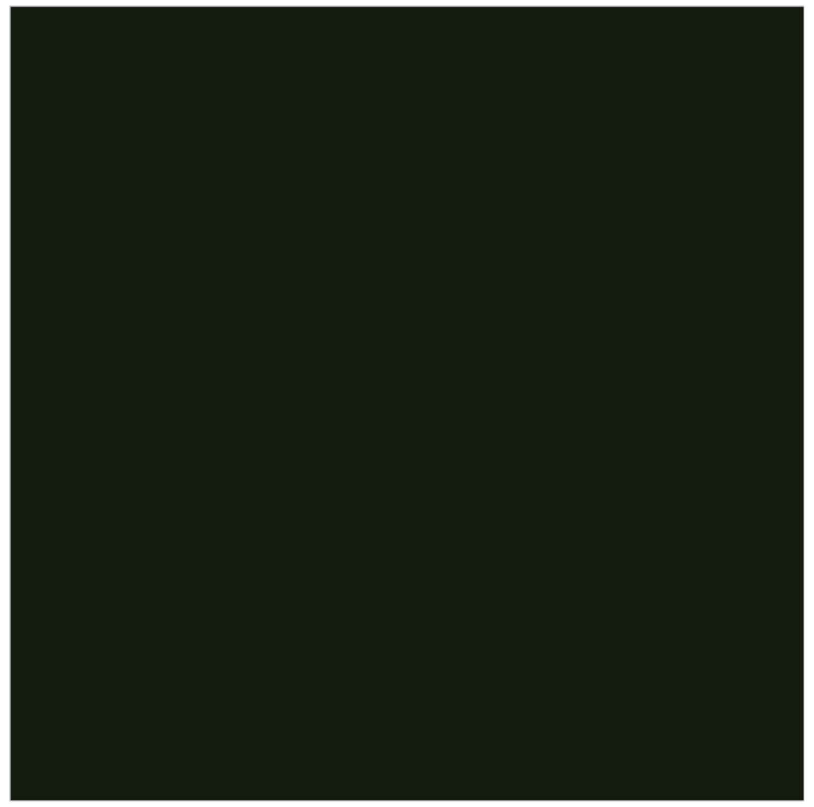
Blurred image as baseline input: $I' = \omega(\sigma) \otimes I$

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

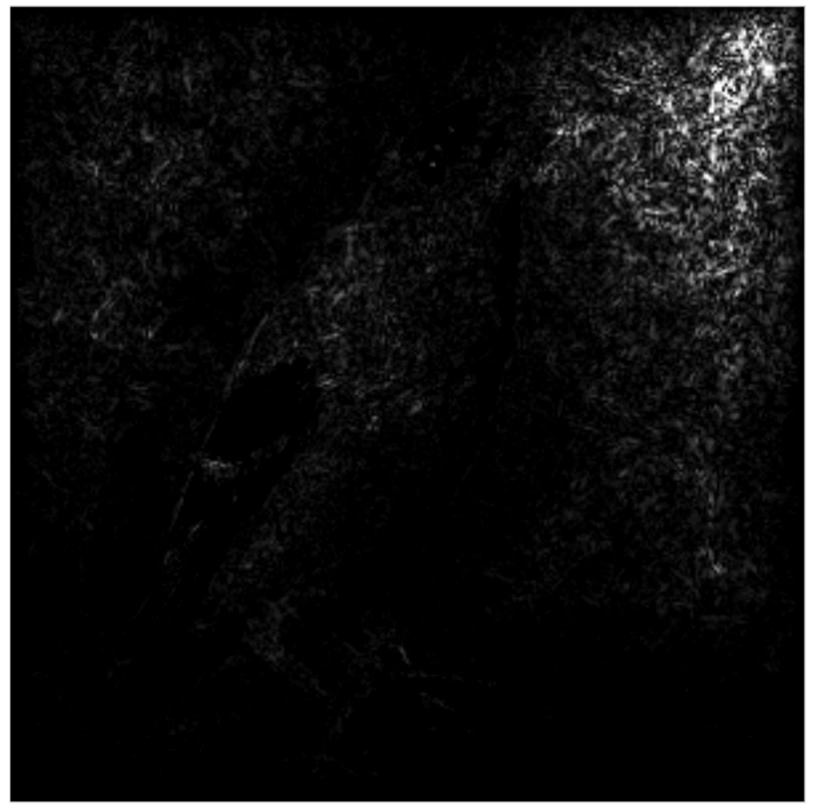
Local Attribution Maps (LAM)

Original Image

Baseline Input x'

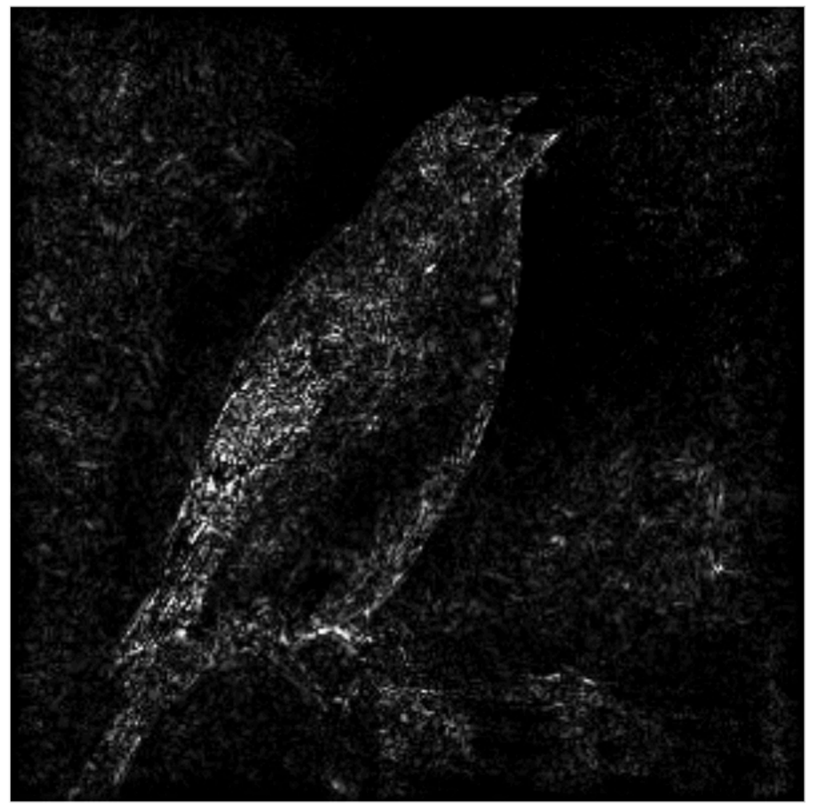


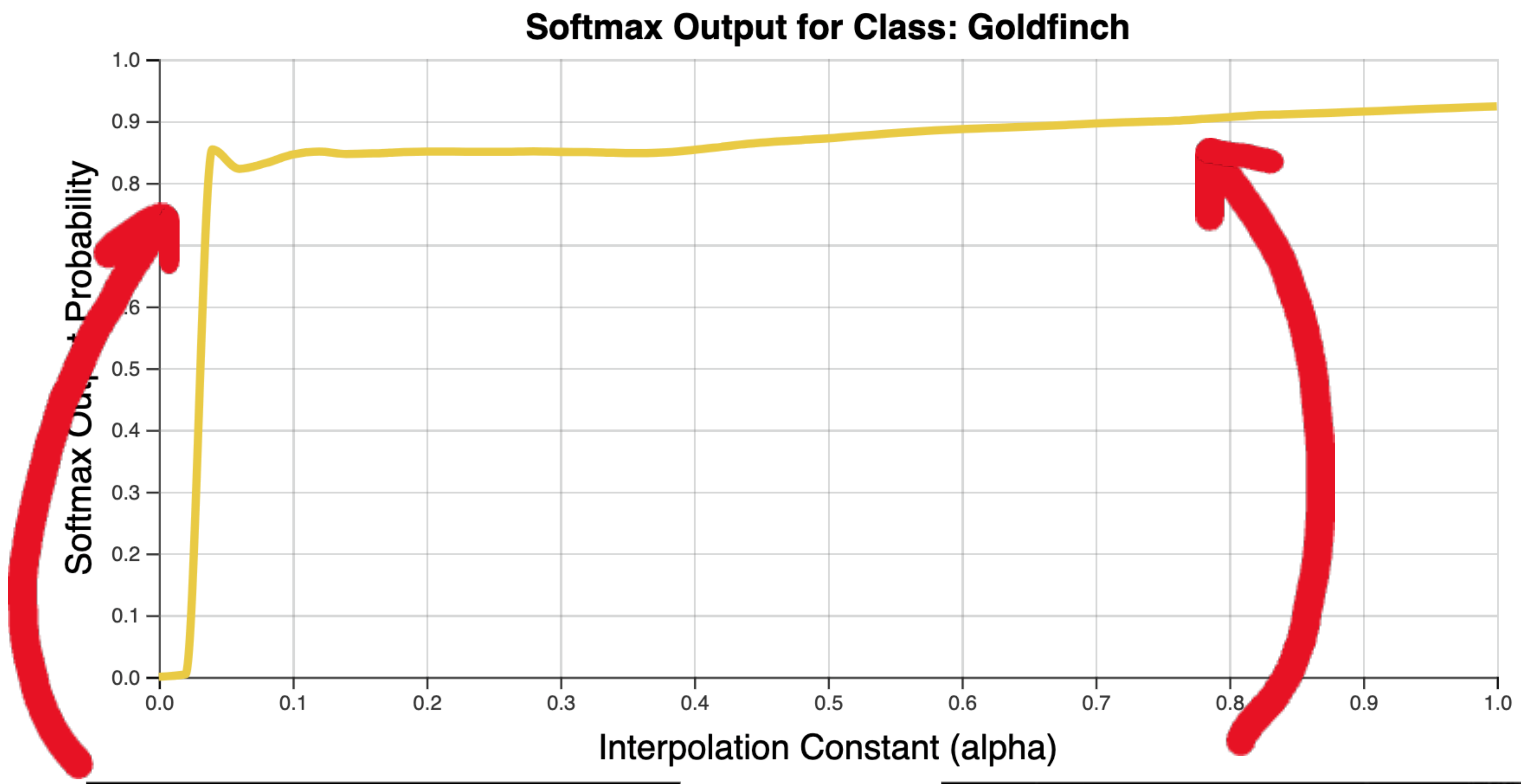
Integrated Gradients Attributions

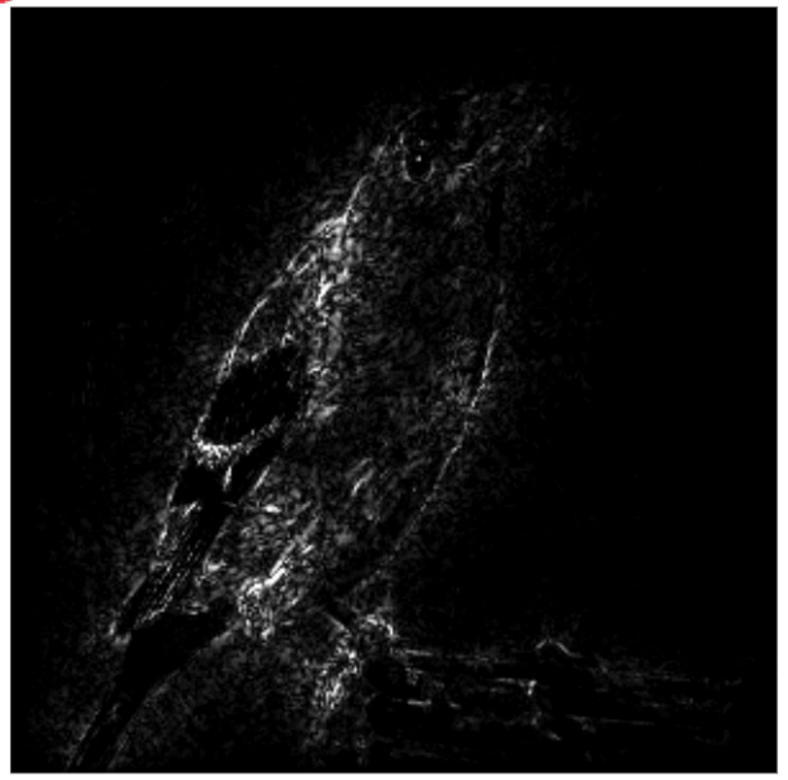


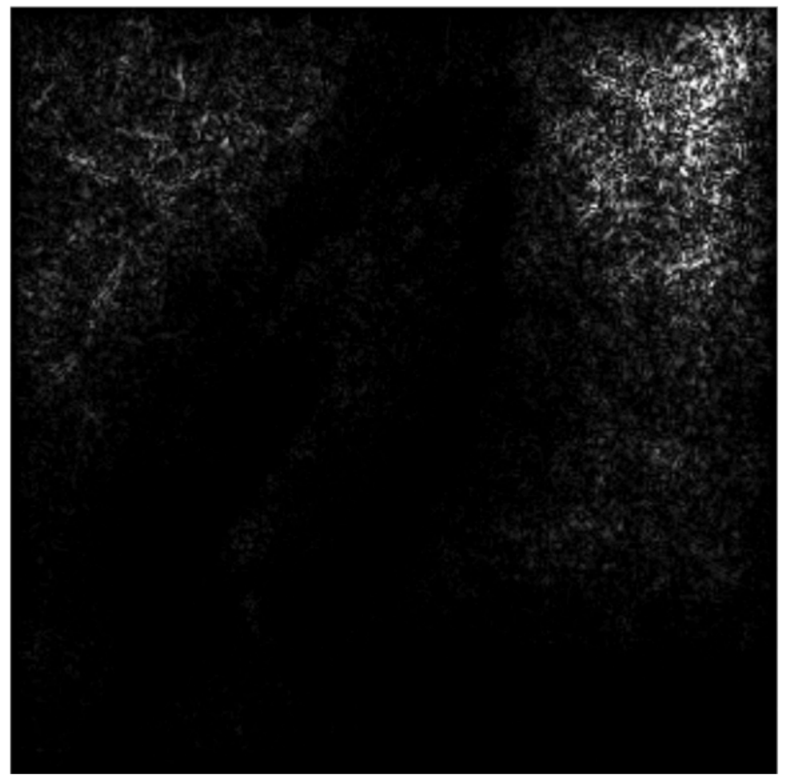
Baseline Input x'

Integrated Gradients Attributions









[Visualizing the Impact of Feature Attribution Baselines]

Local Attribution Maps (LAM)

We employ Path Integral Gradient

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

The Gradient of interpolation

The weight determined by path function

SR Network F

Feature Detector D

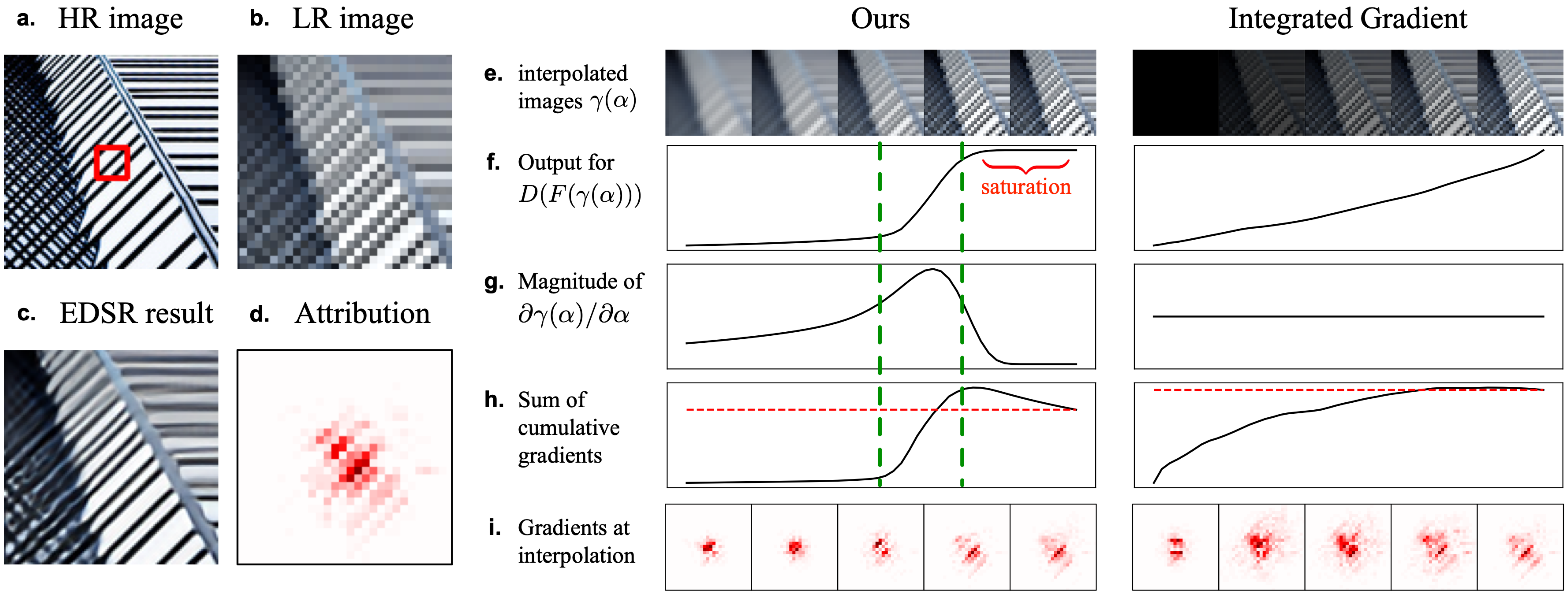
Path function $\gamma(\alpha), \alpha \in \mathbb{R}$

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

Input $\gamma(1) = I$

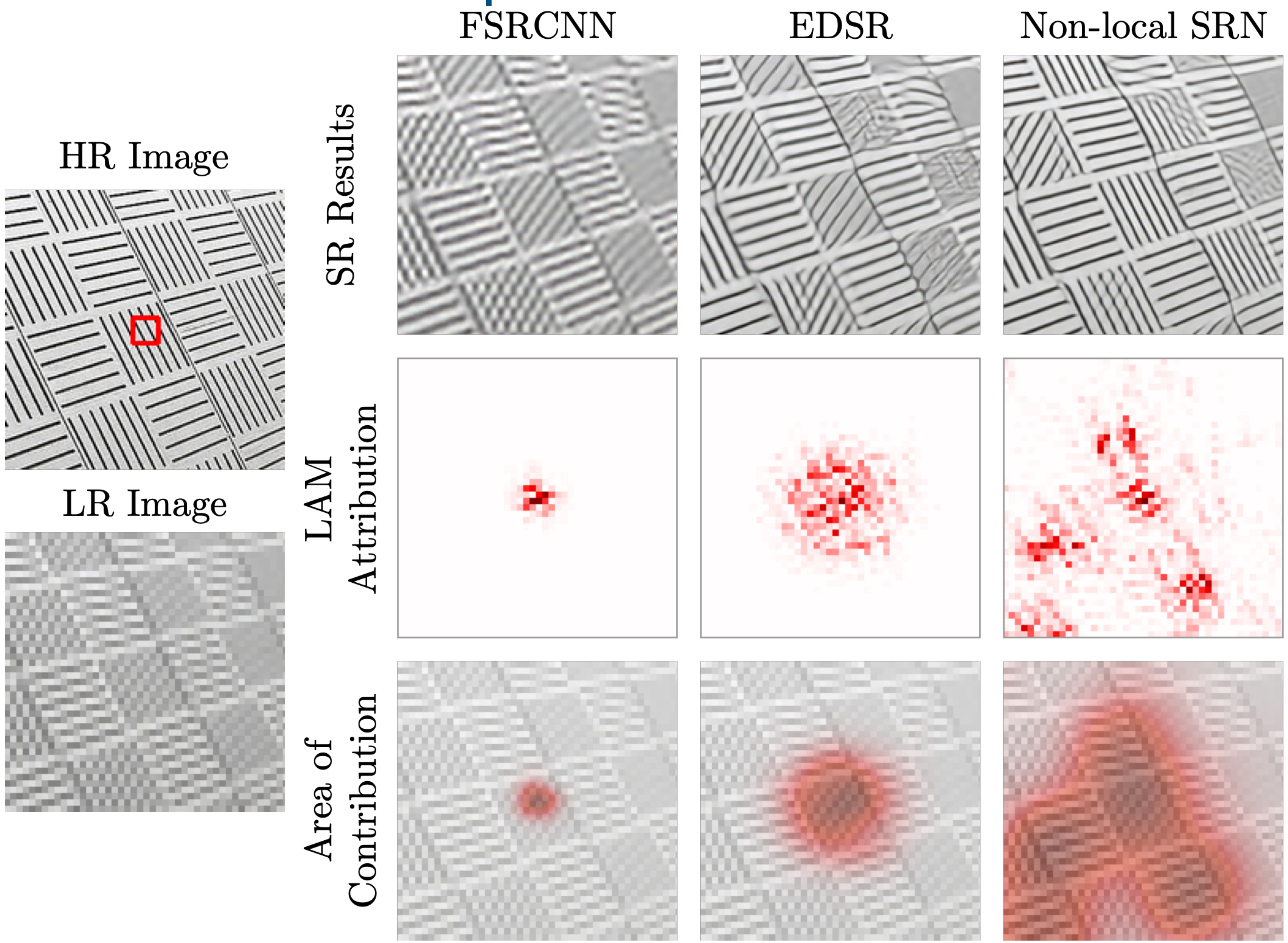
Local Attribution Maps (LAM)

Why using path integral gradient: Gradient Saturation



Local Attribution Maps Results

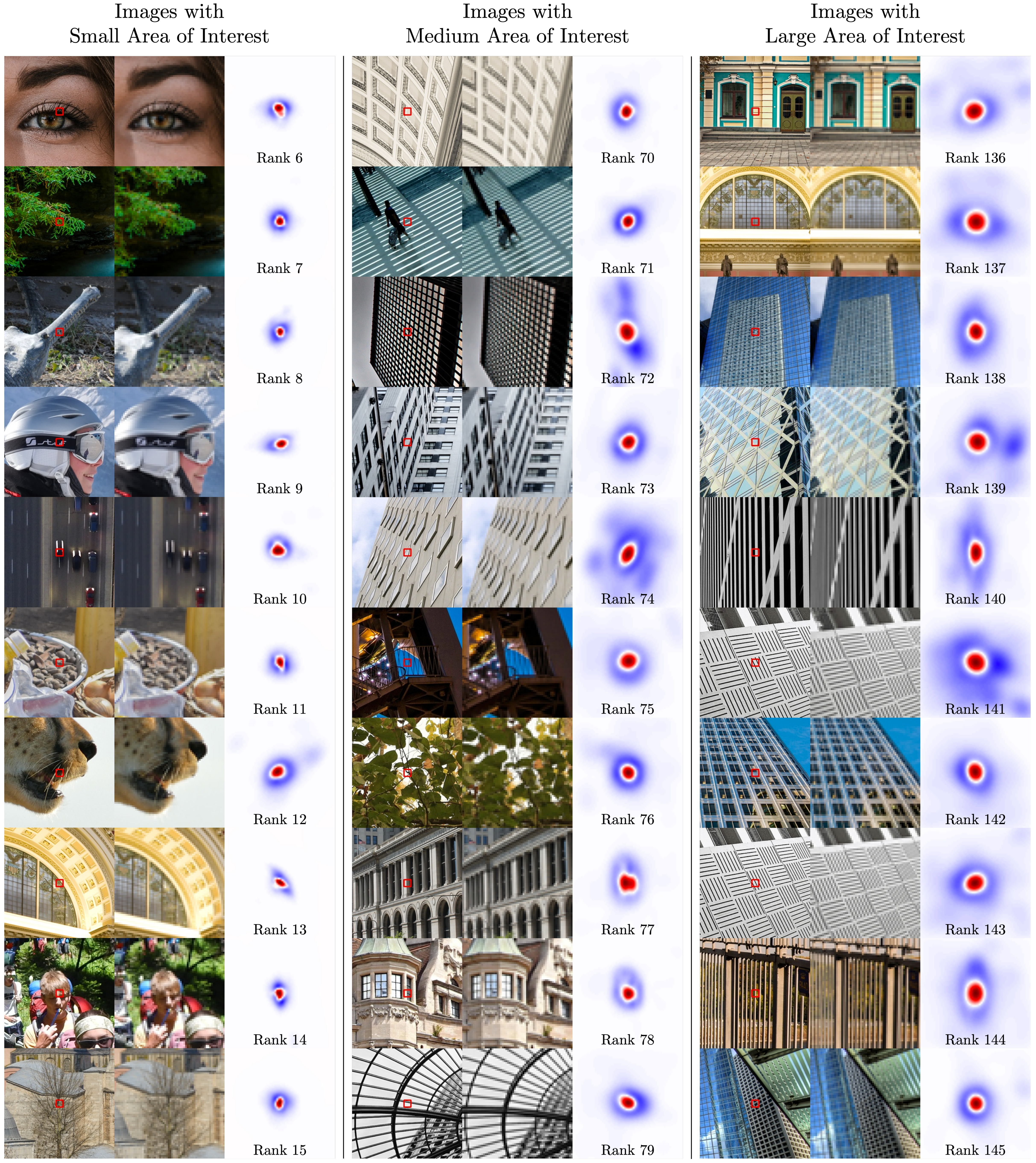
Local Attribution Maps Results



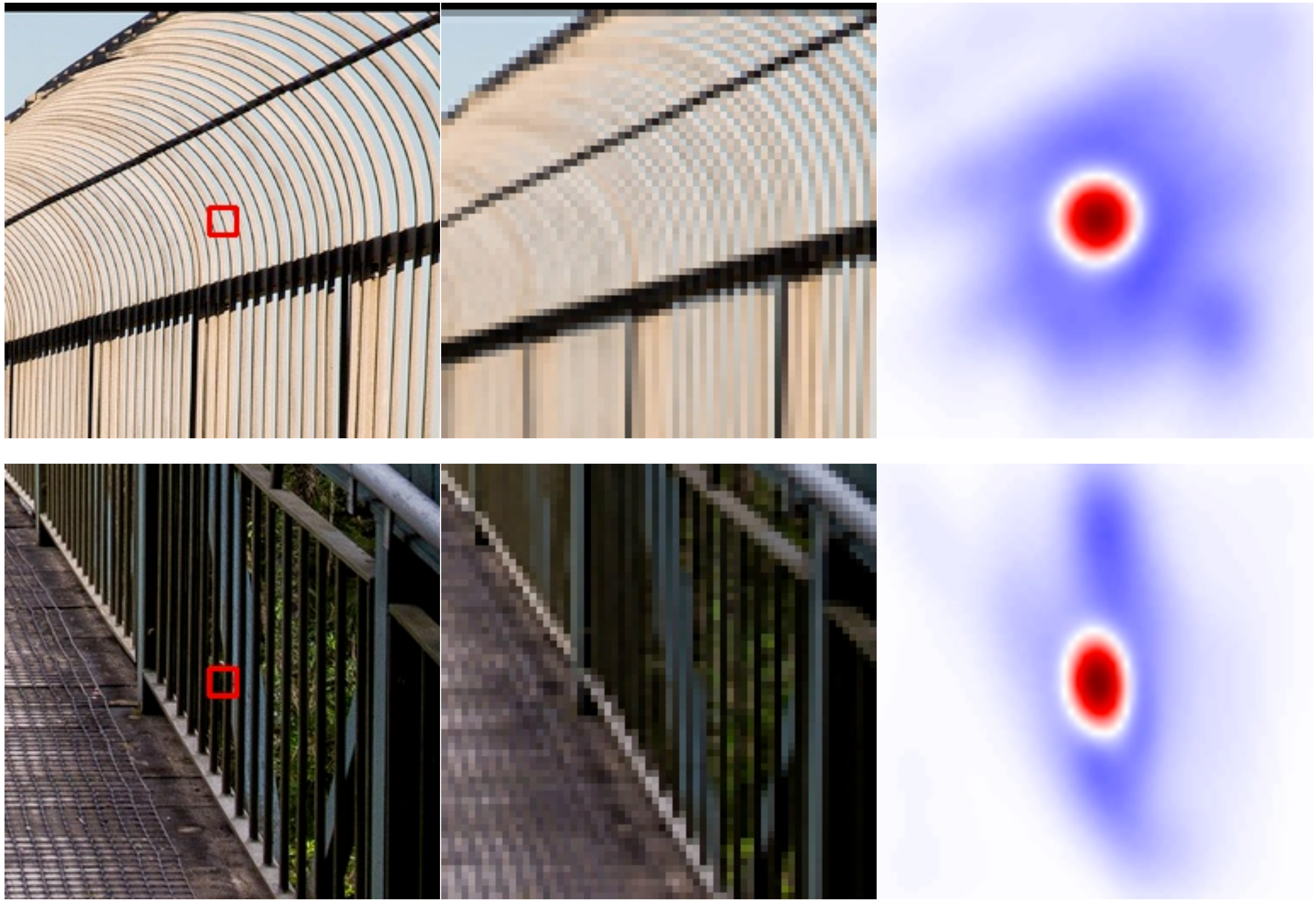
Informative Areas

The similarities and differences of LAM results for different SR networks

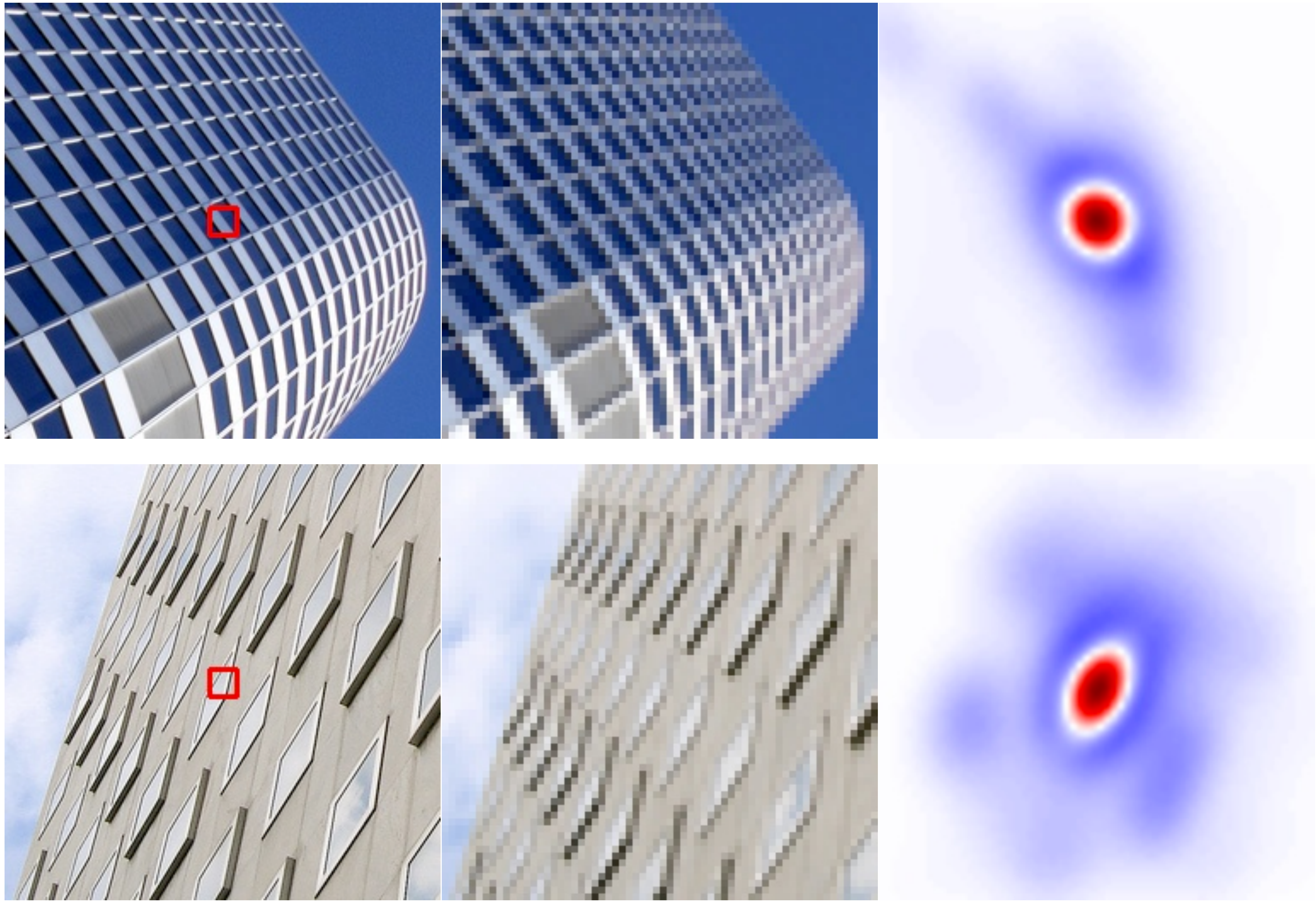
- Red areas can be used for the most preliminary level of SR
- Blue areas show the potential informative areas



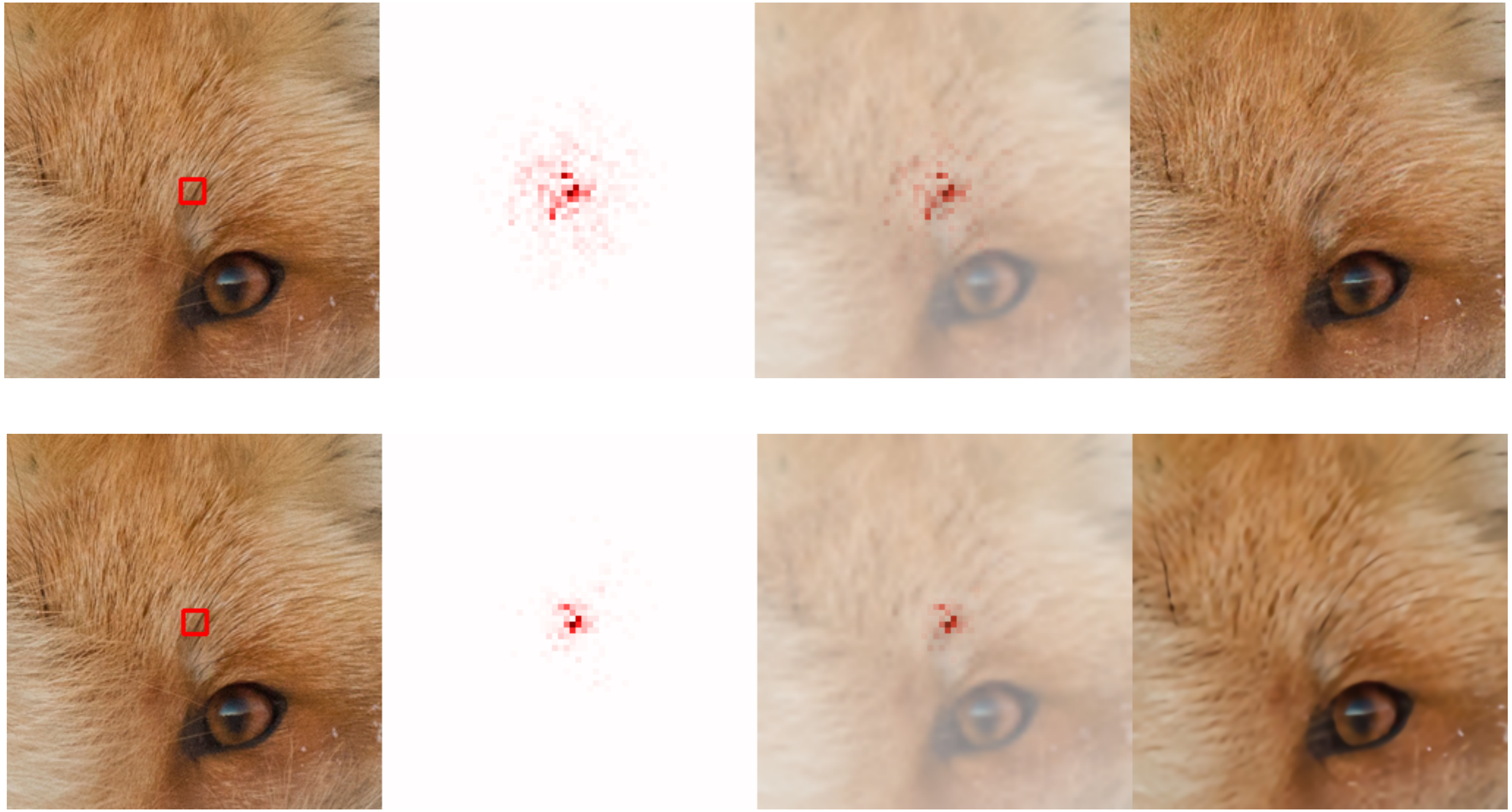
Informative Areas



Informative Areas



SRGANs Learn More Semantics



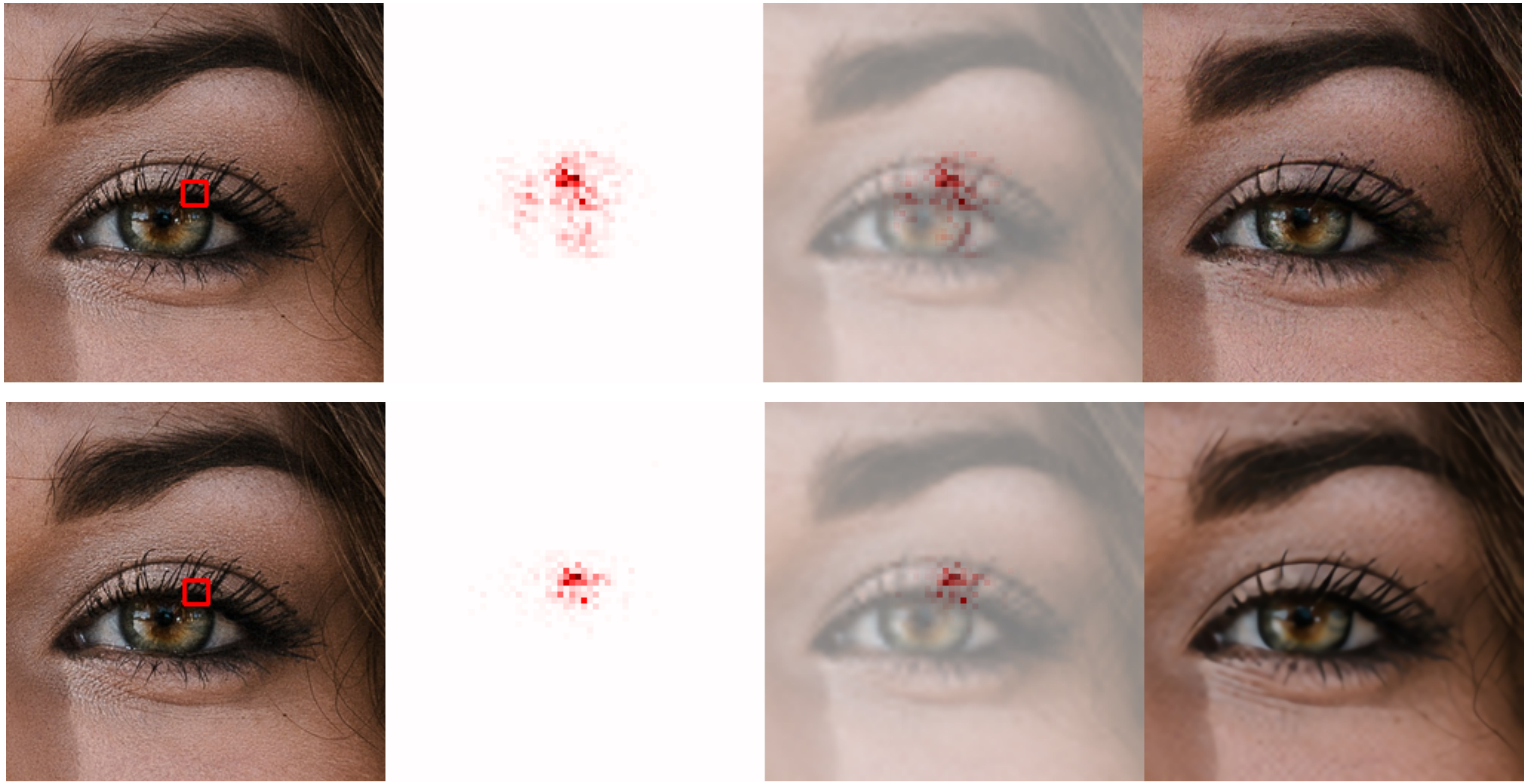
RankSRGAN

RRDBNet

SRGANs Learn More Semantics

RankSRGAN

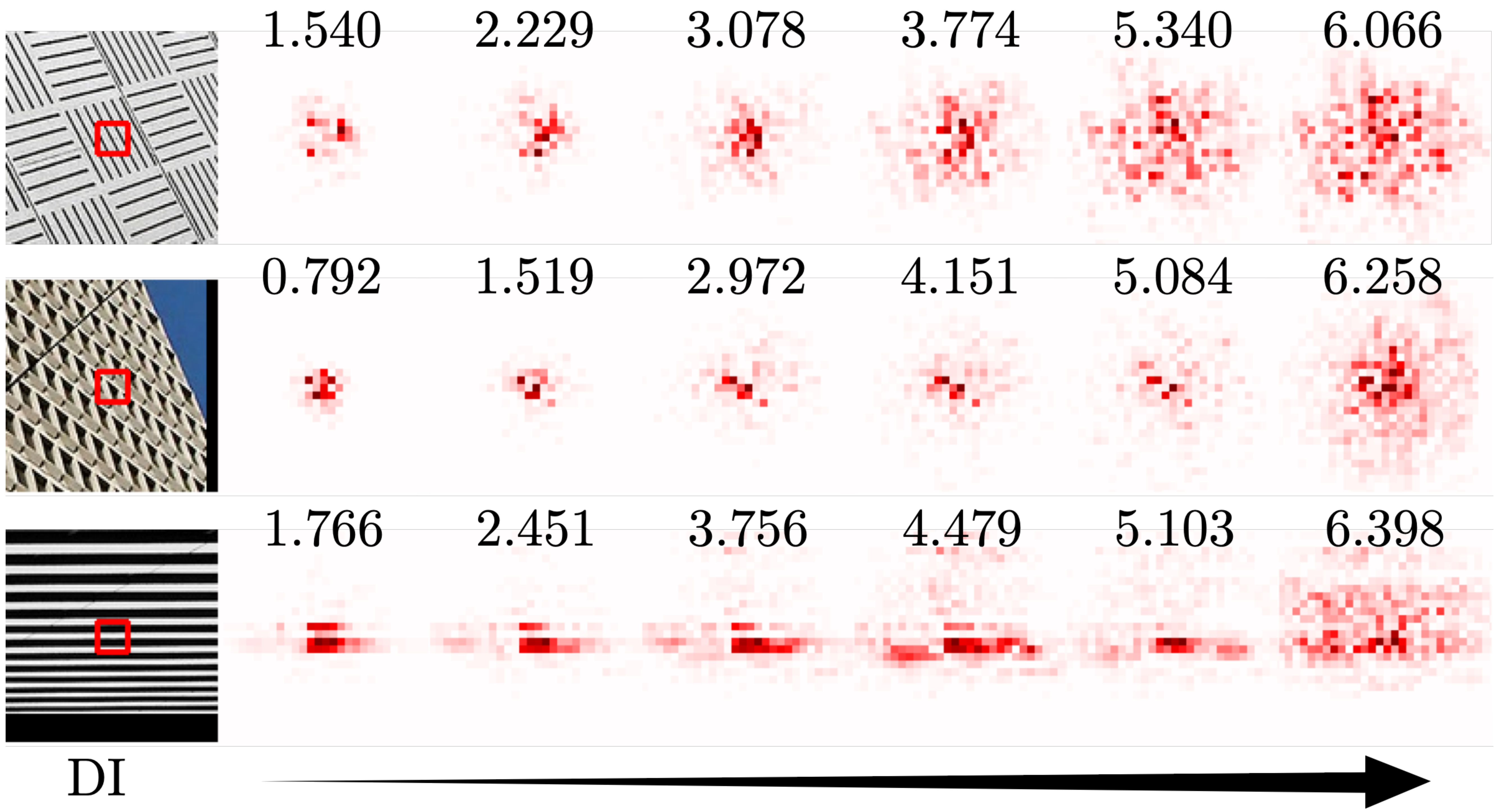
RRDBNet



Exploration with LAM

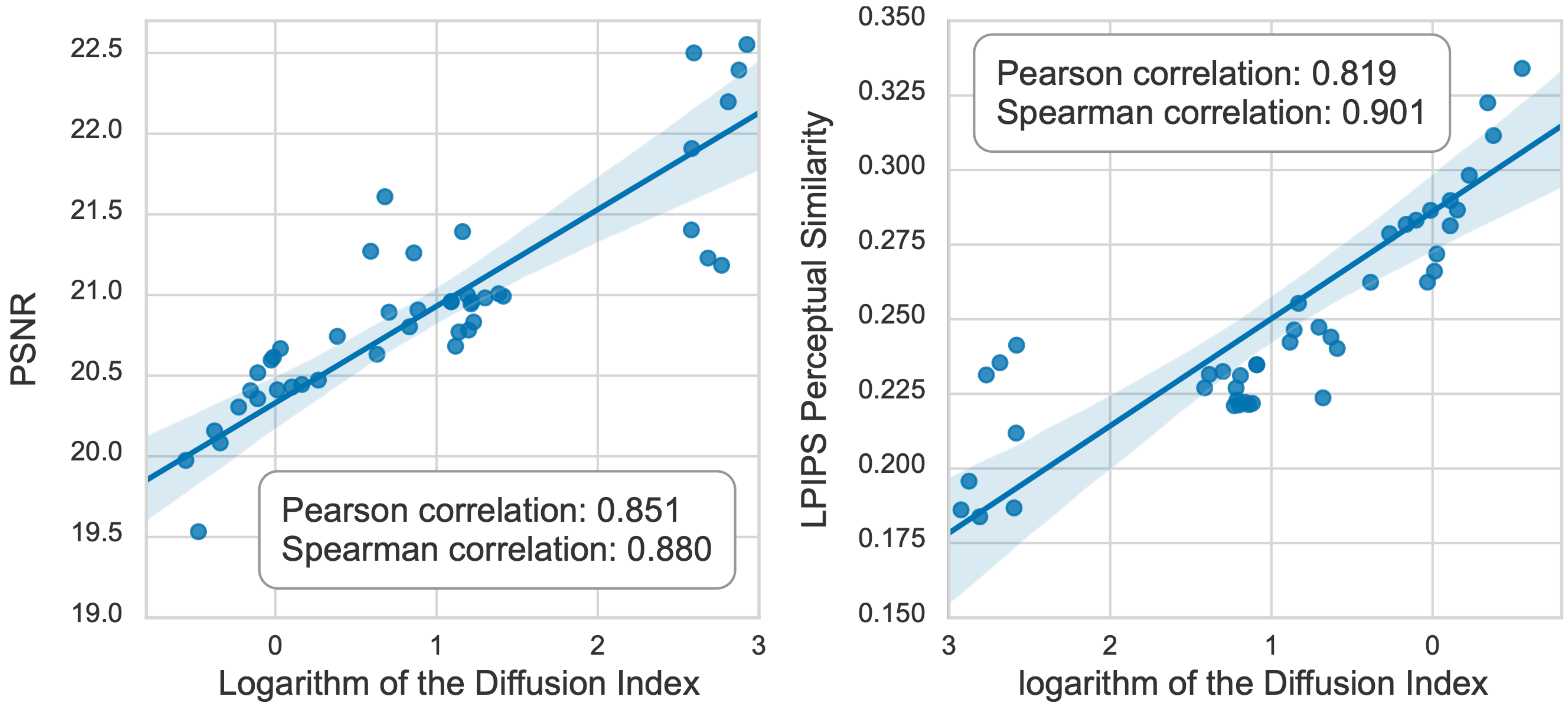
We use Gini Index to indicate the range of involved pixels: $G = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{n}|g_i - g_j|}{2n^2\bar{g}}$

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



Exploration with LAM

Diffusion Index vs. Network Performances.



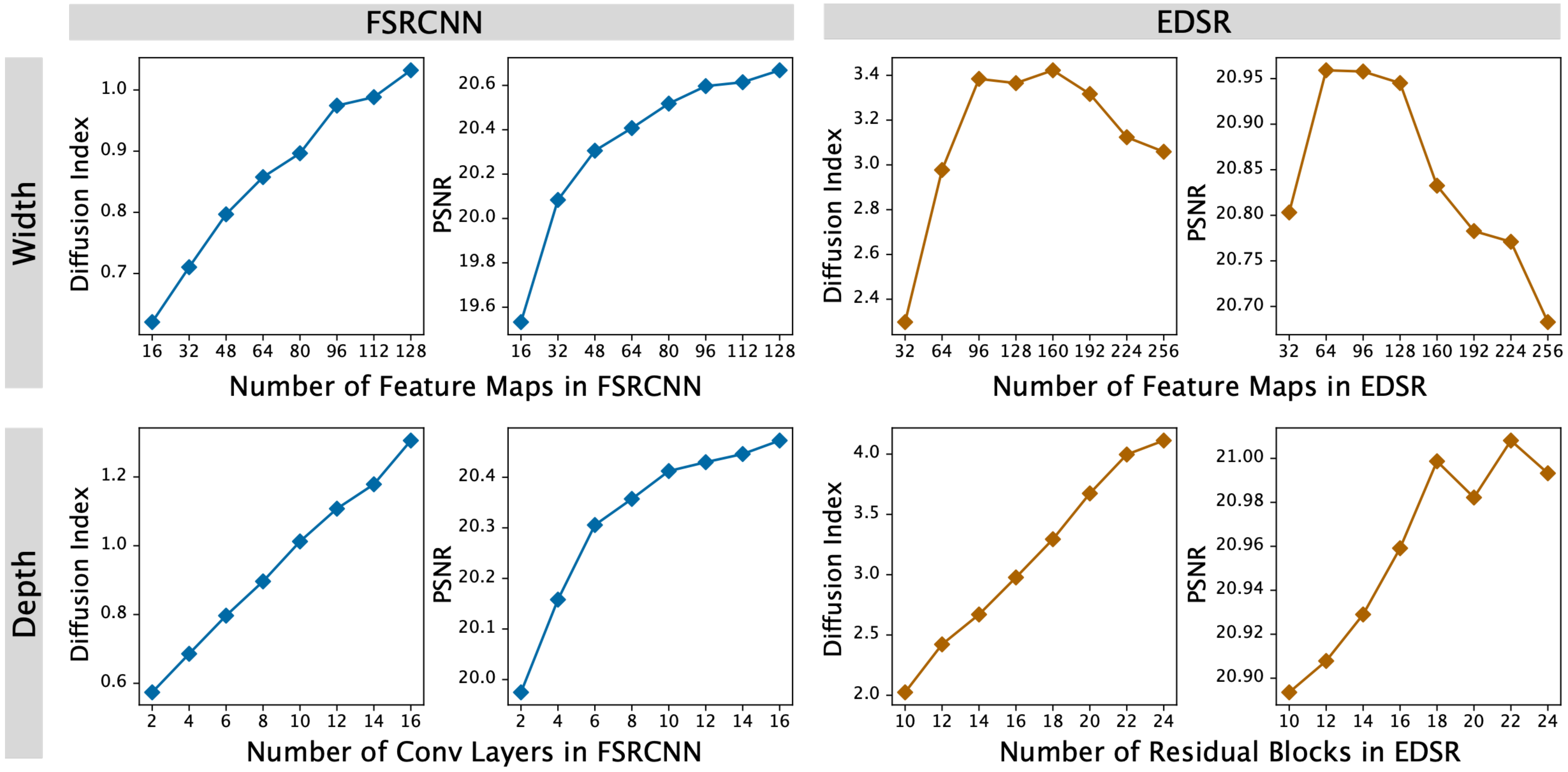
Exploration with LAM

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.

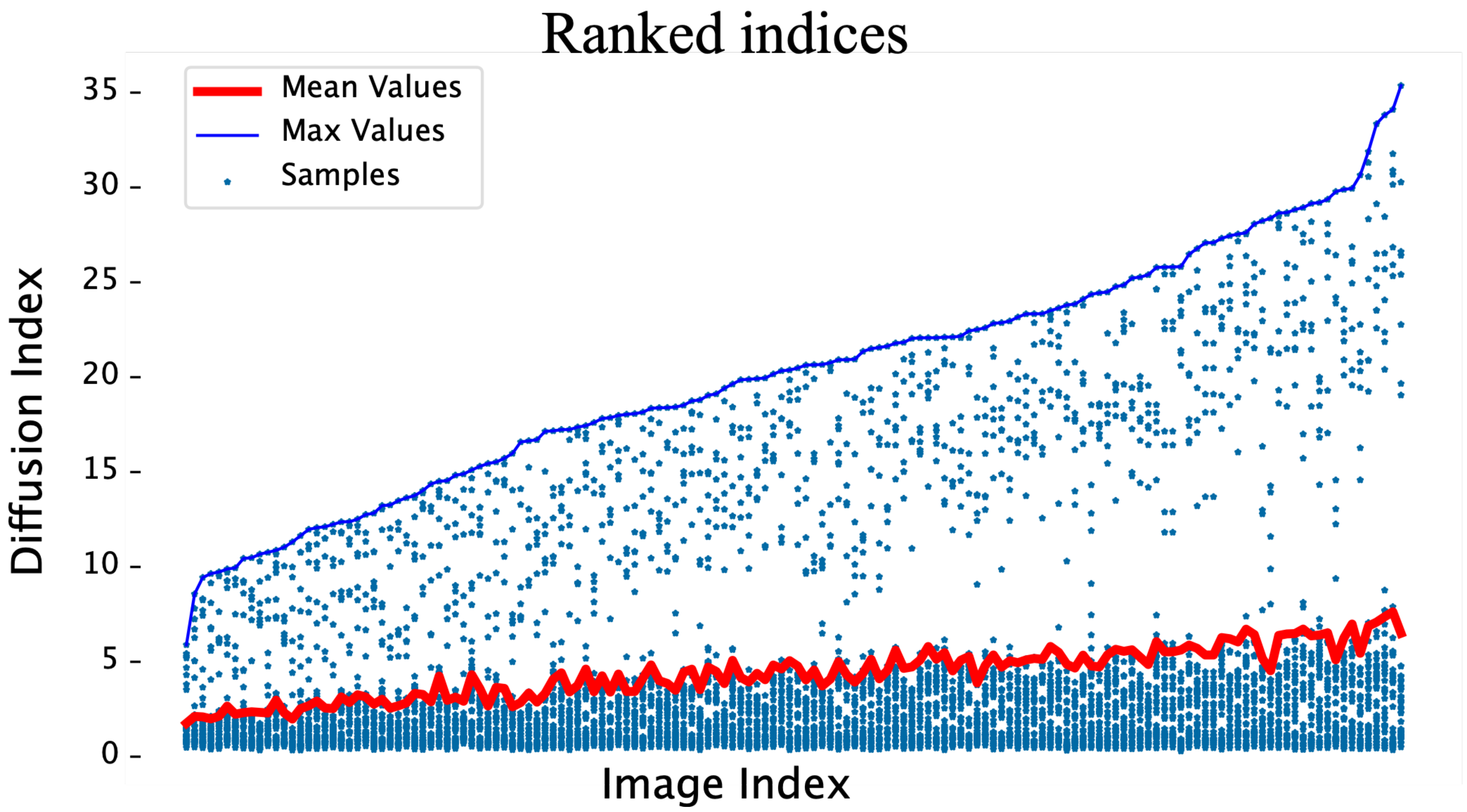
Exploration with LAM

Diffusion Index vs. Network Scale.



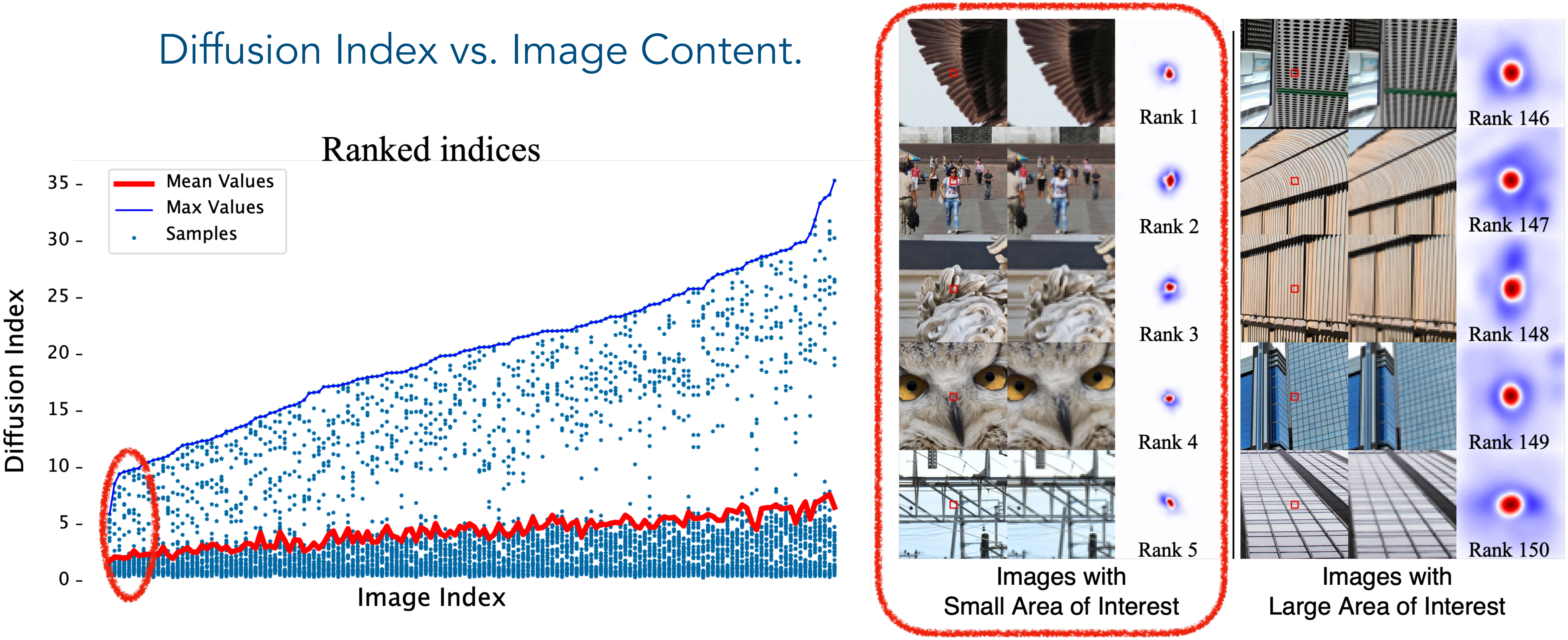
Exploration with LAM

Diffusion Index vs. Image Content.



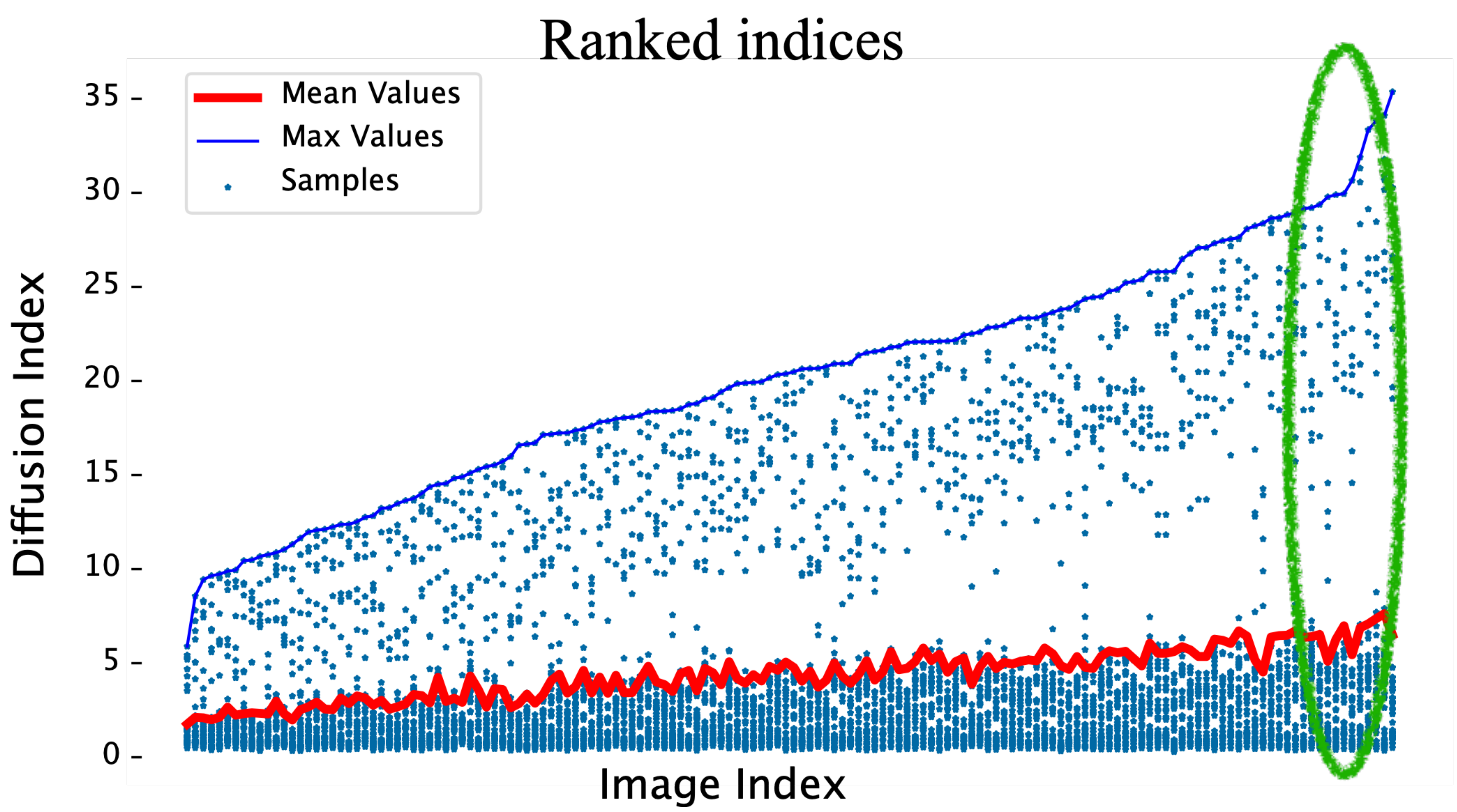
Exploration with LAM

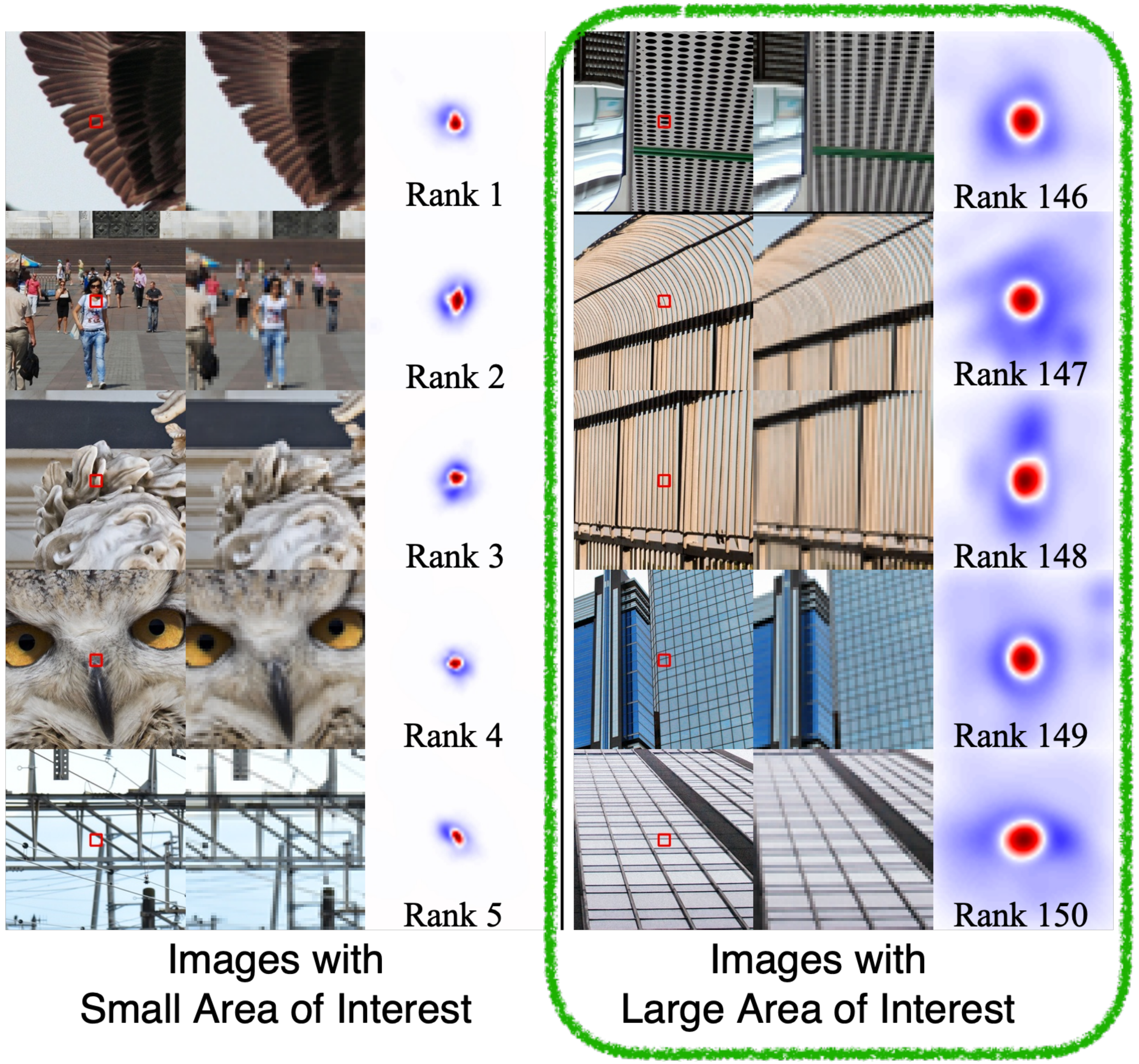
Diffusion Index vs. Image Content.



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

Interpreting Super-Resolution Networks

Interpretability in Low-Level Vision:

- **Pixel**: What pixels contribute most to restoration?
- **Feature**: Where can we find semantics in SR networks?

Discovering "Semantics" in Super-Resolution Networks

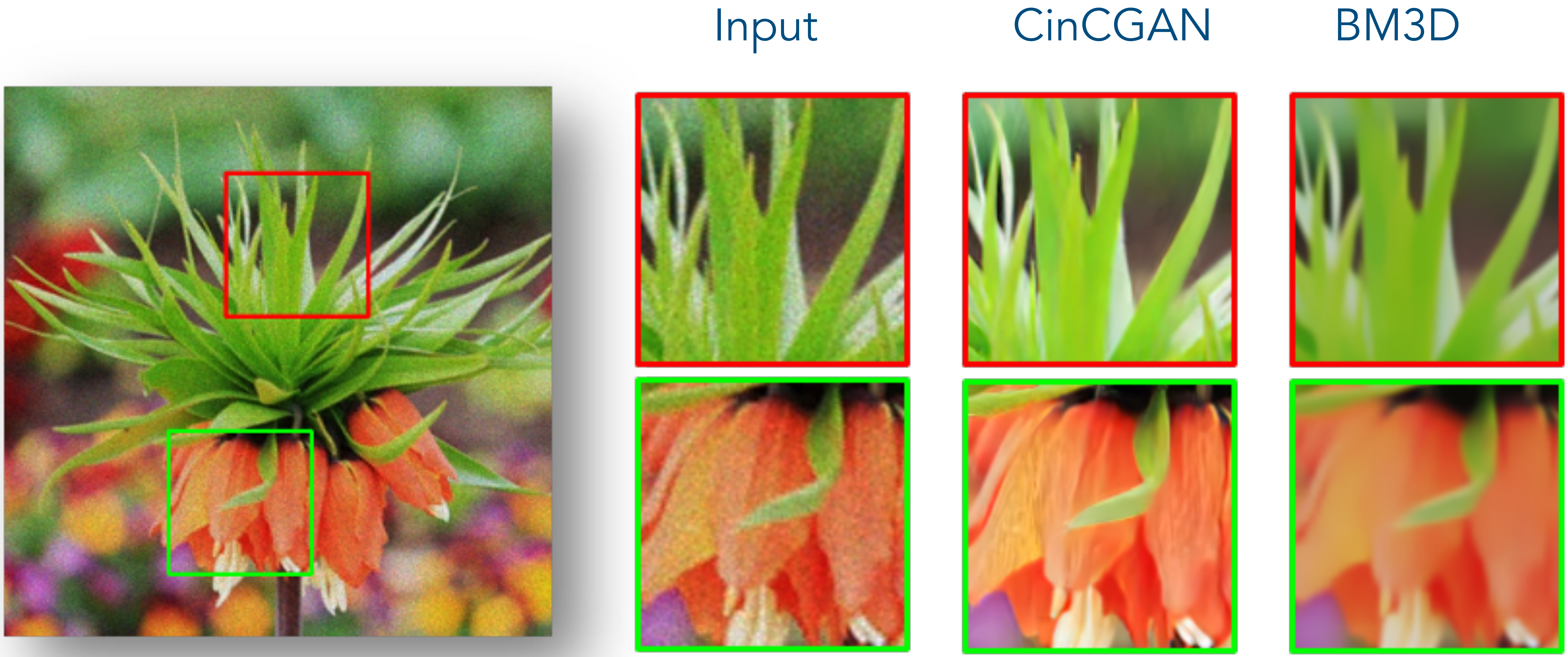
Yahoo Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao, Chao Dong

Shenzhen Institute of Advanced Technology, CAS
The University of Hong Kong,
The University of Sydney,
Shanghai AI Lab,
Institute of Automation, CAS
Baidu Inc.

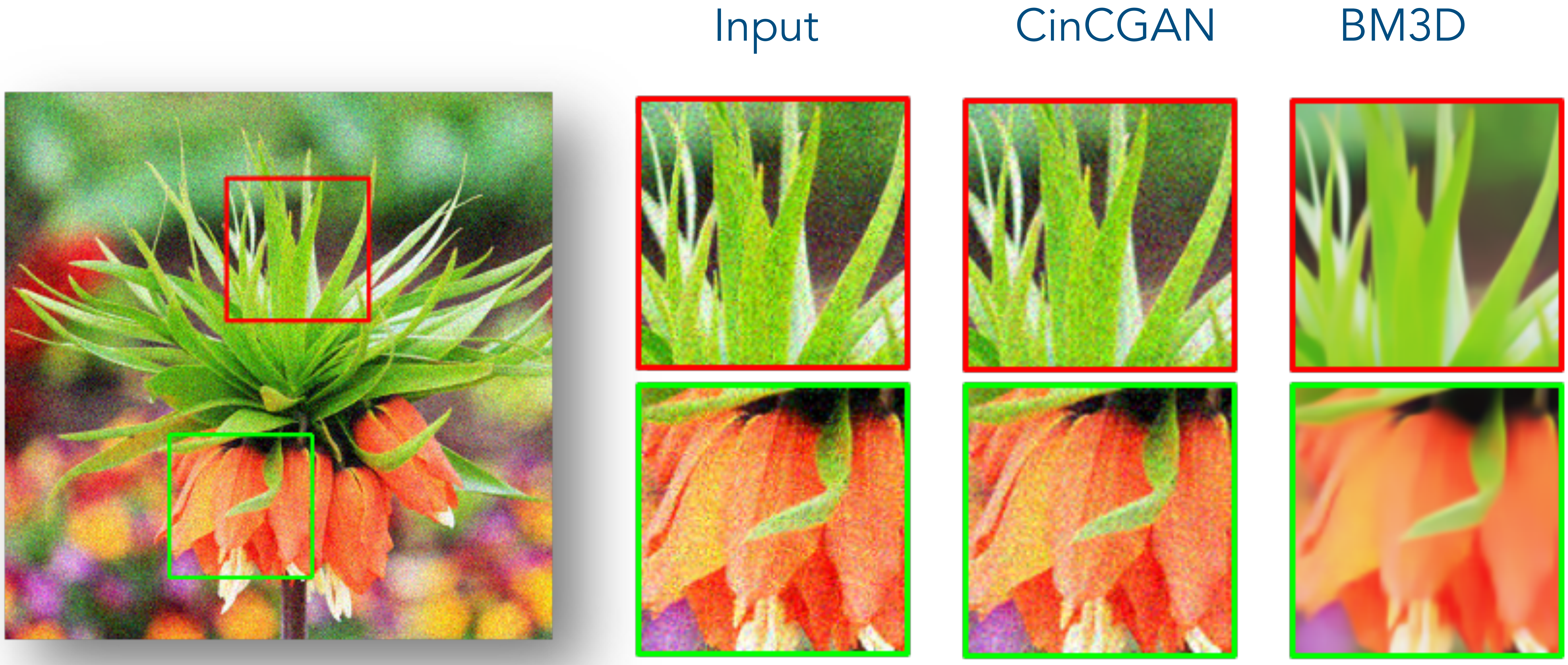
Interpreting Super-Resolution Networks

No Semantics	?? Semantics	Clear Semantics
Traditional Methods such as Interpolation methods	Low-level Vision models such as Super-Resolution Networks	High-level Vision models such as Classification networks

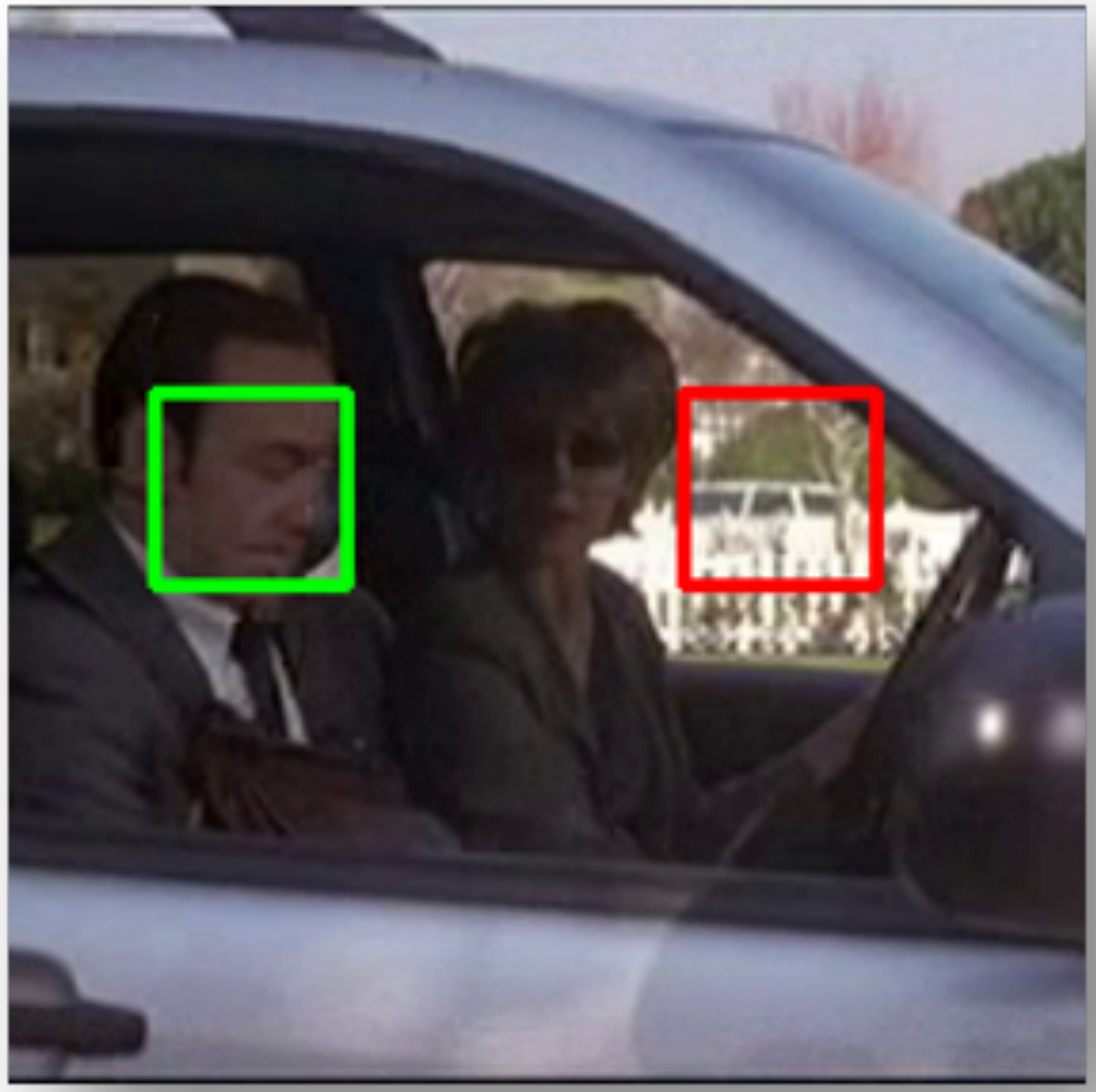
Warm up: An observation



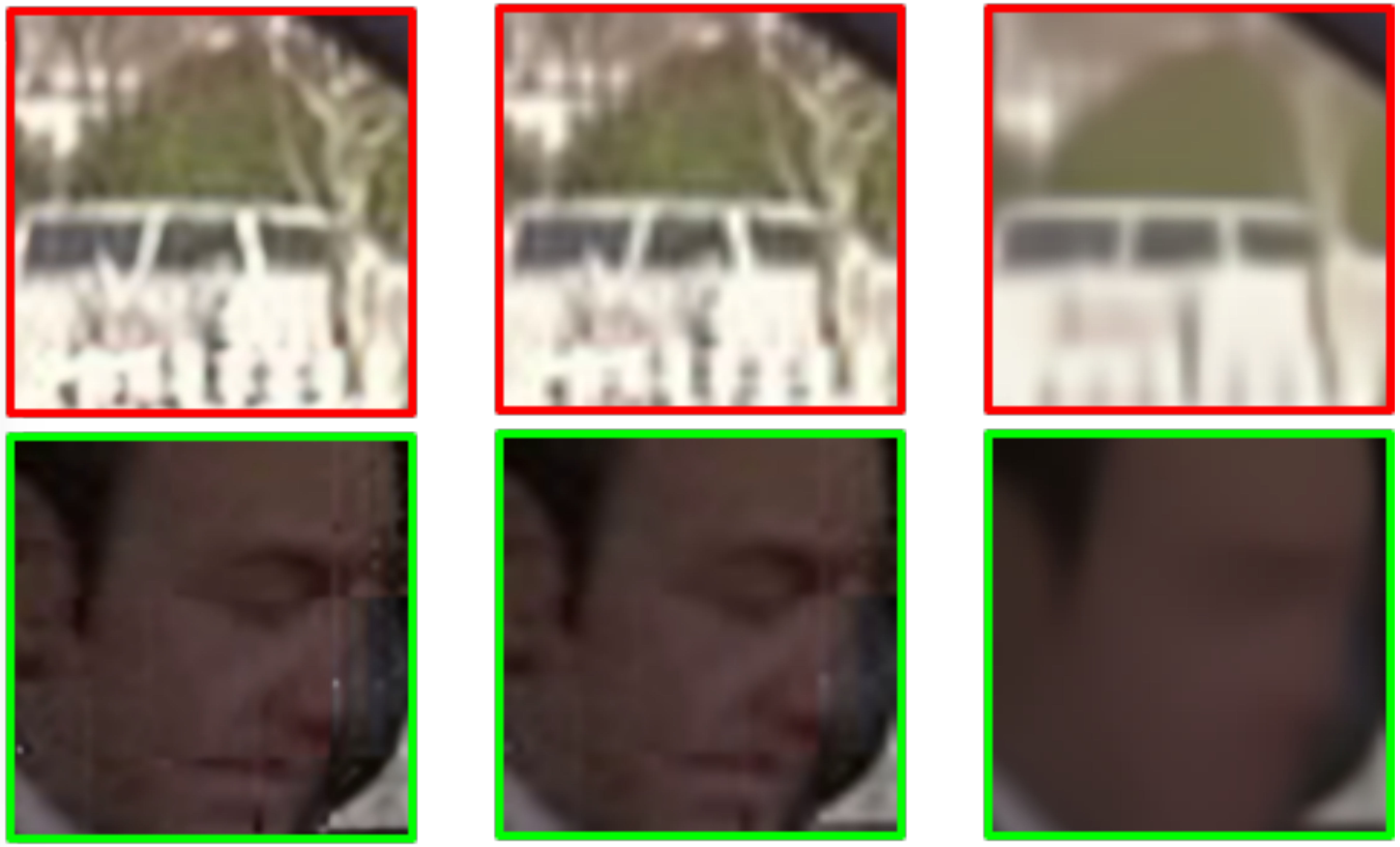
Warm up: An observation



Warm up: An observation



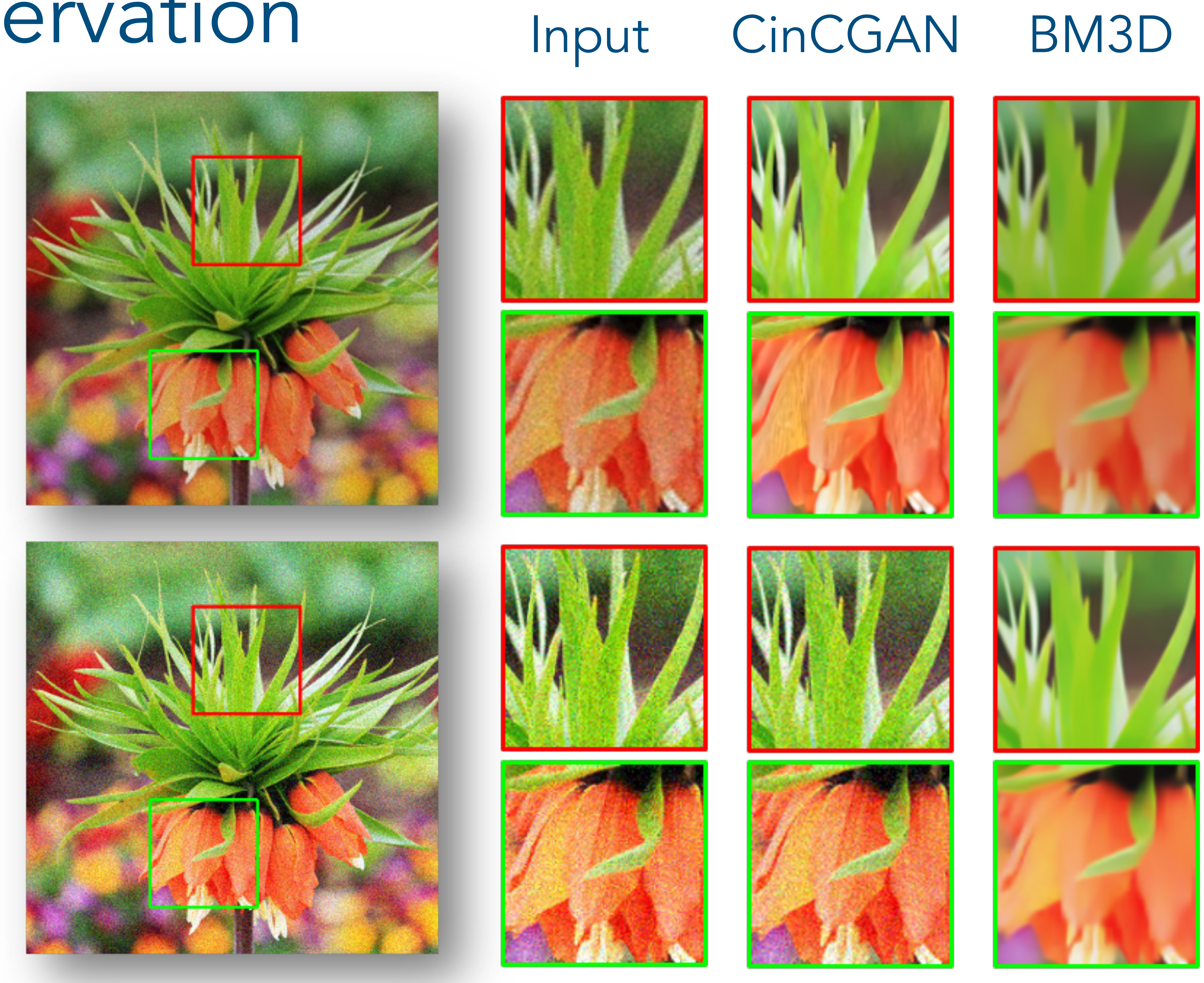
Input CinCGAN BM3D



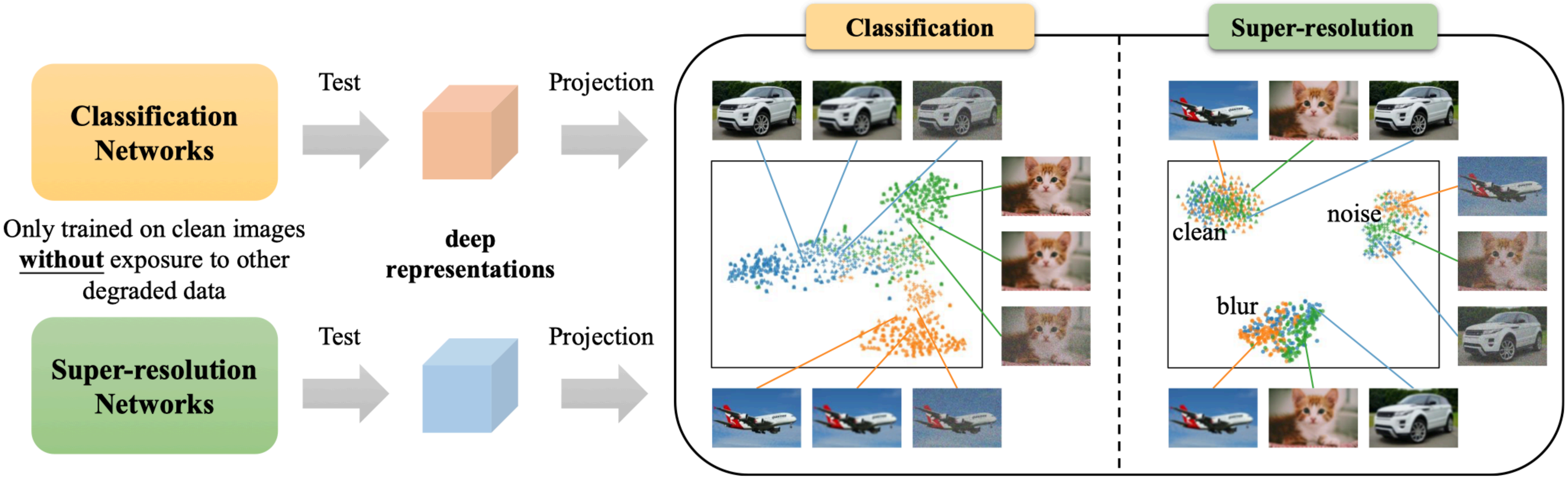
Warm up: An observation

- CinCGAN can figure out the specific degradation within its training data

- The degradation mismatch will make the network "**turn off**" its ability



Methodology



Methodology

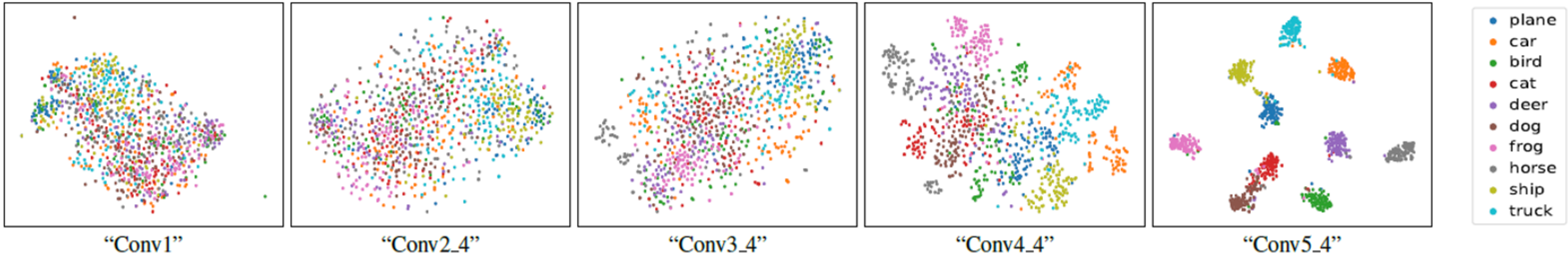
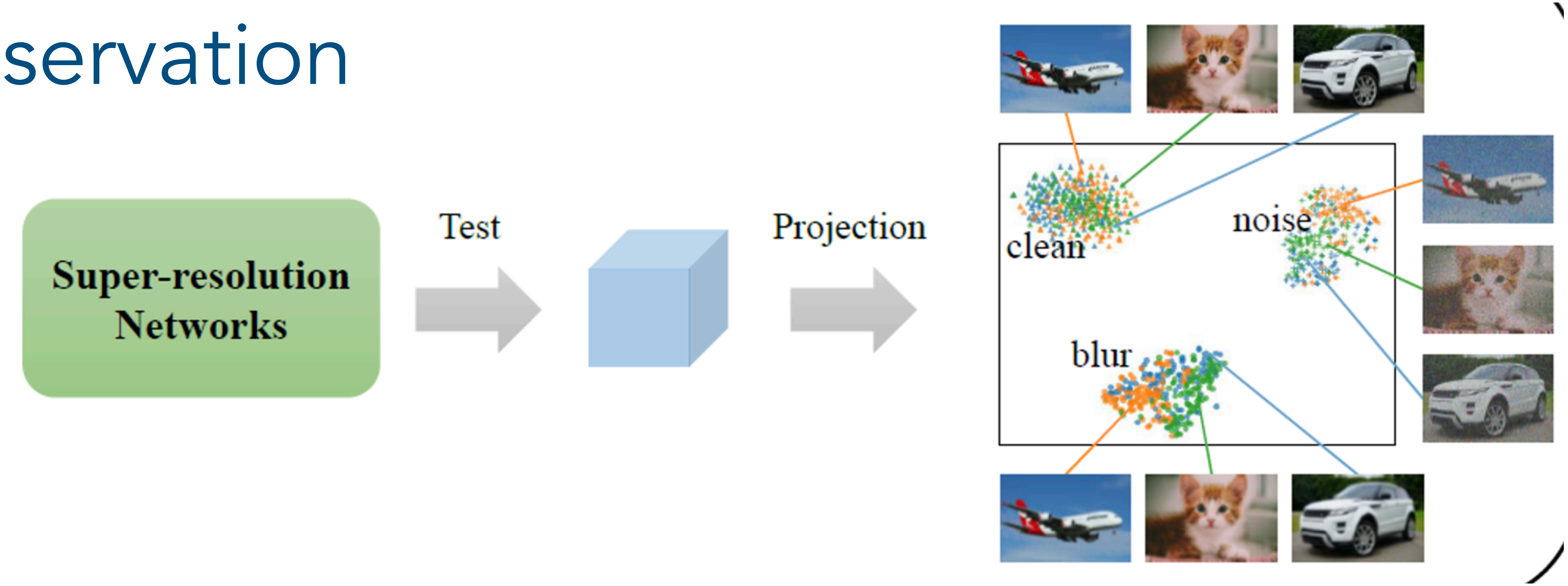
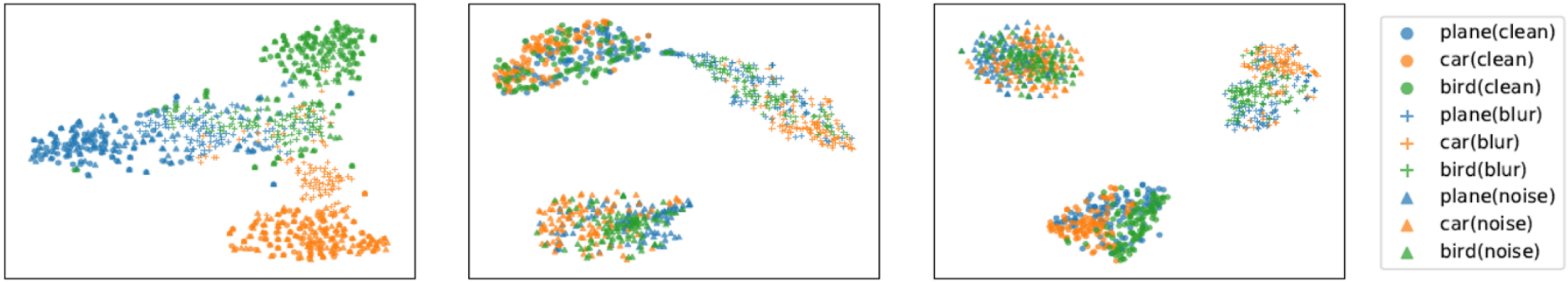


Figure 4. Projected feature representations extracted from different layers of ResNet18 using t-SNE. With the network deepens, the representations become more discriminative to object categories, which clearly shows the semantics of the representations in classification.

Observation

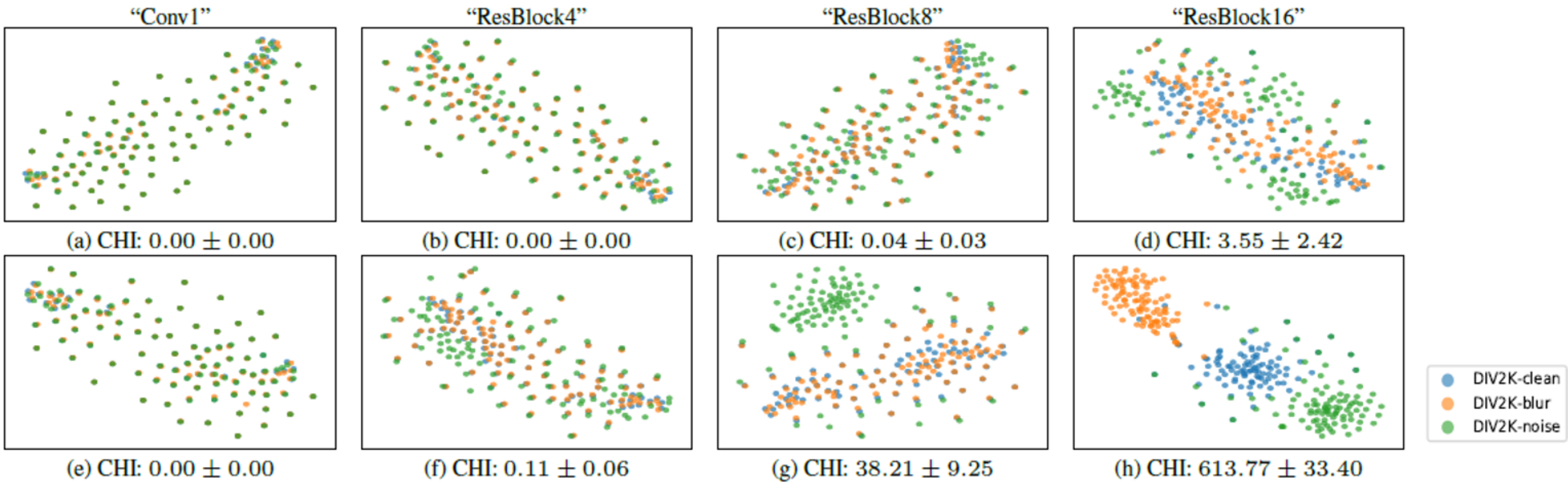




(a) ResNet18 (classification) (b) SRResNet-wGR (c) SRGAN-wGR

Observation

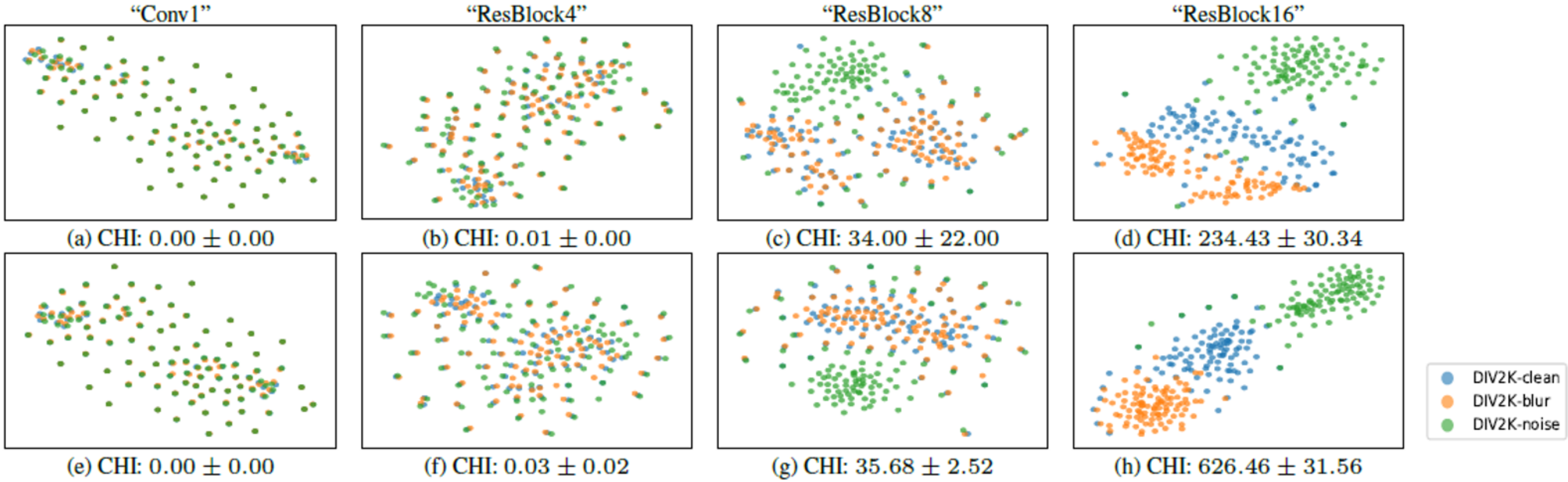
- **SR networks with global residual shows discriminability shows more obvious discriminability to different types.**
- GAN-based SR networks shows more obvious discriminability.



Observation

- SR networks with global residual shows discriminability shows more obvious discriminability to different types.

- **GAN-based SR networks shows more obvious discriminability.**



Inspirations

- Interpreting the Generalization of SR (low-level) Networks
- Developing degradation-adaptive Algorithms
- Disentanglement of Image Content/Degradation
- Degradation Classification/Detection

Thanks

PIPAL Dataset

X Pixel Group

www.jasongt.com