北京交通大學
BEIJING JIAOTONG UNIVERSITY
MePro
数字媒体信息处理研究中心
Center of Digital Media Information Processing

Meeting of Paper Sharing

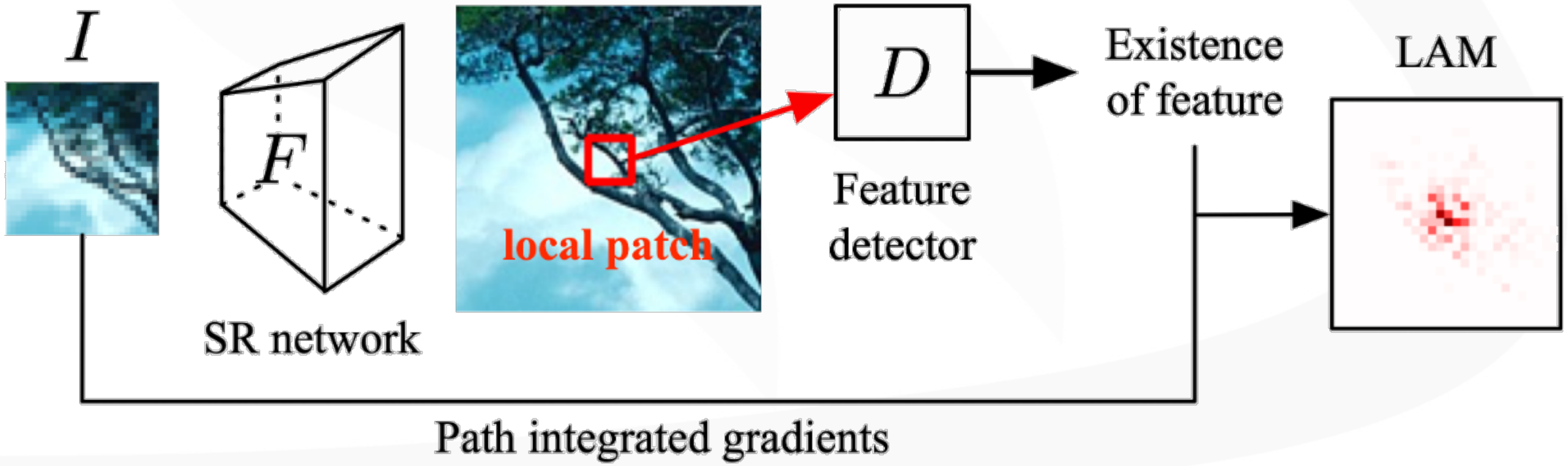
Insight into the Super-Resolution Network

Qi Tang

2023/2/19

Interpreting Super-Resolution Networks with Local Attribution Maps

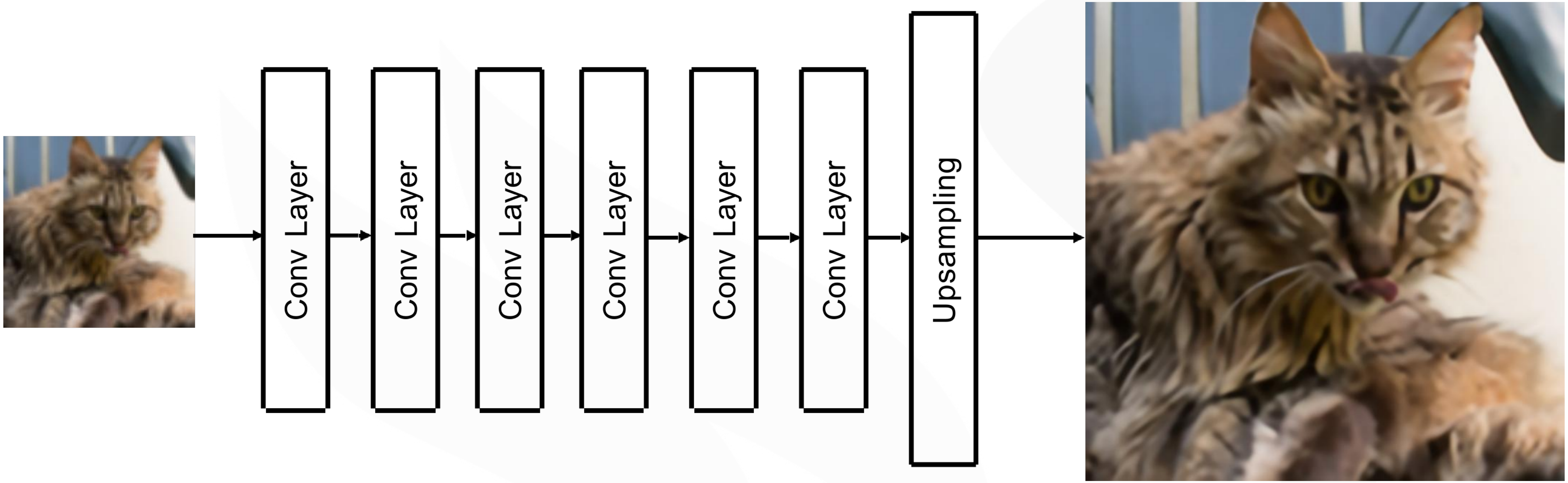
Jinjin Gu[1] **Chao Dong**[2,3]
[1]School of Electrical and Information Engineering, The University of Sydney.
[2]Key Laboratory of Human-Machine Intelligence-Synergy Systems,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.
[3]SIAT Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society



Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

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.

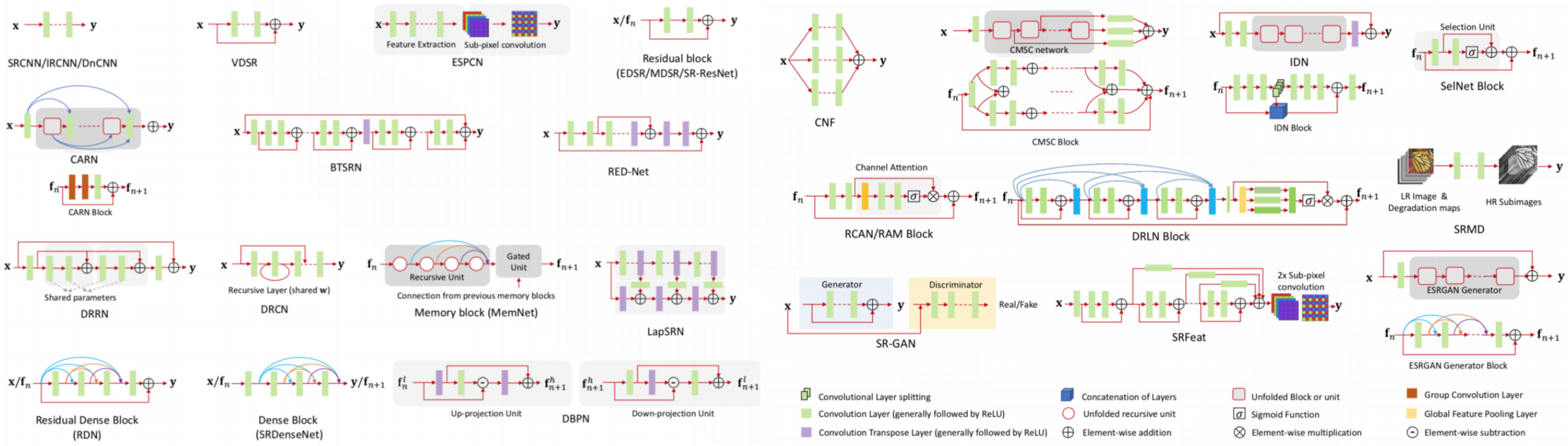
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

Super-Resolution Networks

Many SR network architectures have been proposed.
What makes their different performance?



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

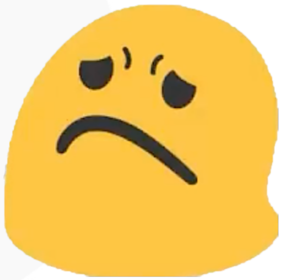
Pixel: What pixels contribute most to restoration?

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?

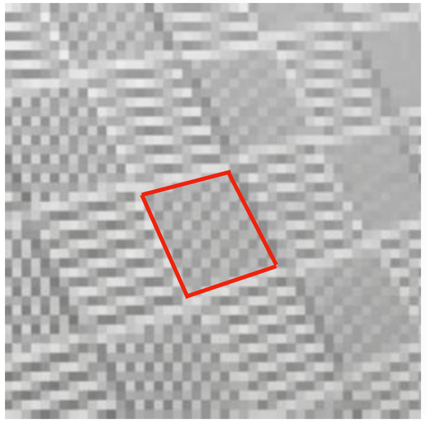
We lack understanding toward these questions
And also research tools



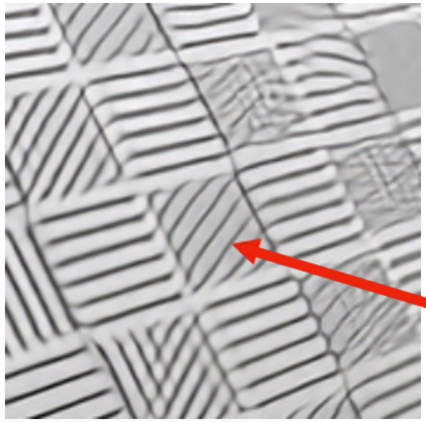
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

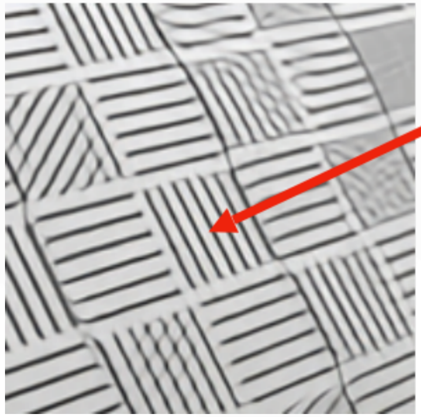
Attribution Analysis



Input Image



EDSR



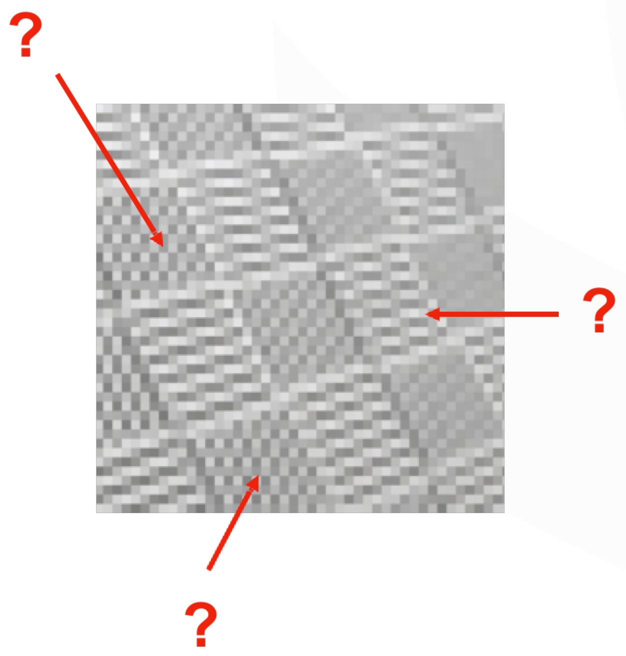
RARN

Why RNAN gives correct results in the center?

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis



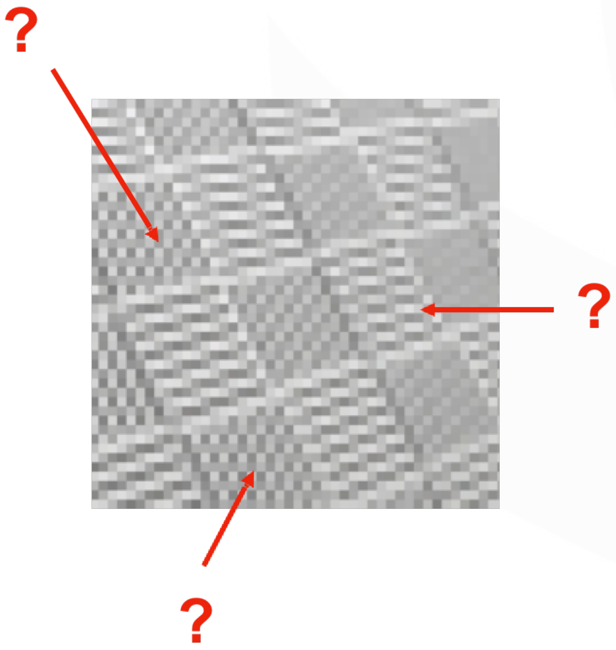
What did RNAN notice from the input that allowed it to make the correct prediction?

Does EDSR notice this information?

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis



Identify input features responsible for SR results.

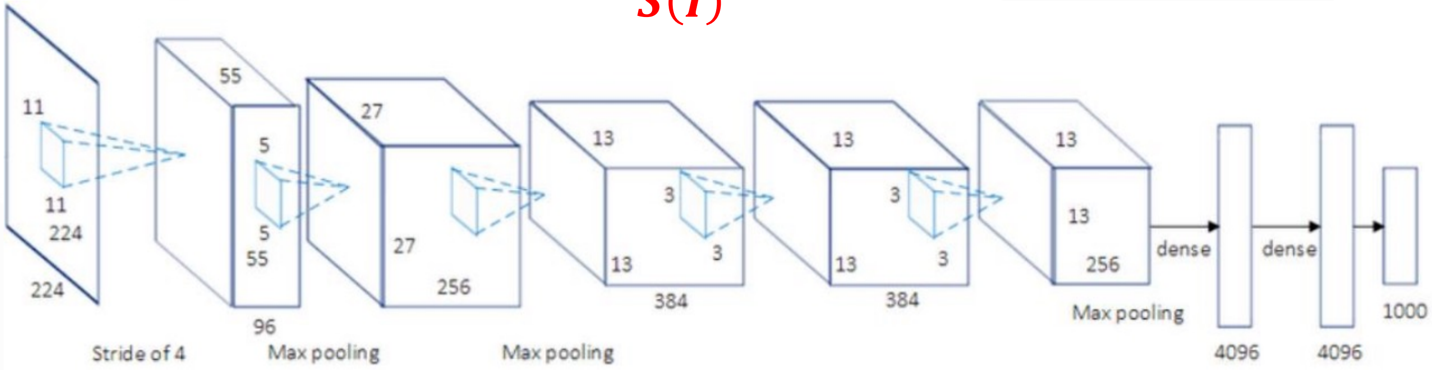
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks

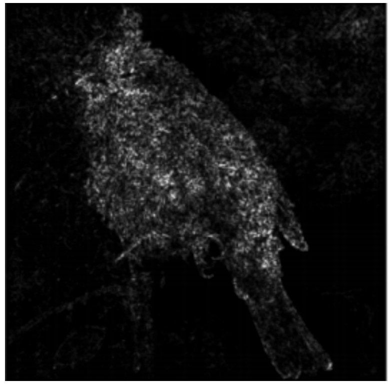
What is S(I) looking at?

$S(I)$

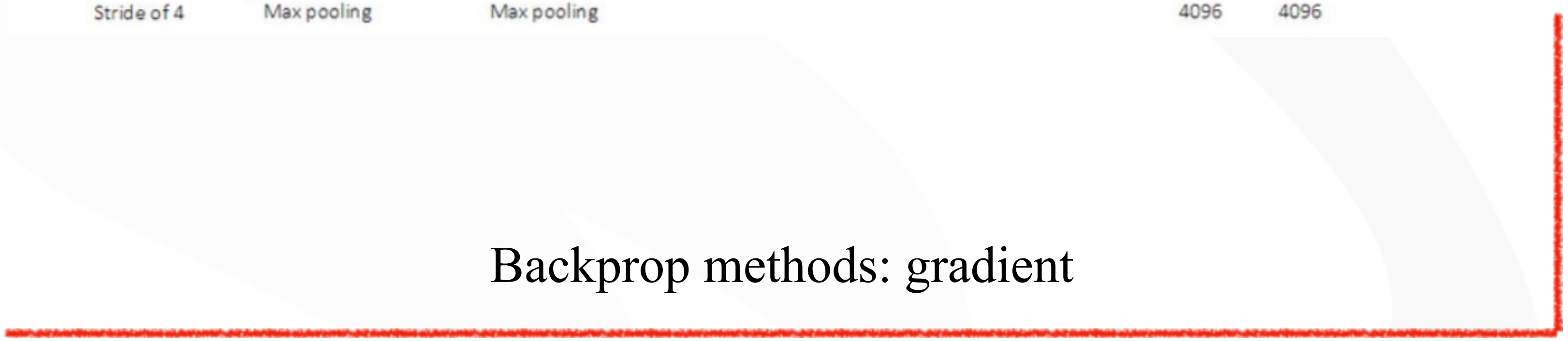


98% house finch
10% bird
1% People

I



Backprop methods: gradient



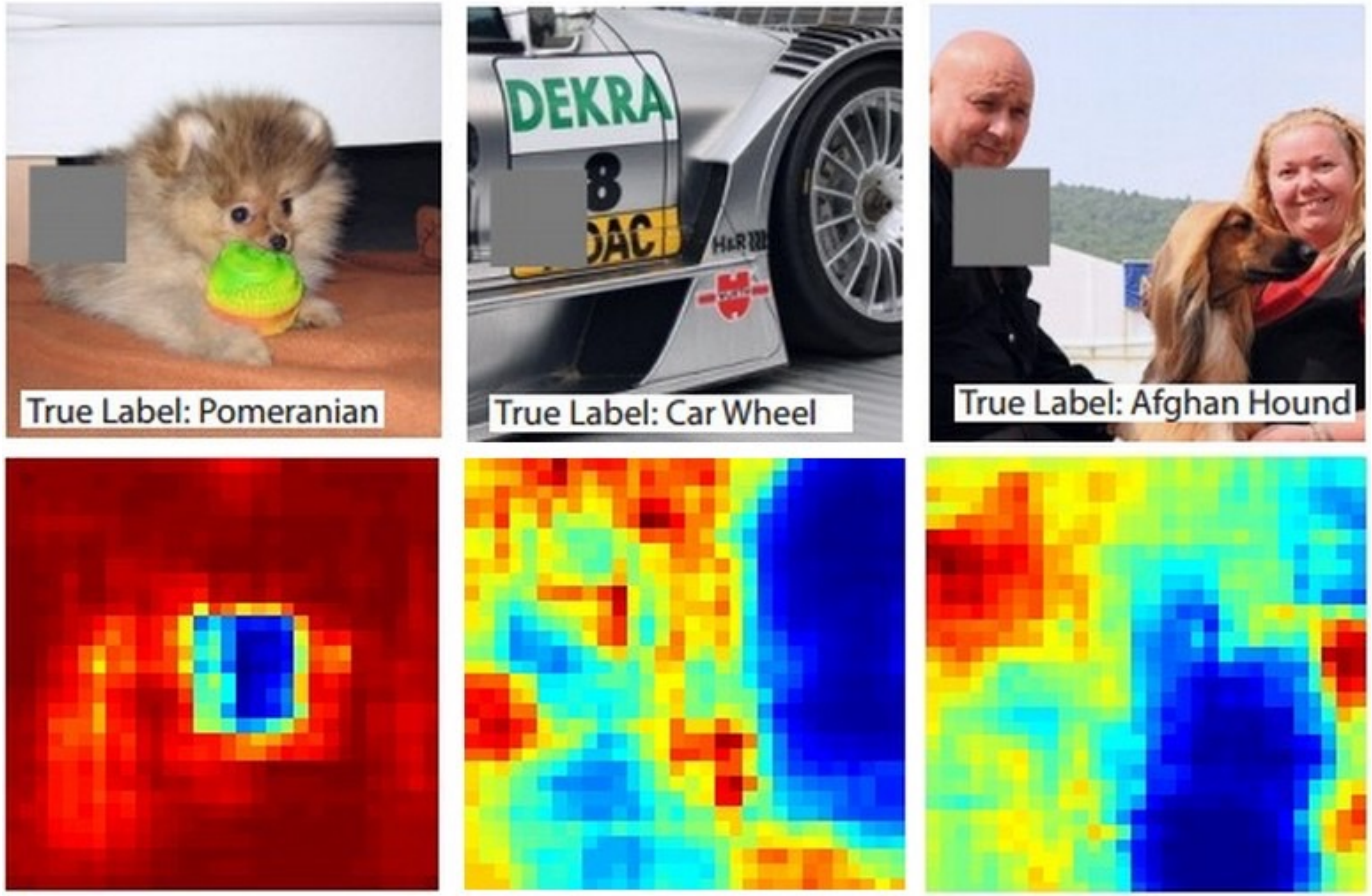
$$\mathrm{Grad}_S(I) = \frac{\partial S(I)}{\partial I}$$

The visualized attribution map

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

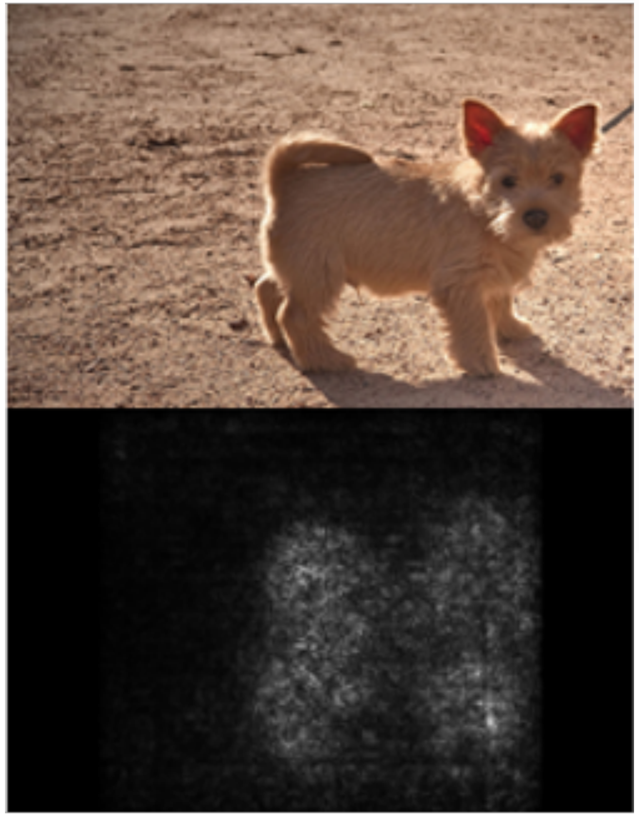
Pixel: What pixels contribute most to restoration?

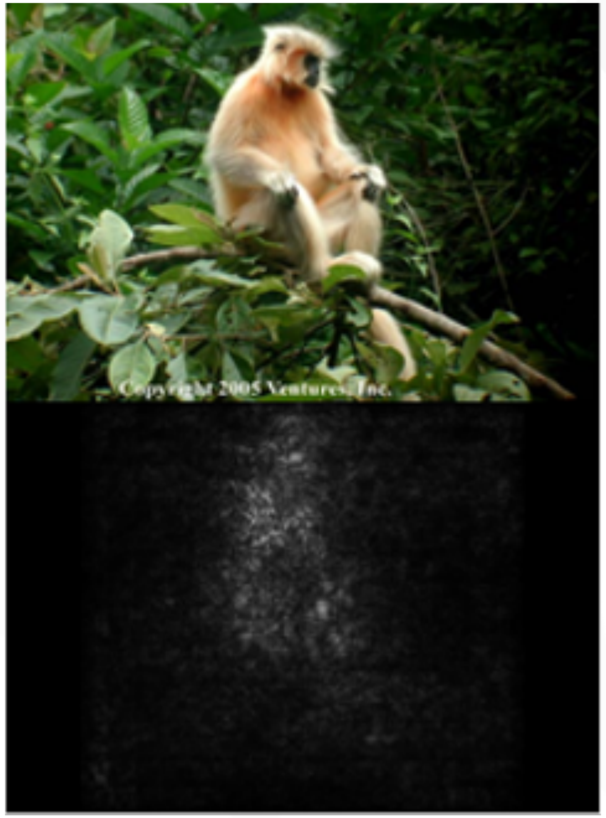
Attribution Analysis for High-level Networks

$$\{x_1, \cdots, x_n, \cdots, x_N\} \rightarrow \{x_1, \cdots, x_n + \Delta x, \cdots, x_N\}$$

$$y_k \rightarrow y_k + \Delta y$$

$$\left|\frac{\Delta y}{\Delta x}\right| \rightarrow \left|\frac{\partial y_k}{\partial x_n}\right|$$

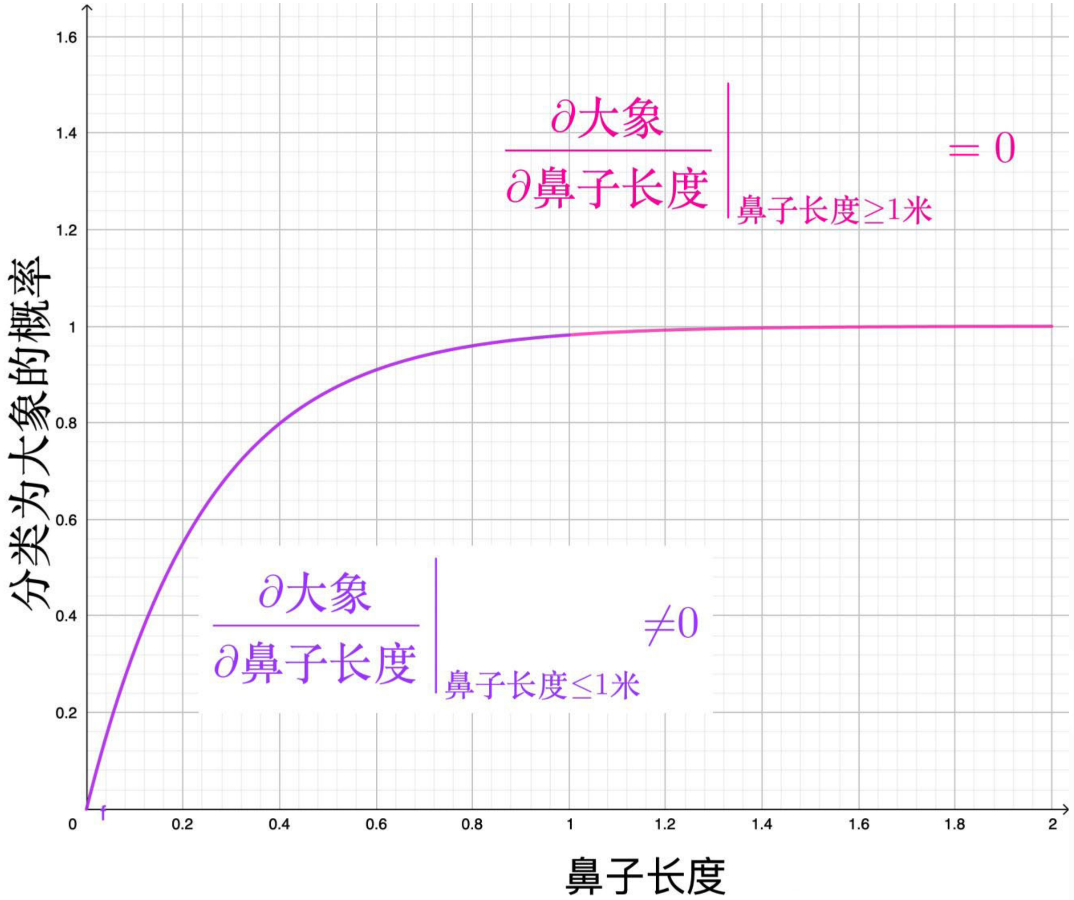




Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks



$$特征重要性 = \int_0^{2m} \frac{\partial 大象}{\partial 鼻子长度} \partial 鼻子长度$$

$$x' + \alpha(x - x')$$

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks

$$\phi_i^{IG}(f, x, x') = \overbrace{(x_i - x'_i)}^{\text{Difference from baseline}} \times \int_{\alpha=0}^{1} \frac{\delta f(x' + \alpha(x - x'))}{\delta x_i} d\alpha$$

$$\gamma(\alpha) = x' + \alpha(x - x')$$

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

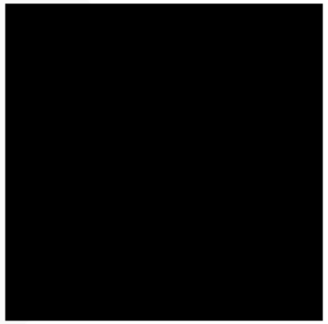
Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks

$$\phi_i^{IG}(f, x, x') = \overbrace{(x_i - x_i')}^{\text{Difference from baseline}} \times \int_{\alpha=0}^{1} \frac{\delta f(x' + \alpha(x - x'))}{\delta x_i} d\alpha$$

$$\gamma(\alpha) = x' + \alpha(x - x')$$

- Generate the baseline input. In case of image, we generate all-zero image to as the baseline



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

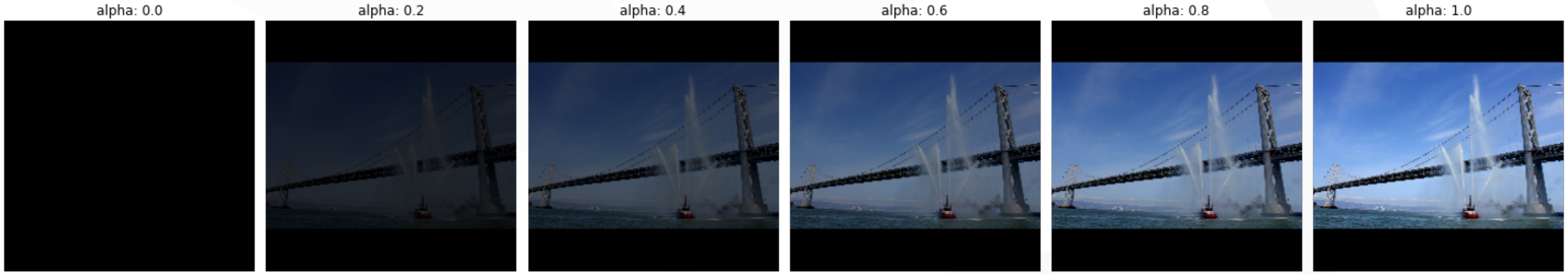
Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks

$$\phi_i^{IG}(f, x, x') = \overbrace{(x_i - x'_i)}^{\text{Difference from baseline}} \times \int_{\alpha=0}^{1} \frac{\delta f(x' + \alpha(x - x'))}{\delta x_i} d\alpha$$

$$\gamma(\alpha) = x' + \alpha(x - x')$$

- Generate the baseline input. In case of image, we generate all-zero image to as the baseline
- Compute the α-blended between the baseline input and the actual input.



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

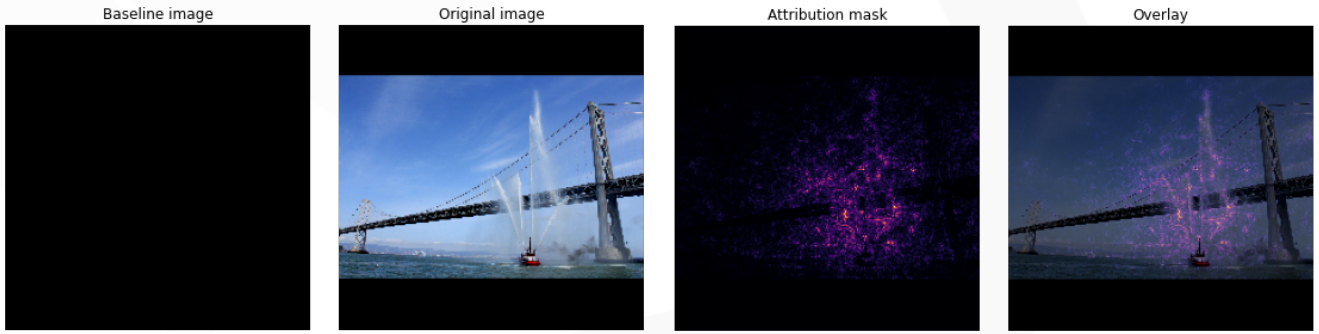
Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks

$$\phi_i^{IG}(f, x, x') = \overbrace{(x_i - x_i')}^{\text{Difference from baseline}} \times \int_{\alpha=0}^{1} \frac{\delta f(x' + \alpha(x - x'))}{\delta x_i} d\alpha$$

$$\gamma(\alpha) = x' + \alpha(x - x')$$

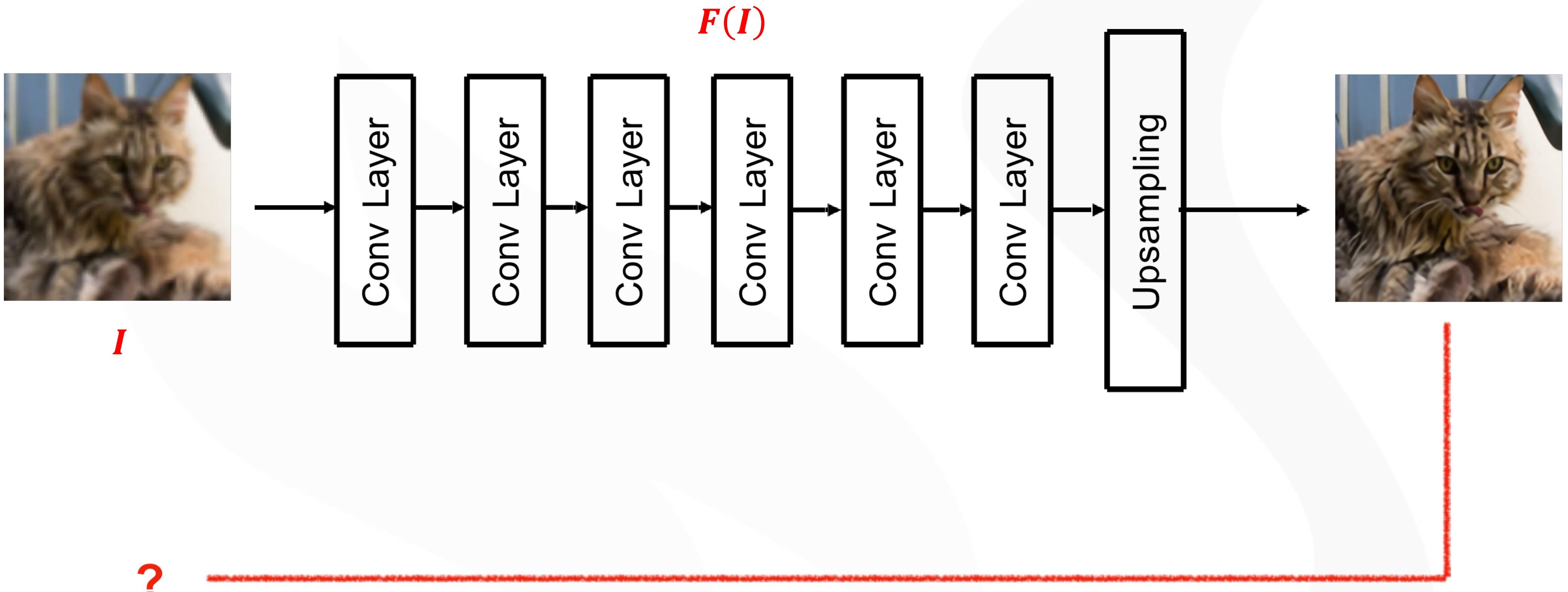
- Generate the baseline input. In case of image, we generate all-zero image to as the baseline
- Compute the α-blended between the baseline input and the actual input.
- Compute the gradient for all α-blended images. Then estimate the attribute from the gradient and visualize with the original image.



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Attribution Analysis for High-level Networks



How to calculate gradient for low-level networks?

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

Auxiliary Principles

We introduce auxiliary principles for interpreting low-level networks:

- Interpreting local not global

SR networks can not be interpreted globally

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

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

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

北京交通大學
BEIJING JIAOTONG UNIVERSITY

Auxiliary Principles

We introduce auxiliary principles for interpreting low-level networks:

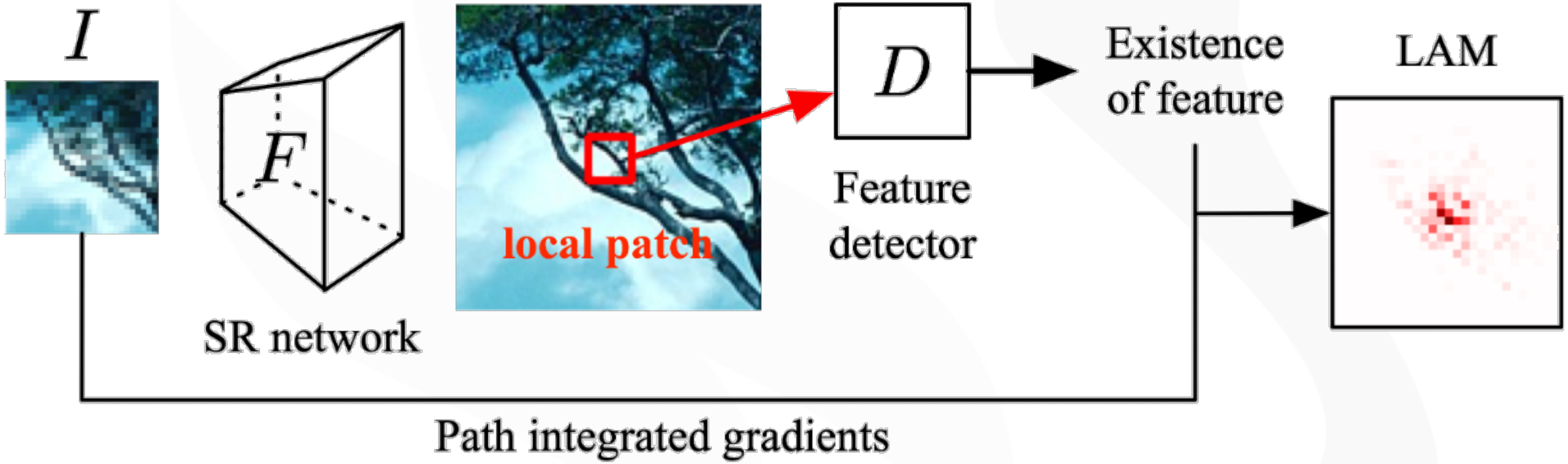
- Interpreting local not global
- Interpreting hard not simple
- Interpreting features not pixels

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

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Local Attribution Maps (LAM)



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

北京交通大學 BEIJING JIAOTONG UNIVERSITY

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 $\gamma(\alpha), \alpha \in R$

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

Input $\gamma(1) = I$

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Local Attribution Maps (LAM)

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

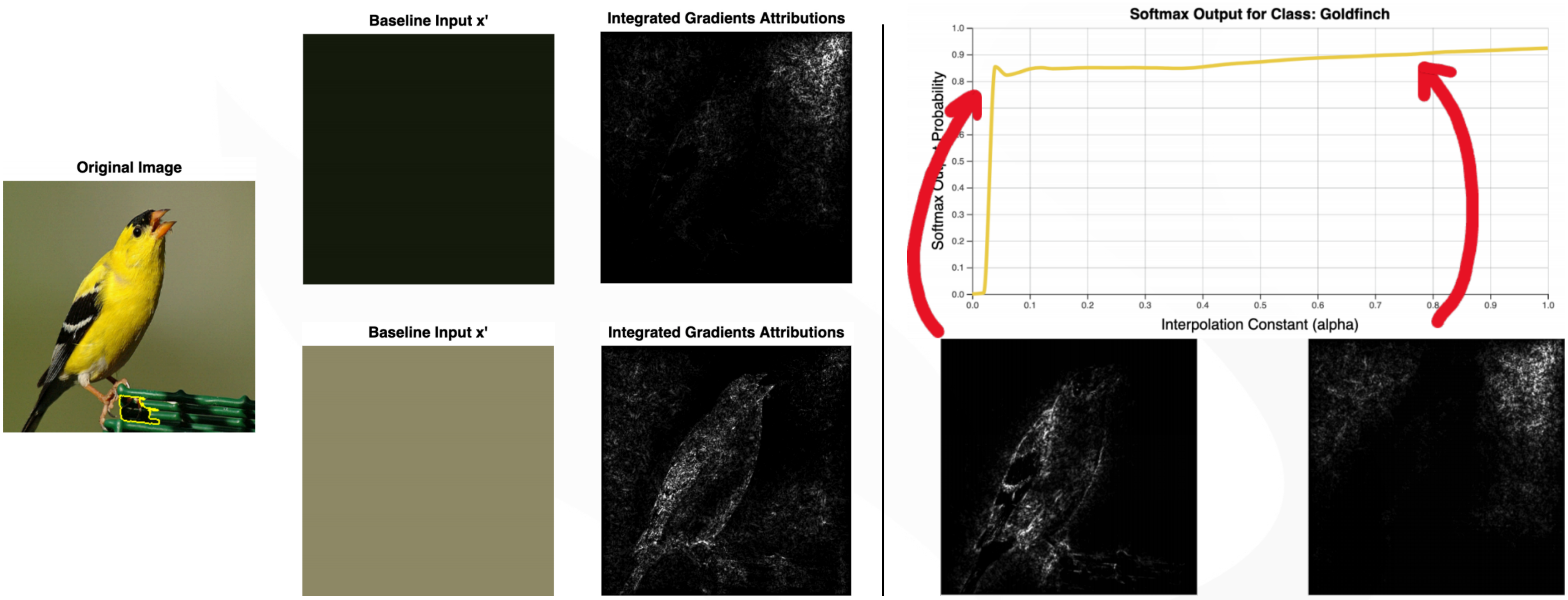
Blurred image as baseline input : $I' = \omega(\sigma) \otimes I$

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

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Local Attribution Maps (LAM)



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

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 $\gamma(\alpha), \alpha \in R$

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

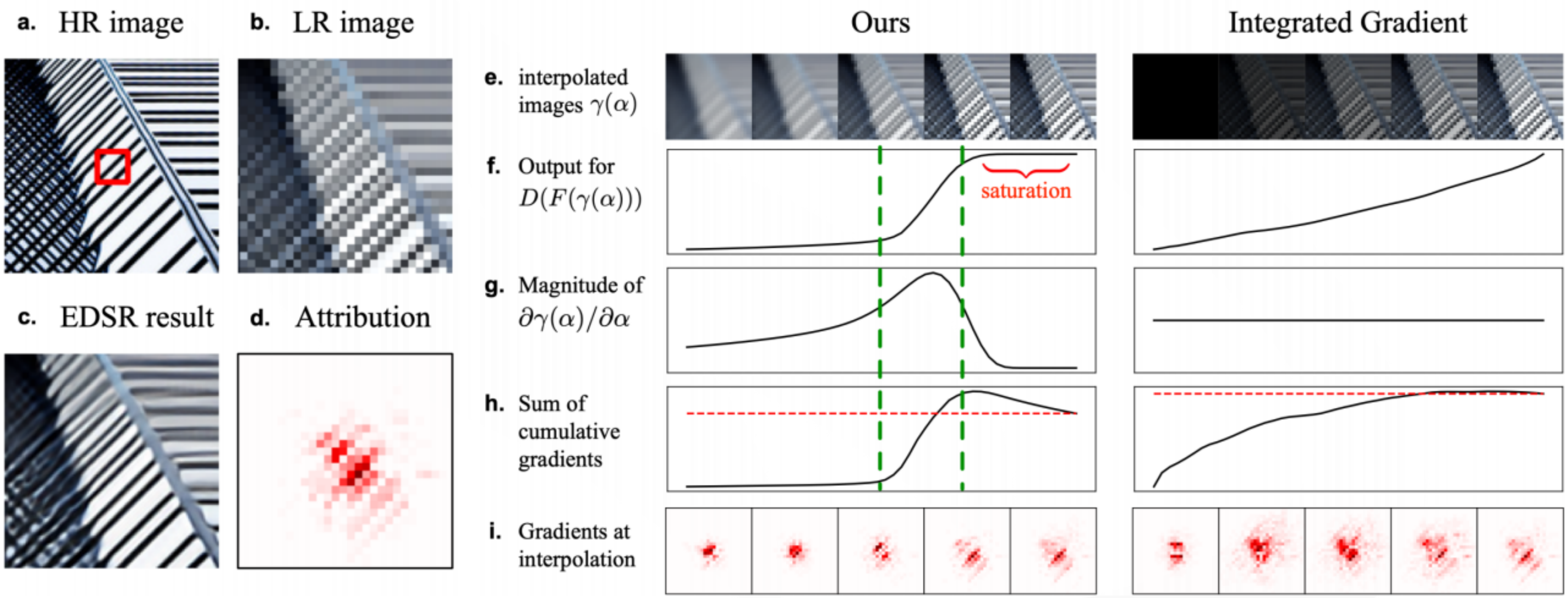
Input $\gamma(1) = I$

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Local Attribution Maps (LAM)

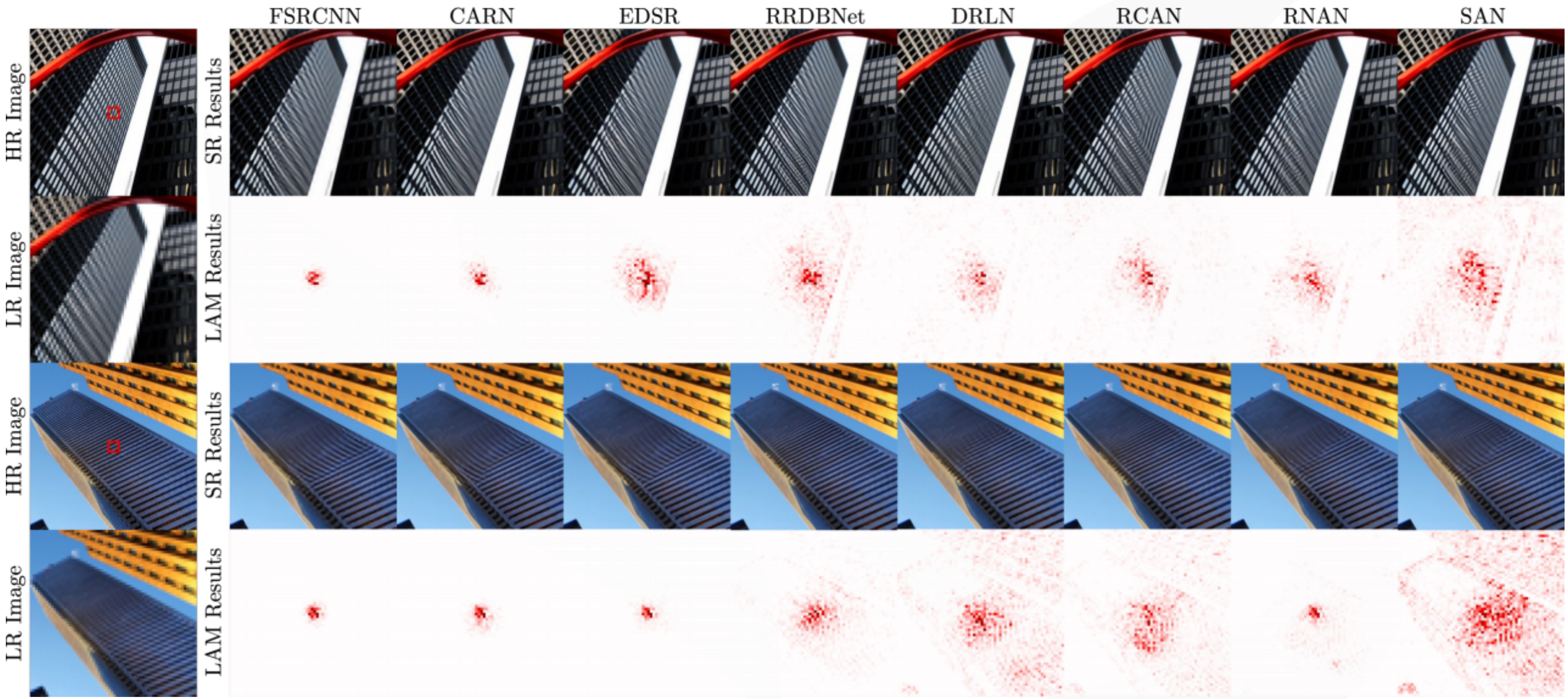
Why using path integral gradient: Gradient Saturation



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

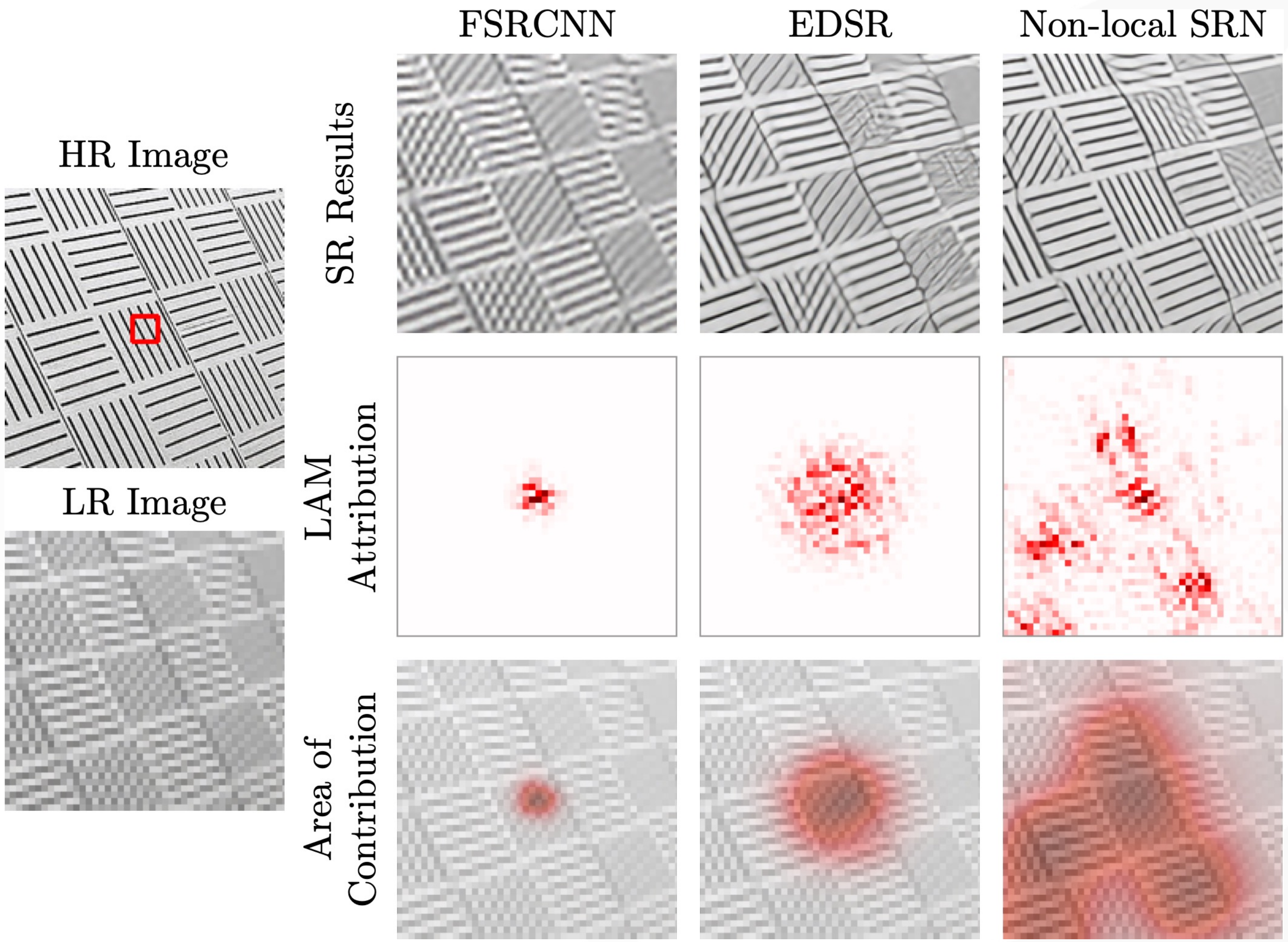
Local Attribution Maps (LAM)



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Local Attribution Maps (LAM)



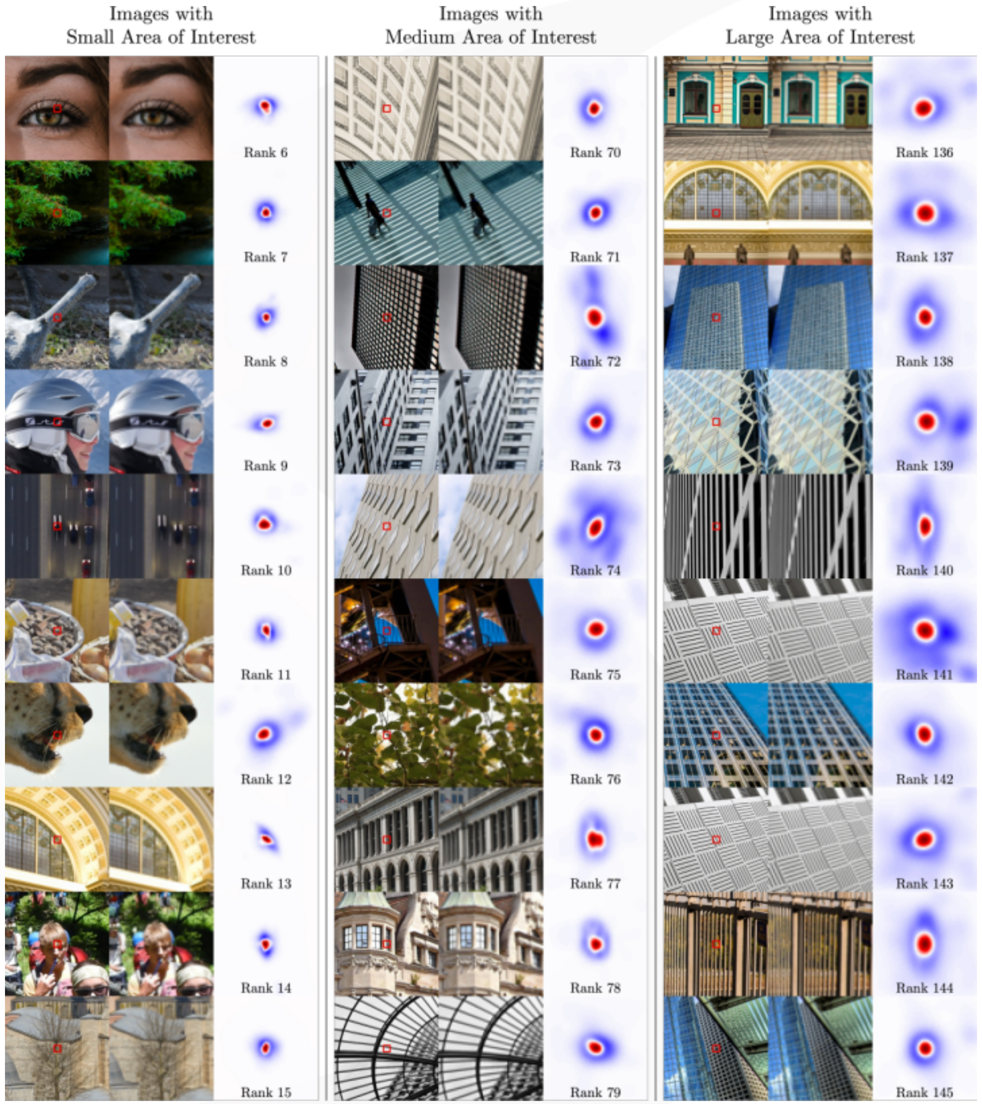
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Informative Areas

The similarities and differences of LAM results for different SR networks

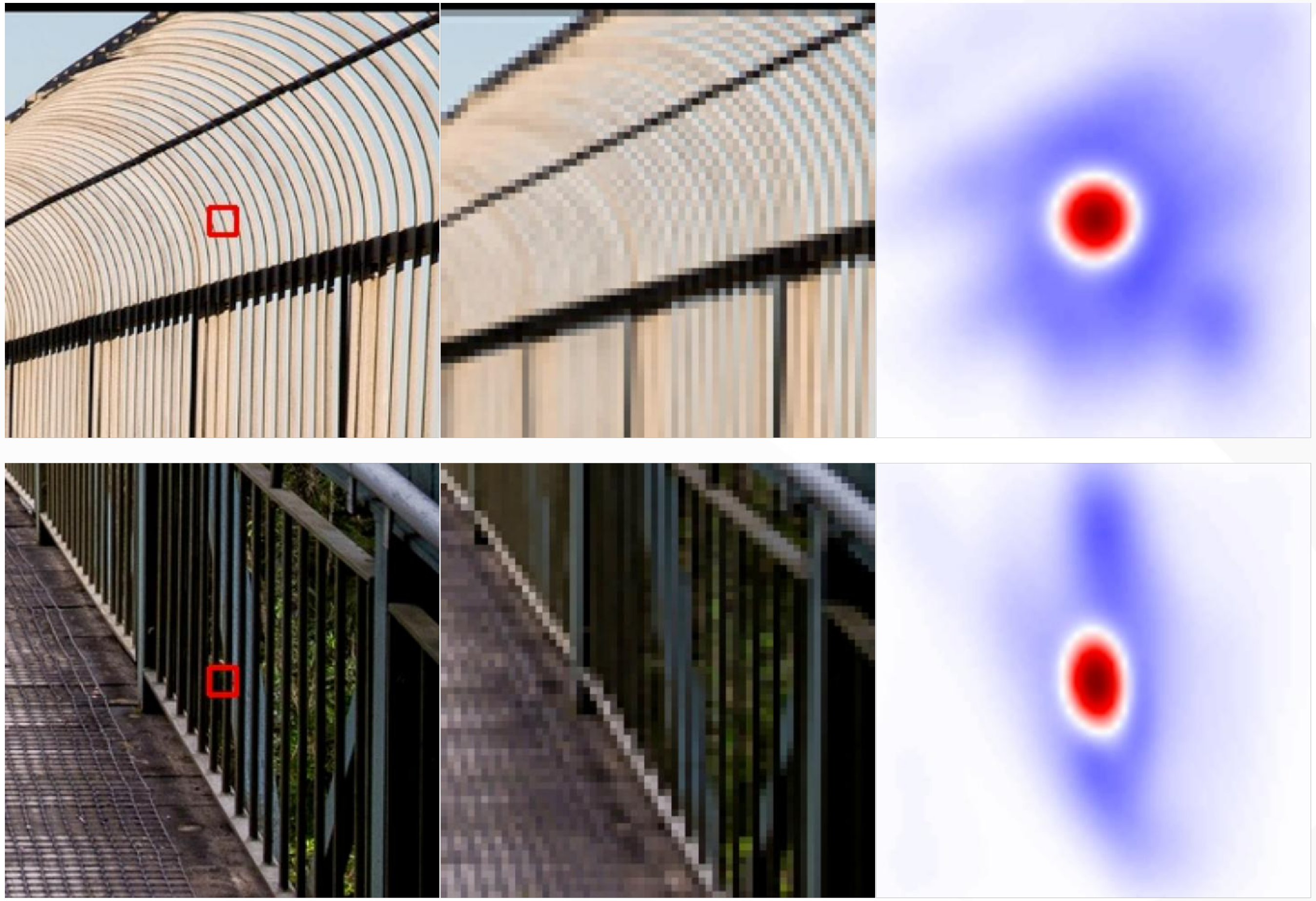
- Red areas can be used for the most preliminary level of SR
- Blue areas show the potential informative areas



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

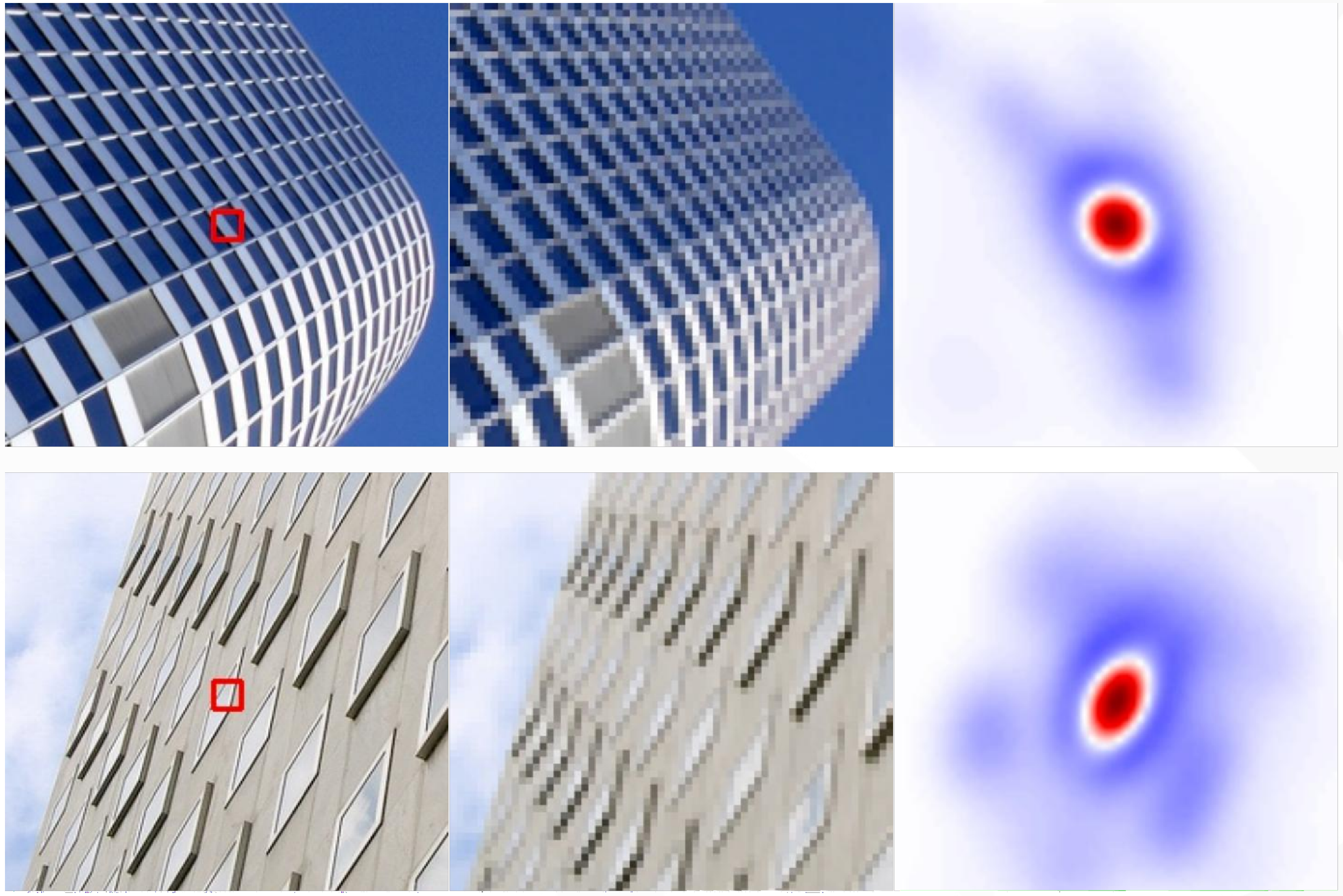
Informative Areas



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Informative Areas



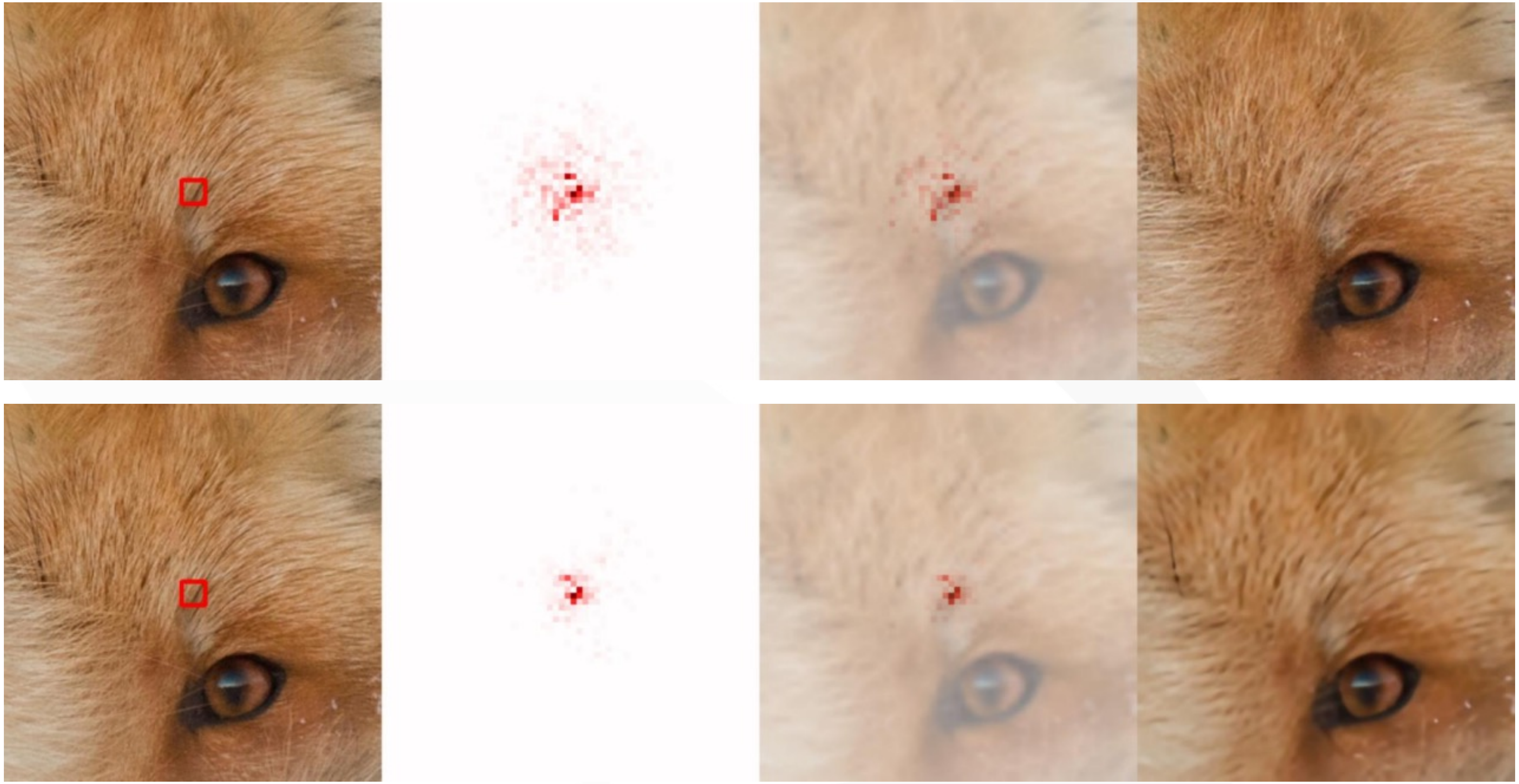
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

SRGANs Learn More Semantics

RankSRGAN

RRDBNet



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

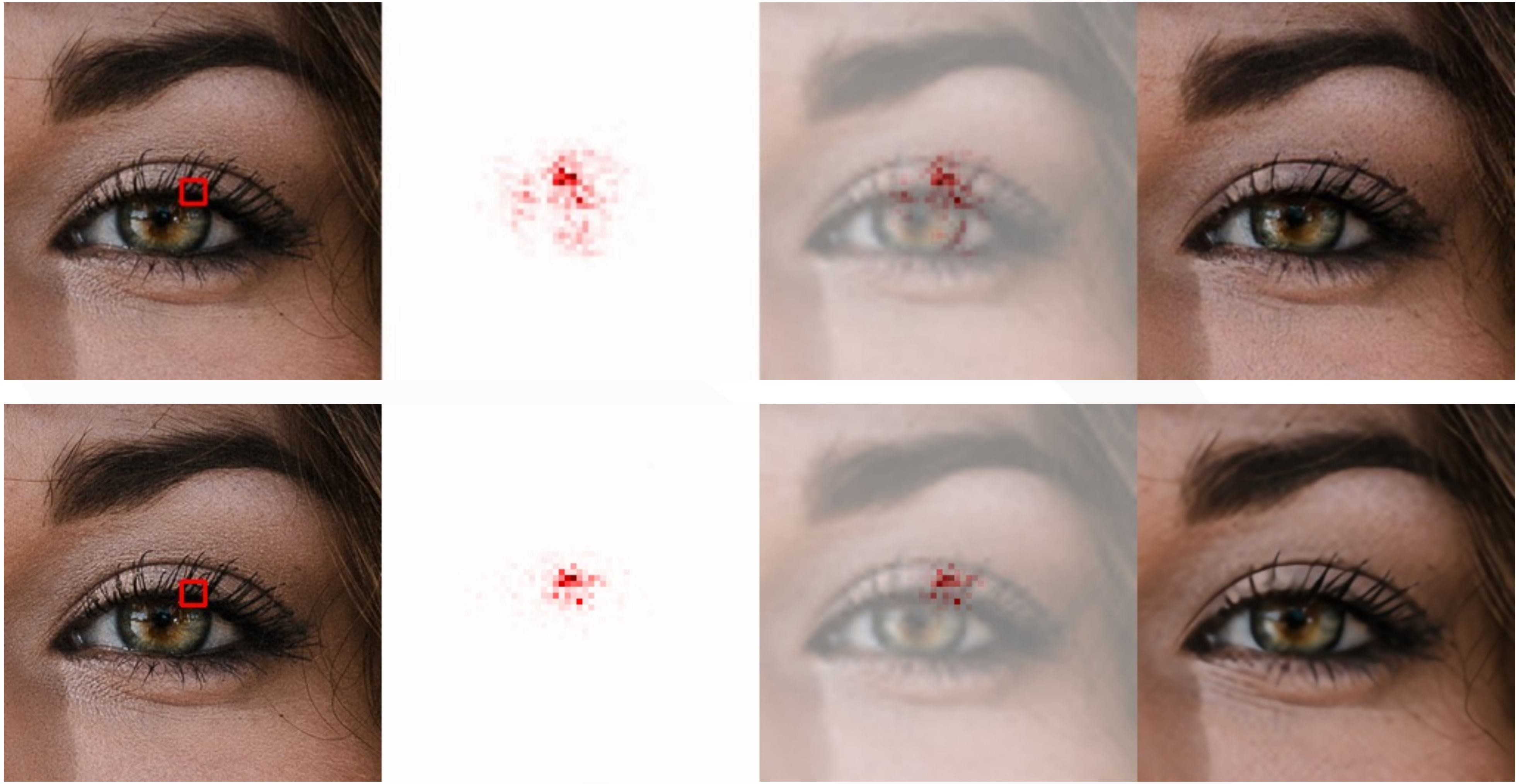
Pixel: What pixels contribute most to restoration?

北京交通大學
BEIJING JIAOTONG UNIVERSITY

SRGANs Learn More Semantics

RankSRGAN

RRDBNet



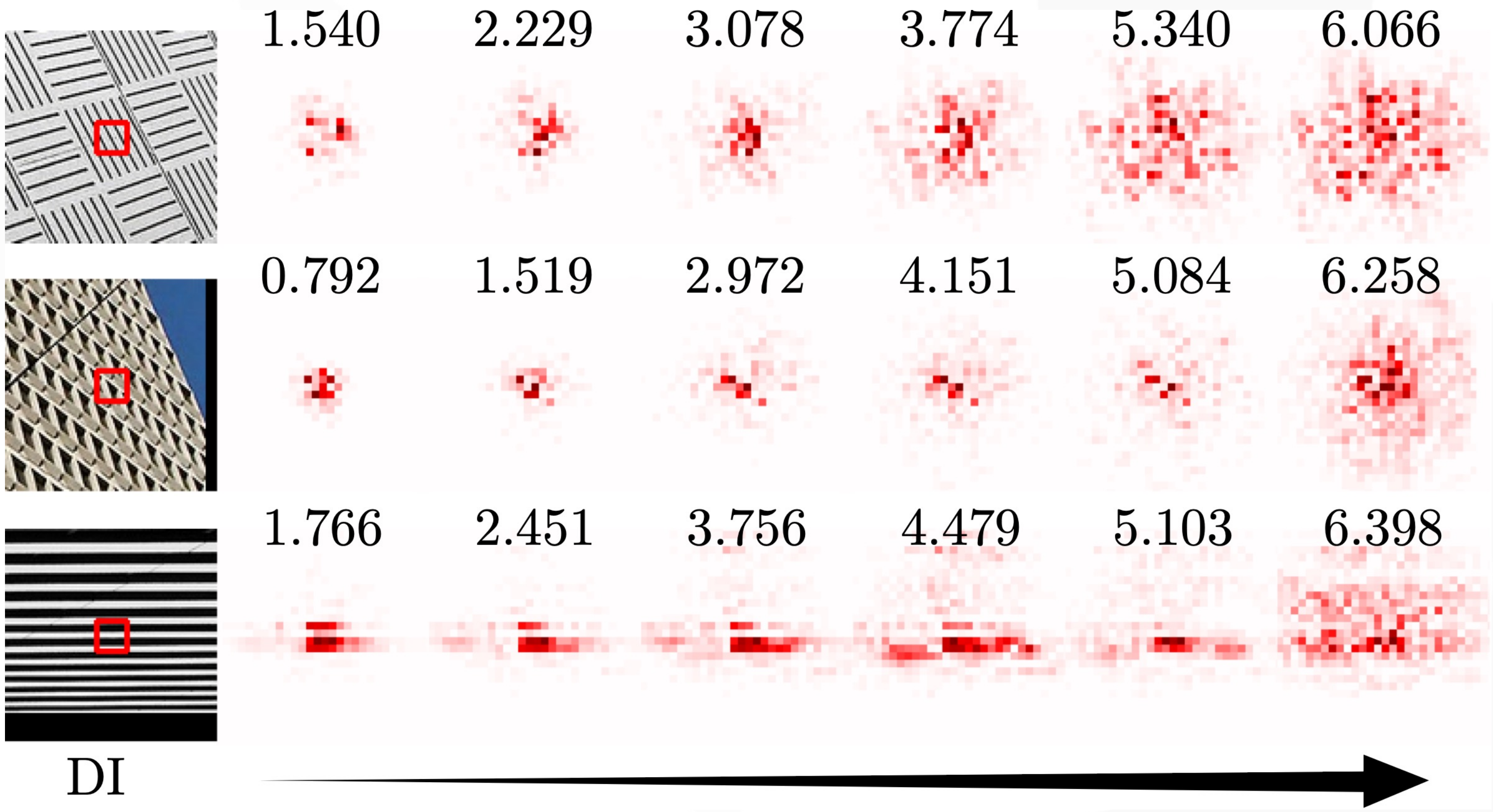
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

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

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

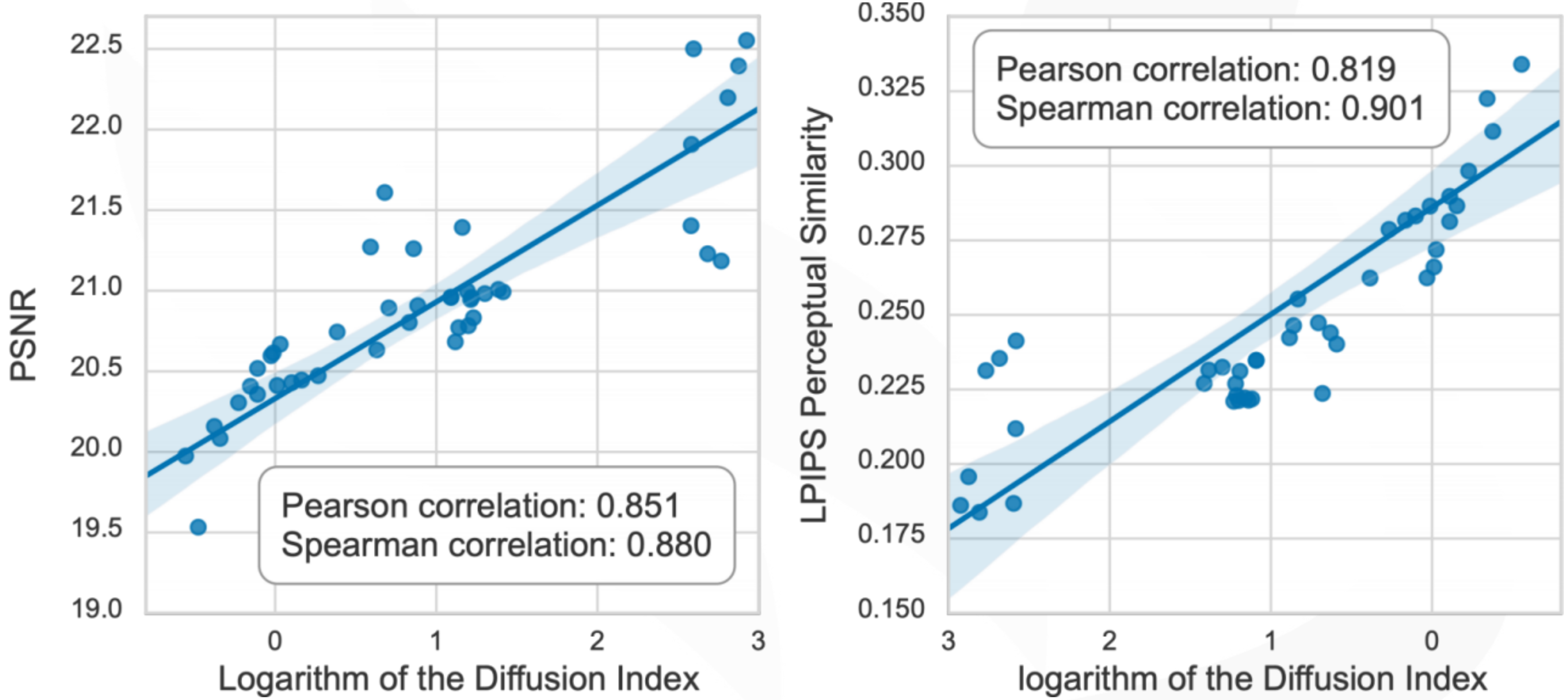


Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

Diffusion Index vs. Network Performances.



Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

Diffusion Index vs. Receptive Field.

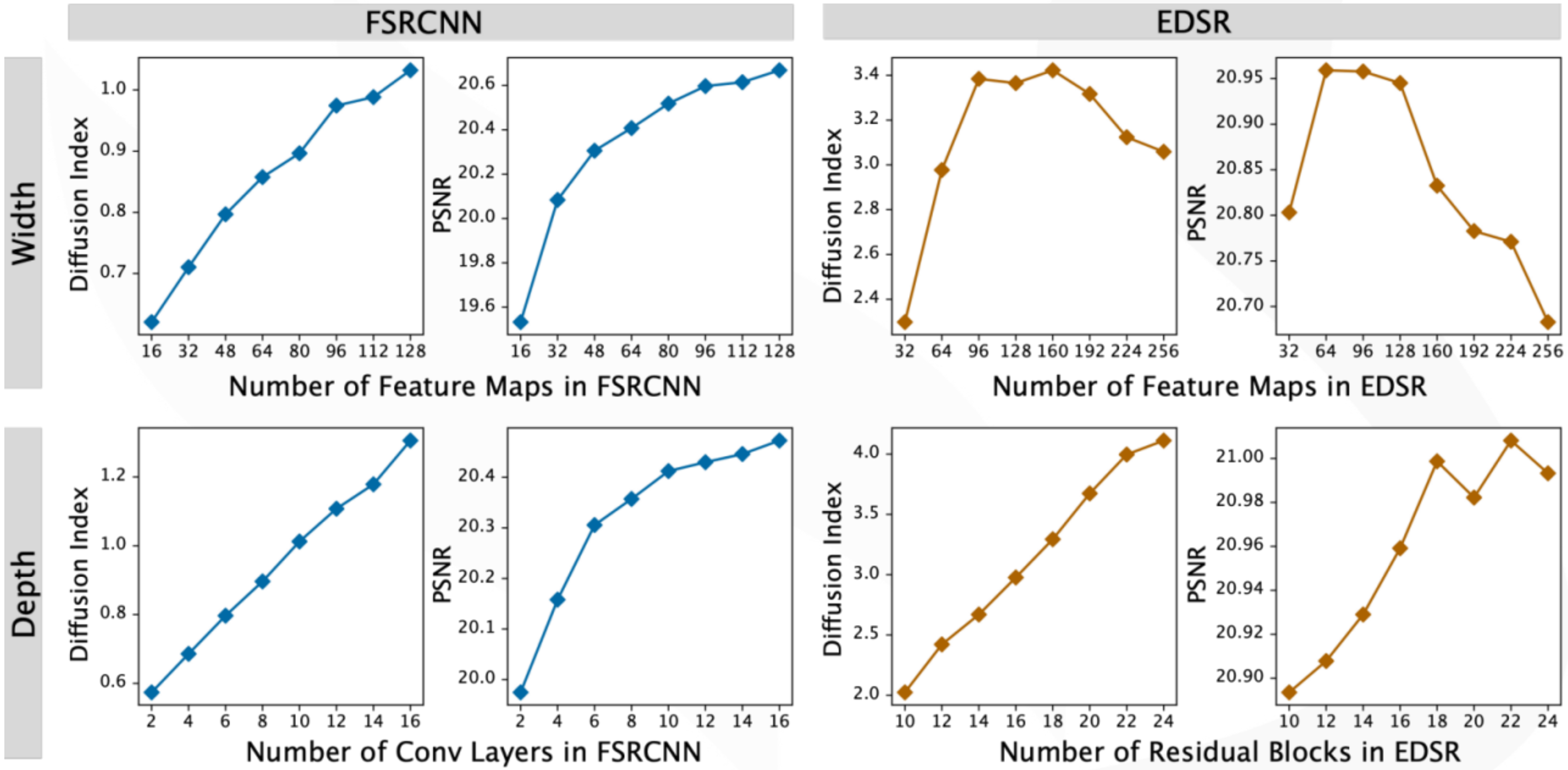
Model	Recpt. Field	PSNR	DI	Remark
FSRCNN	17×17	20.30	0.797	Fully convolution network.
CARN	45×45	21.27	1.807	Residual network.
EDSR	75×75	20.96	2.977	Residual network.
MSRN	107×107	21.39	3.194	Residual network.
RRDBNet	703×703	20.96	13.417	Residual network.
IMDN	global	21.23	14.643	Global pooling.
RFDN	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.

Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

Diffusion Index vs. Network Scale.

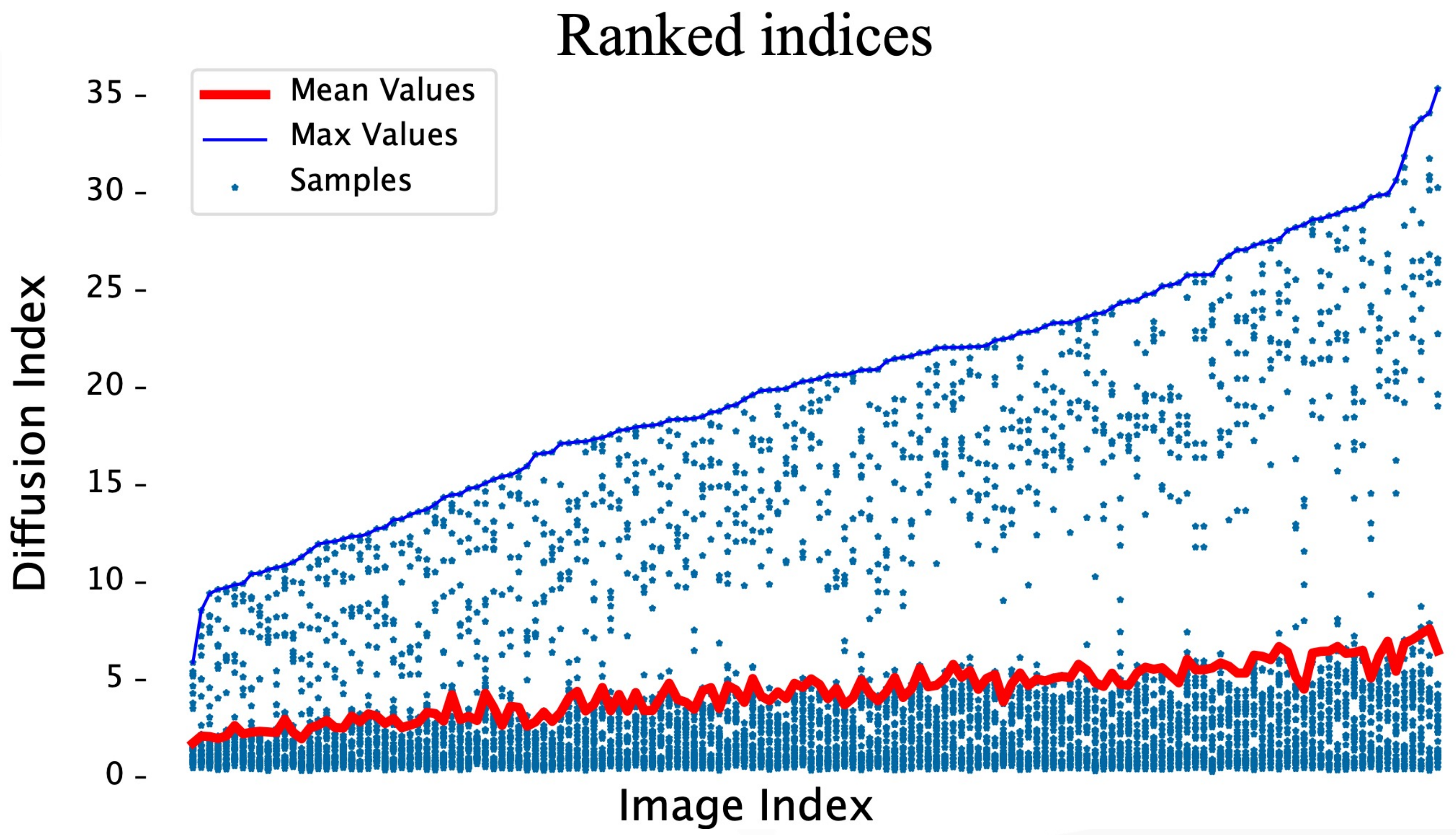


Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

Diffusion Index vs. Image Content.

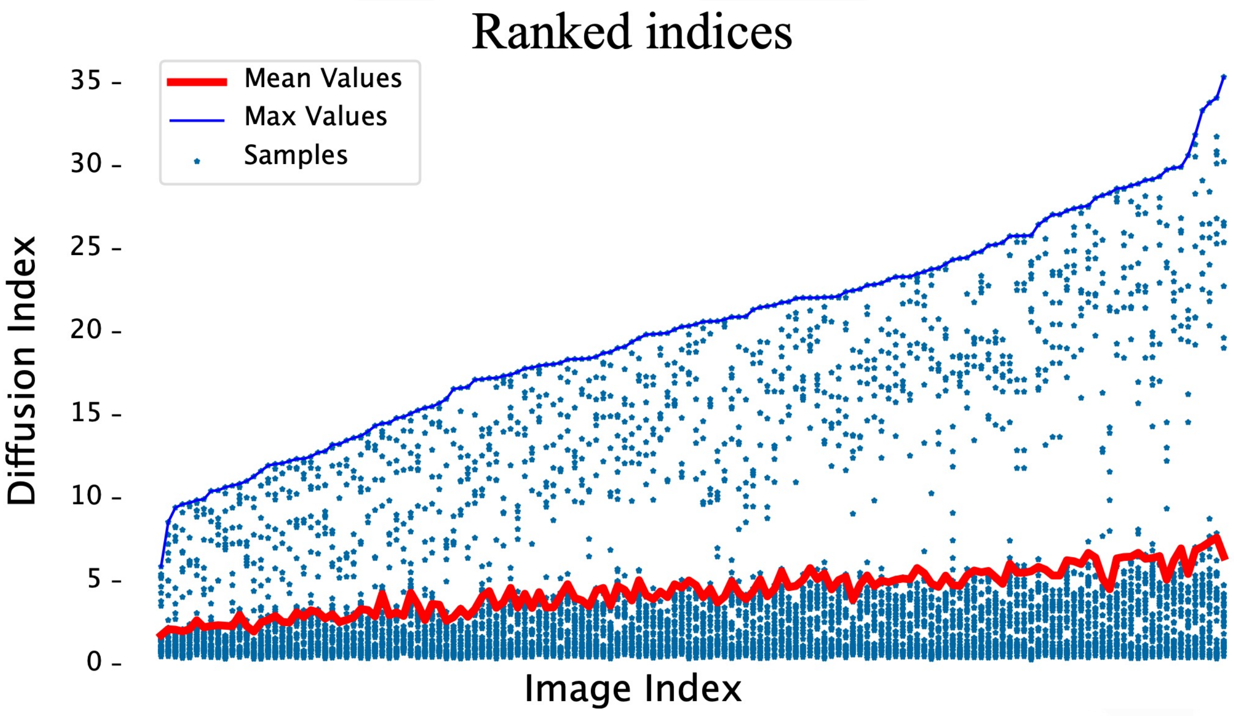


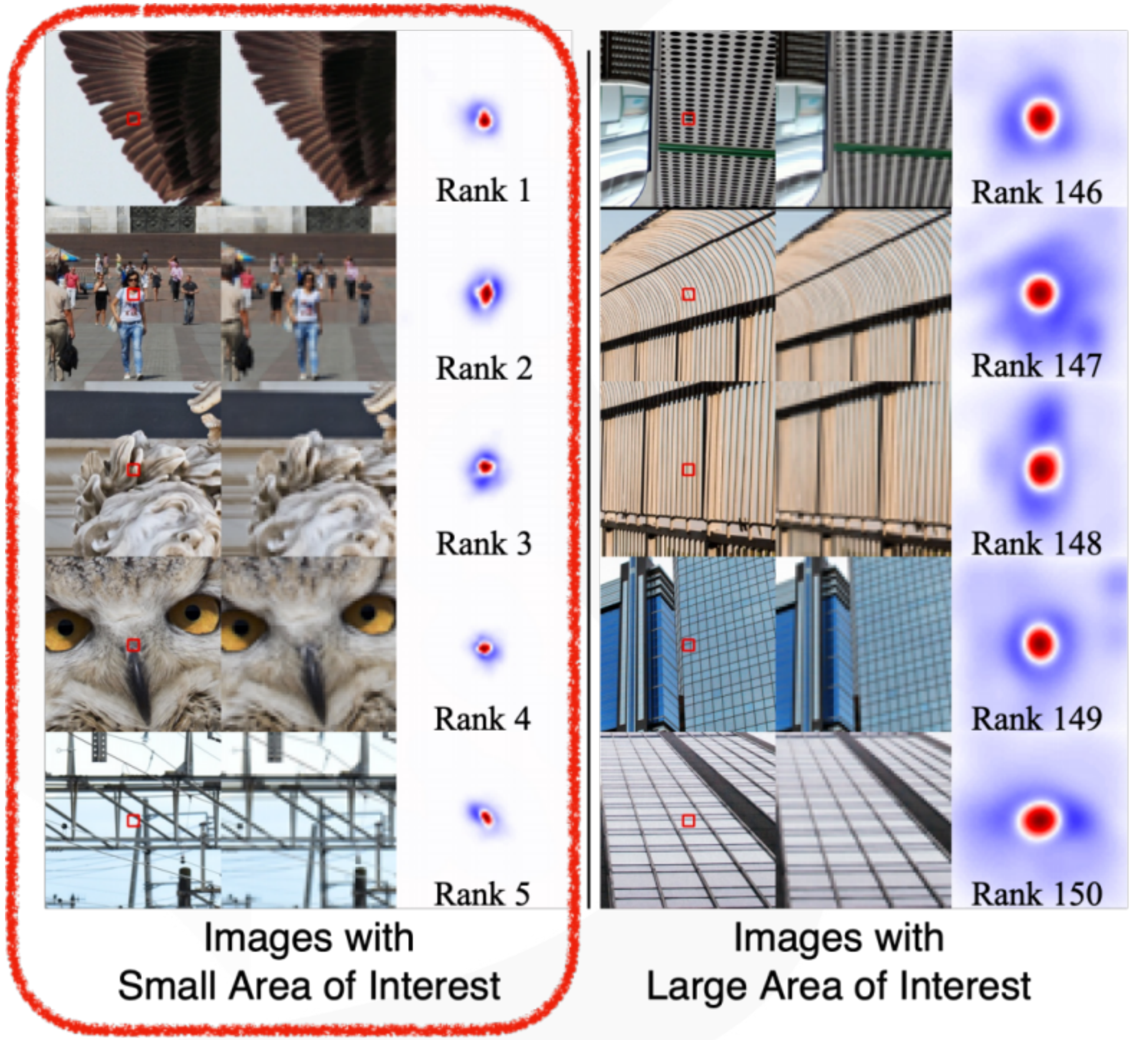
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

Exploration with LAM

Diffusion Index vs. Image Content.





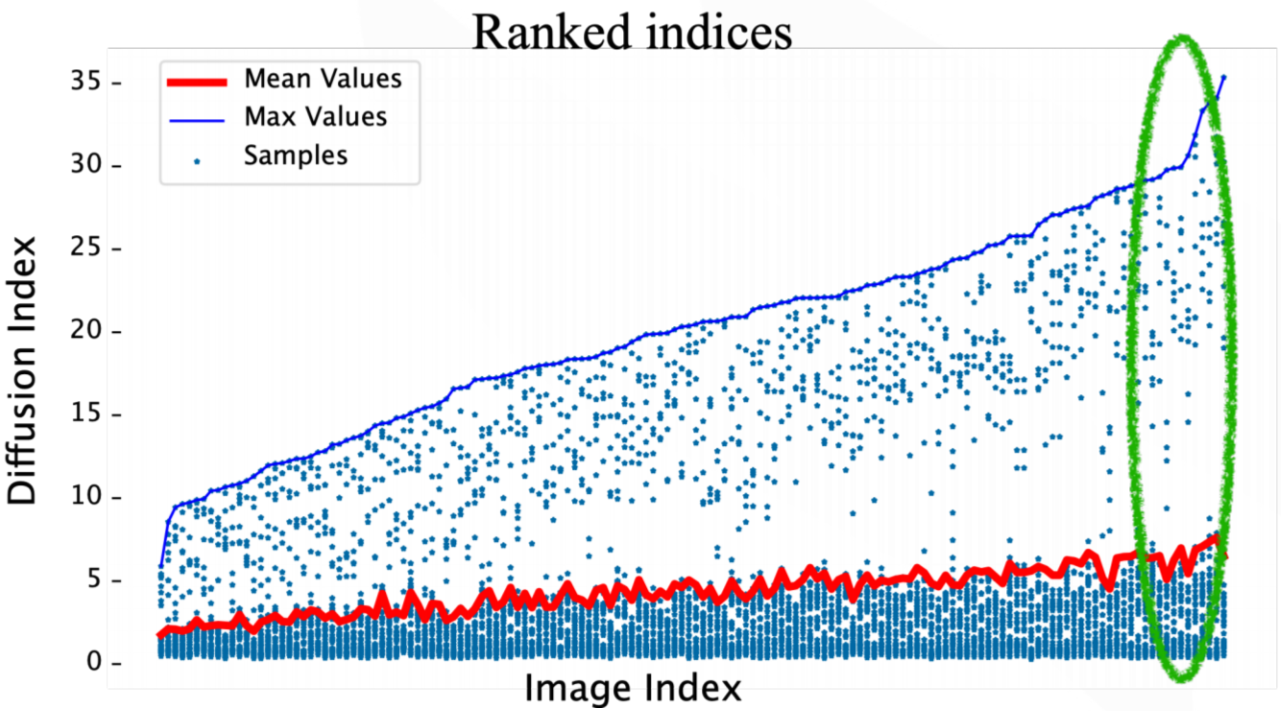
Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

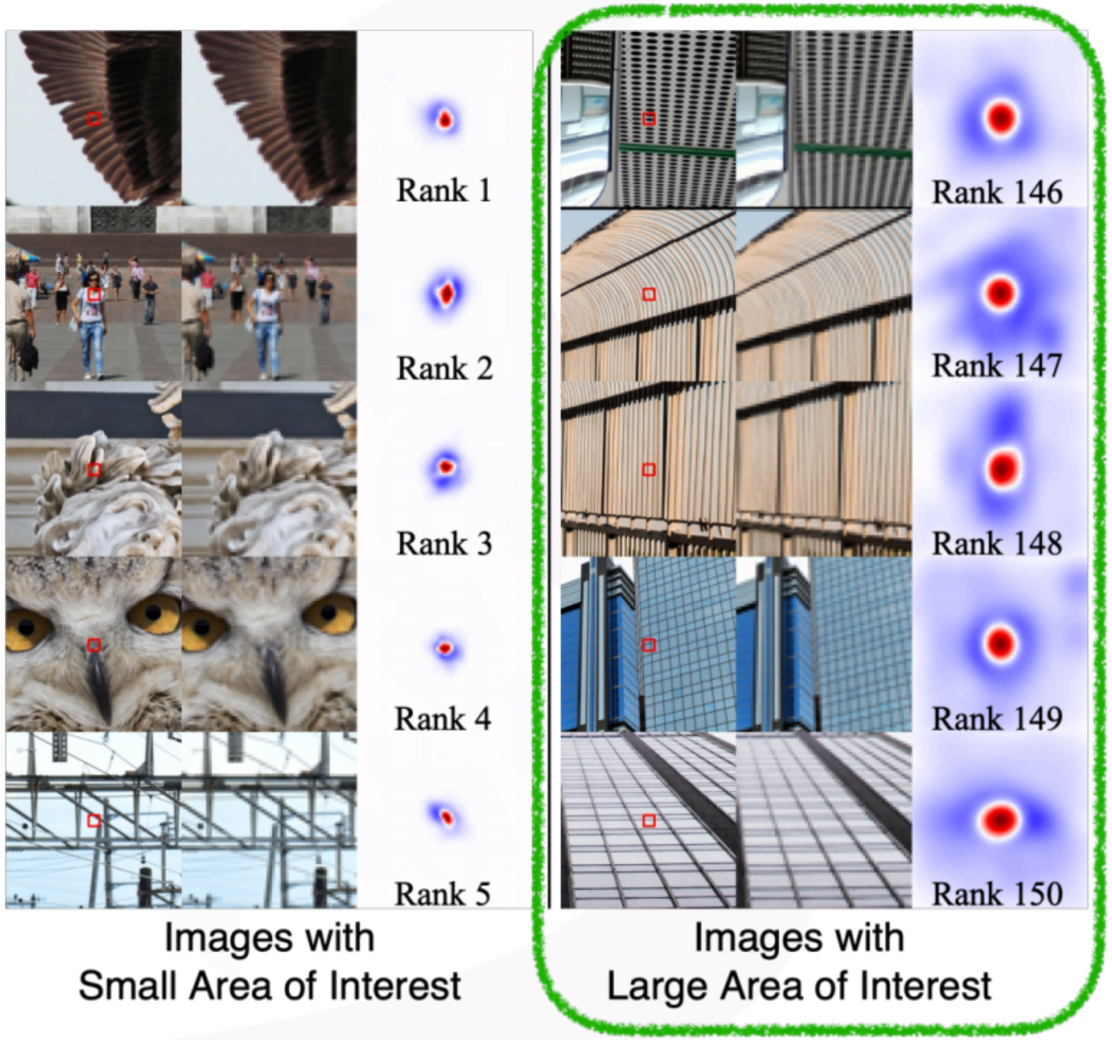
Pixel: What pixels contribute most to restoration?

北京交通大學
BEIJING JIAOTONG UNIVERSITY

Exploration with LAM

Diffusion Index vs. Image Content.





Jinjin Gu and Chao Dong. 2021. Interpreting Super-Resolution Networks With Local Attribution Maps. In IEEE Conference on Computer Vision and Pattern Recognition. 9199–9208.

Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

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

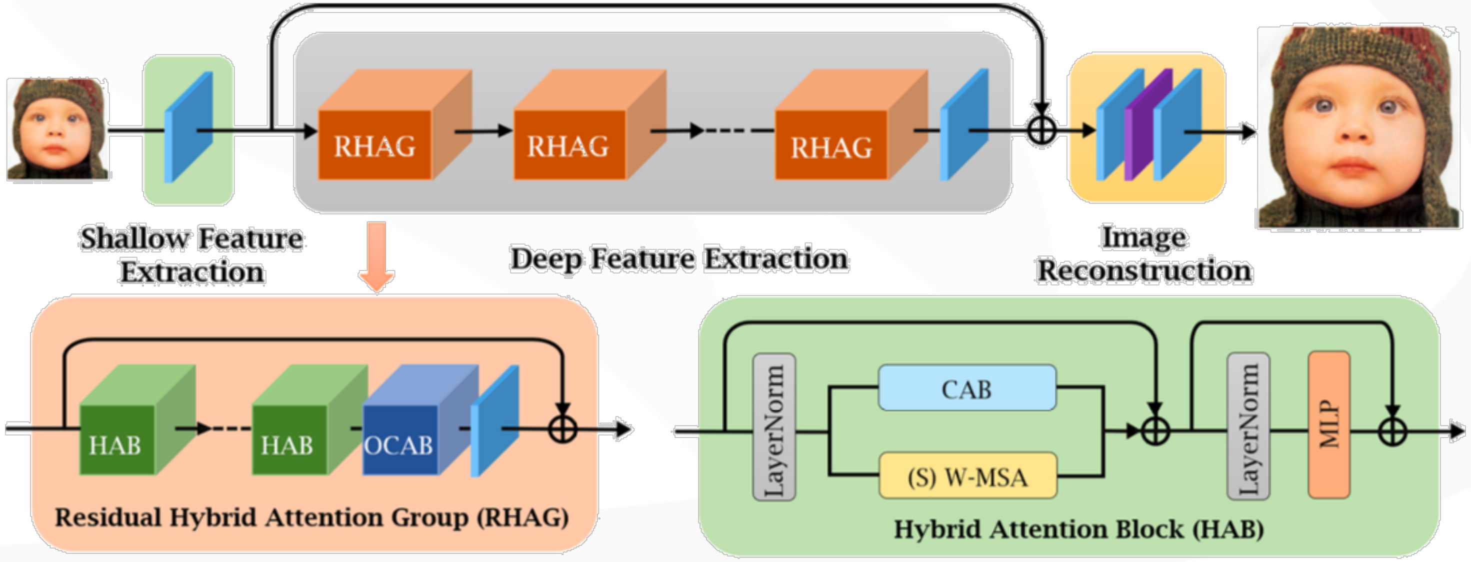
```
import torch, cv2, os, sys, numpy as np, matplotlib.pyplot as plt
from PIL import Image
```

Load model codes and model files

This may take a while

Activating More Pixels in Image Super-Resolution Transformer

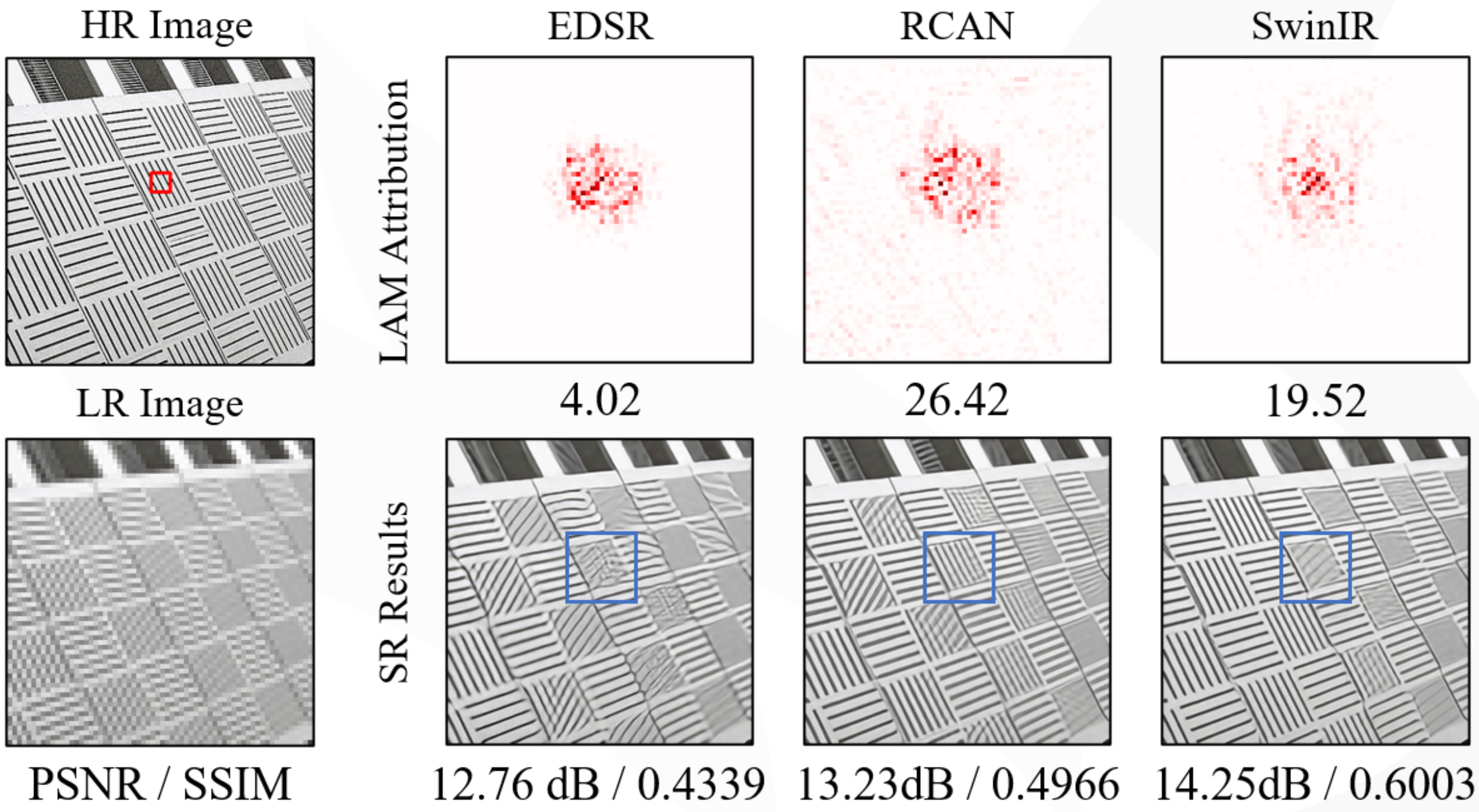
Jinjin Gu[1] **Chao Dong**[2,3]
[1]School of Electrical and Information Engineering, The University of Sydney.
[2]Key Laboratory of Human-Machine Intelligence-Synergy Systems,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.
[3]SIAT Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society



Pixel: What pixels contribute most to restoration?

BEIJING JIAOTONG UNIVERSITY

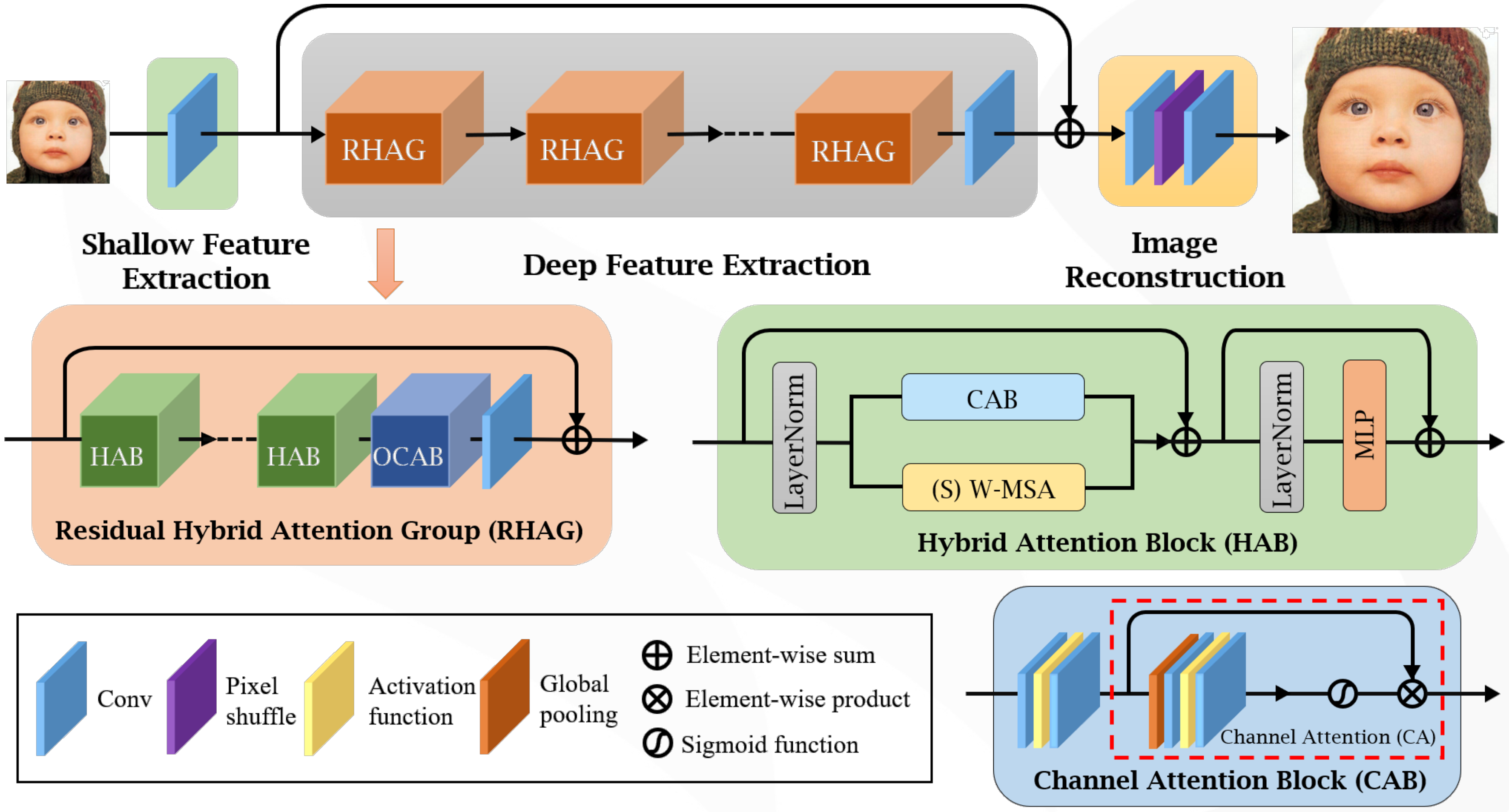
How to activate more pixels?



Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2022. Activating More Pixels in Image Super-Resolution Transformer. arXiv preprint arXiv:2205.04437.

Pixel: What pixels contribute most to restoration?

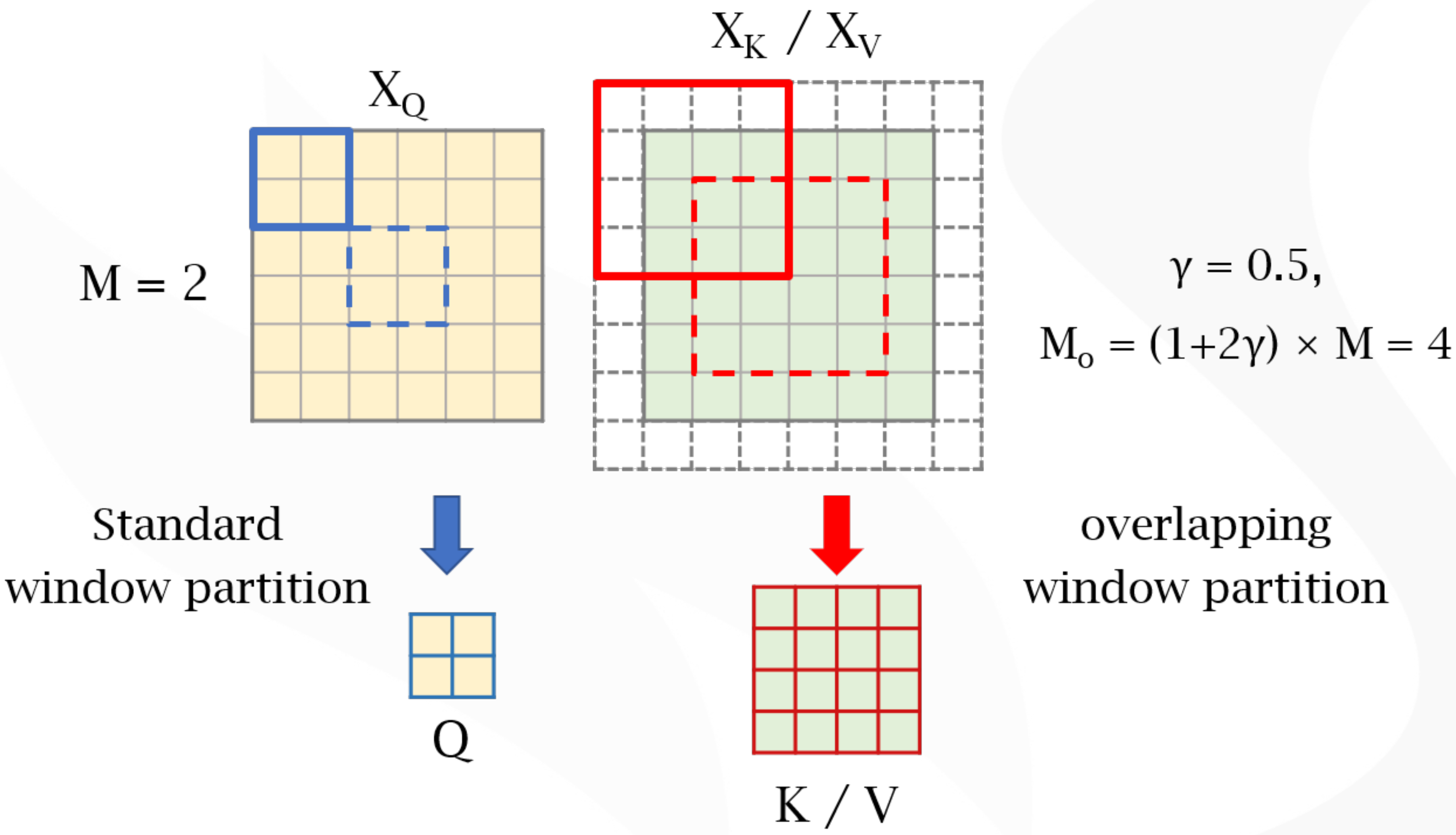
How to activate more pixels?



Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2022. Activating More Pixels in Image Super-Resolution Transformer. arXiv preprint arXiv:2205.04437.

Pixel: What pixels contribute most to restoration?

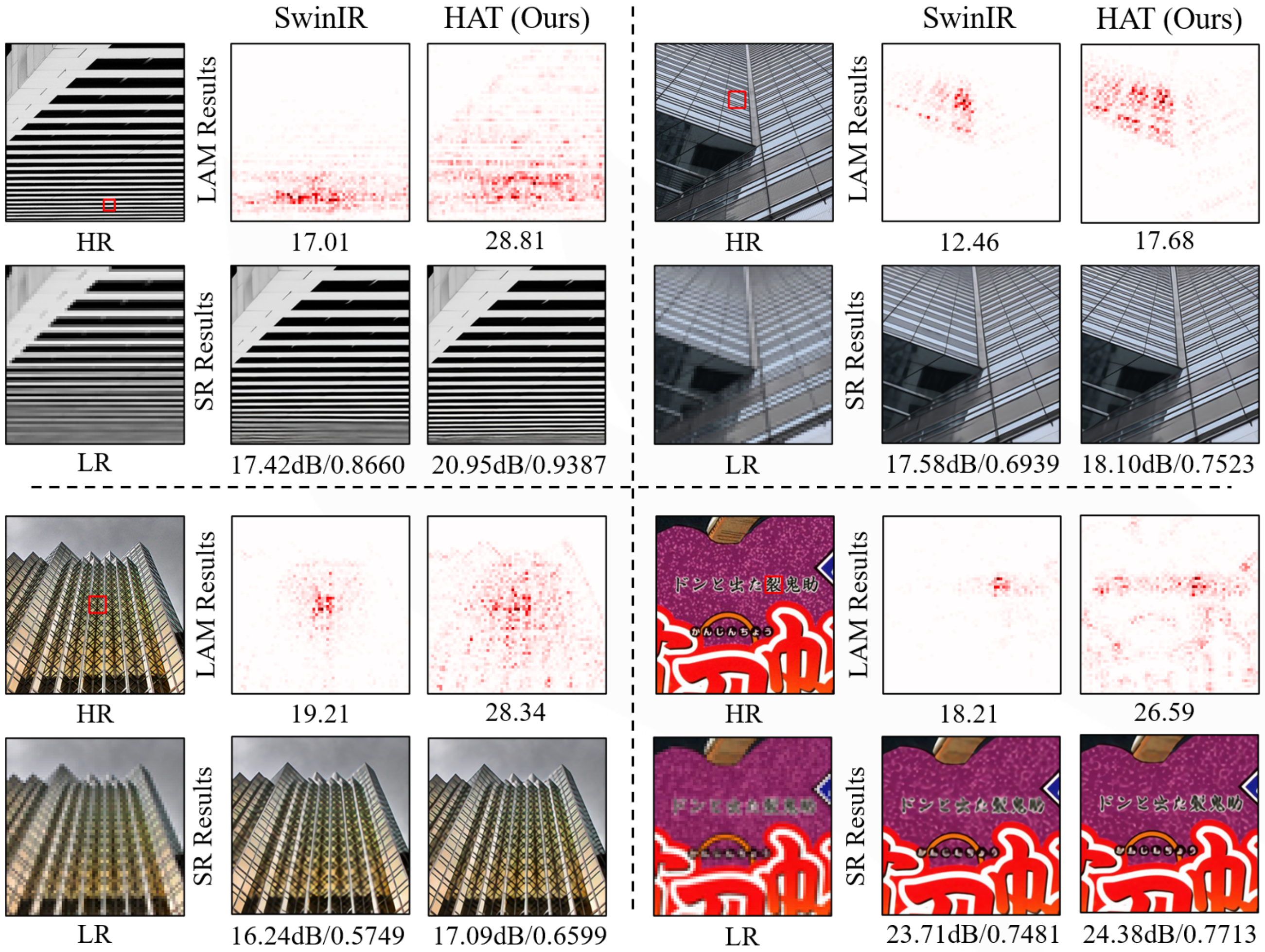
How to activate more pixels?

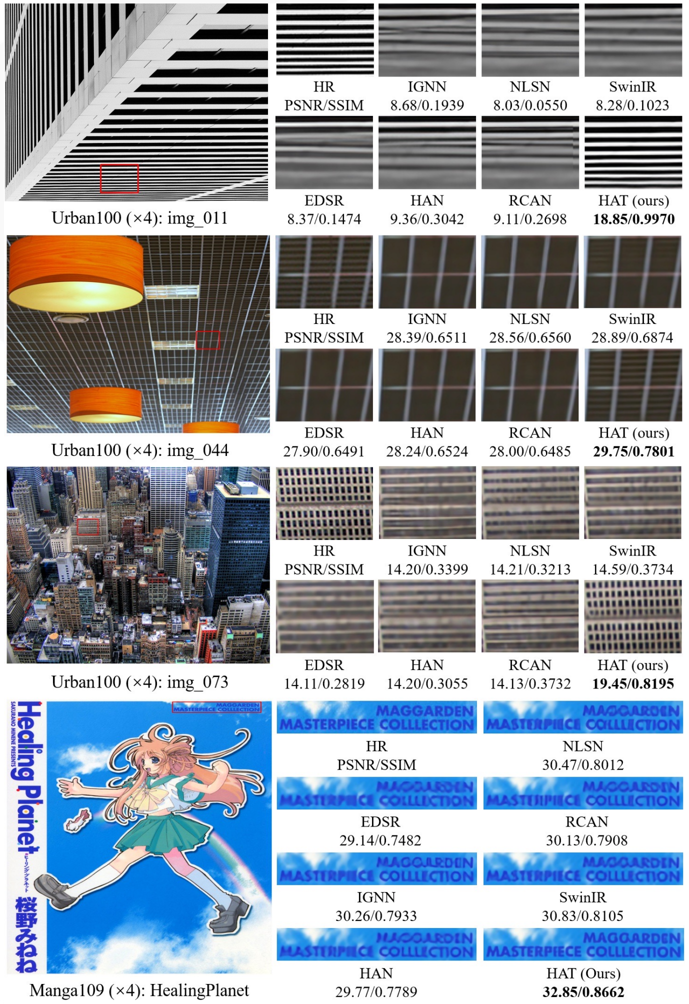


Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2022. Activating More Pixels in Image Super-Resolution Transformer. arXiv preprint arXiv:2205.04437.

Pixel: What pixels contribute most to restoration?

How to activate more pixels?





Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2022. Activating More Pixels in Image Super-Resolution Transformer. arXiv preprint arXiv:2205.04437.

Discovering "Semantics" in Super-Resolution Networks

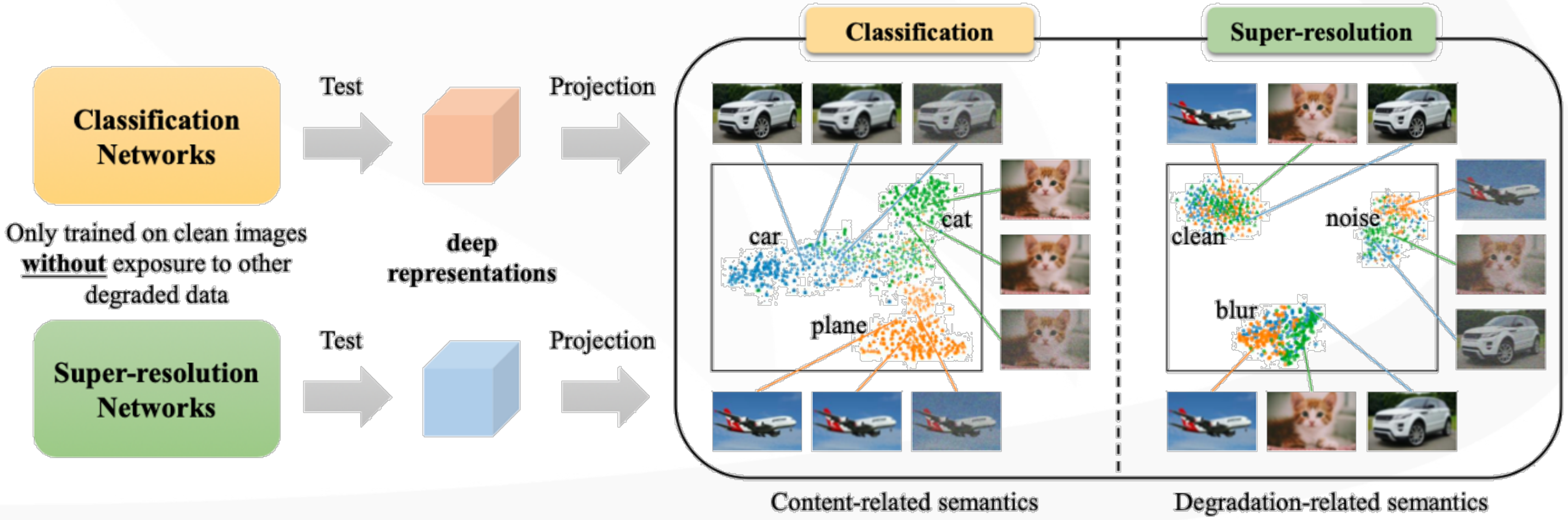
Yihao Liu[1 2*] Anran Liu[1 4*] Jinjin Gu[1 5] Zhipeng Zhang[2 6] Wenhao Wu[7] Yu Qiao[1 3] Chao Dong[1 3†]

[1]Shenzhen Institute of Advanced Technology, CAS

[2]University of Chinese Academy of Sciences

[3]Shanghai AI Lab [4]The University of Hongkong

[5]University of Sydney [6]Institute of Automation, CAS [7]Baidu Inc.



Feature: Where can we find semantics in SR networks?

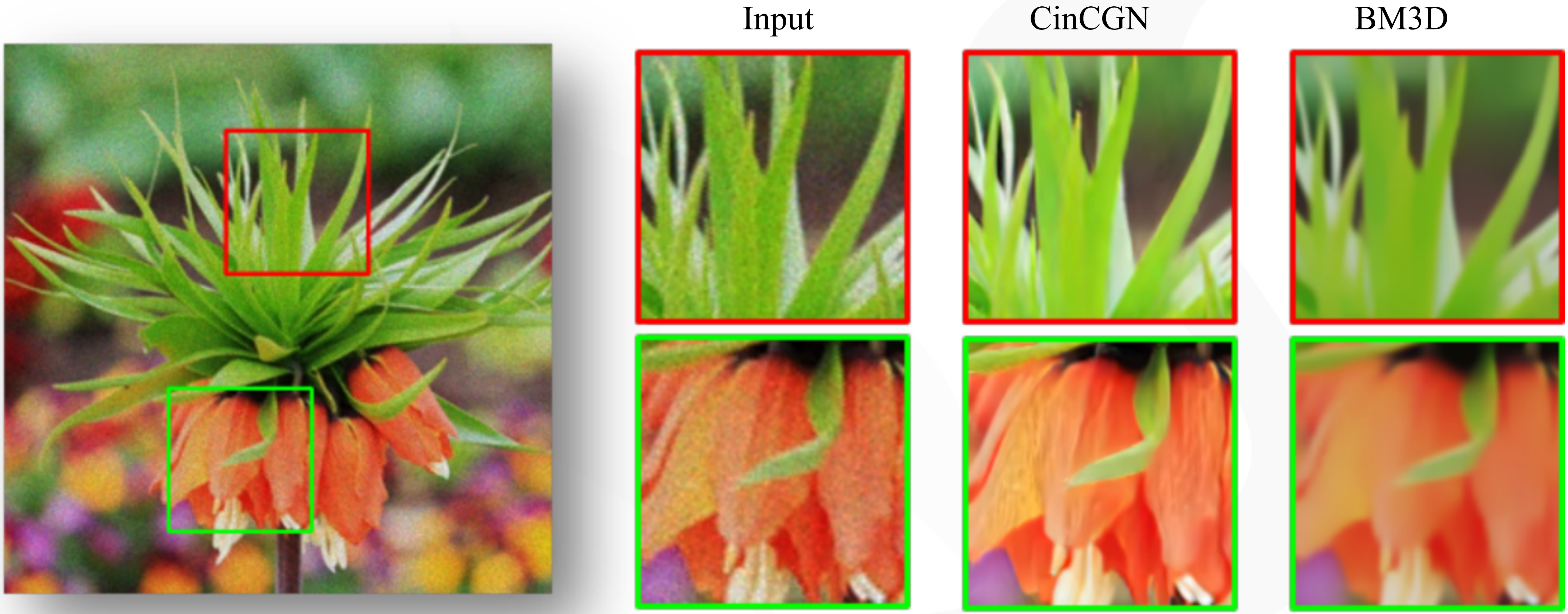
Interpreting Super-Resolution Networks

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**

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

Feature: Where can we find semantics in SR networks?

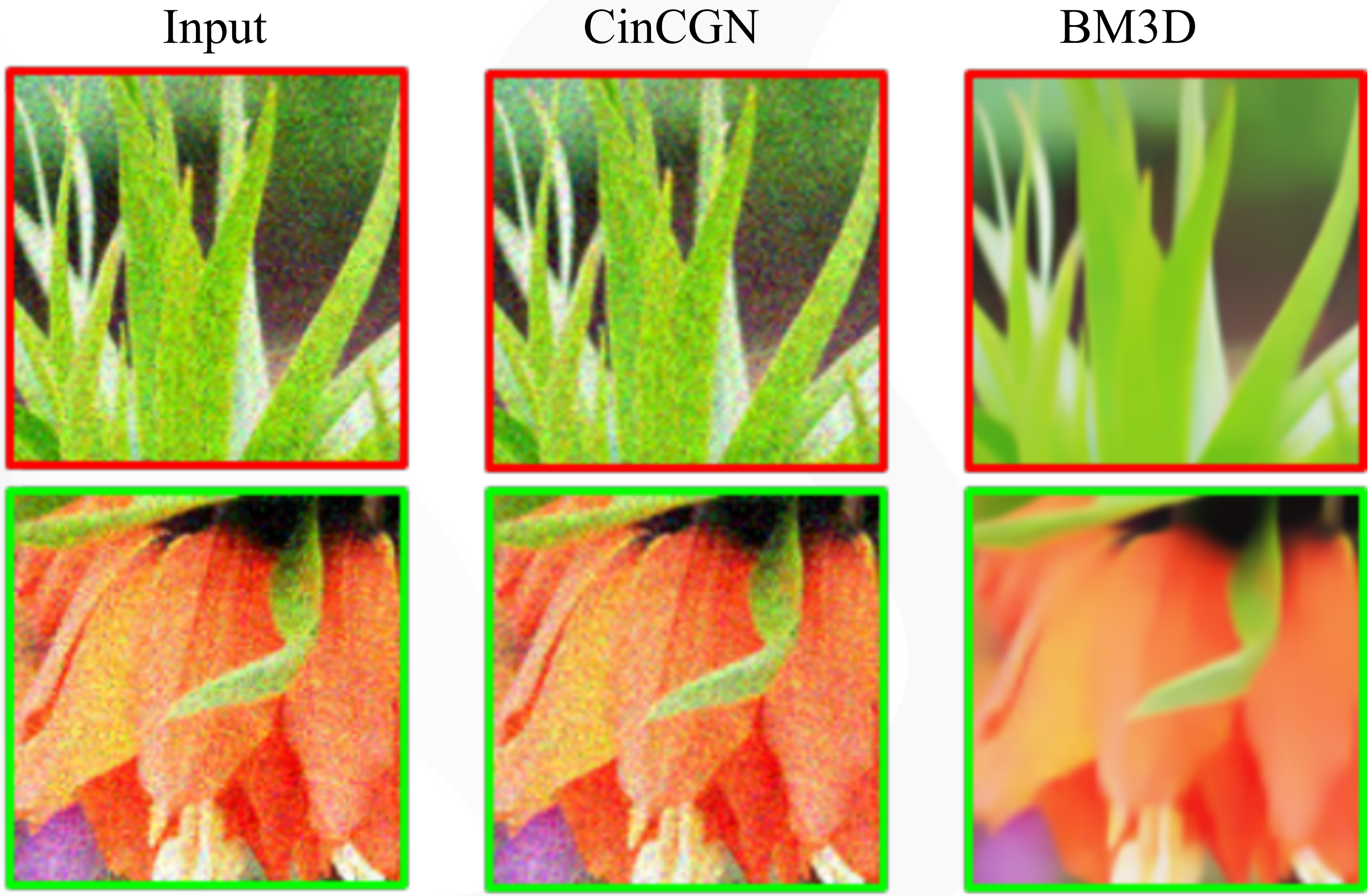
Warm up: An observation



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

BEIJING JIAOTONG UNIVERSITY

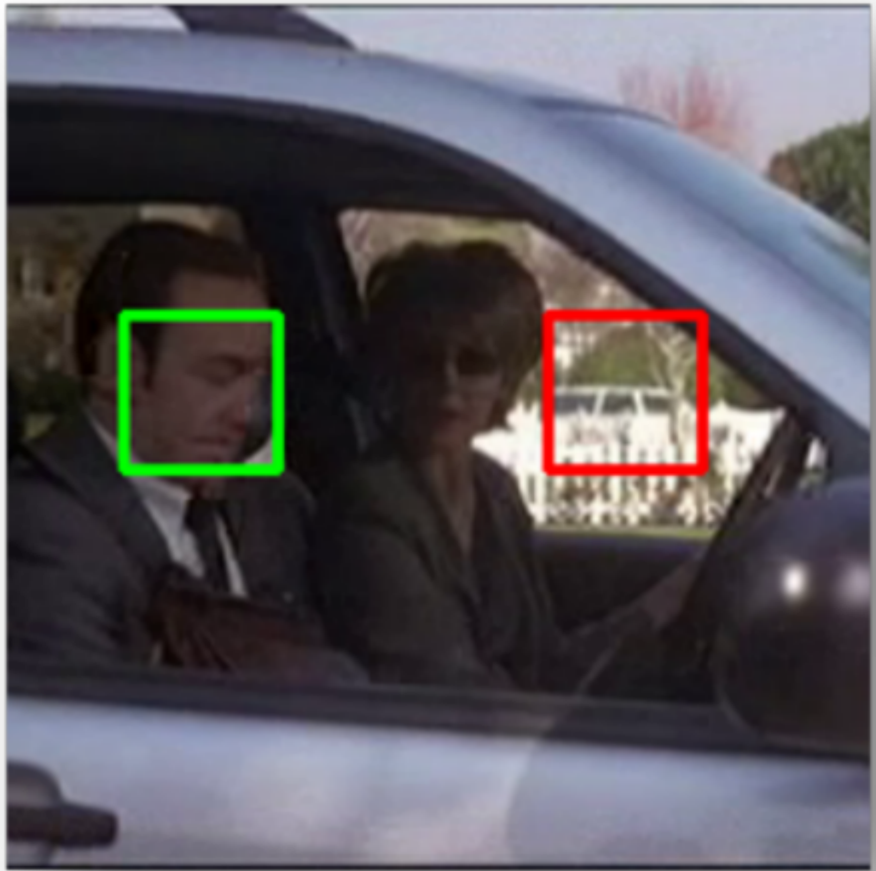
Warm up: An observation

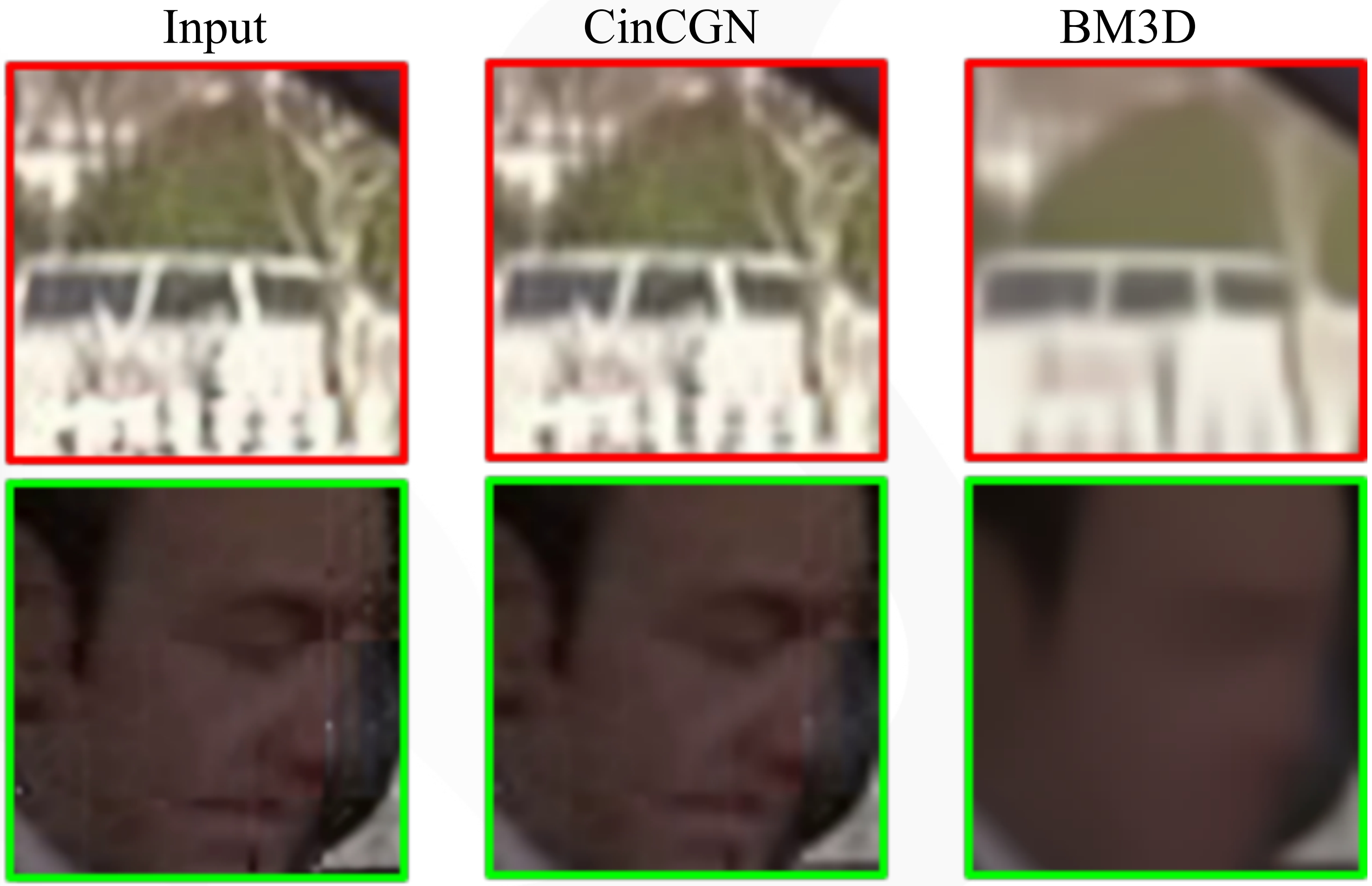


Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

Feature: Where can we find semantics in SR networks?

Warm up: An observation





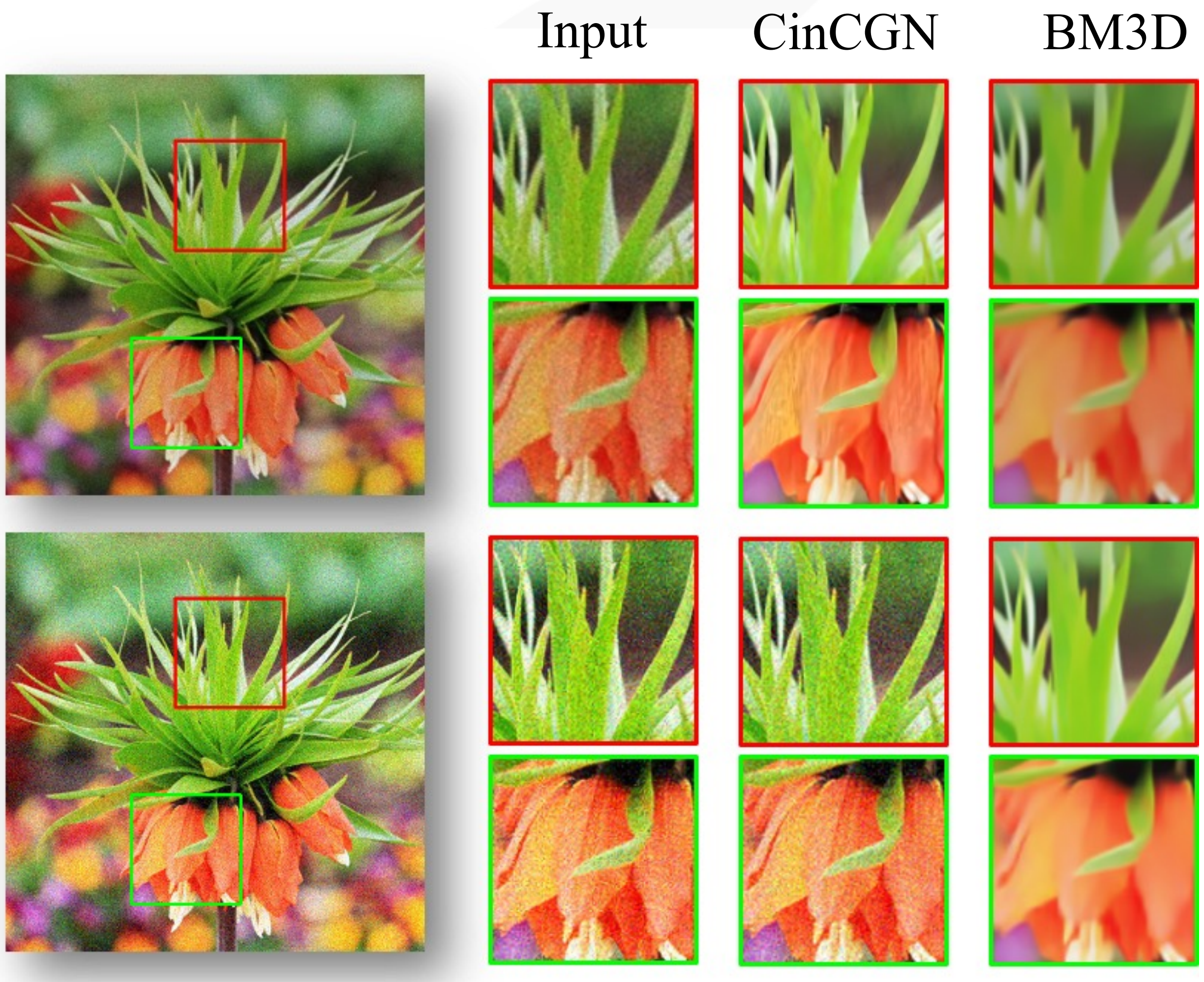
Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

Feature: Where can we find semantics in SR networks?

BEIJING JIAOTONG UNIVERSITY

Warm up: An observation

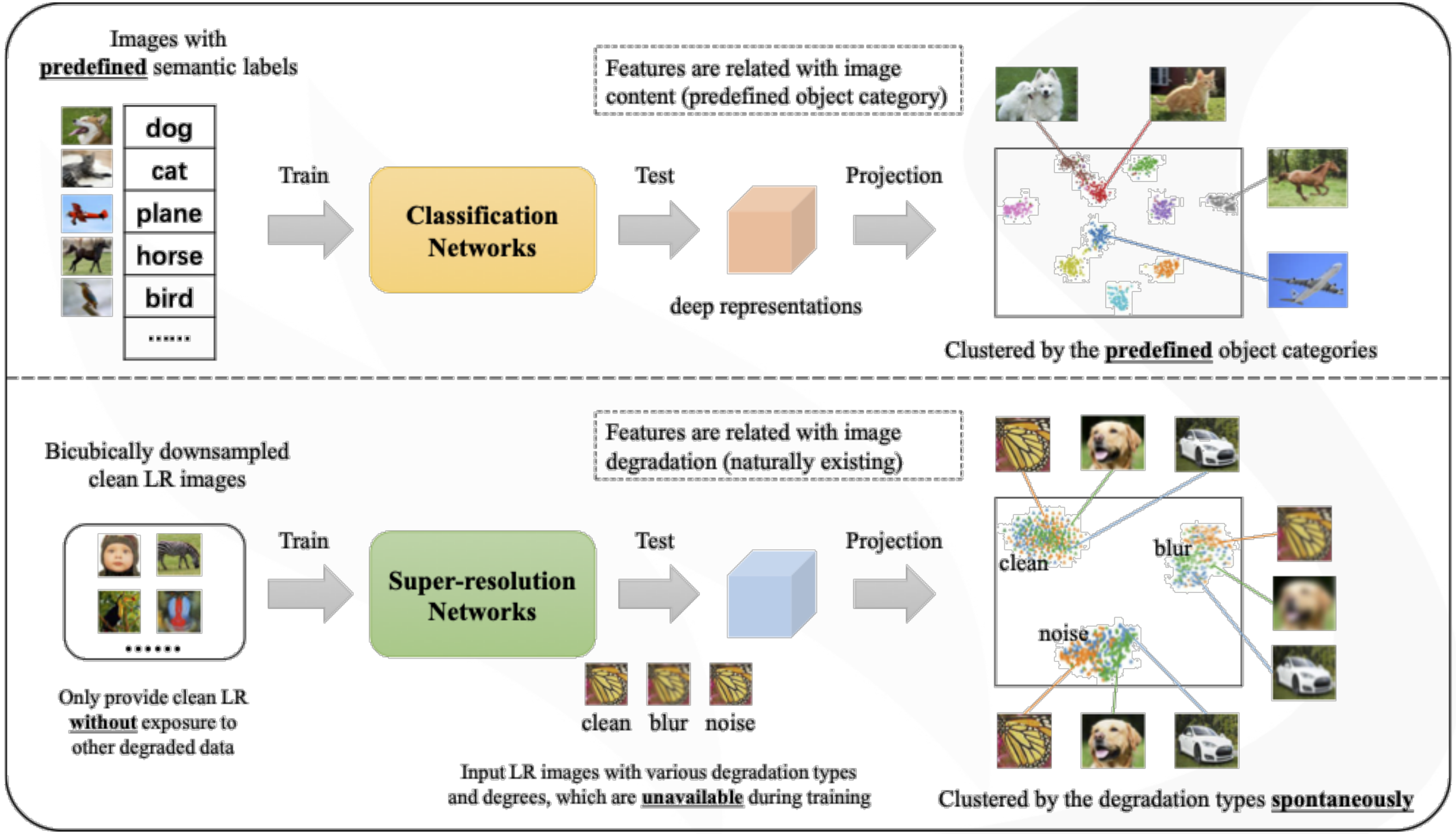
- CinCGAN can figure out the specific degradation within its training data
- The degradation mismatch will make the network "**turn off**" its ability



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

Feature: Where can we find semantics in SR networks?

Methodology



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

Feature: Where can we find semantics in SR networks?

Methodology

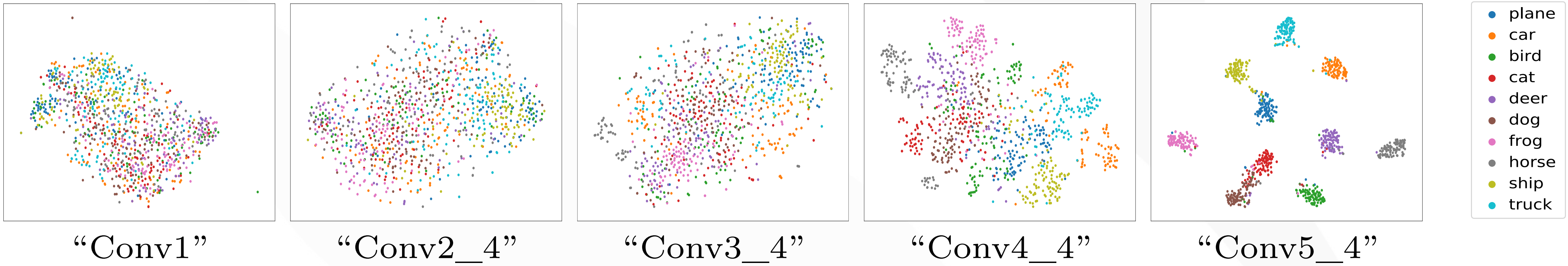


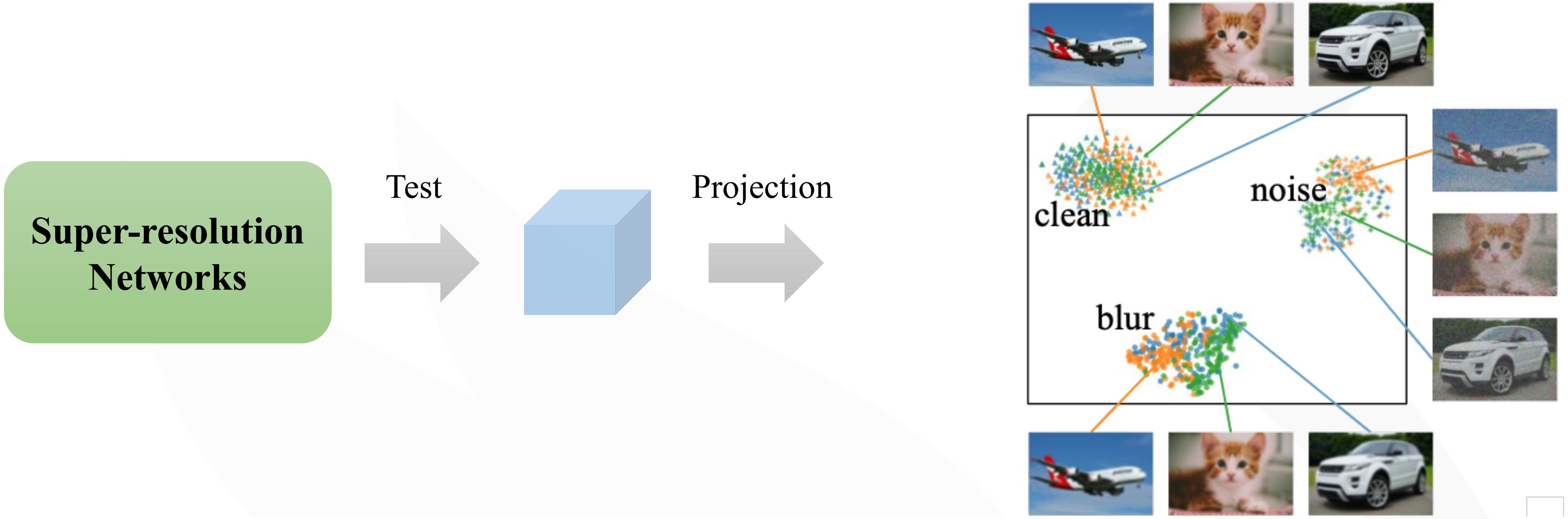
Figure 1: Projected feature representations extracted from different layers of ResNet18 using t-SNE. With the network deepens, the representations become more discriminative to object categories, which clearly shows the semantics of the representations in classification.

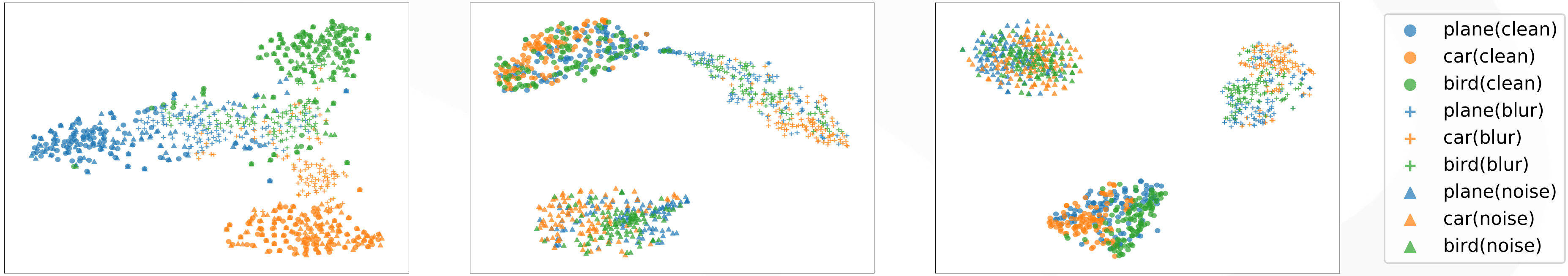
Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

Feature: Where can we find semantics in SR networks?

BEIJING JIAOTONG UNIVERSITY

Observation





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

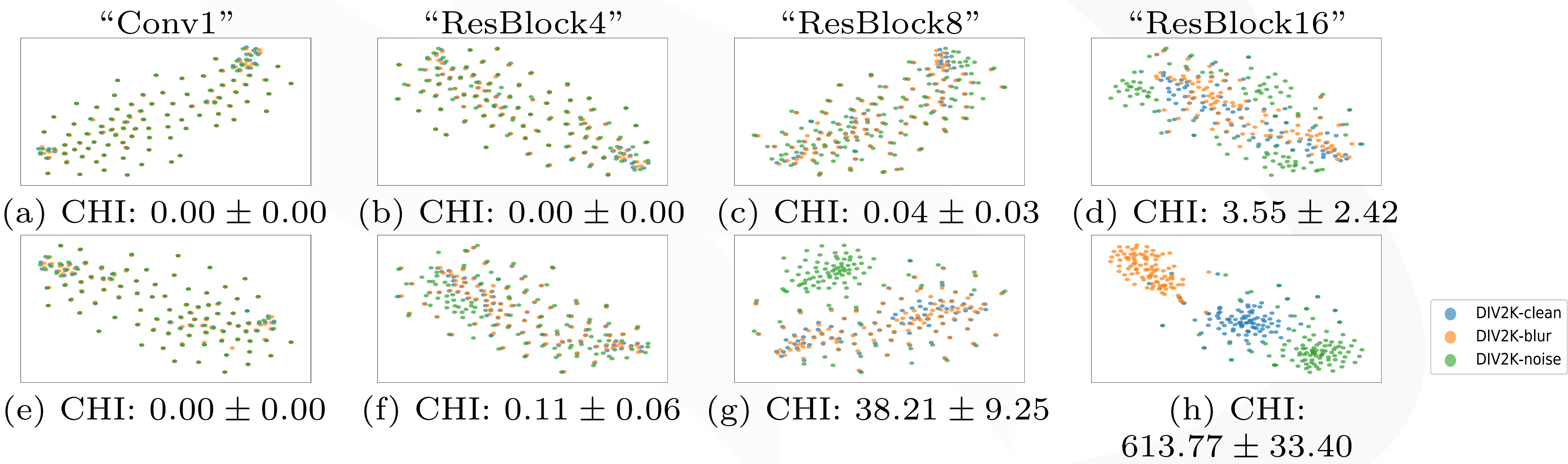
Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

Feature: Where can we find semantics in SR networks?

Observation

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

GAN-based SR networks shows more obvious discriminability.



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

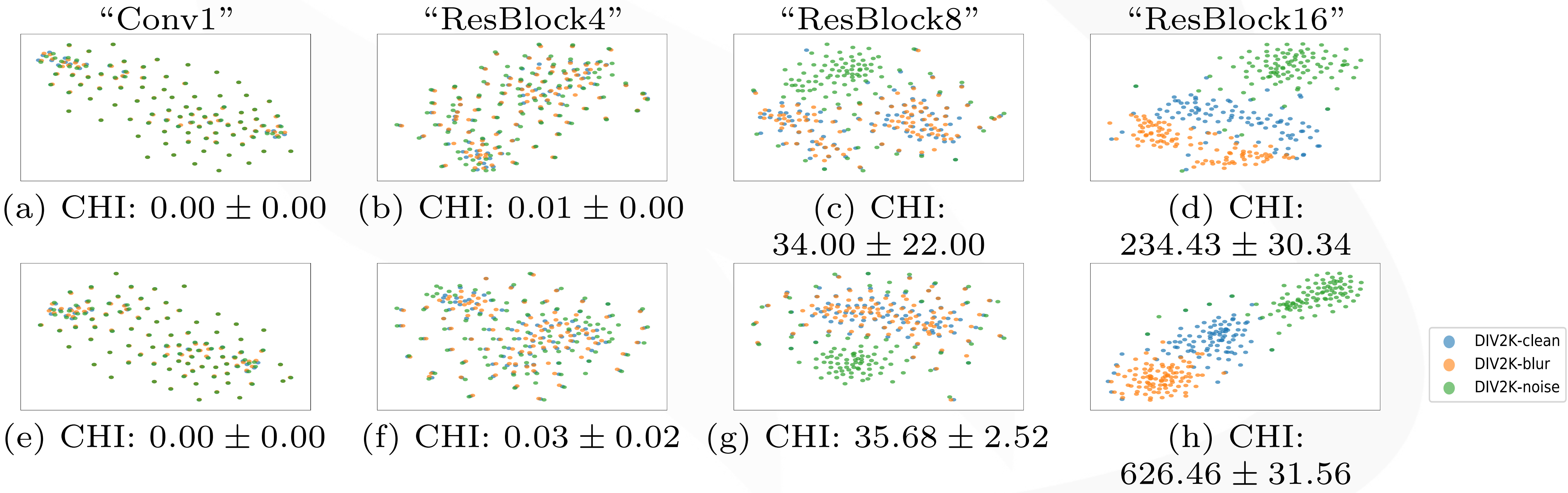
Feature: Where can we find semantics in SR networks?

BEIJING JIAOTONG UNIVERSITY

Observation

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

GAN-based SR networks shows more obvious discriminability.

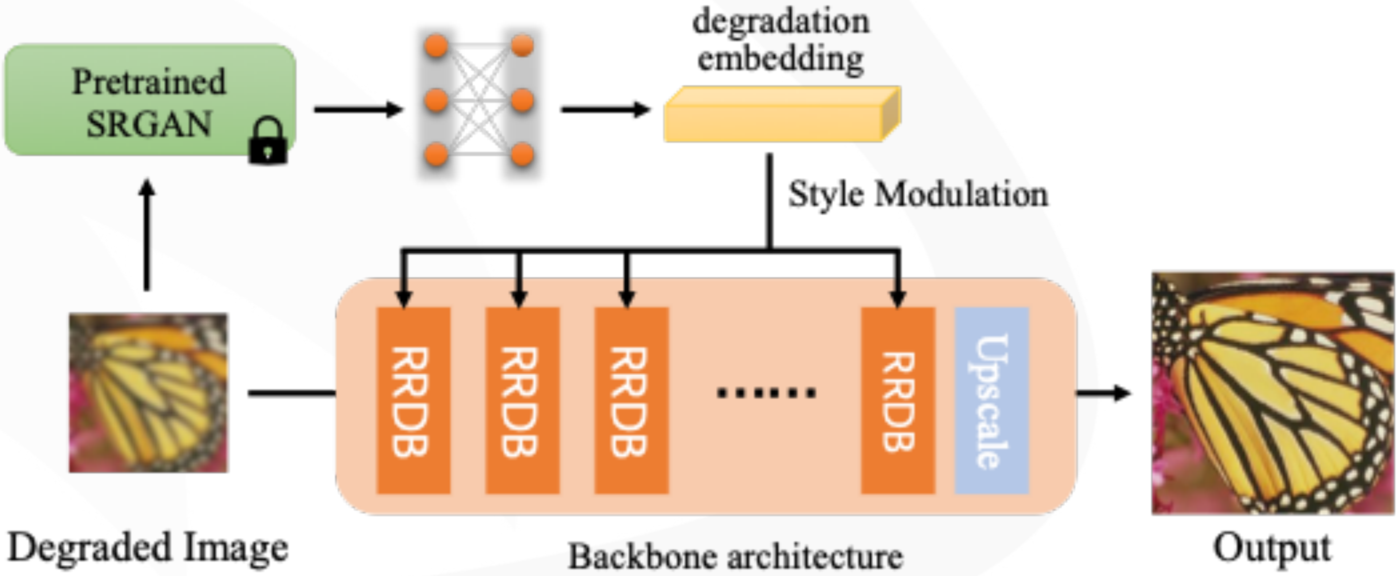


Yihao Liu, Anran Liu, Jinjin Gu, Zhipeng Zhang, Wenhao Wu, Yu Qiao and Chao Dong. 2021. Discovering Distinctive" Semantics" in Super-Resolution Networks. arXiv preprint arXiv:2108.00406.

Feature: Where can we find semantics in SR networks?

Inspirations

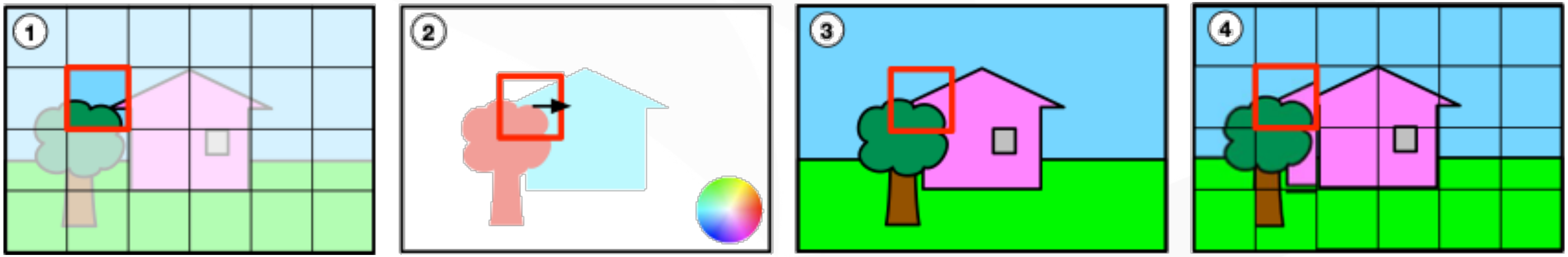
- Interpreting the Generalization of SR (low-level) Networks
- Developing degradation-adaptive Algorithms
- Disentanglement of Image Content/Degradation
- Degradation Classification/Detection



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

Rethinking Alignment in Video Super-Resolution Transformers

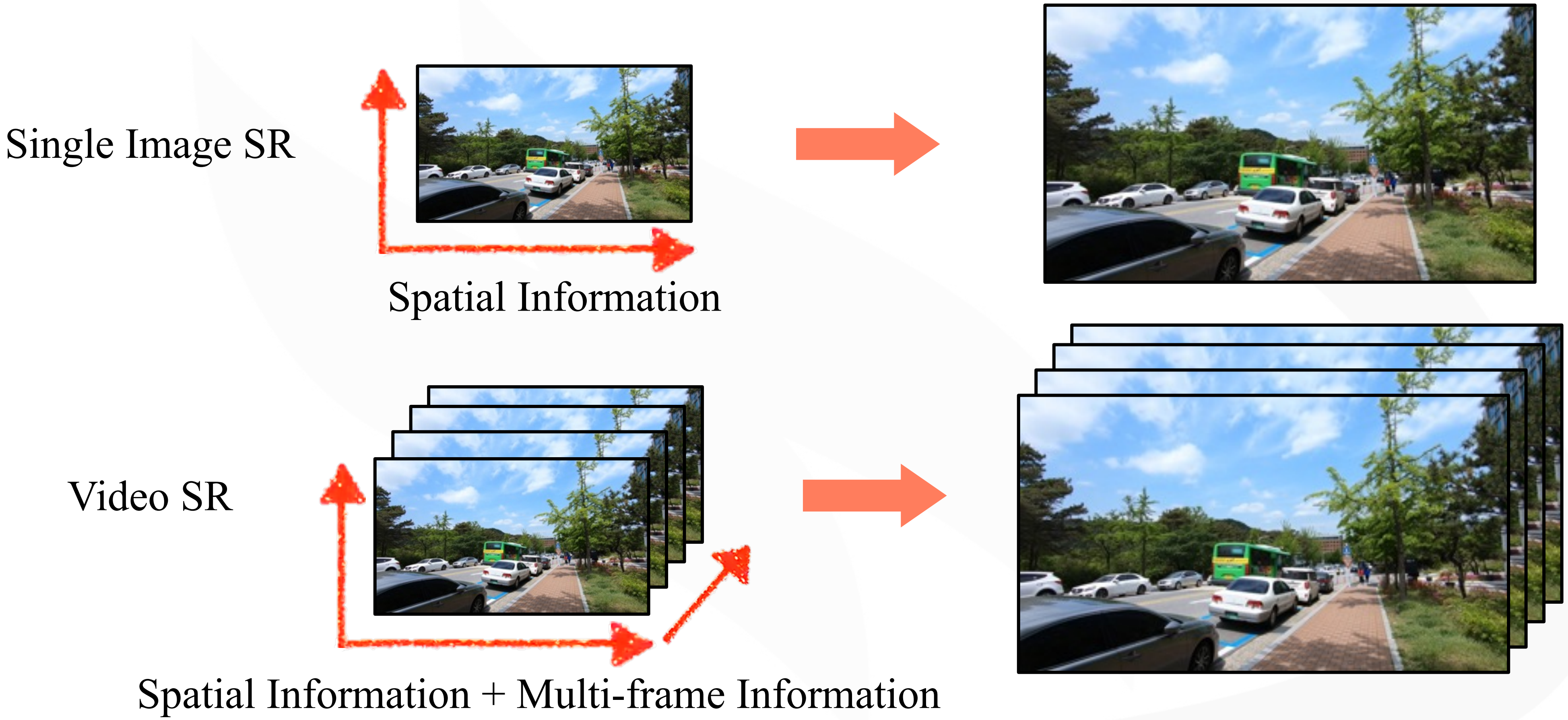
Shuwei Shi[1,2,*], Jinjin Gu[3,4,*], Liangbin Xie[2,5,6], Xintao Wang[6], Yujiu Yang[1], Chao Dong[2,3,†]
[1] Shenzhen International Graduate School, Tsinghua University
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3] Shanghai AI Laboratory [4] The University of Sydney
[5] University of Chinese Academy of Sciences [6] ARC Lab, Tencent PCG



Alignment: Which method benefit to VSR Transformer?

Video Super-Resolution

Video SR exploit the complementary sub-pixel information from multiple frames.

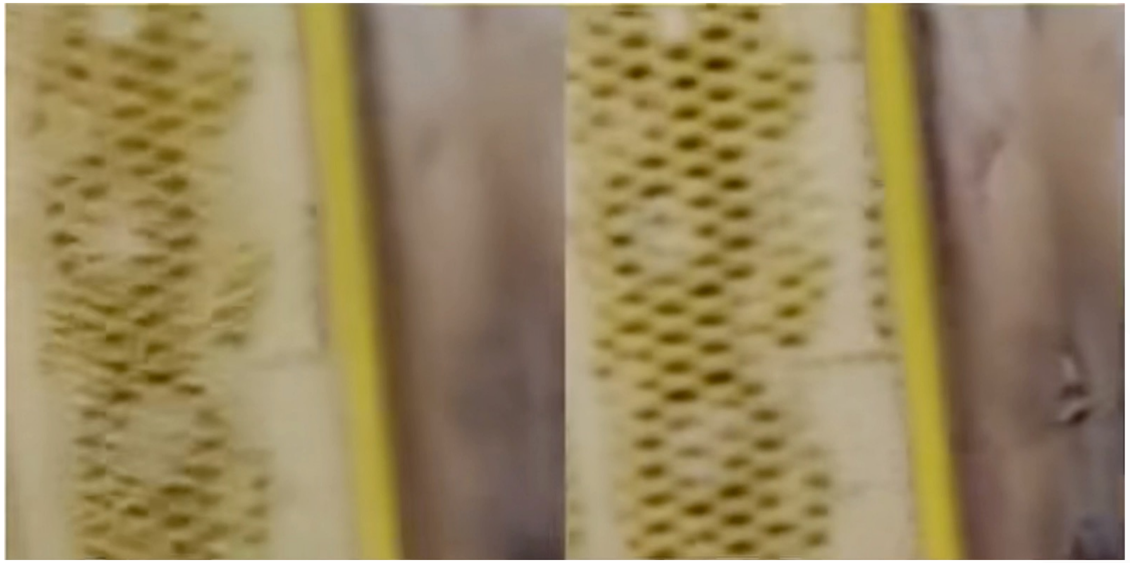


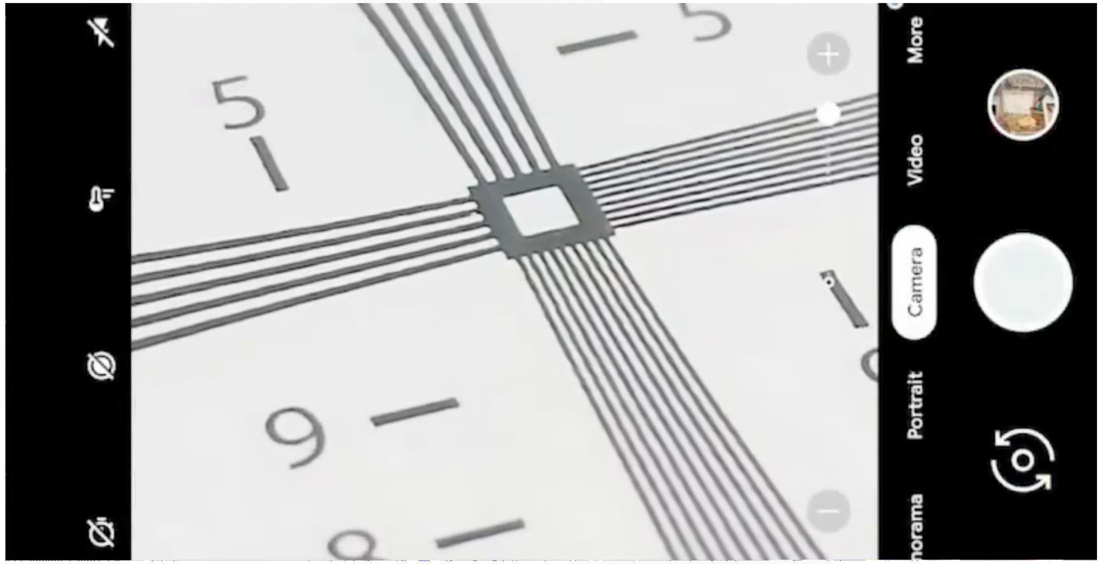
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Video Super-Resolution

Video SR exploit the complementary sub-pixel information from multiple frames.

SISR VSR



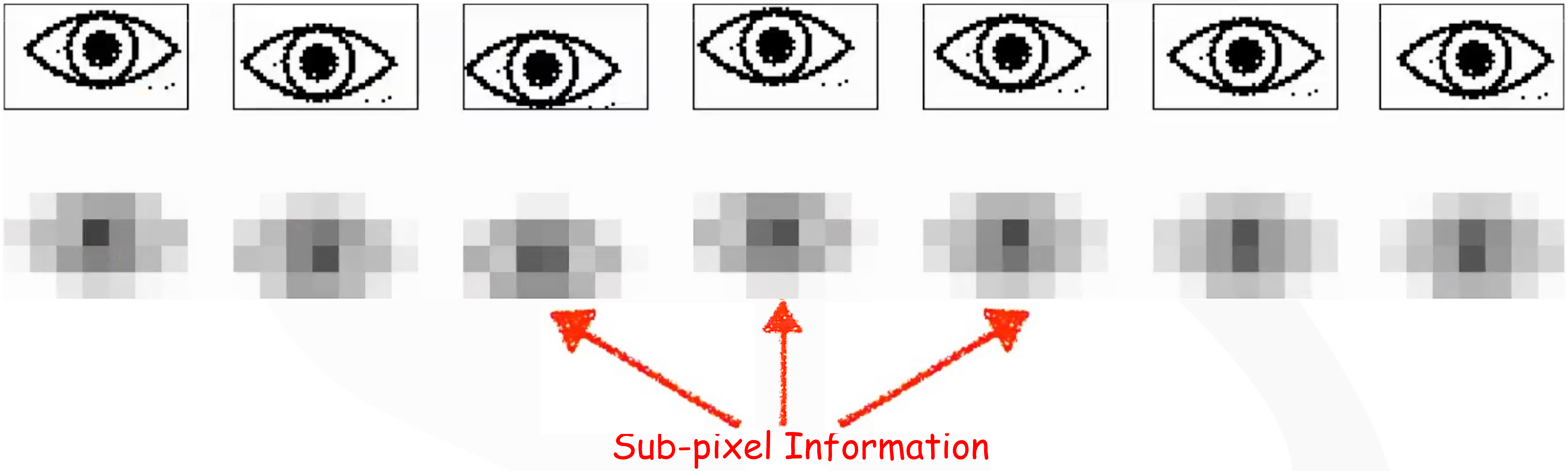


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Video Super-Resolution



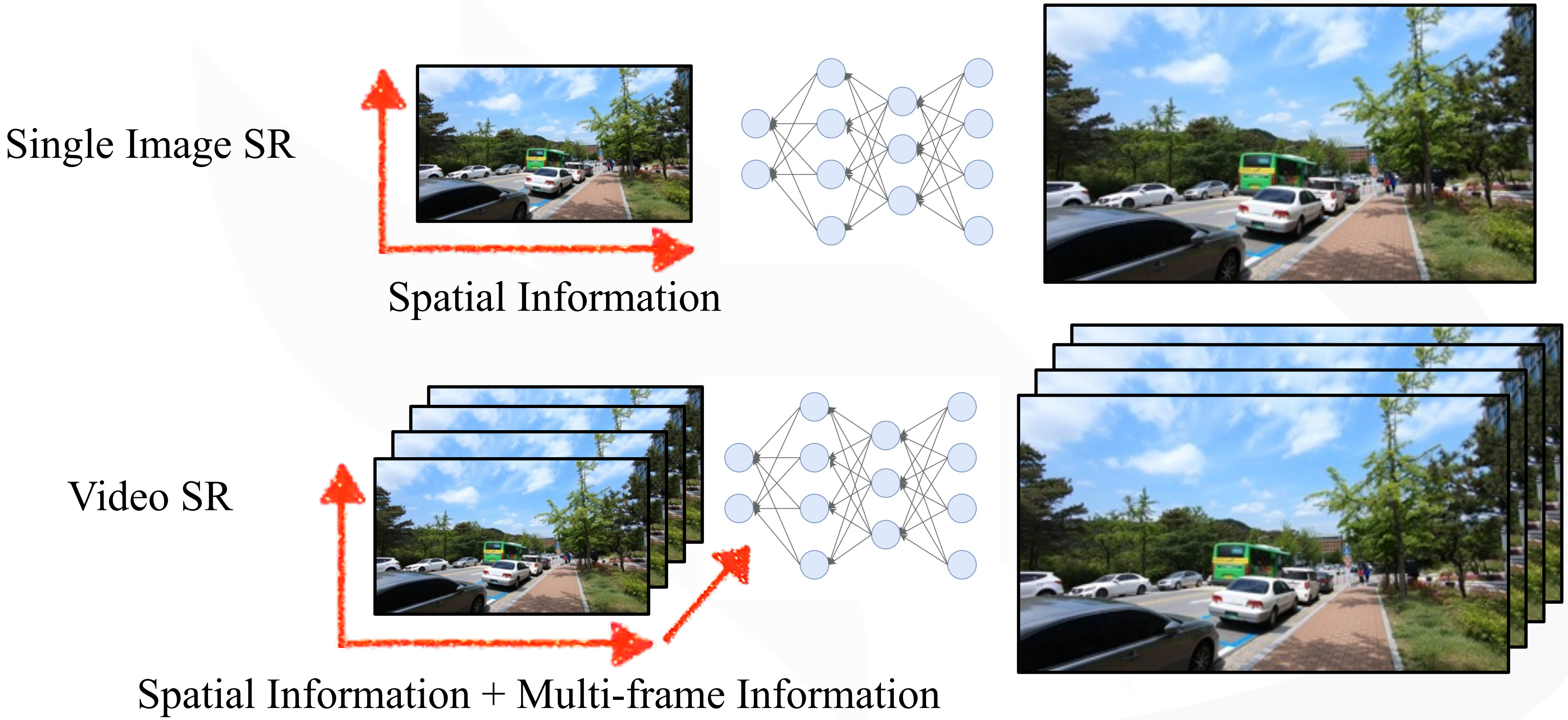
Different downsampled observations of the same object across frames provide additional constraints/information for SR

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

BEIJING JIAOTONG UNIVERSITY

Video Super-Resolution

Video SR exploit the complementary sub-pixel information from multiple frames.

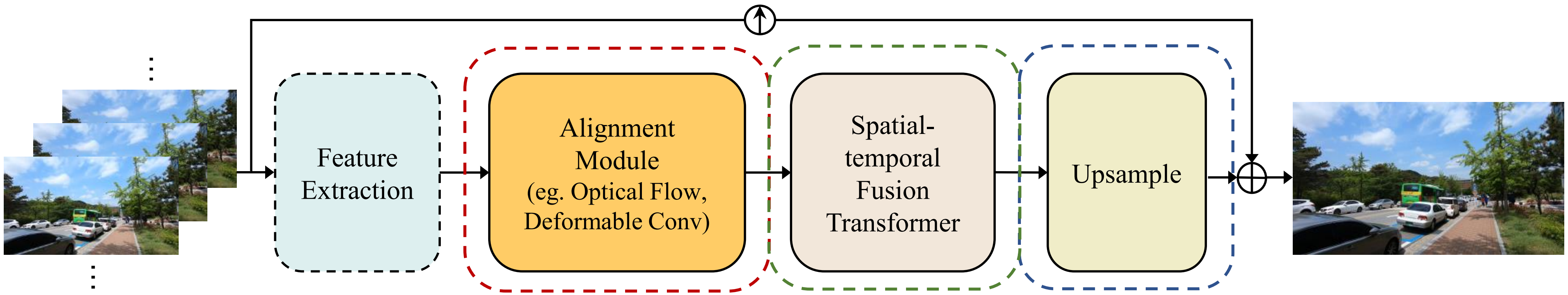


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Framework design

Existing methods can be roughly divided into sliding window-based and recurrent methods.



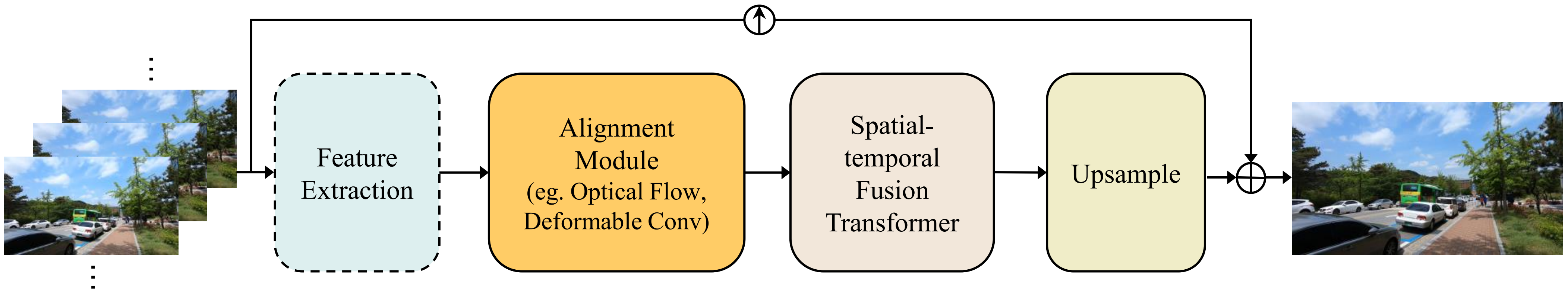
	Sliding-Window			Recurrent				
	EDVR	MuCAN	TDAN	BRCN	FRVSR	RSDN	**BasicVSR**	**IconVSR**
Propagation	Local	Local	Local	**Bidirectional**	Unidirectional	Unidirectional	**Bidirectional**	**Bidirectional** (coupled)
Alignment	**Yes** (DCN)	**Yes** (correlation)	**Yes** (DCN)	No	**Yes** (flow)	No	**Yes** (flow)	**Yes** (flow)
Aggregation	Concatenate + **TSA**	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate + **Refill**
Upsampling	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle

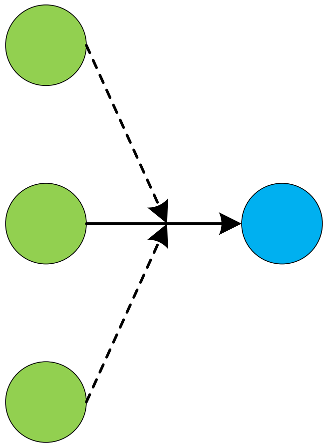
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

BEIJING JIAOTONG UNIVERSITY

Framework design

Existing methods can be roughly divided into sliding window-based and recurrent methods.





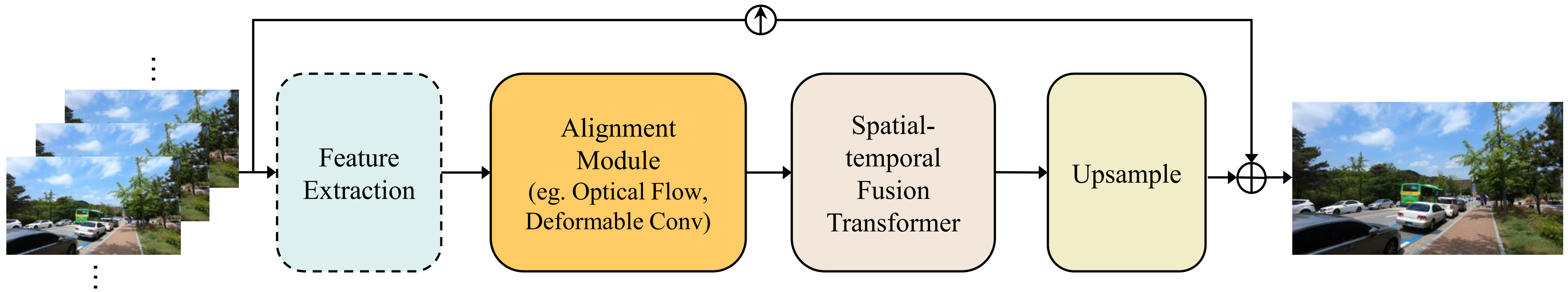
	Sliding-Window			Recurrent				
	EDVR	MuCAN	TDAN	BRCN	FRVSR	RSDN	**BasicVSR**	**IconVSR**
Propagation	Local	Local	Local	**Bidirectional**	Unidirectional	Unidirectional	**Bidirectional**	**Bidirectional** (coupled)
Alignment	**Yes** (DCN)	**Yes** (correlation)	**Yes** (DCN)	No	**Yes** (flow)	No	**Yes** (flow)	**Yes** (flow)
Aggregation	Concatenate + **TSA**	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate + **Refill**
Upsampling	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle

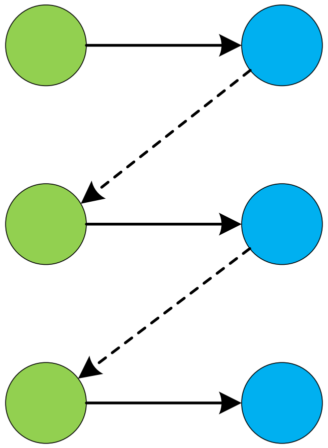
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Framework design

Existing methods can be roughly divided into sliding window-based and recurrent methods.





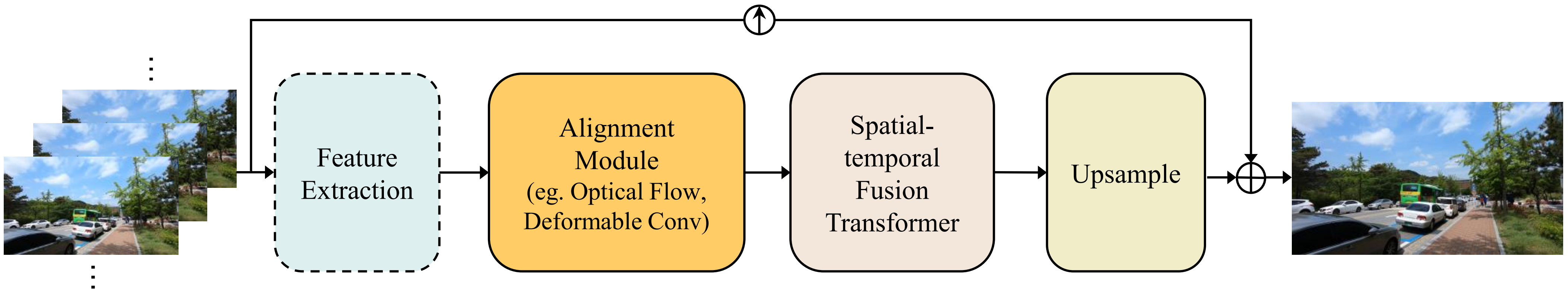
	Sliding-Window			Recurrent				
	EDVR	MuCAN	TDAN	BRCN	FRVSR	RSDN	**BasicVSR**	**IconVSR**
Propagation	Local	Local	Local	**Bidirectional**	Unidirectional	Unidirectional	**Bidirectional**	**Bidirectional** (coupled)
Alignment	**Yes** (DCN)	**Yes** (correlation)	**Yes** (DCN)	No	**Yes** (flow)	No	**Yes** (flow)	**Yes** (flow)
Aggregation	Concatenate + **TSA**	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate + **Refill**
Upsampling	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle

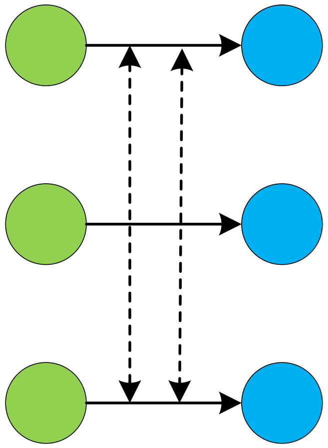
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Framework design

Existing methods can be roughly divided into sliding window-based and recurrent methods.



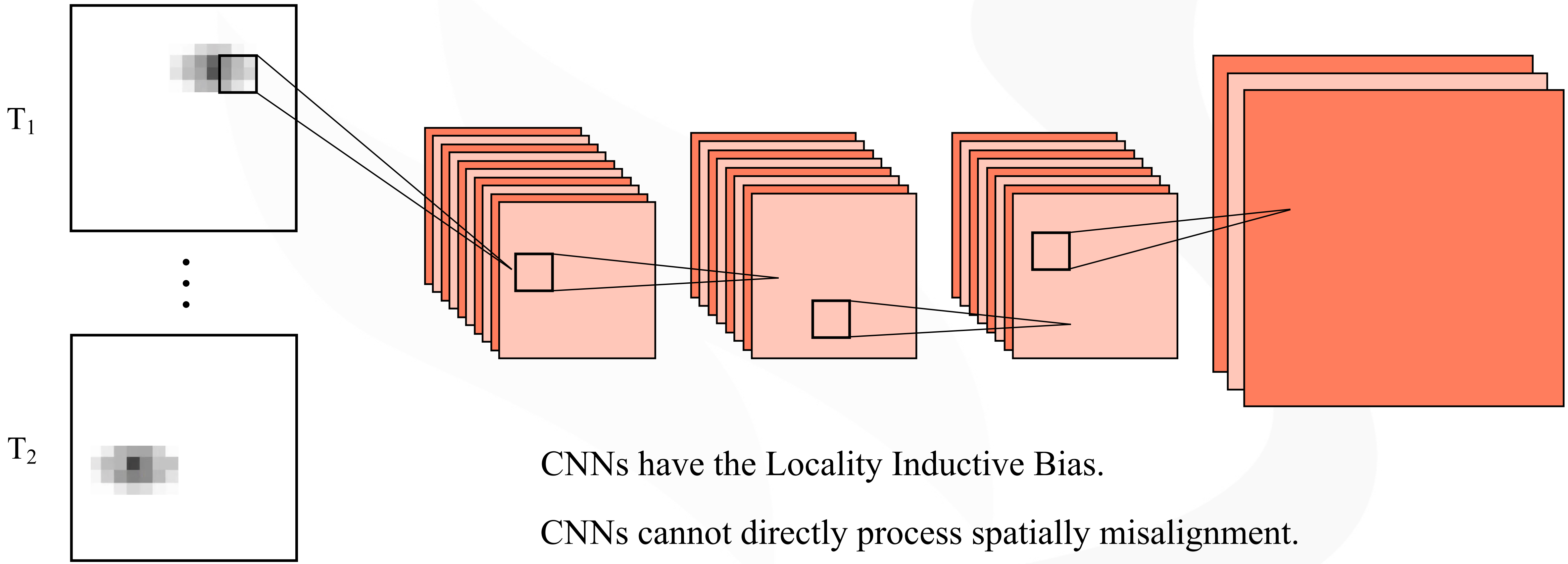


	Sliding-Window			Recurrent				
	EDVR	MuCAN	TDAN	BRCN	FRVSR	RSDN	**BasicVSR**	**IconVSR**
Propagation	Local	Local	Local	**Bidirectional**	Unidirectional	Unidirectional	**Bidirectional**	**Bidirectional** (coupled)
Alignment	**Yes** (DCN)	**Yes** (correlation)	**Yes** (DCN)	No	**Yes** (flow)	No	**Yes** (flow)	**Yes** (flow)
Aggregation	Concatenate + **TSA**	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate	Concatenate + **Refill**
Upsampling	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle	Pixel-Shuffle

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment

Why we should conduct alignment in a VSR convolutional network.



CNNs have the Locality Inductive Bias.

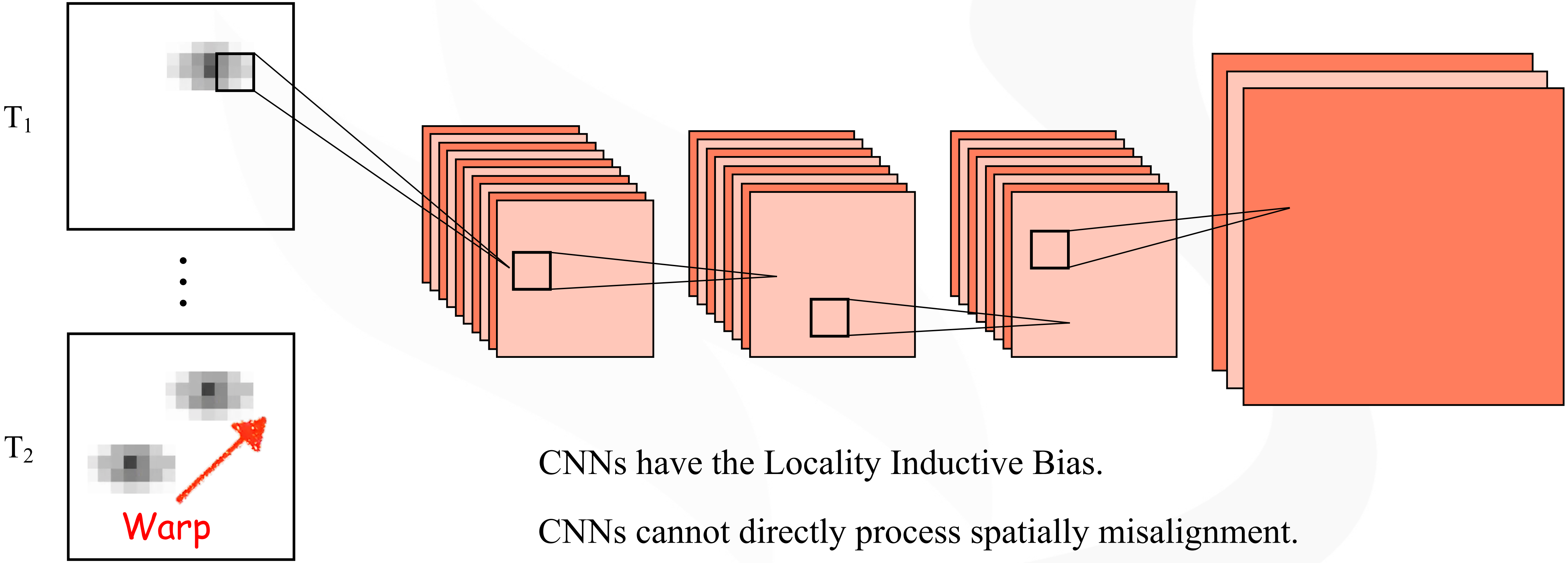
CNNs cannot directly process spatially misalignment.

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Alignment

Why we should conduct alignment in a VSR convolutional network.



CNNs have the Locality Inductive Bias.

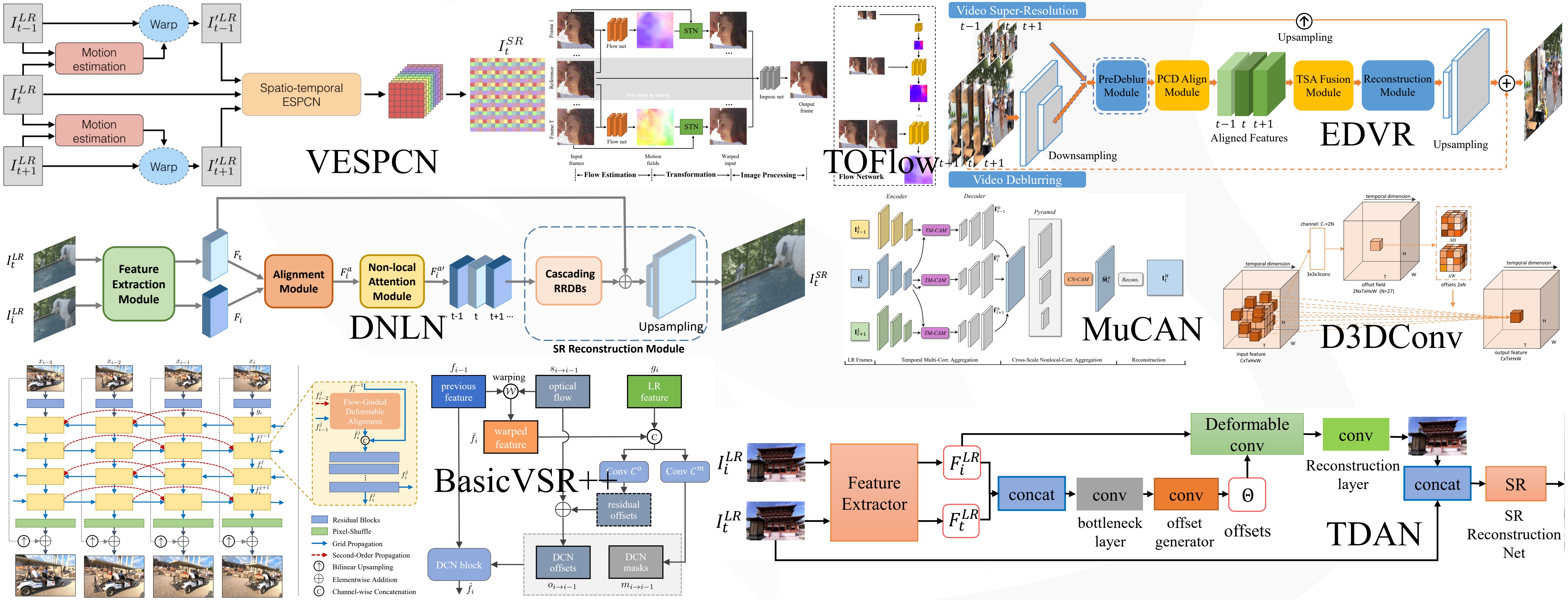
CNNs cannot directly process spatially misalignment.

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Alignment

Alignment is an important module and is the core of VSR method development.

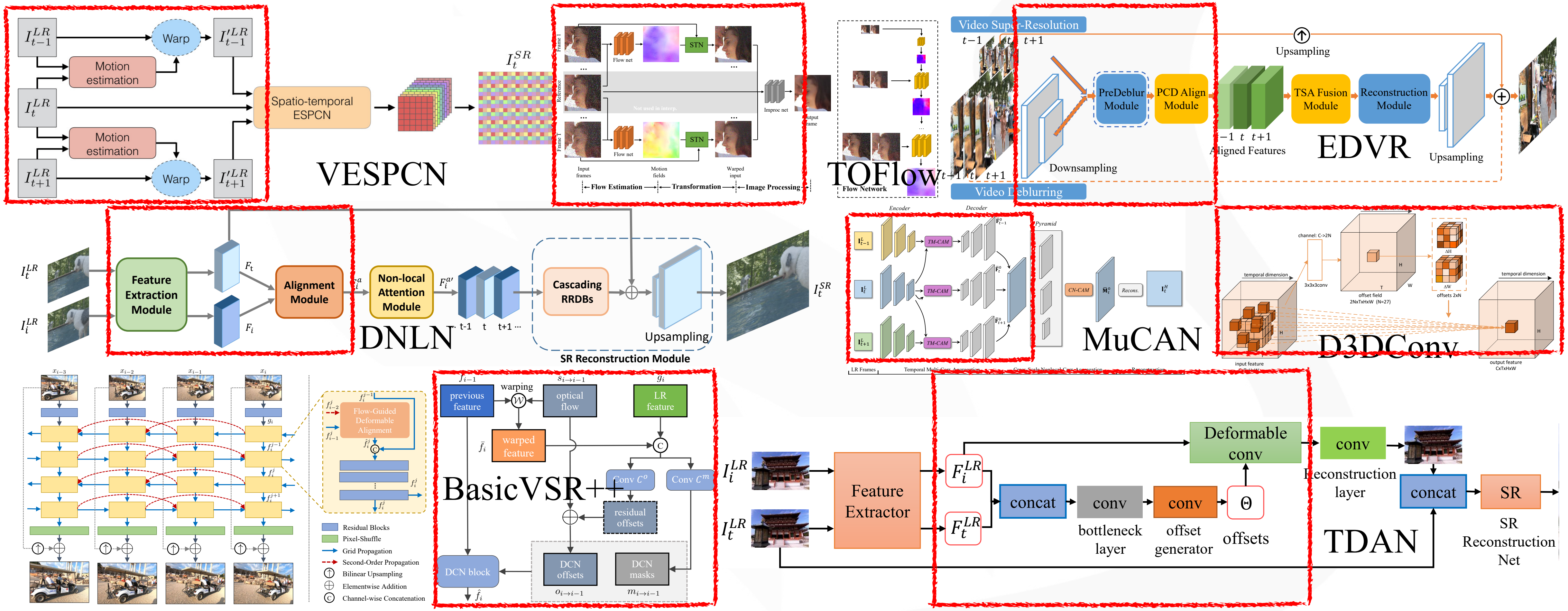


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Alignment

Alignment is an important module and is the core of VSR method development.

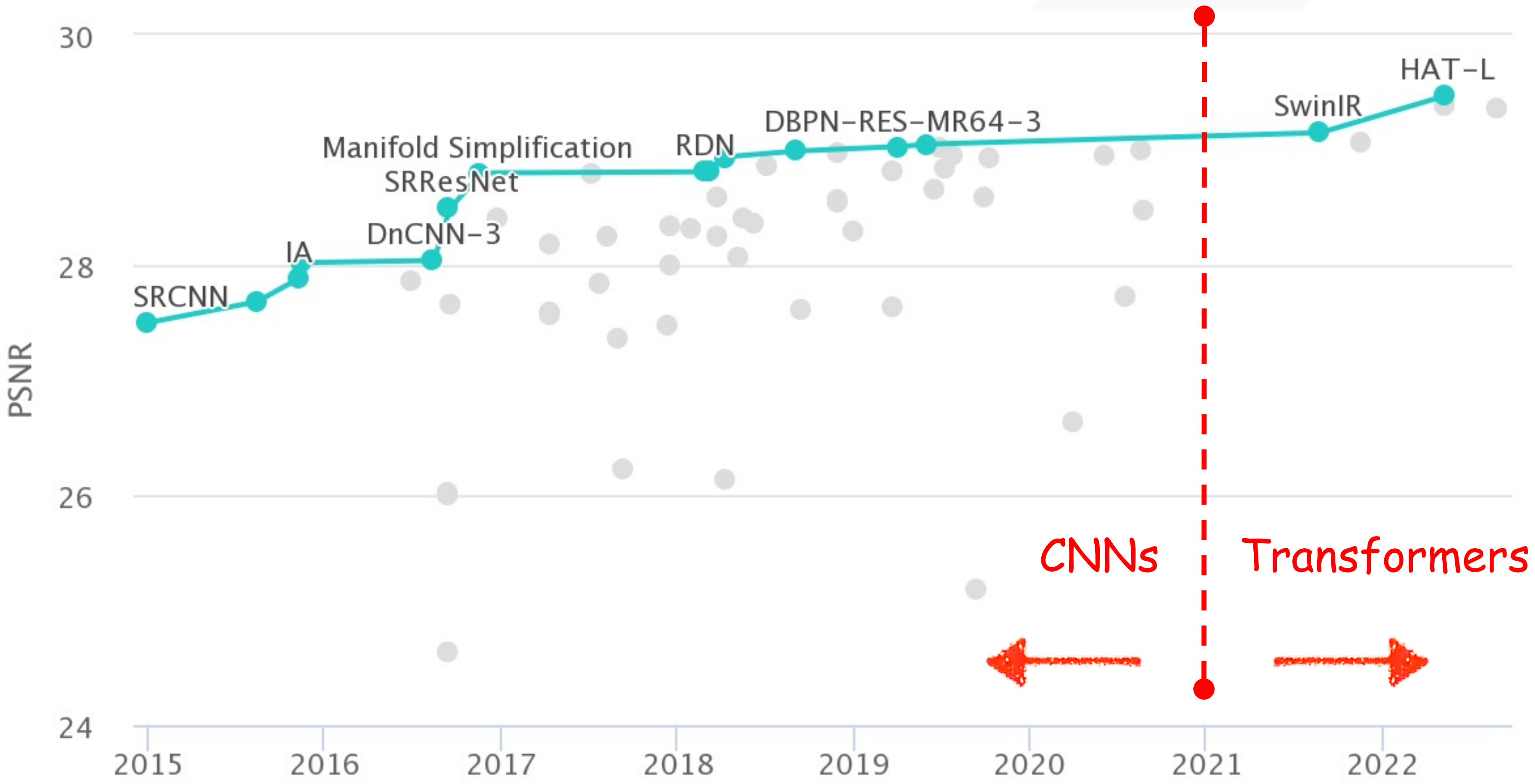


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Image Restoration Transformers

Transformers refresh the state-of-the-art in Network designs.



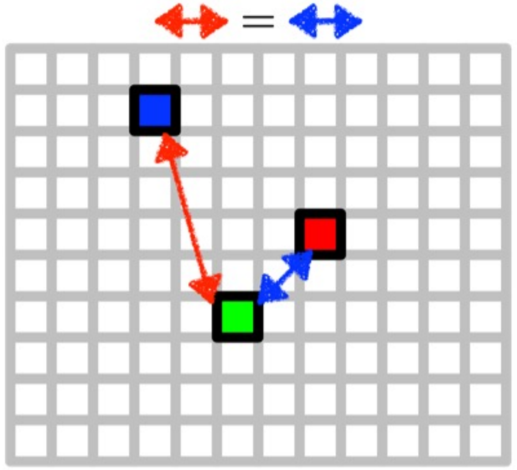
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Image Restoration Transformers

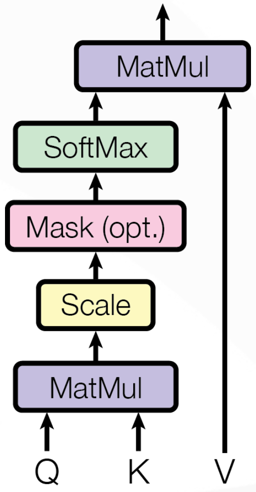
Transformers:

- Treat the input signal as tokens. In image restoration, one pixel is one token.
- Using self-attention to process spatial information, instead of convolutions.
- Self-attention is efficient for spatially long-term distributed elements.
- Do not assume the locality inductive bias.

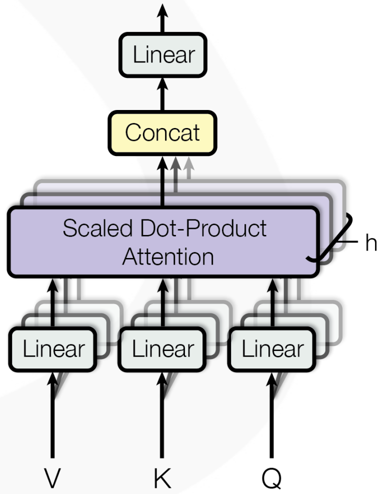


Transformer

Scaled Dot-Product Attention



Multi-Head Attention

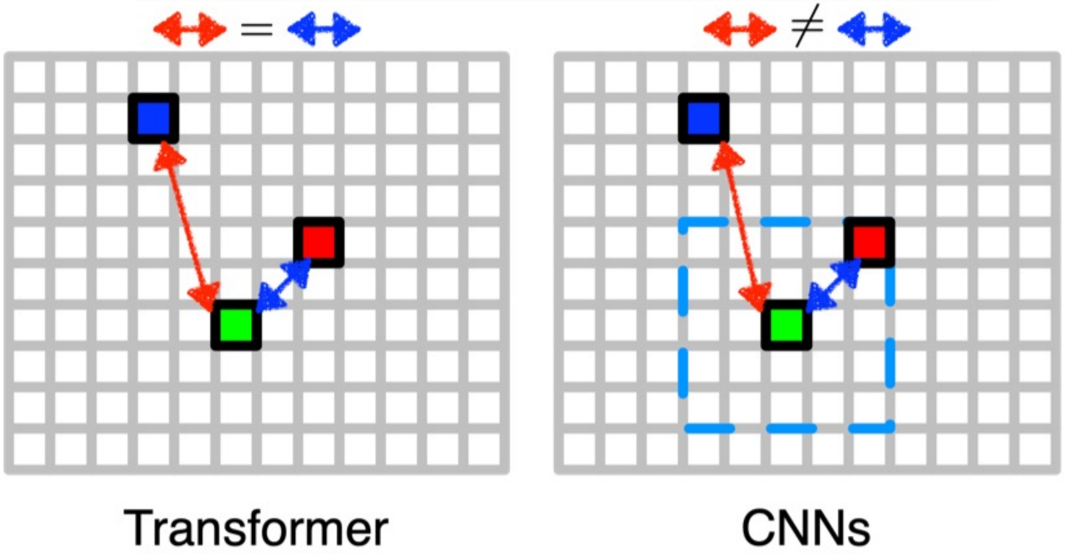


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

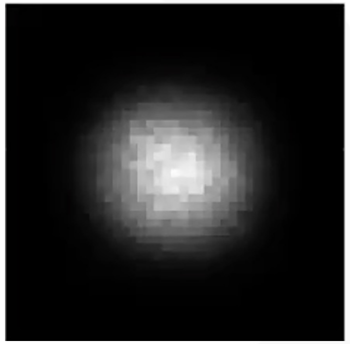
Image Restoration Transformers

Transformers:

- Treat the input signal as tokens. In image restoration, one pixel is one token.
- Using self-attention to process spatial information, instead of convolutions.
- Self-attention is efficient for spatially long-term distributed elements.
- **Do not assume the locality inductive bias.**



CNNs' locality inductive bias



Luo, Wenjie, et al. "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks." NIPS2016.

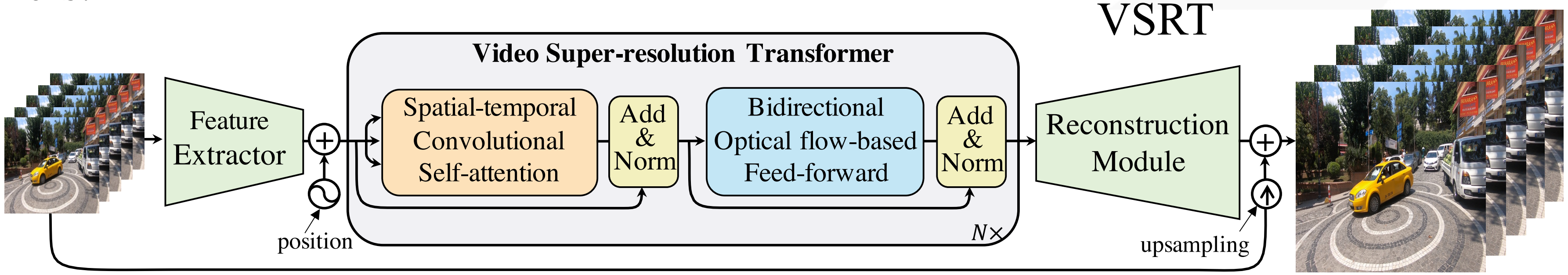
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

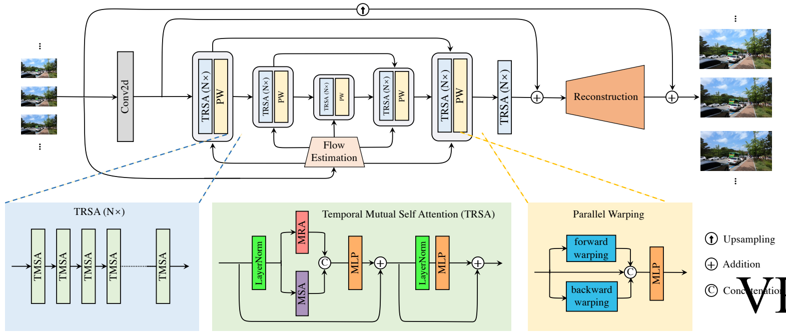
Alignment: Which method benefit to VSR Transformer?

Video Restoration Transformers

Transformers:

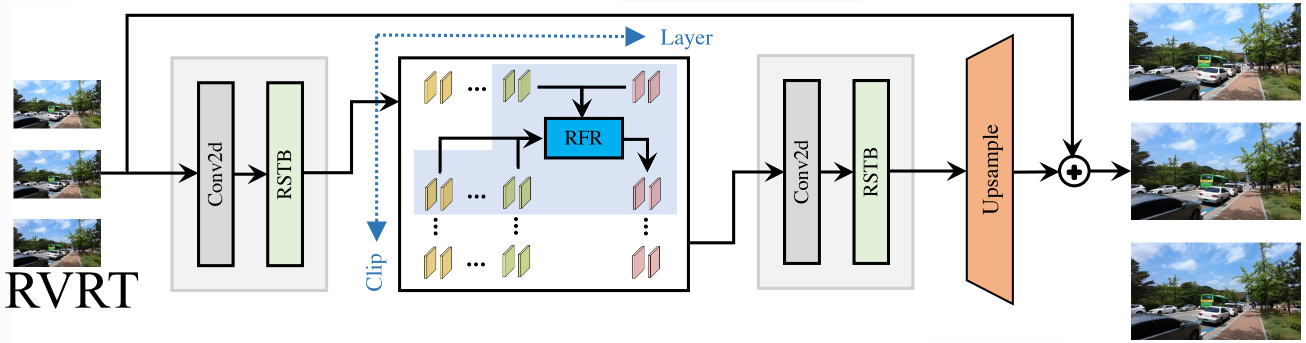
VSRT

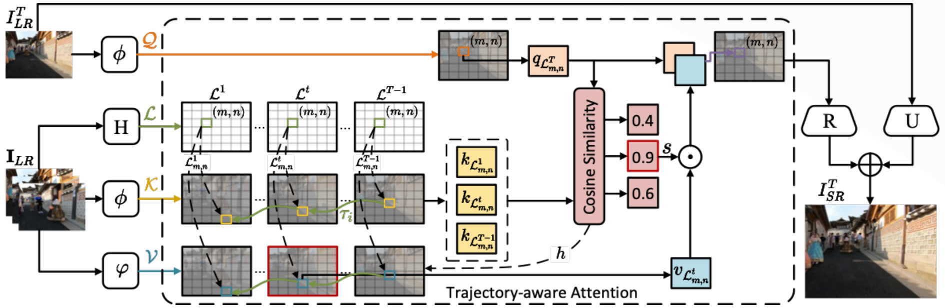




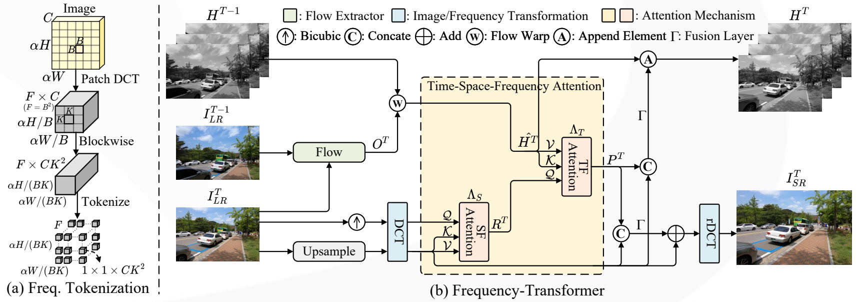
VRT

RVRT





TTVSR



FTVSR

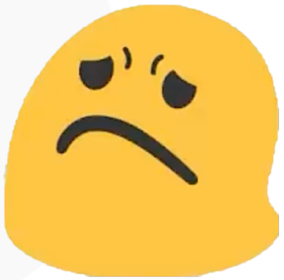
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Rethinking

Question 1:

- The VSR model needs alignment because CNN has locality inductive bias.
- Transformers have no locality inductive bias.
- **Do we still need alignment for VSR Transformers?**



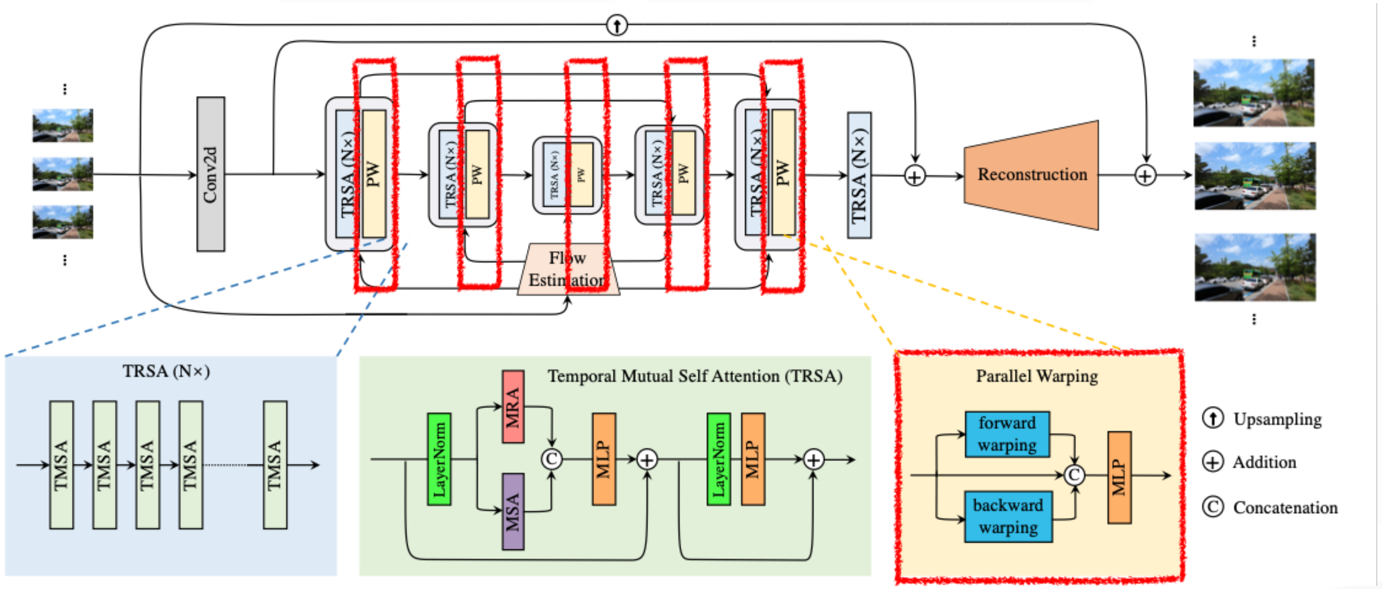
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Rethinking

Question 1:

- ➢ The VSR model needs alignment because CNN has locality inductive bias.
- ➢ Transformers have no locality inductive bias.
- ➢ **Do we still need alignment for VSR Transformers?**



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

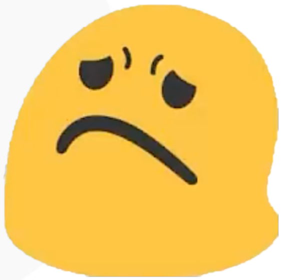
Rethinking

Question 1:

- The VSR model needs alignment because CNN has locality inductive bias.
- Transformers have no locality inductive bias.
- **Do we still need alignment for VSR Transformers?**

Question 2:

- If we do not need alignment in VSR Transformer,
- **What will happen if we use alignment in it?**

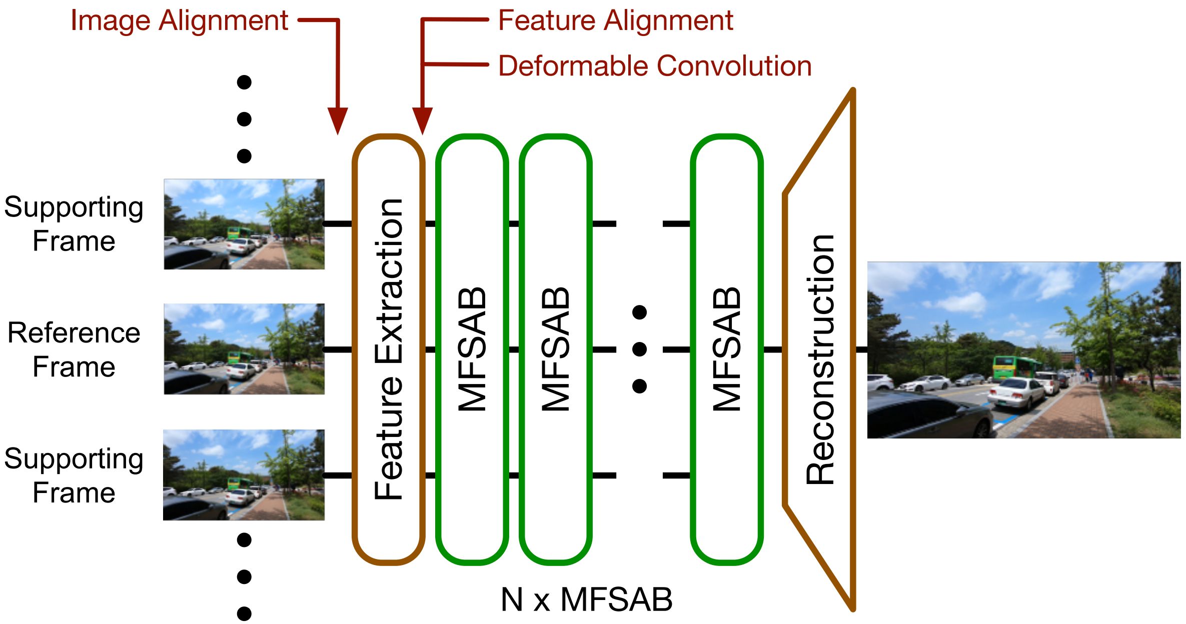


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

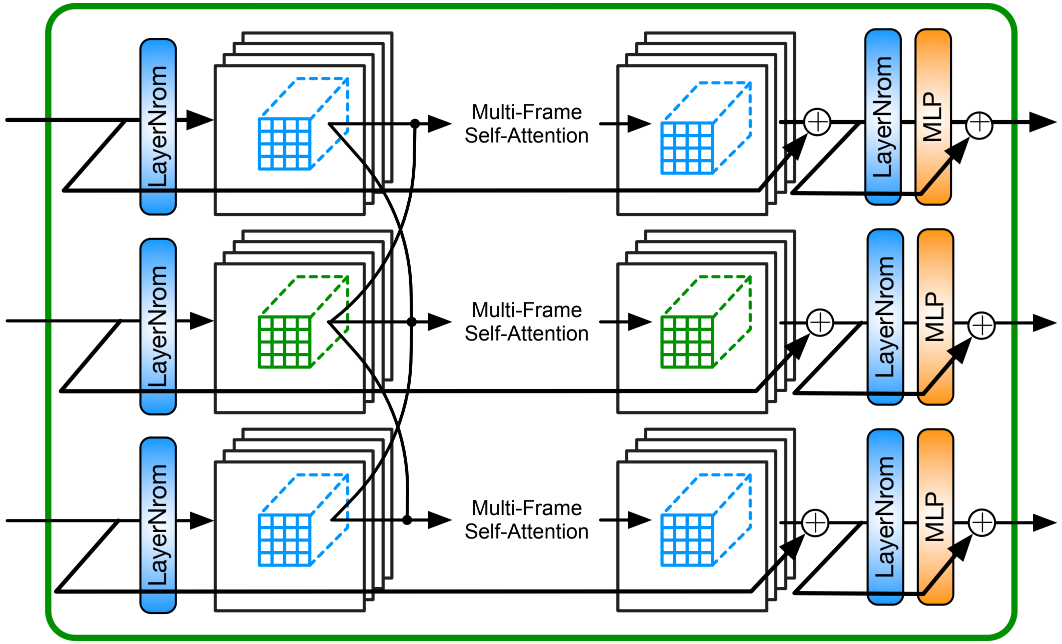
Alignment: Which method benefit to VSR Transformer?

Preliminary Settings

We build the basic VSR Transformer model using multi-frame self-attention blocks. This is an example basic on the sliding window strategy.



Video Super-Resolution Transformer



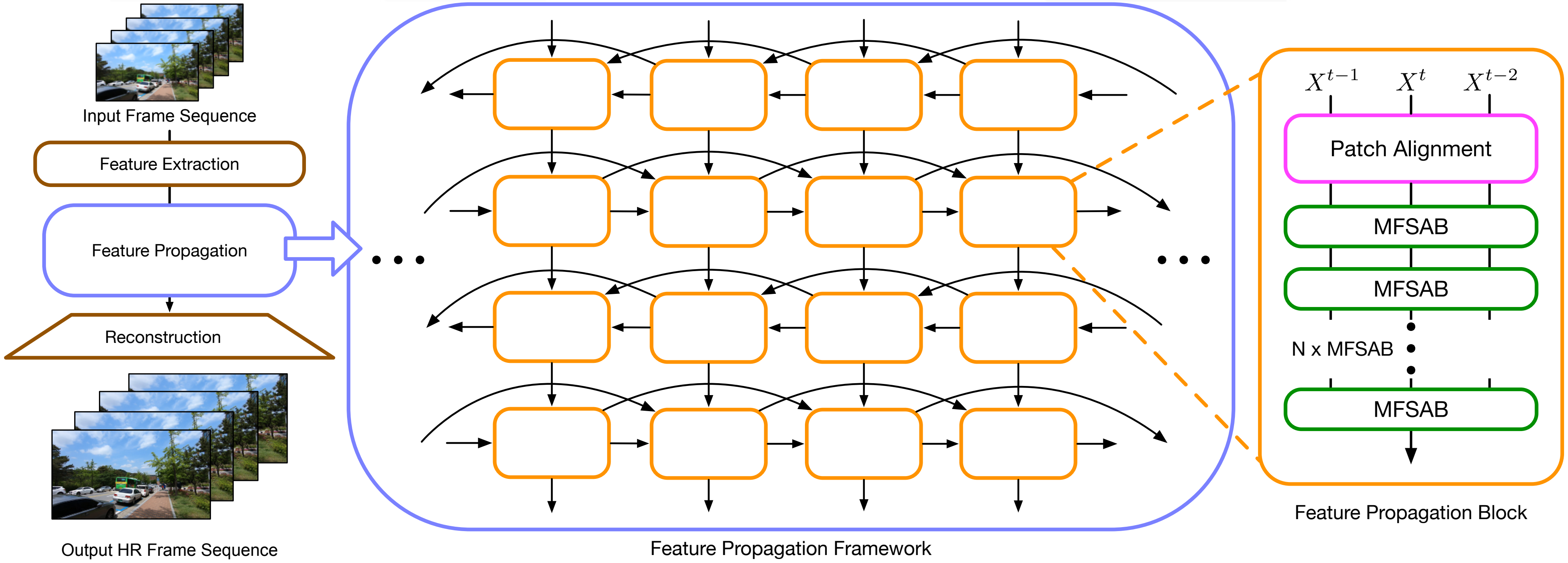
Multi-Frame Self-Attention Block (MFSAB)

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Preliminary Settings

We build the basic VSR Transformer model using multi-frame self-attention blocks. This is an example basic on the sliding window strategy.



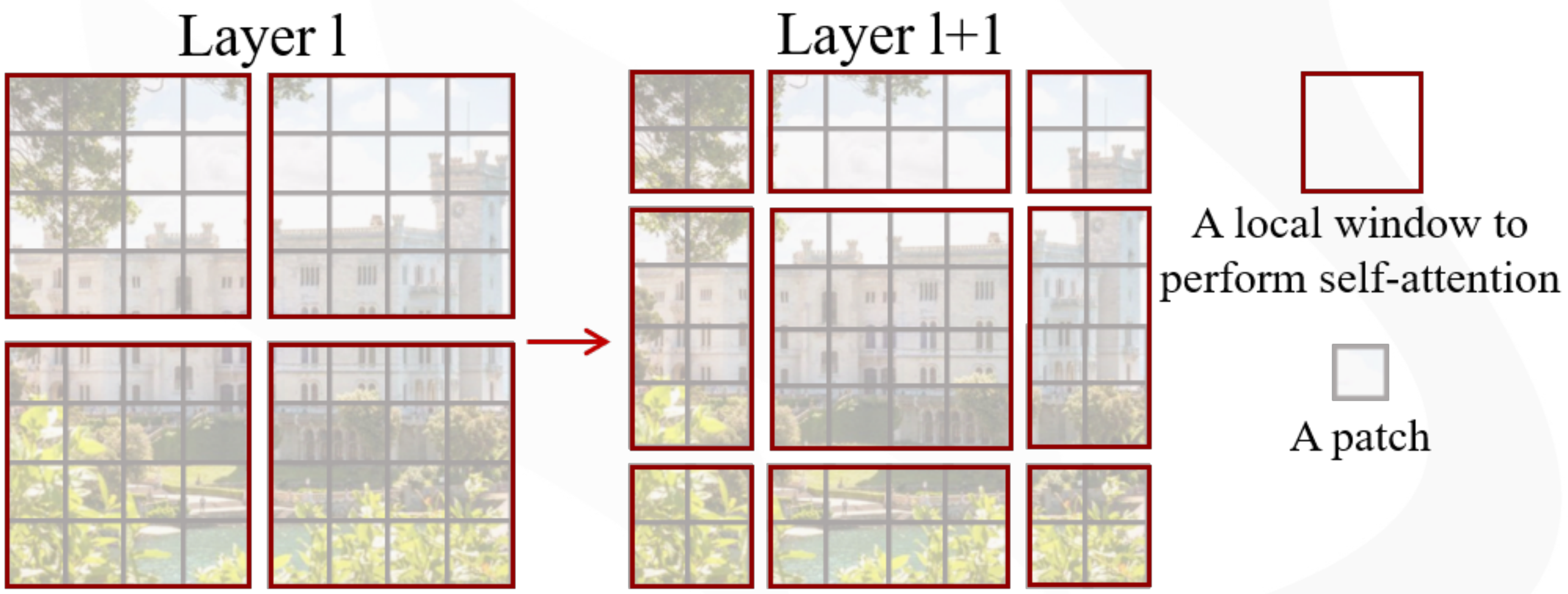
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Preliminary Settings

We build the basic VSR Transformer model using multi-frame self-attention blocks. This is an example basic on the sliding window strategy.



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Preliminary Settings

Alignment Methods:

1. Image Alignment.

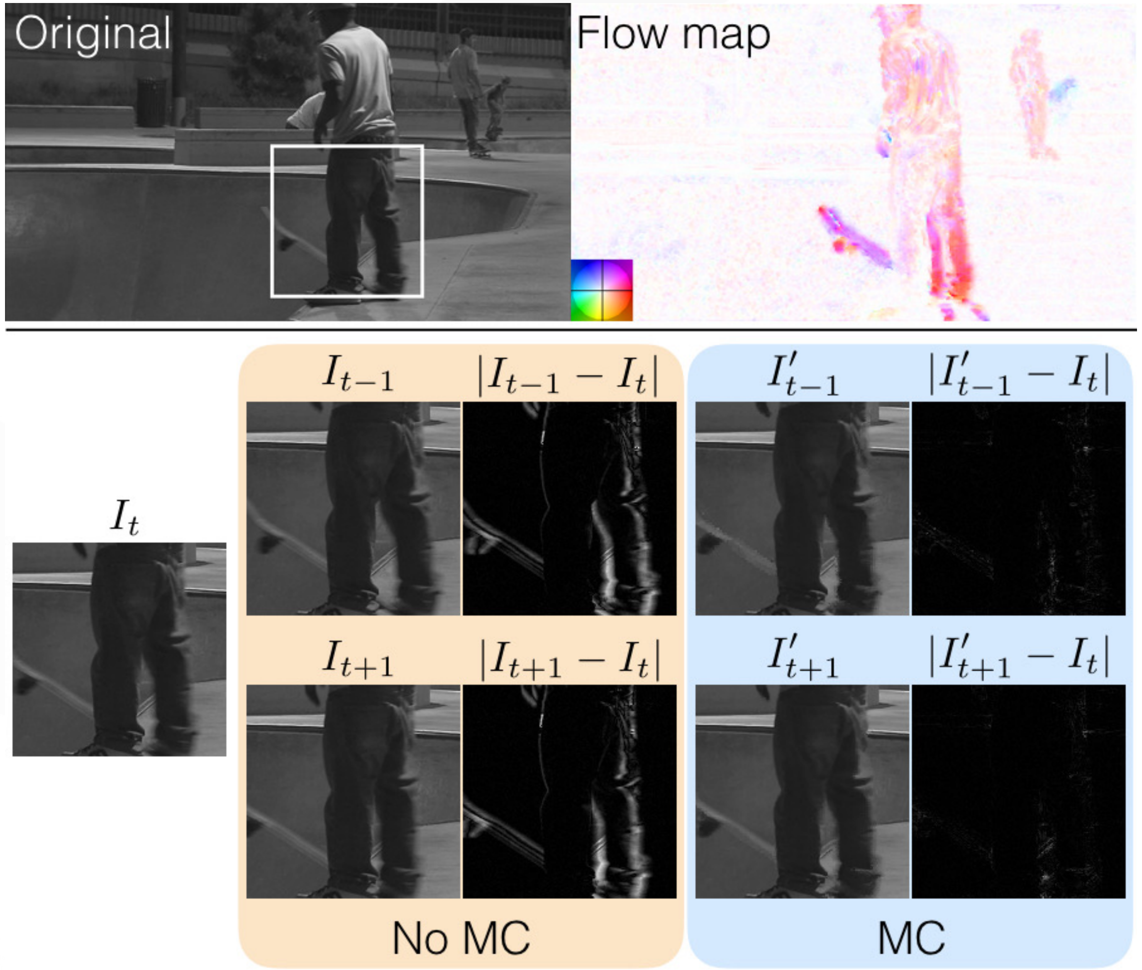
Flow estimation

↓

Warp/Resampling supporting frames using flow

↓

Obtain aligned frames



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

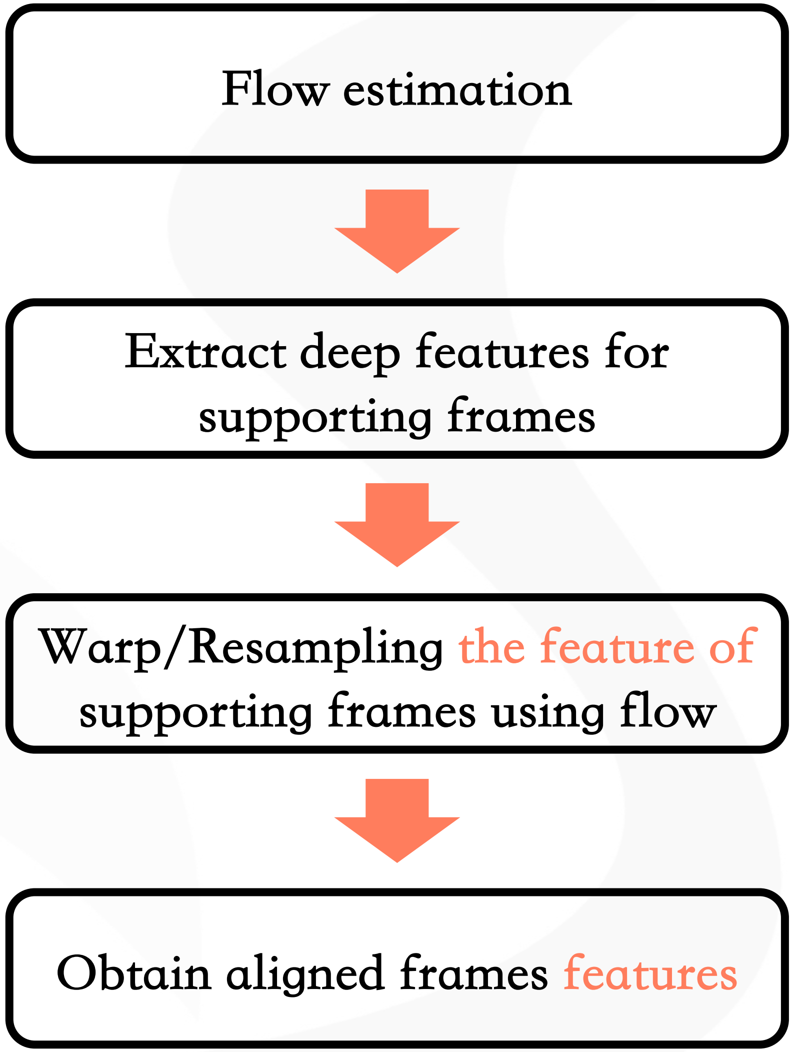
Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Preliminary Settings

Alignment Methods:

1. Image Alignment.
2. Feature Alignment.



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

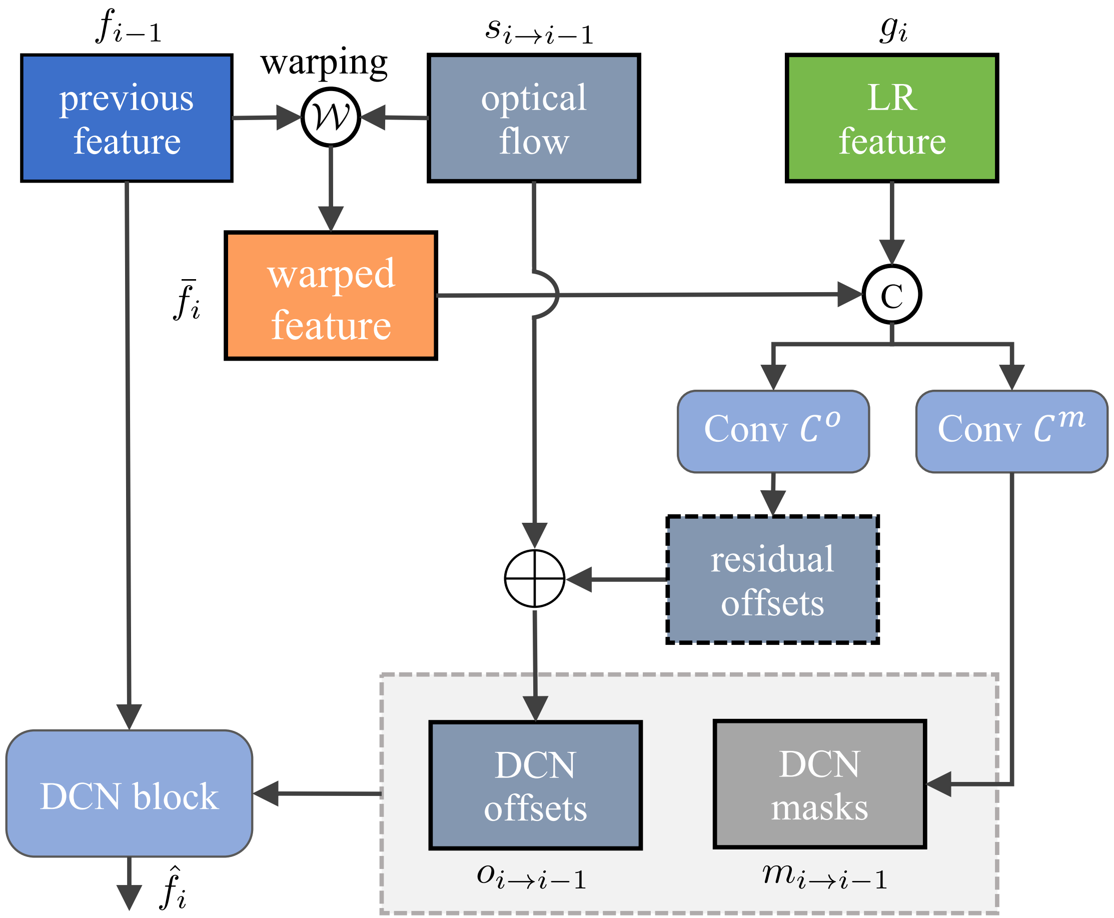
Preliminary Settings

Alignment Methods:

1. Image Alignment.

2. Feature Alignment.

3. Flow Guided Deformable Convolution.



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Preliminary Settings

Alignment Methods:

1. Image Alignment.

2. Feature Alignment.

3. Flow Guided Deformable Convolution.

4. No Alignment.

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Preliminary Settings

Dataset and Benchmarks:

- Setting One:

  Training: REDS dataset, 266 sequences

  Testing: READS4 test sequences

- Setting Two:

  Training: Vimeo-90K dataset, 64,612 sequences

  Testing:

  1. Vimeo-90K testing set, 7,824 video sequences
  2. Vid4 testing set, 4 video sequences

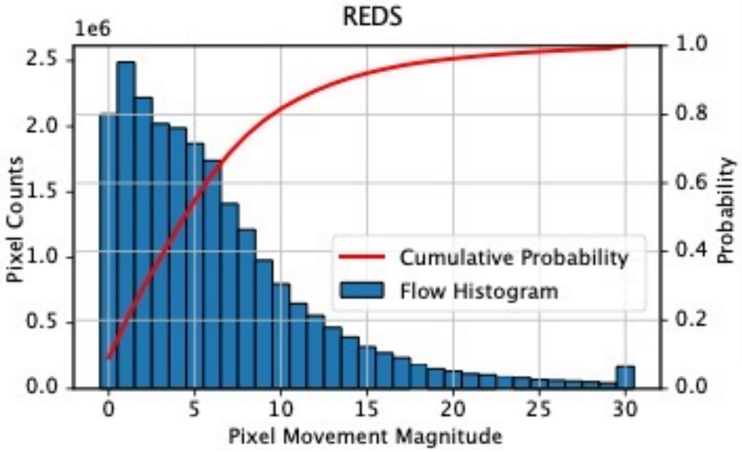
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

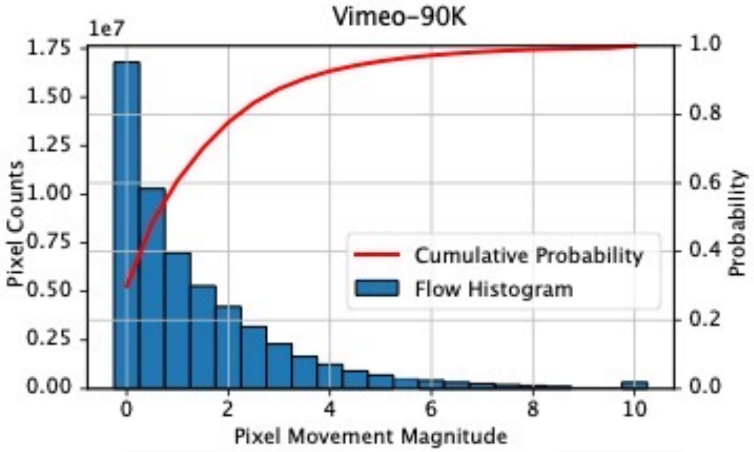
Preliminary Settings

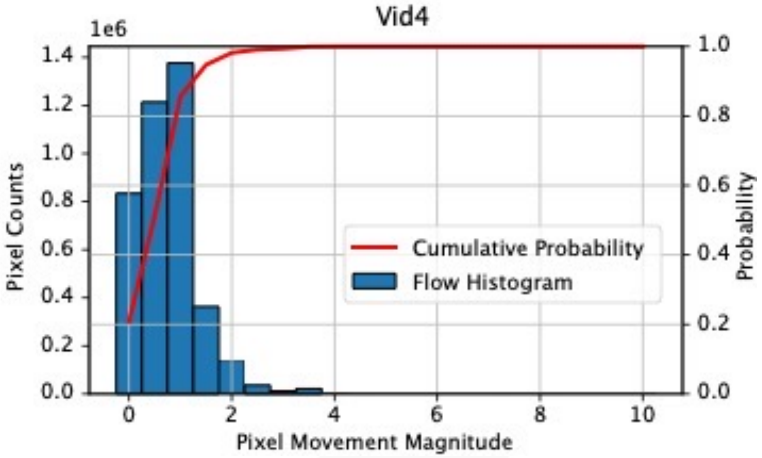
The distribution of movement:

Large Movement



Small Movement





Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

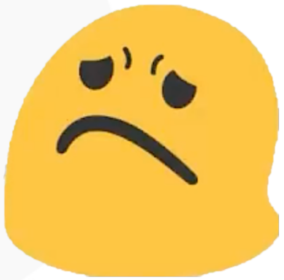
Rethinking

Question 1:

- The VSR model needs alignment because CNN has locality inductive bias.
- Transformers have no locality inductive bias.
- **Do we still need alignment for VSR Transformers?**

Question 2:

- If we do not need alignment in VSR Transformer,
- **What will happen if we use alignment in it?**



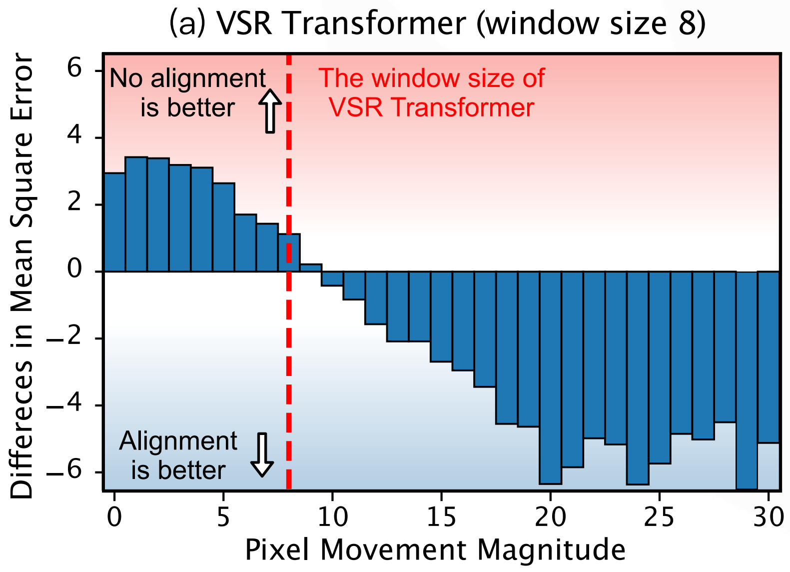
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

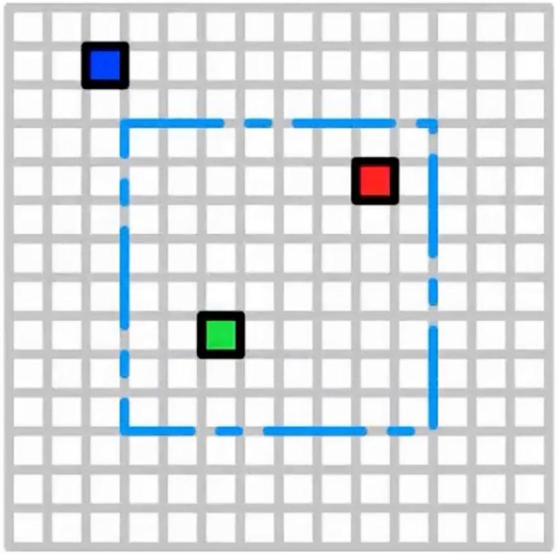
Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Does alignment benefit VSR Transformers?

Differences in pixel processing effects for different movement conditions.





Transformer with 8x8 attention window:

Only pixels inside the window can have direct interactions.

Can not process movement lager than the window size.

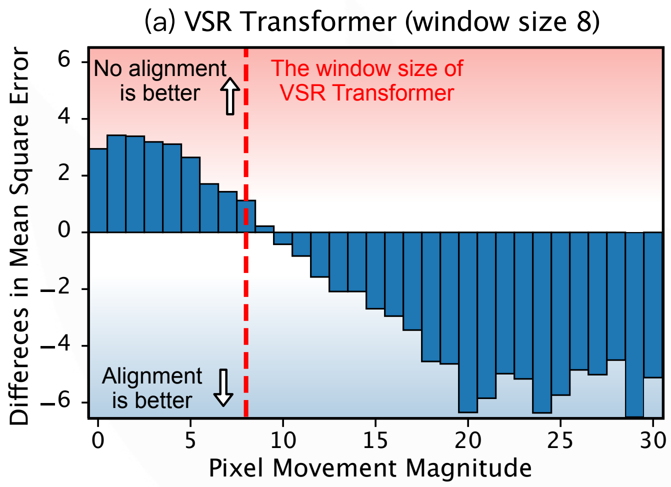
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

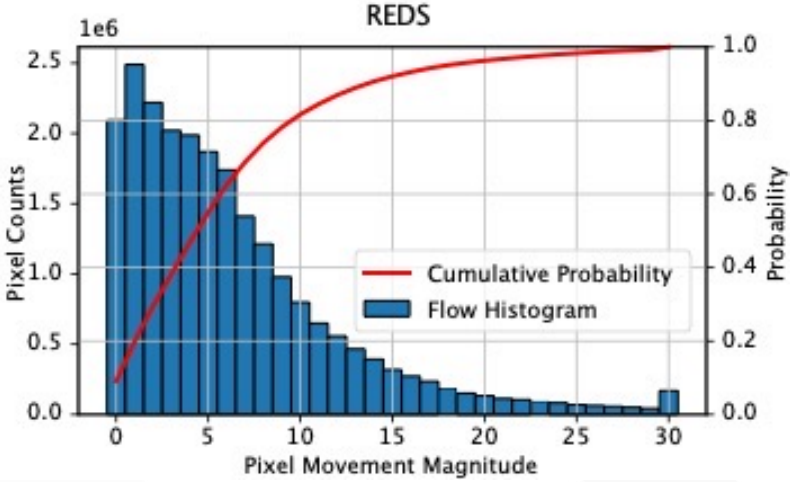
Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Does alignment benefit VSR Transformers?

Differences in pixel processing effects for different movement conditions.





Exp. Index	Method	Alignment	Remark	Vimeo90K-T PSNR	Vimeo90K-T SSIM	REDS4 PSNR	REDS4 SSIM
1	VSR-CNN	Image alignment	Finetune flow	36.13	0.9342	29.81	0.8541
2	VSR-CNN	No alignment		36.24	0.9359	28.95	0.8280
3	VSR Transformer	Image alignment	Fix flow	36.87	0.9429	30.25	0.8637
4	VSR Transformer	Image alignment	Finetune flow	37.44*	0.9472*	30.43	0.8677
5	VSR Transformer	Feature alignment	Finetune flow	37.36	0.9468	30.74	0.8740
6	VSR Transformer	No alignment	Window size 8	37.43	0.9470	30.56	0.8696
7	VSR Transformer	No alignment	Window size 16	37.46	0.9474	30.81	0.8745

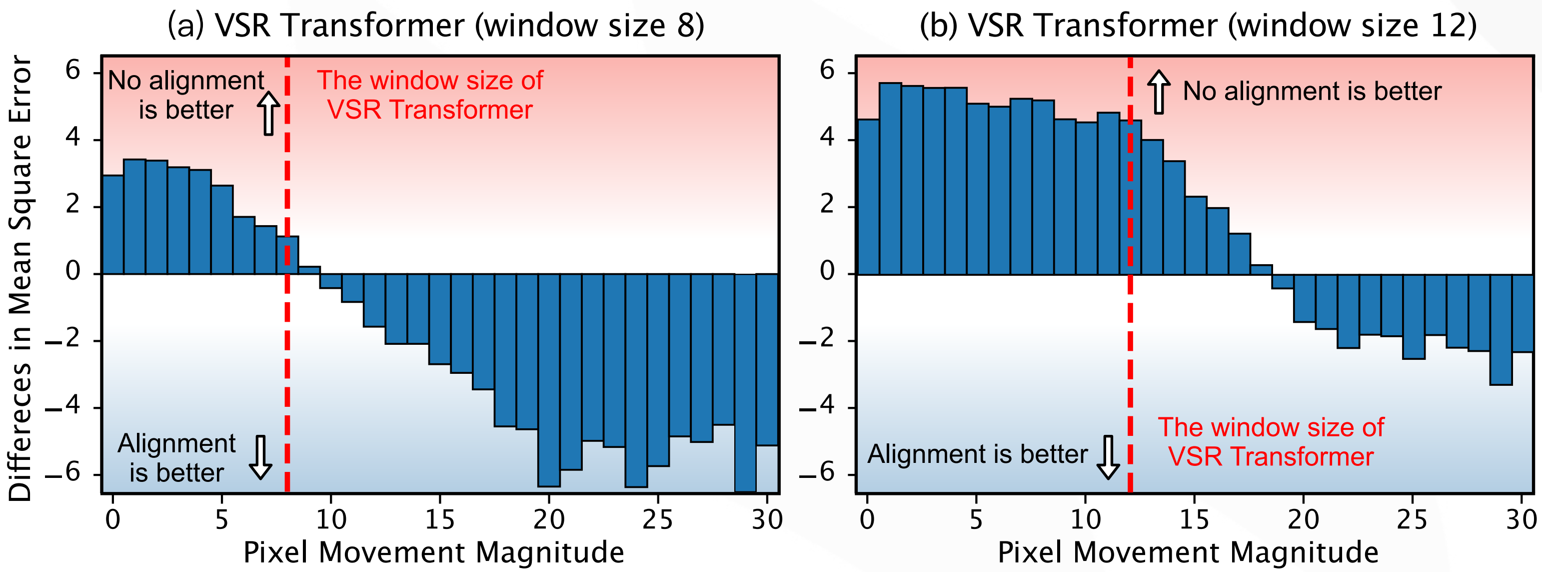
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Does alignment benefit VSR Transformers?

Differences in pixel processing effects for different movement conditions.

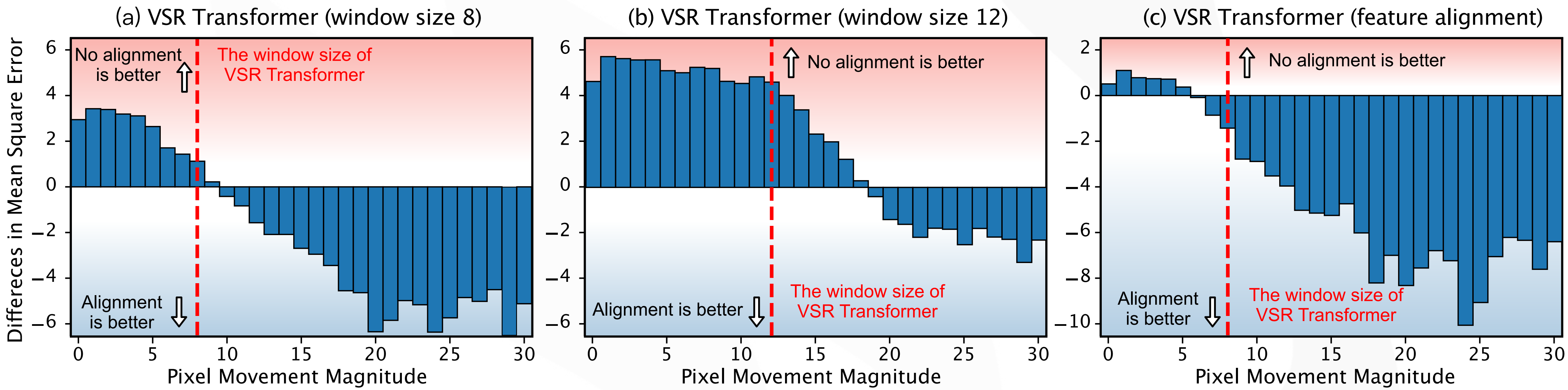


Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Does alignment benefit VSR Transformers?

Differences in pixel processing effects for different movement conditions.



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Does alignment benefit VSR Transformers?

Conclusions:

1. The VSR Transformer can handle misalignment within a certain range, and using alignment at this range will bring negative effects.
2. This range is closely related to the window size of the VSR Transformer.
3. Alignment is necessary for motions beyond the VSR Transformer's processing range.

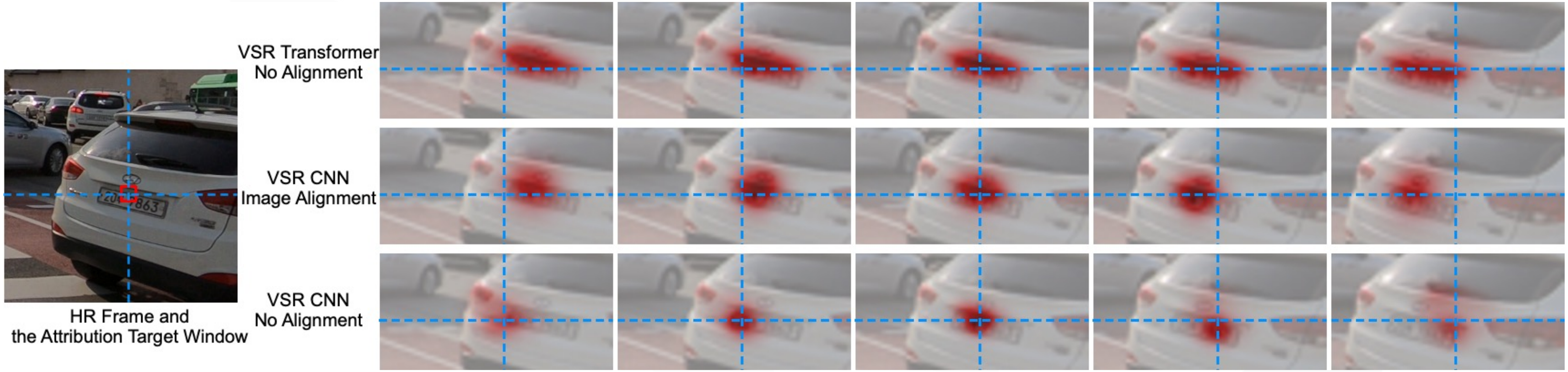
- **Do we still need alignment for VSR Transformers?**
- **To a certain extent, it is not necessary.**

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Does Transformer implicitly track the motion between unaligned frames?

Can an alignment-like function be done inside the VSR Transformers?



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

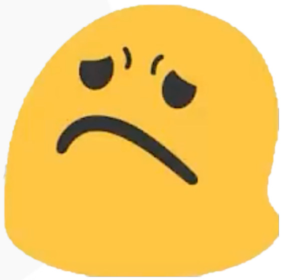
Rethinking

Question 1:

- The VSR model needs alignment because CNN has locality inductive bias.
- Transformers have no locality inductive bias.
- **Do we still need alignment for VSR Transformers?**

Question 2:

- If we do not need alignment in VSR Transformer,
- **What will happen if we use alignment in it?**



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Do alignment methods have negative effects? And Why?

Exp. Index	Method	Alignment	Remark	Vimeo90K-T PSNR	Vimeo90K-T SSIM	REDS4 PSNR	REDS4 SSIM
1	VSR-CNN	Image alignment	Finetune flow	36.13	0.9342	29.81	0.8541
2	VSR-CNN	No alignment		36.24	0.9359	28.95	0.8280
3	VSR Transformer	Image alignment	Fix flow	36.87	0.9429	30.25	0.8637
4	VSR Transformer	Image alignment	Finetune flow	37.44*	0.9472*	30.43	0.8677
5	VSR Transformer	Feature alignment	Finetune flow	37.36	0.9468	30.74	0.8740
6	VSR Transformer	No alignment	Window size 8	37.43	0.9470	30.56	0.8696
7	VSR Transformer	No alignment	Window size 16	37.46	0.9474	30.81	0.8745

Two Interesting Observation:

1. Optimizing the flow estimator during training will bring better results. Because the flow estimator at this time learns the optimized flow for VSR.

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Do alignment methods have negative effects? And Why?

Exp. Index	Method	Alignment	Remark	Vimeo90K-T PSNR	Vimeo90K-T SSIM	REDS4 PSNR	REDS4 SSIM
1	VSR-CNN	Image alignment	Finetune flow	36.13	0.9342	29.81	0.8541
2	VSR-CNN	No alignment		36.24	0.9359	28.95	0.8280
3	VSR Transformer	Image alignment	Fix flow	36.87	0.9429	30.25	0.8637
4	VSR Transformer	Image alignment	Finetune flow	37.44*	0.9472*	30.43	0.8677
5	VSR Transformer	Feature alignment	Finetune flow	37.36	0.9468	30.74	0.8740
6	VSR Transformer	No alignment	Window size 8	37.43	0.9470	30.56	0.8696
7	VSR Transformer	No alignment	Window size 16	37.46	0.9474	30.81	0.8745

Two Interesting Observation:

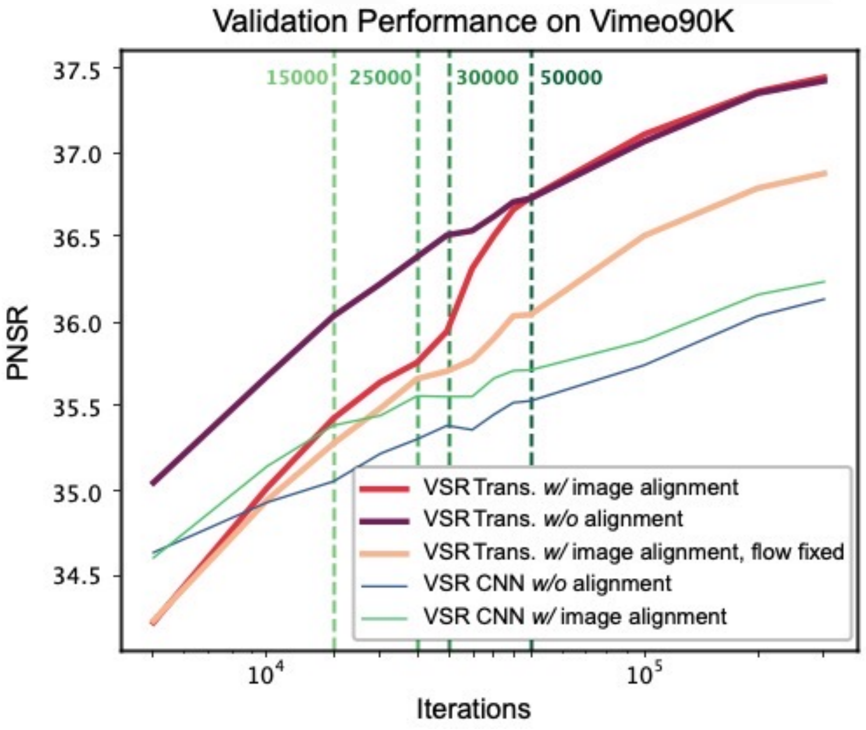
1. Optimizing the flow estimator during training will bring better results. Because the flow estimator at this time learns the optimized flow for VSR.
2. We observe different results on Vimeo-90K dataset: image-alignment with flow fine-tuning is almost identical to no alignment.

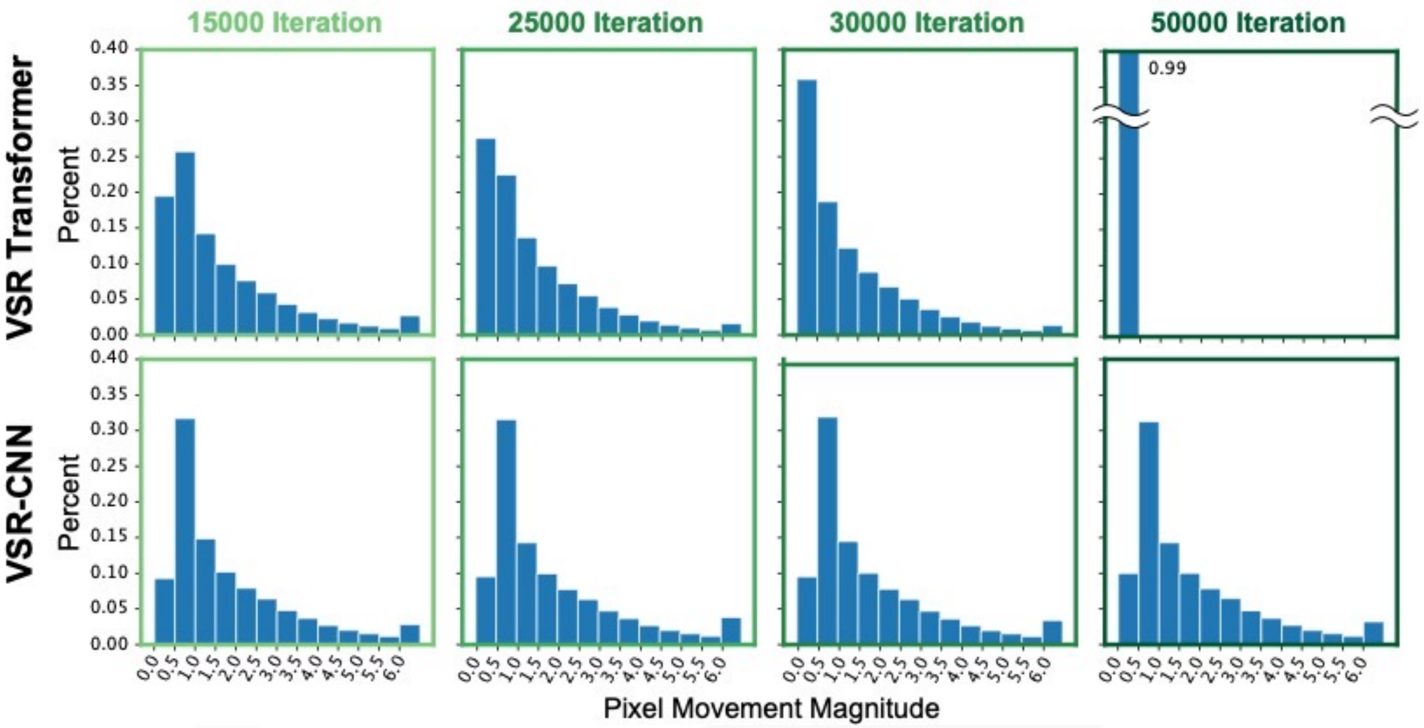
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Do alignment methods have negative effects? And Why?

We observe different results on Vimeo-90K dataset: image-alignment with flow fine-tuning is almost identical to no alignment.





Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Do alignment methods have negative effects? And Why?

At least two reasons:

1. The flow is noisy. And this noise introduces uncertainty to the mode between frames. And harm the performance.
2. The resampling operation also causes the sub-pixel information loss.

#	Alignment Method: No Ali.	Alignment Method: Img. Ali.	Alignment Method: Feat. Ali.	Alignment Method: FGDC	Position: Img.	Position: Feat.	Resampling: BI	Resampling: NN	Params. (M)	REDS4 PSNR / SSIM
1	✓								12.9	30.92 / 0.8759
2		✓			✓		✓		12.9	30.84 / 0.8752
3			✓			✓	✓		14.8	31.06 / 0.8792
4			✓			✓		✓	14.8	31.11 / 0.8801
5				✓		✓			16.1	31.11 / 0.8804

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

北京交通大学 BEIJING JIAOTONG UNIVERSITY

Do alignment methods have negative effects? And Why?

At least two reasons:

1. The flow is noisy. And this noise introduces uncertainty to the mode between frames. And harm the performance.
2. The resampling operation also causes the sub-pixel information loss.

#	Alignment Method				Position		Resampling		Params.	REDS4
	No Ali.	Img. Ali.	Feat. Ali.	FGDC	Img.	Feat.	BI	NN	(M)	PSNR / SSIM
1	✓								12.9	30.92 / 0.8759
2		✓			✓		✓		12.9	30.84 / 0.8752
3			✓			✓	✓		14.8	31.06 / 0.8792
4			✓			✓		✓	14.8	31.11 / 0.8801
5				✓		✓			16.1	31.11 / 0.8804

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

北京交通大学
BEIJING JIAOTONG UNIVERSITY

Does alignment benefit VSR Transformers?

Conclusions:

1. The VSR Transformer can handle misalignment within a certain range, and using alignment at this range will bring negative effects.
2. This range is closely related to the window size of the VSR Transformer.
3. Alignment is necessary for motions beyond the VSR Transformer’s processing range.

Why alignment hurts VSR Transformer?

1. Inaccurate flow
2. Resampling Operation

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

How to do better?

We want better Transformer:

1. Increasing the Transformer's window size (Too expensive)
2. A new alignment method.

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

How to do better?

We want better Transformer:

1. Increasing the Transformer's window size (Too expensive)
2. A new alignment method.

We propose Patch Alignment, that:

1. Only rely on approximate flow information, ignoring flow inaccuracies.
2. Cut and move the target position as a whole without changing the relative relationship between pixels.

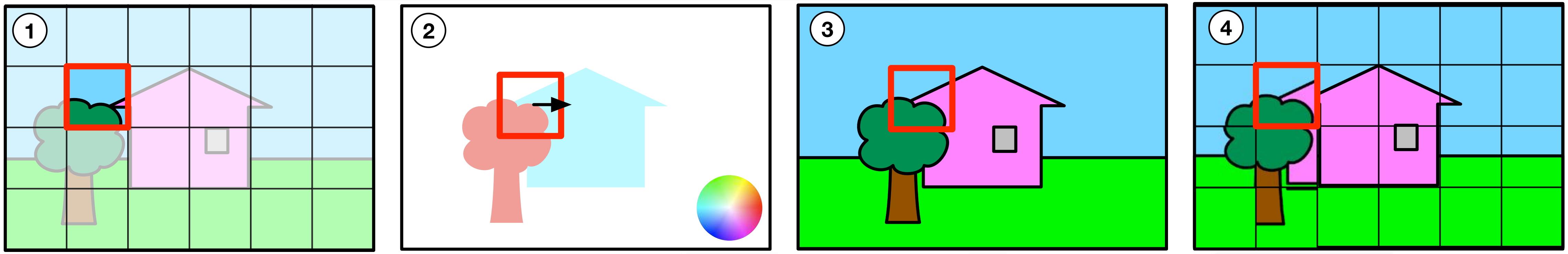
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Patch Alignment

We propose Patch Alignment, that:

1. Only rely on approximate flow information, ignoring flow inaccuracies.
2. Cut and move the target position as a whole without changing the relative relationship between pixels.



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Patch Alignment

We propose Patch Alignment, that:

1. Only rely on approximate flow information, ignoring flow inaccuracies.
2. Cut and move the target position as a whole without changing the relative relationship between pixels.

Reference Frame   Image Alignment   Patch Alignment

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Experimental Results

Compare to other alignment methods:

#	Alignment Method: No Ali.	Alignment Method: Img. Ali.	Alignment Method: Feat. Ali.	Alignment Method: FGDC	Position: Img.	Position: Feat.	Resampling: BI	Resampling: NN	Params. (M)	REDS4 PSNR / SSIM
1	✓								12.9	30.92 / 0.8759
2		✓			✓		✓		12.9	30.84 / 0.8752
3			✓			✓	✓		14.8	31.06 / 0.8792
4			✓			✓		✓	14.8	31.11 / 0.8801
5				✓		✓			16.1	31.11 / 0.8804

Method	Position: Img.	Position: Feat.	Resampling: BI	Resampling: NN	REDS4 PSNR	REDS4 SSIM
Patch Alignment	✓			✓	31.11	0.8800
Patch Alignment		✓	✓		31.00	0.8781
Patch Alignment		✓		✓	31.17	0.8810

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Experimental Results

Compare to state-of-the-art:

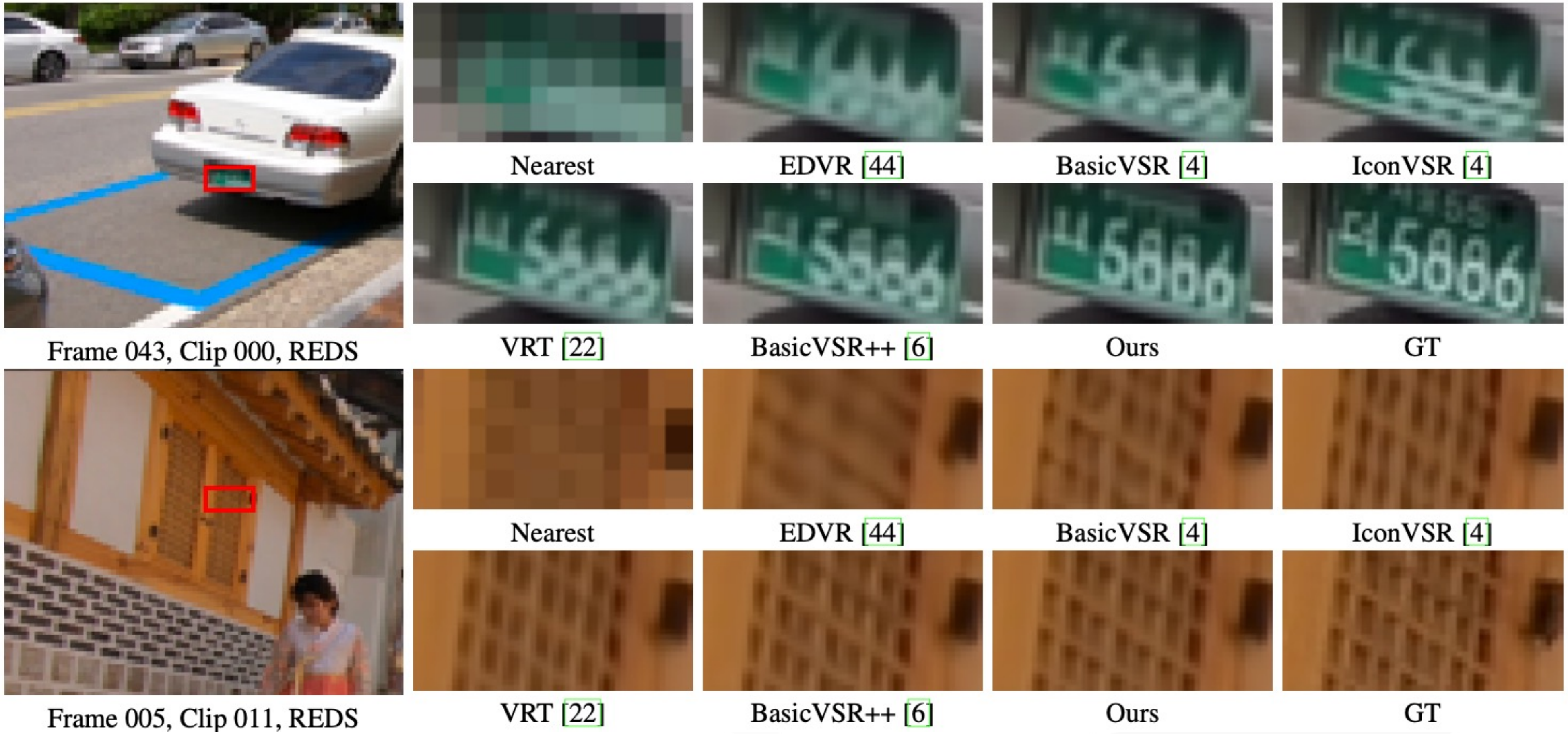
Method	Frames REDS/Vimeo	Params (M)	REDS4 PSNR	REDS4 SSIM	Vimeo-90K-T PSNR	Vimeo-90K-T SSIM	Vid4 PSNR	Vid4 SSIM
Bicubic	-/-	-	26.14	0.7292	31.32	0.8684	23.78	0.6347
RCAN	-/-	-	28.78	0.8200	35.35	0.9251	25.46	0.7395
SwinIR	-/-	11.9	29.05	0.8269	35.67	0.9287	25.68	0.7491
TOFlow	5/7	-	27.98	0.7990	33.08	0.9054	25.89	0.7651
DUF	7/7	5.8	28.63	0.8251	-	-	27.33	0.8319
PFNL	7/7	3.0	29.63	0.8502	36.14	0.9363	26.73	0.8029
RBPN	7/7	12.2	30.09	0.8590	37.07	0.9435	27.12	0.8180
EDVR	5/7	20.6	31.09	0.8800	37.61	0.9489	27.35	0.8264
MuCAN	5/7	-	30.88	0.8750	37.32	0.9465	-	-
VSR-T	5/7	32.6	31.19	0.8815	37.71	0.9494	27.36	0.8258
PSRT-sliding	5/-	14.8	31.32	0.8834	-	-	-	-
VRT	6/-	30.7	31.60	0.8888	-	-	-	-
PSRT-recurrent	6/-	10.8	31.88	0.8964	-	-	-	-
BasicVSR	15/14	6.3	31.42	0.8909	37.18	0.9450	27.24	0.8251
IconVSR	15/14	8.7	31.67	0.8948	37.47	0.9476	27.39	0.8279
BasicVSR++	30/14	7.3	32.39	0.9069	37.79	0.9500	27.79	0.8400
VRT	16/7	35.6	32.19	0.9006	38.20	0.9530	27.93	0.8425
RVRT	30/14	10.8	32.75	0.9113	38.15	0.9527	27.99	0.8462
PSRT-recurrent	16/14	13.4	32.72	0.9106	38.27	0.9536	28.07	0.8485

Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

Experimental Results

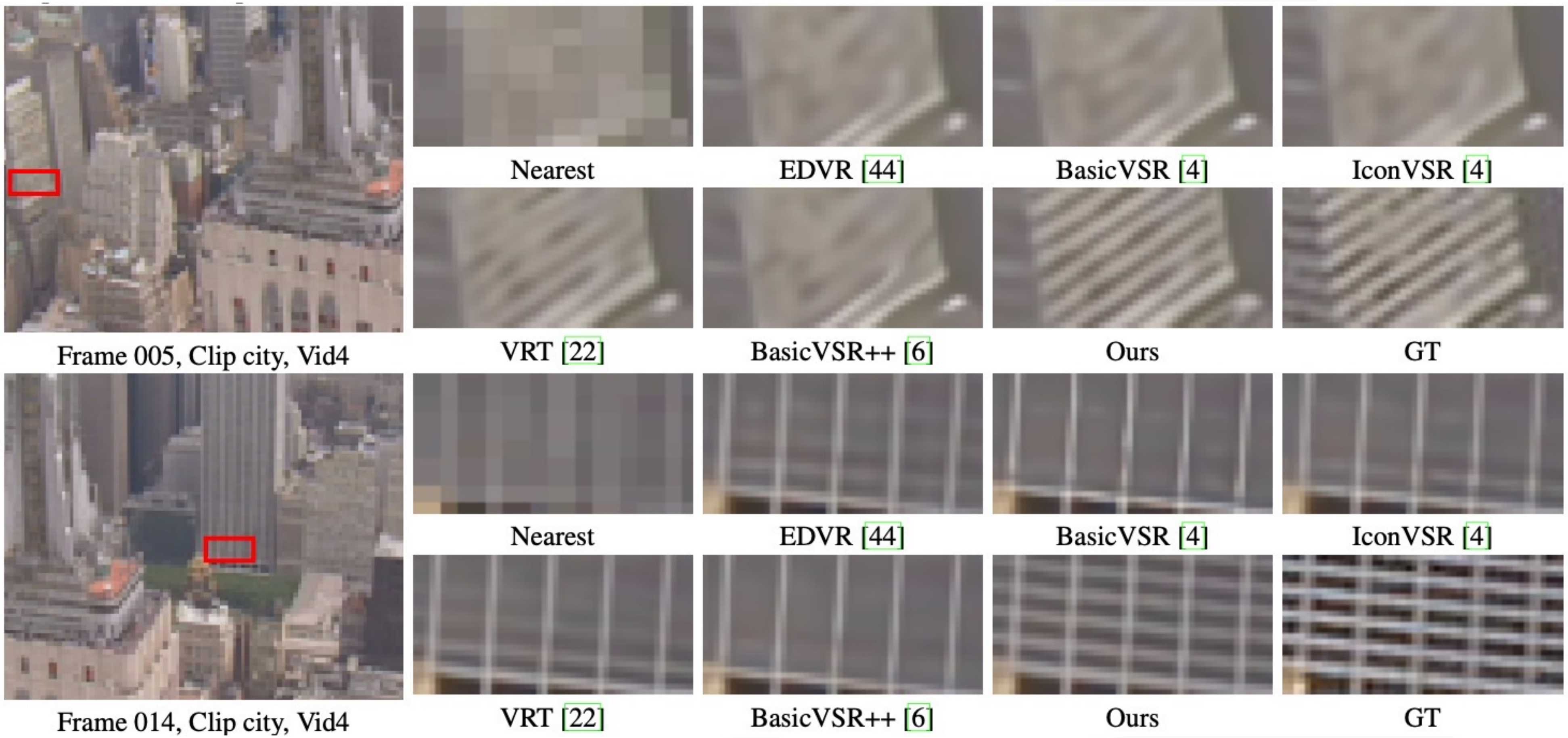
Compare to state-of-the-art:



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Experimental Results

Compare to state-of-the-art:



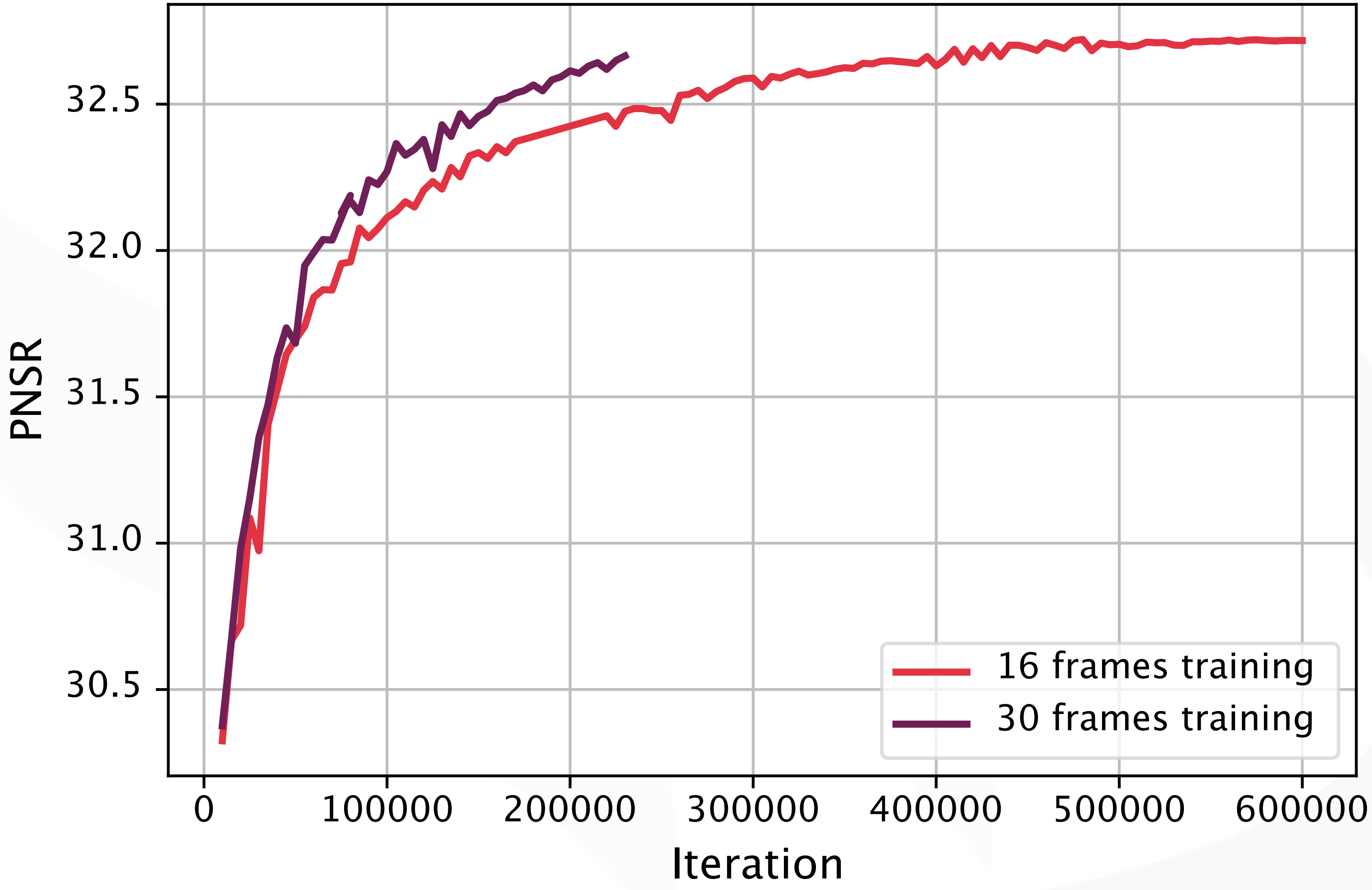
Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Alignment: Which method benefit to VSR Transformer?

BEIJING JIAOTONG UNIVERSITY

Experimental Results

Compare to state-of-the-art:



Shuwei Shi, Jinjin Gu, Liangbin Xie, Xintao Wang, Yujiu Yang and Chao Dong. 2022. Rethinking Alignment in Video Super-Resolution Transformers. In Advances in Neural Information Processing Systems.

Thank you