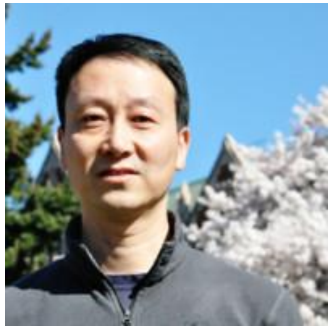
Finding *Discriminative Filters* for Specific Degradations in Blind Super-Resolution







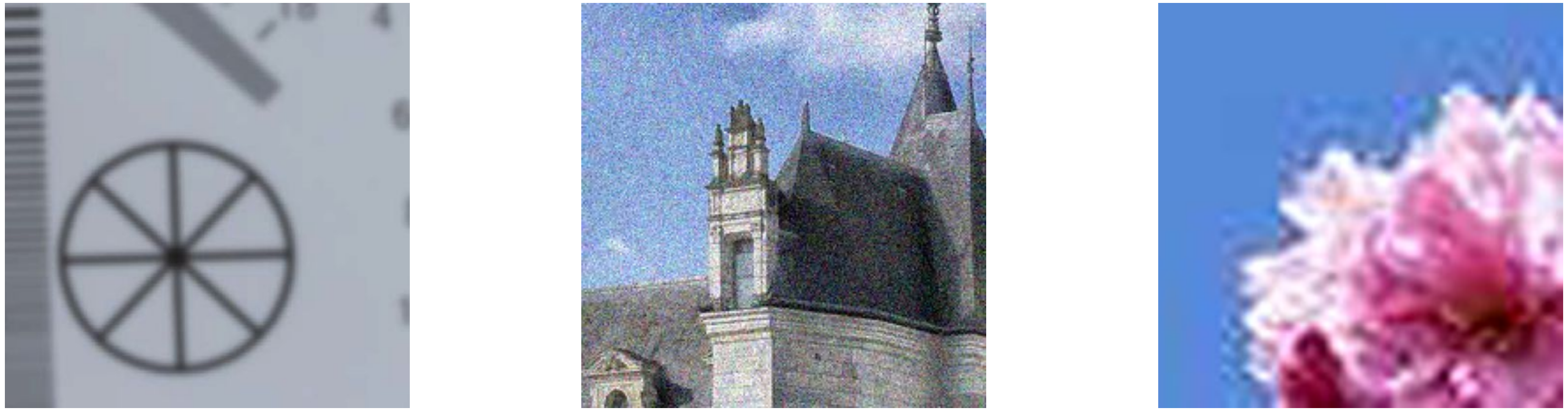
Liangbin Xie*[1,2] Xintao Wang*[2] Chao Dong[1] Zhongang Qi[2] Ying Shan[2]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]Applied Research Center (ARC), Tencent PCG

* Equal contributions

Background – Blind Super-Resolution

- Reconstruct a high-resolution image from its low-resolution counterpart which contains ***unknown and complex degradations***, for example:

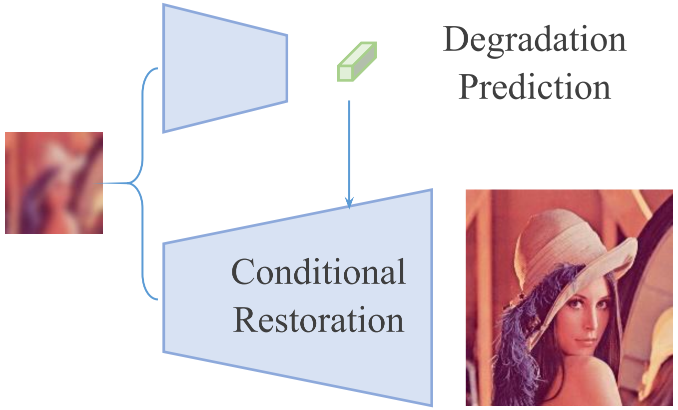


blur

noise

JPEG Compression

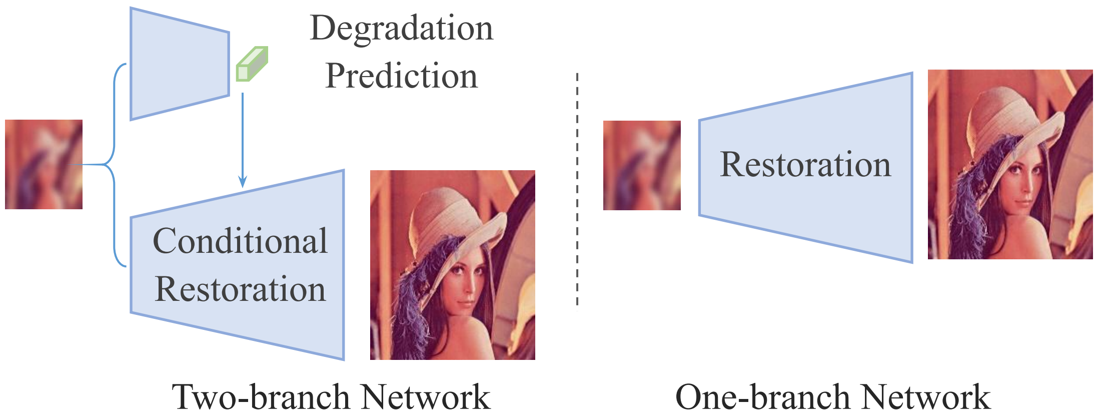
- Blind SR methods typically consist of two branches
  - one for degradation prediction
  - the other for conditional restorations

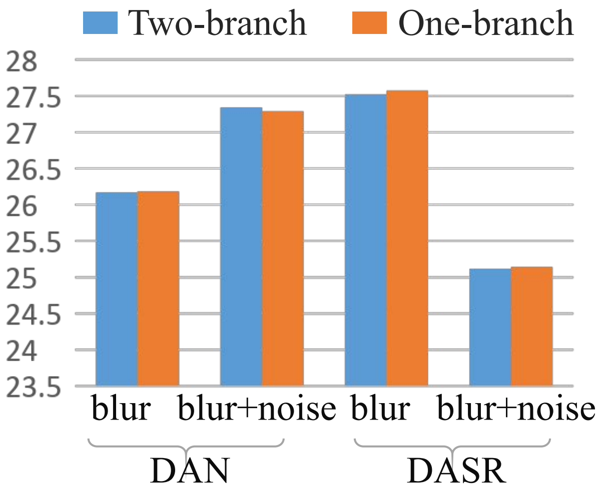


Two-branch Network

Wonder: How about a unified one-branch network?

- We conduct preliminary experiments on several state-of-the-art methods: DAN and DASR.





PSNR (dB)	DAN [3]		DASR [5]	
	blur	blur+noise	blur	blur+noise
Official two-branch	26.168±0.009	27.341±0.072	27.518±0.034	25.116±0.012
SRResNet one-branch	26.182±0.011	27.288±0.027	27.573±0.010	25.143±0.013

- A unified one-branch network could achieve **comparable** performance under **similar computation budgets** for state-of-the-art blind SR methods

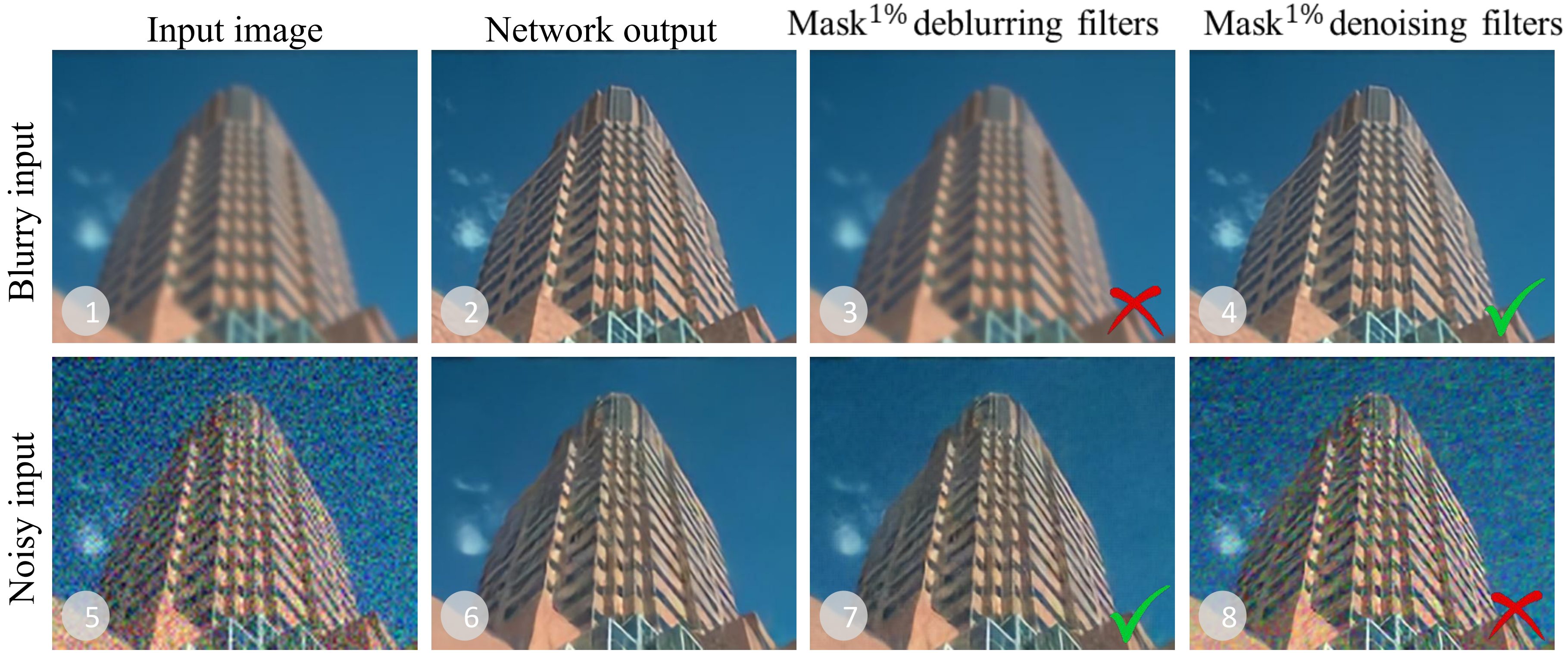
Motivations

- Find connections between:
  - Two-branch network - delicate designs with higher interpretability
  - One-branch network - more like a ‘black-box’

- Two key questions:
  - Could one-branch networks automatically learn to distinguish degradations as what we specially design in two-branch methods?
  - Are there any small sub-network (*i.e.*, a set of filters) existing for a specific degradation?

Our Findings

1. In one-branch blind SR networks, we are able to find a very small number of (*at least to 1%*) discriminative filters for each specific degradation (*e.g.*, blur, noise).



Our Findings

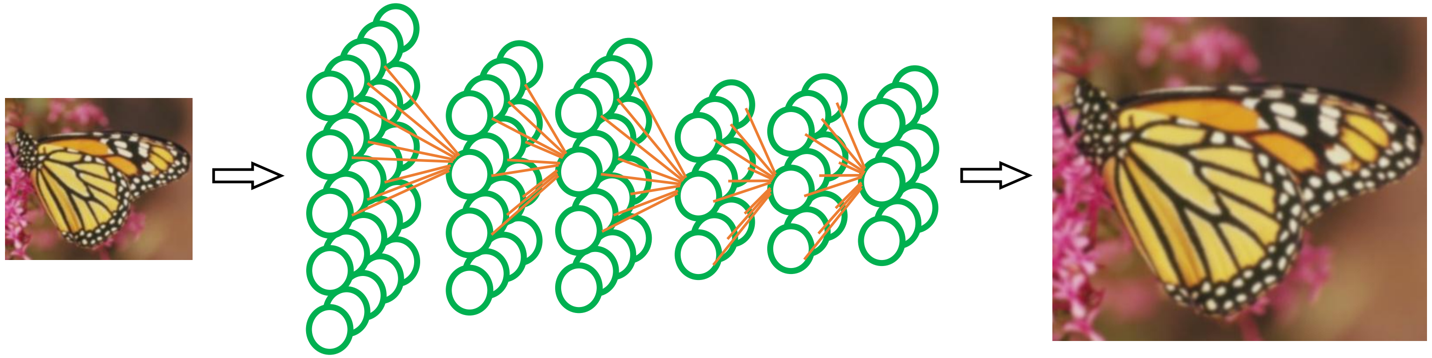
2. Expect for the *filter weights*, **the locations and connections** of the discovered filters also play an essential role in the network function for a specific degradation

3. Based on these discovered filters, we could easily **predict the degradation** of input images without training in the supervision of degradation labels

a unified one-branch network is similar to a well-designed two-branch network in the working mechanism

Methods – Filter Attribution Integrated Gradients (FAIG)

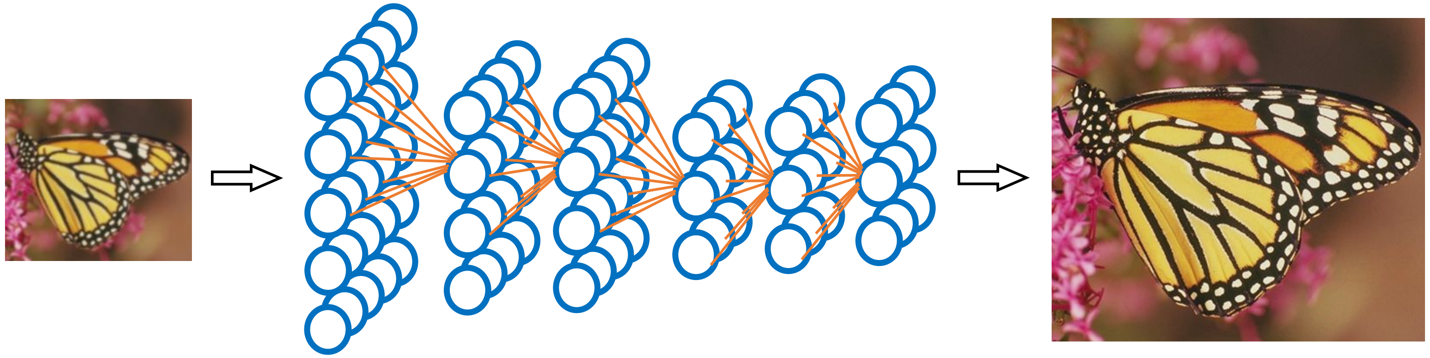
Baseline model



cannot remove blur

The ***baseline network*** $\boldsymbol{F}(\overline{\boldsymbol{\theta}})$ is a pure SR network that cannot remove any degradations.

Target model



can remove blur

The ***target network*** $\boldsymbol{F}(\boldsymbol{\theta})$ is a re-trained network that can deal with complex degradations.

Key idea: Given the same input, the changes of the network output can be attributed to the changes of network parameters (*i.e.*, filters).

Methods –

We need to find important filters in a network that make the greatest contribution to the function of a specific degradation removal

One may consider to directly calculate the absolute value changes of parameters to determine the most important filters for the network functional alteration.

- ✗ The filter changes may result in other dimensions of transition, e.g., color changes
- ✗ Cannot discover different filter sets for various degradations

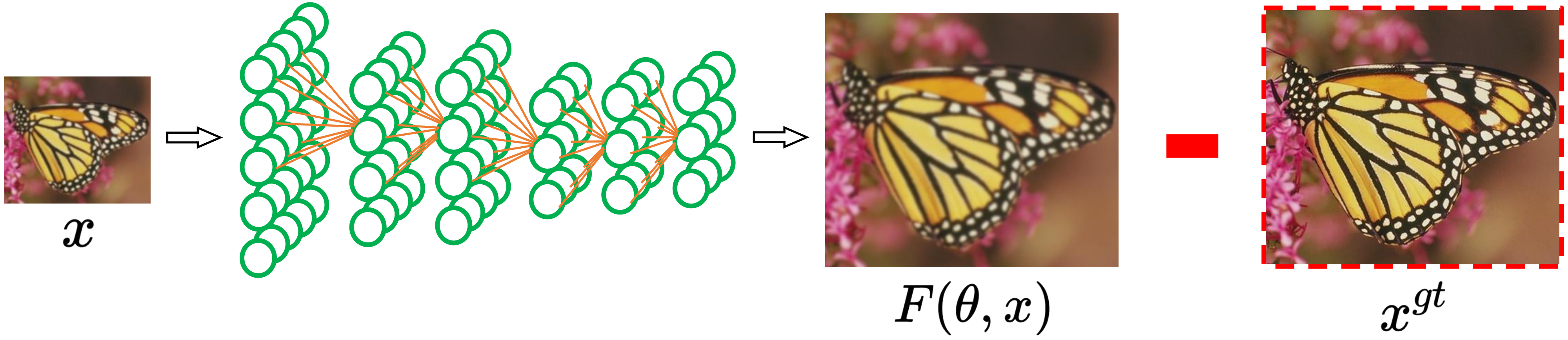
Methods –

We propose to calculate the gradient of each filter

Gradient – The fastest changing direction

We first quantify the network function of degradation removal by

$$\mathcal{L}(\theta, x) = \|F(\theta, x) - x^{gt}\|_2^2$$



Methods –

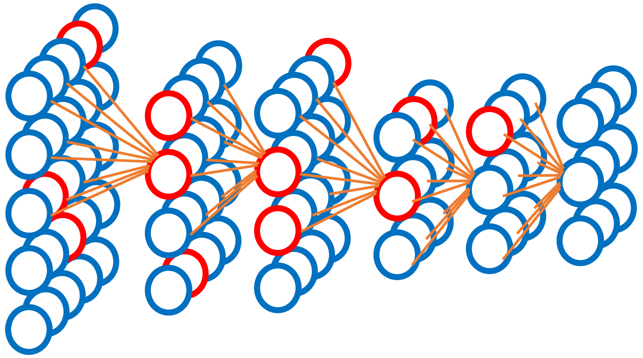
For an input image x , we then attribute the changes of network functions to the changes of filters.

Motivated by Integrated Gradient(IG) that accumulates the gradients at all points along a straight-line path, we accumulate gradients along a path

$$\gamma(\alpha)=\bar{\theta}+\alpha\times(\theta-\bar{\theta})$$

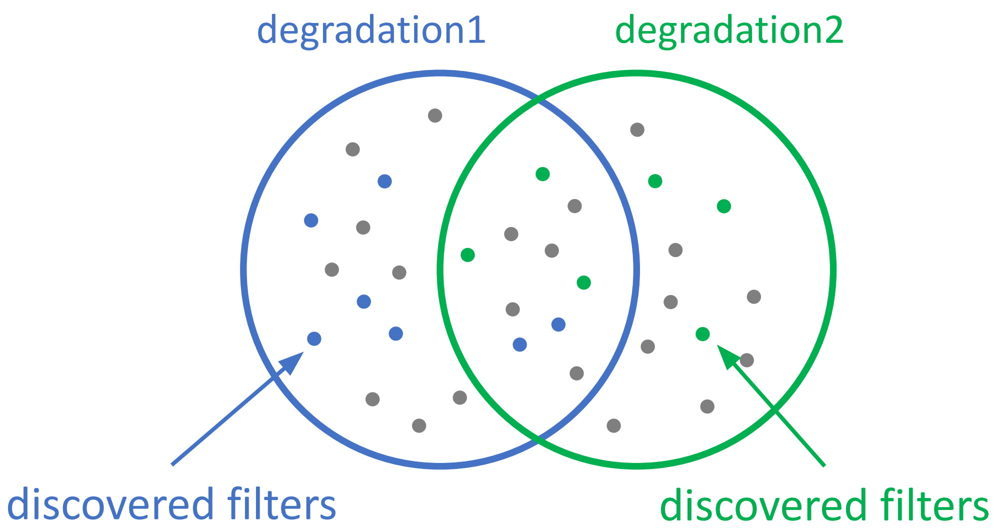
$$\mathtt{FAIG}_i(\theta, x) = \int_{\alpha=0}^{1} \frac{\partial \mathcal{L}(\gamma(\alpha), x)}{\partial \gamma(\alpha)_i} \times \frac{\gamma(\alpha)_i}{\partial \alpha} d\alpha$$

The filters with highest gradient are the most discriminative filters

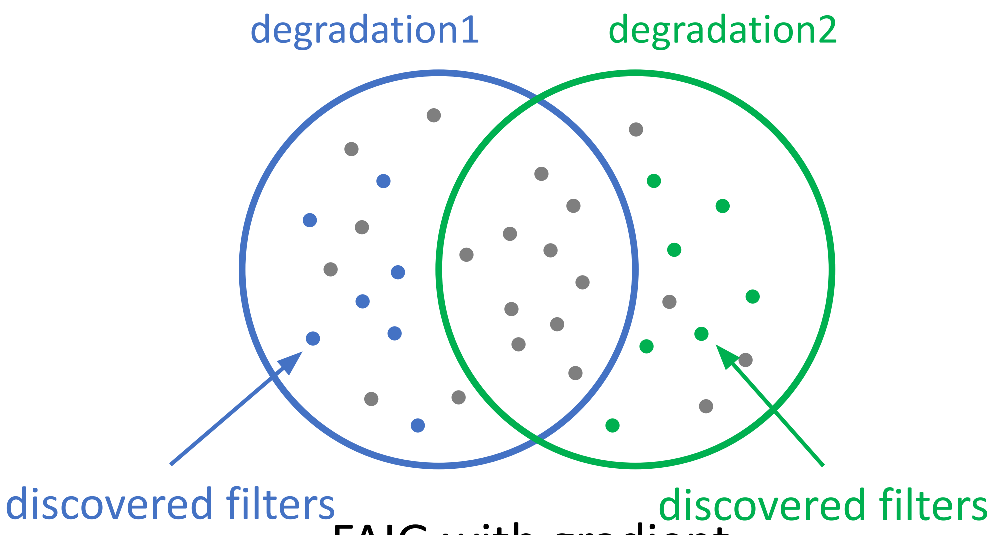


Output changes

Methods– FAIG with gradient elimination



FAIG



FAIG with gradient elimination

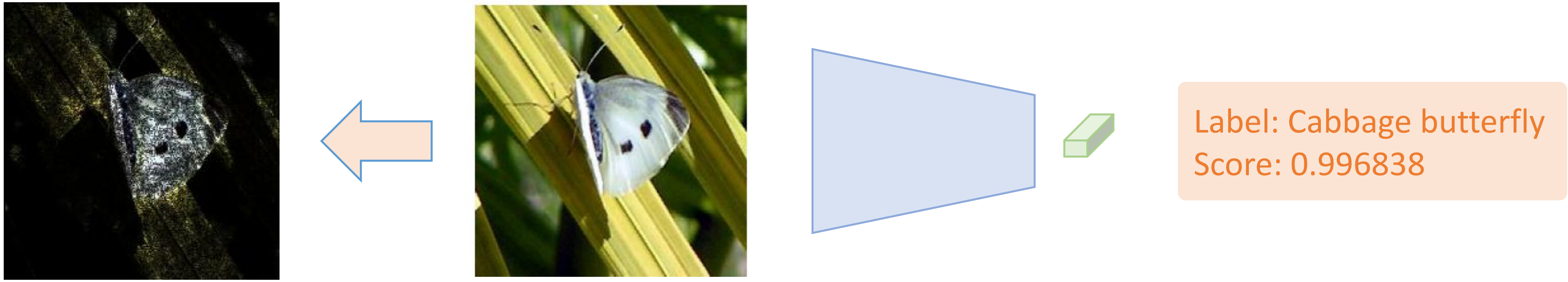
The discovered filters do not guarantee to be only responsible for the specific degradation

We suppress the gradients to other degradations to eliminate their influence while retaining the interested gradients

$$\mathtt{FAIG}_i^{\mathcal{D}}(\theta) = \frac{1}{\|\mathcal{X}\|}\Big(\underbrace{\sum_{x\in\mathcal{X}} |\mathtt{FAIG}_i(\theta, x^{\mathcal{D}})|}_{\text{attribution for degradtion}\mathcal{D}} - \underbrace{\sum_{x\in\mathcal{X}} |\mathtt{FAIG}_i(\theta, x^{\sim\mathcal{D}})|}_{\text{attribution for other degradtions}}\Big)$$

Integrated Gradient in high-level

In classification task, *Integrated Gradient (IG)* is used to attributes the most important input components (*e.g.*, pixels in input images) that affect the network predictions.



The difference between IG and FAIG

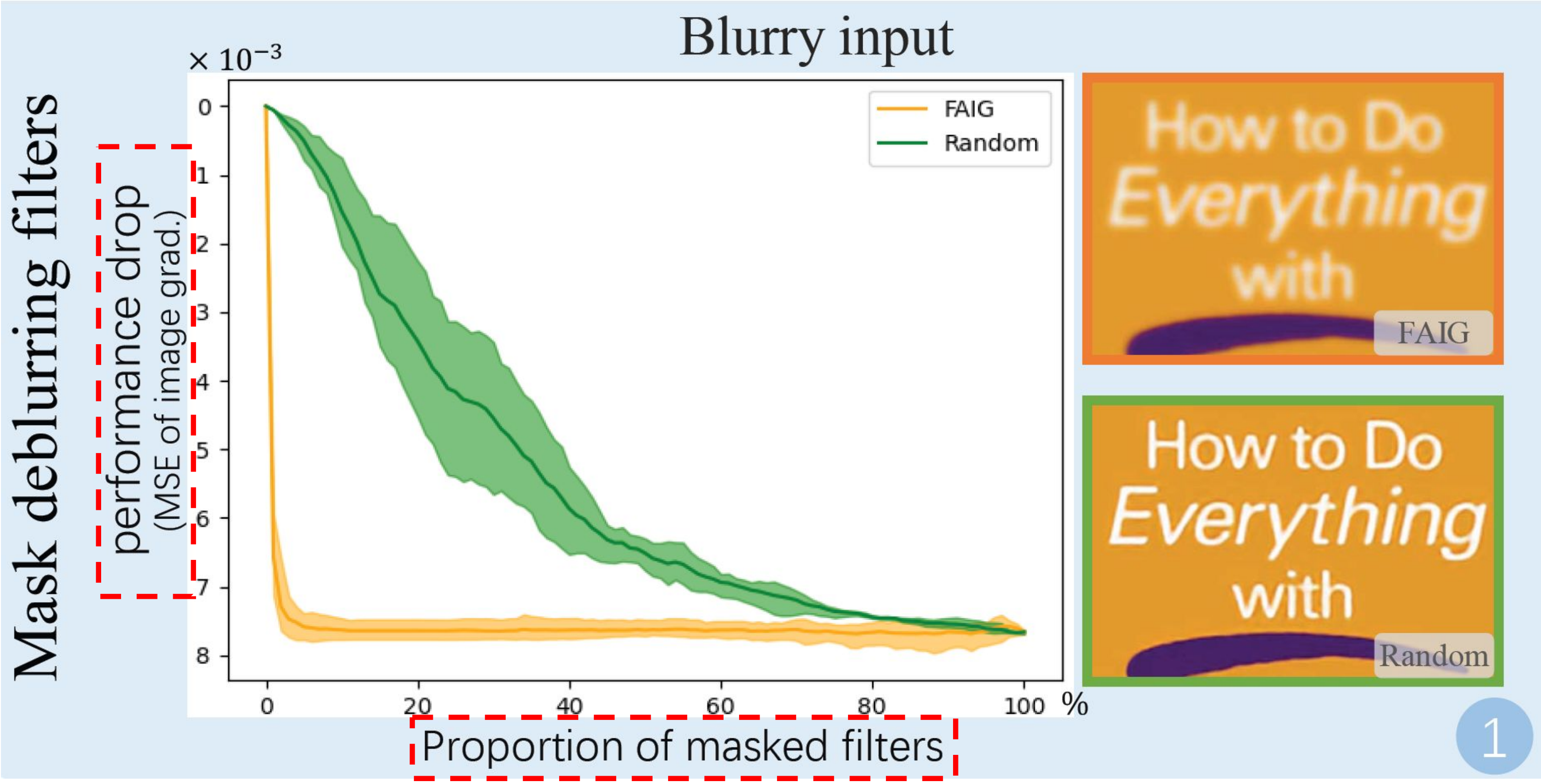
Method	Filter Attribution Integrated Gradients (FAIG)	Integrated Gradient
Task	Restoration (blind SR)	High-level (classification)
Purpose	Fine core filters that explain **degradation removal**	Find input pixels that explain **network prediction**
Attribute to	network parameters (**filters**)	Input **pixels**
Integral path	Parameter space	Input space

Experiments

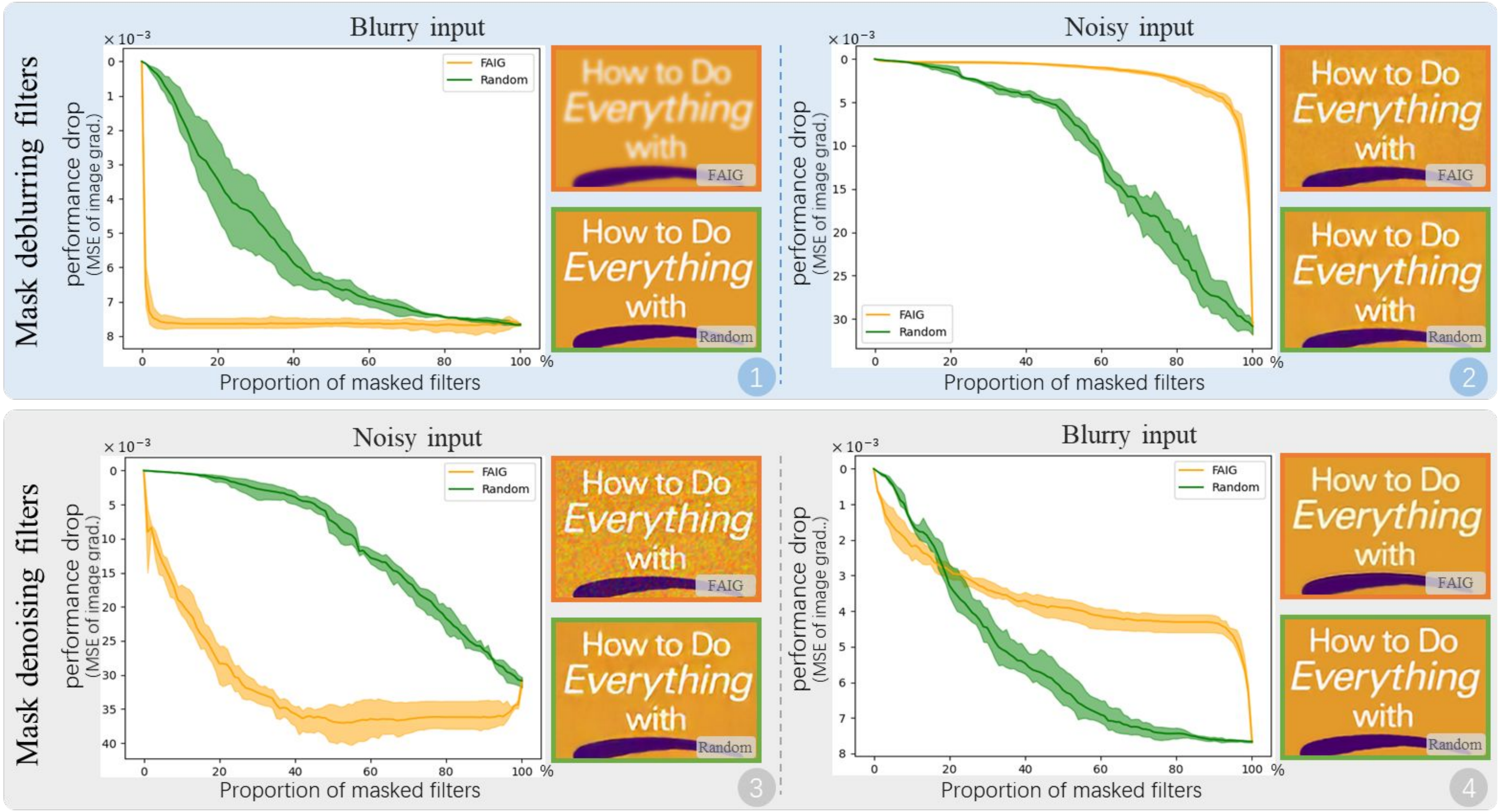
- Masking Discovered Filters
  - We measure the importance of discovered filters by replacing them with the filters in the baseline model (at the same location)
- Re-training only Discovered Filters
  - We investigate whether the locations and connections of the discovered filters are important
- Distribution of Discovered Filters in a Network
- Degradation Classification

Experiments — Masking Discovered Filters

We measure the importance of discovered filters by replacing them with the filters in the baseline model (at the same locations).



Experiments — Masking Discovered Filters



FAIG VS Random

Experiments — Masking Discovered Filters

Compare the performance drop with other methods.

(Larger values indicates a large performance drop)

(10^{-3}) Input	mask 1% discovered filters				mask 5% discovered filters			
	FAIG (ours)	IG	$\lvert\theta - \bar{\theta}\rvert$	Random	FAIG (ours)	IG	$\lvert\theta - \bar{\theta}\rvert$	Random
Blurry image	**6.68**$\pm$0.63	4.31$\pm$1.54	0.18$\pm$0.13	0.07$\pm$0.01	**7.53**$\pm$0.24	6.41$\pm$0.88	2.16$\pm$0.61	0.55$\pm$0.32
Noisy image	**6.62**$\pm$0.54	4.22$\pm$0.44	0.49$\pm$0.10	0.04$\pm$0.01	**16.28**$\pm$3.84	8.01$\pm$1.04	3.25$\pm$1.85	0.19$\pm$0.05

SRResNet Network (on Set14 dataset)

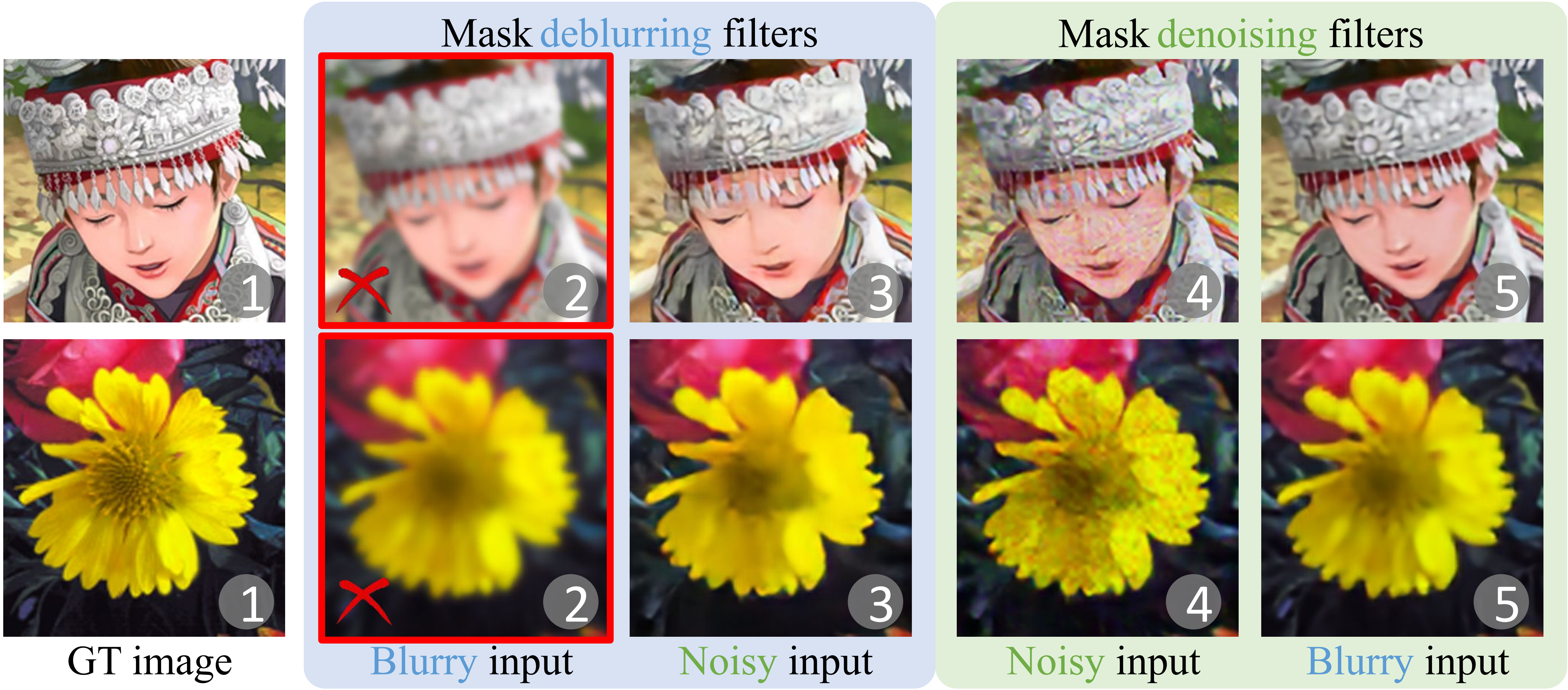
(10^{-3}) Input	mask 1% discovered filters				mask 5% discovered filters			
	FAIG (ours)	IG	$\lvert\theta - \bar{\theta}\rvert$	Random	FAIG (ours)	IG	$\lvert\theta - \bar{\theta}\rvert$	Random
Blurry image	**5.02**$\pm$0.36	0.74$\pm$0.42	0.63$\pm$0.65	0.05$\pm$0.02	**5.74**$\pm$0.29	2.53$\pm$1.02	2.18$\pm$0.25	0.45$\pm$0.07
Noisy image	**1.40**$\pm$0.37	0.28$\pm$0.11	0.63$\pm$0.35	0.02$\pm$0.00	**4.86**$\pm$0.90	0.93$\pm$1.05	2.01$\pm$0.12	0.11$\pm$0.01

SRCNN-style Network (on BSDS100 dataset)

FAIG can discover more important filters for corresponding degradations.

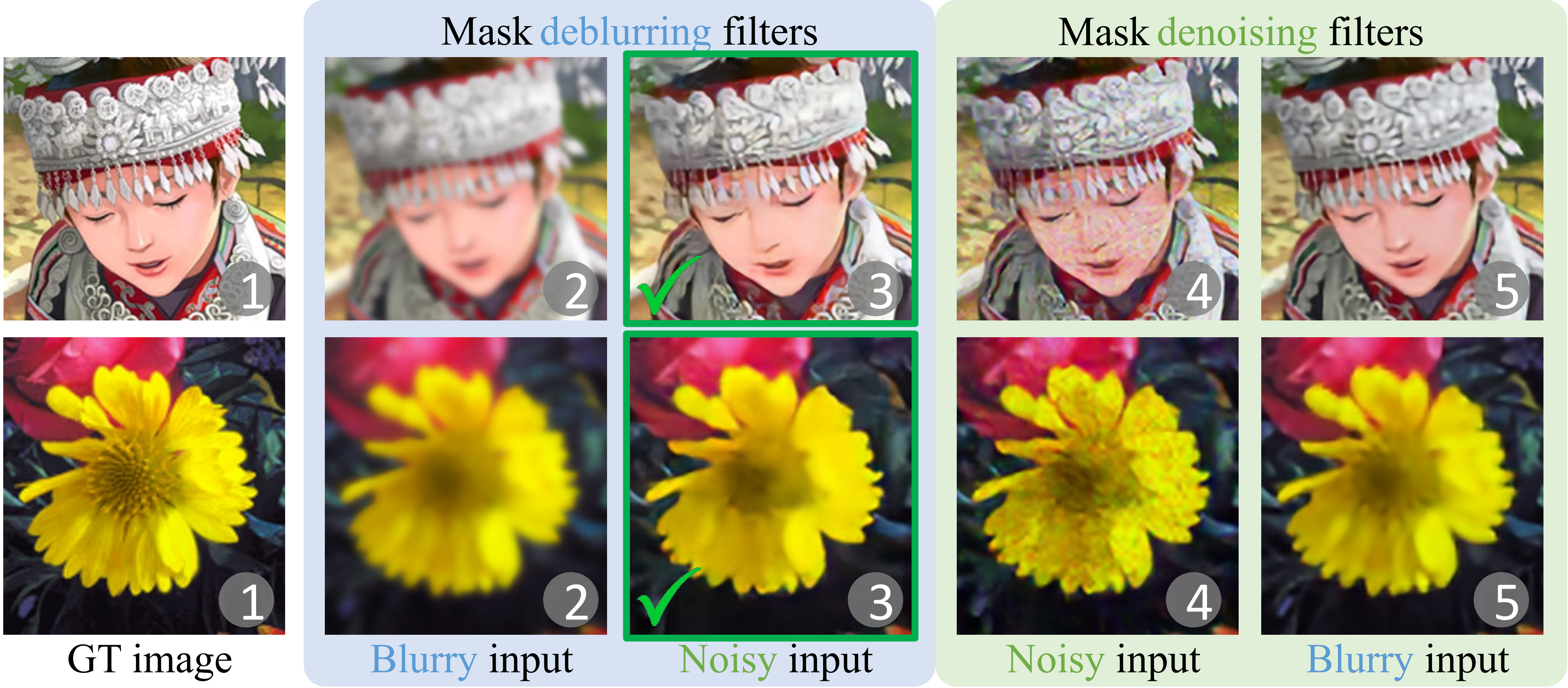
Experiments — Masking Discovered Filters

More qualitative results – Mask 1% filters



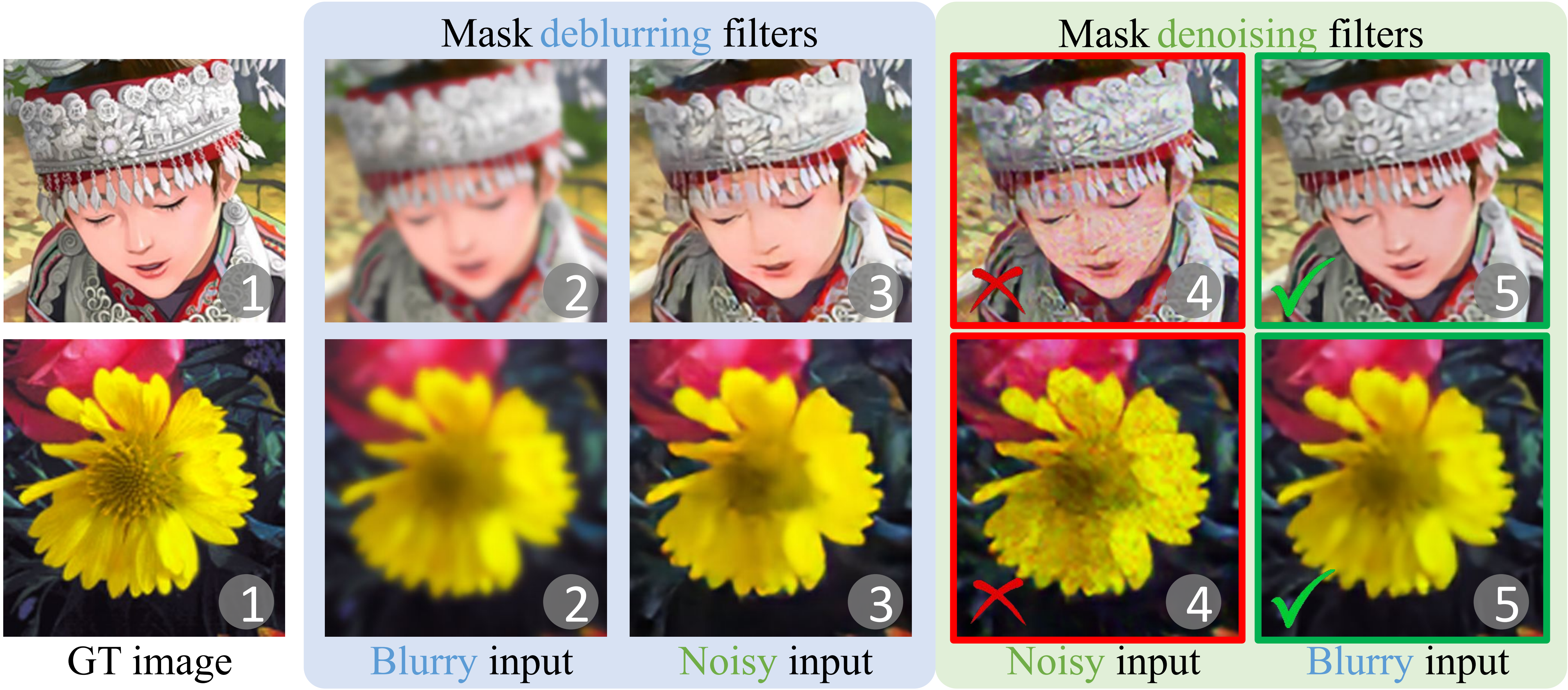
Experiments — Masking Discovered Filters

More qualitative results – Mask 1% filters



Experiments — Masking Discovered Filters

More qualitative results – Mask 1% filters

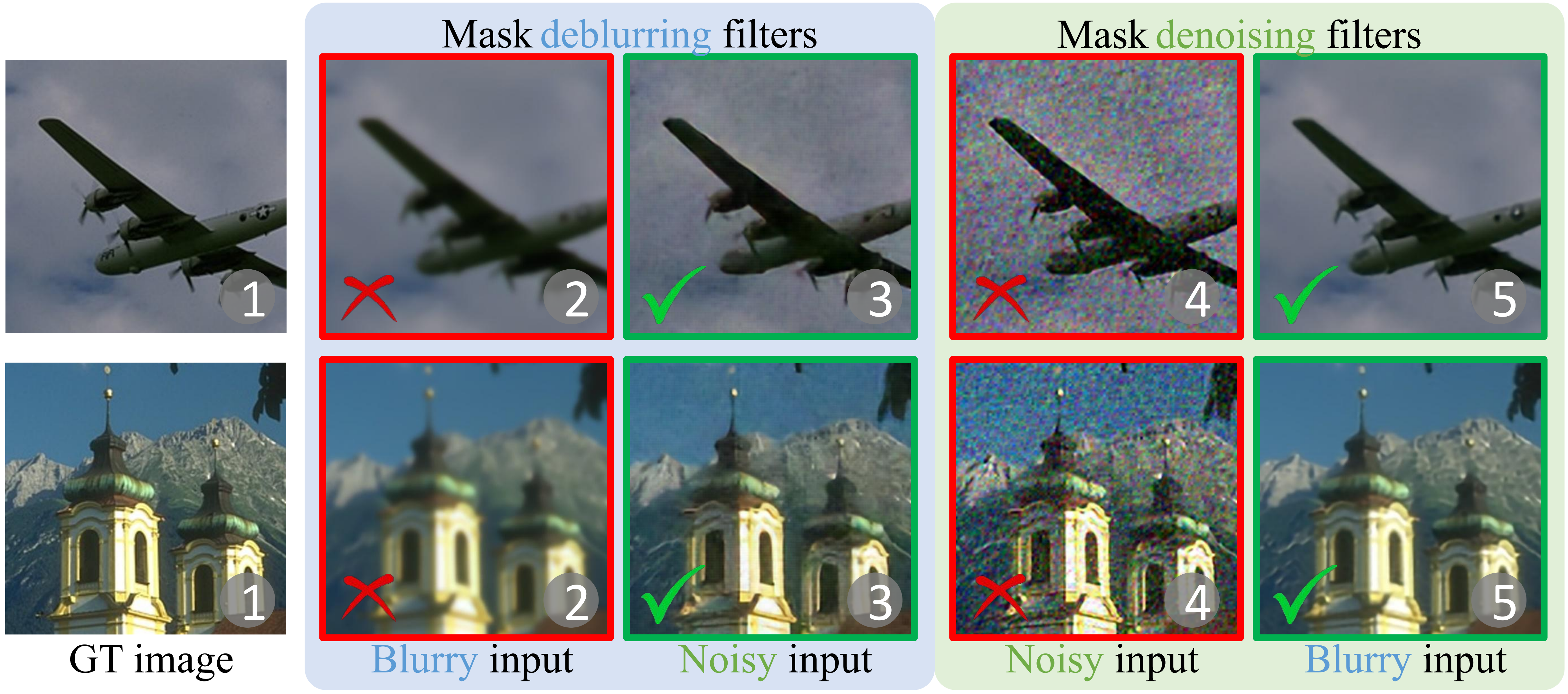


Experiments — Masking Discovered Filters

More qualitative results – Mask 5% filters

Experiments — Masking Discovered Filters

More qualitative results – Mask 10% filters



Experiments – Retrain only Discovered Filters

Steps:

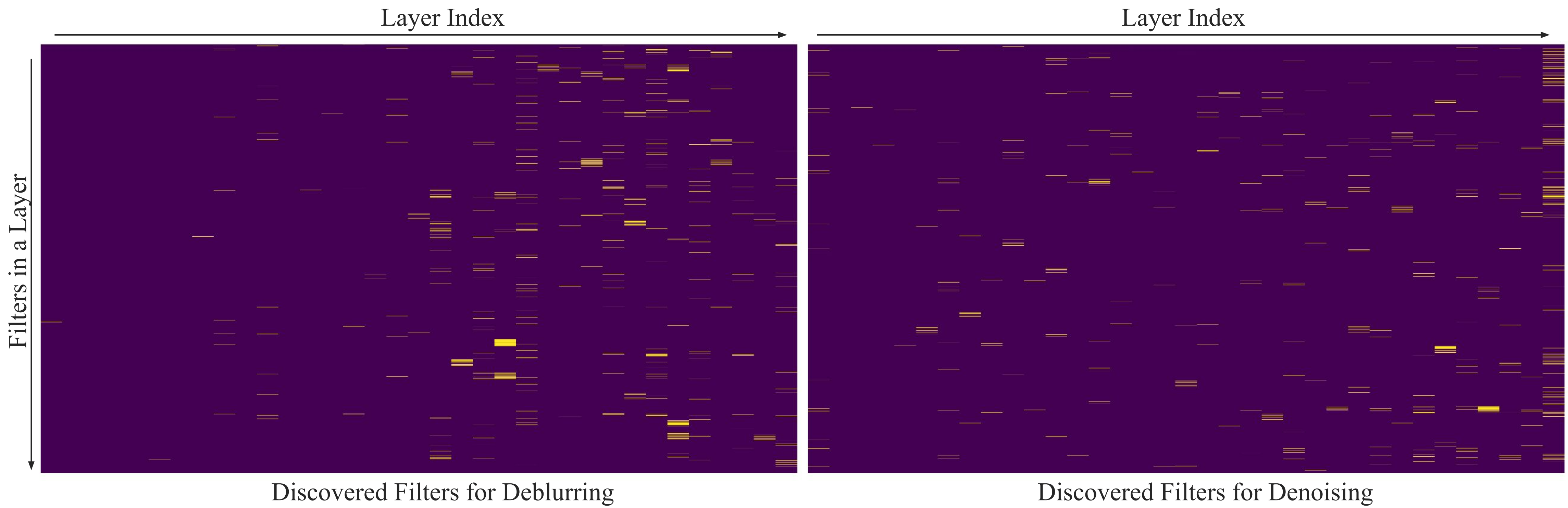
1. Find the discriminative filters for a degradation on the target model
2. Record the ***locations*** of those filters
3. Re-train the corresponding filters *with the same locations* in the baseline model on the desired degradation

PSNR(dB)		Re-train 1% filters for deblurring				Re-train 1% filters for denoising			
Input	Upper bound	FAIG	IG	$\|\theta - \bar{\theta}\|$	Random	FAIG	IG	$\|\theta - \bar{\theta}\|$	Random
Blurry	29.203 (±0.021)	**28.047** (±0.023)	26.474 (±0.295)	26.758 (±0.103)	27.028 (±0.154)	27.463 (±0.133)	26.656 (±0.265)	26.746 (±0.139)	27.041 (±0.149)
Noisy	26.712 (±0.008)	25.590 (±0.071)	25.233 (±0.051)	25.444 (±0.008)	25.525 (±0.037)	**25.793** (±0.021)	25.465 (±0.014)	25.448 (±0.009)	25.526 (±0.010)

Results of re-training baseline models with 1% filters for deblurring and denoising

The ***weights, locations and connections*** of the discovered filters are all important to determine the network function for a specific degradation

Experiments — Distribution of Discovered



The deblurring filters are more located in the back part of the network while denoising filters locate more uniformly

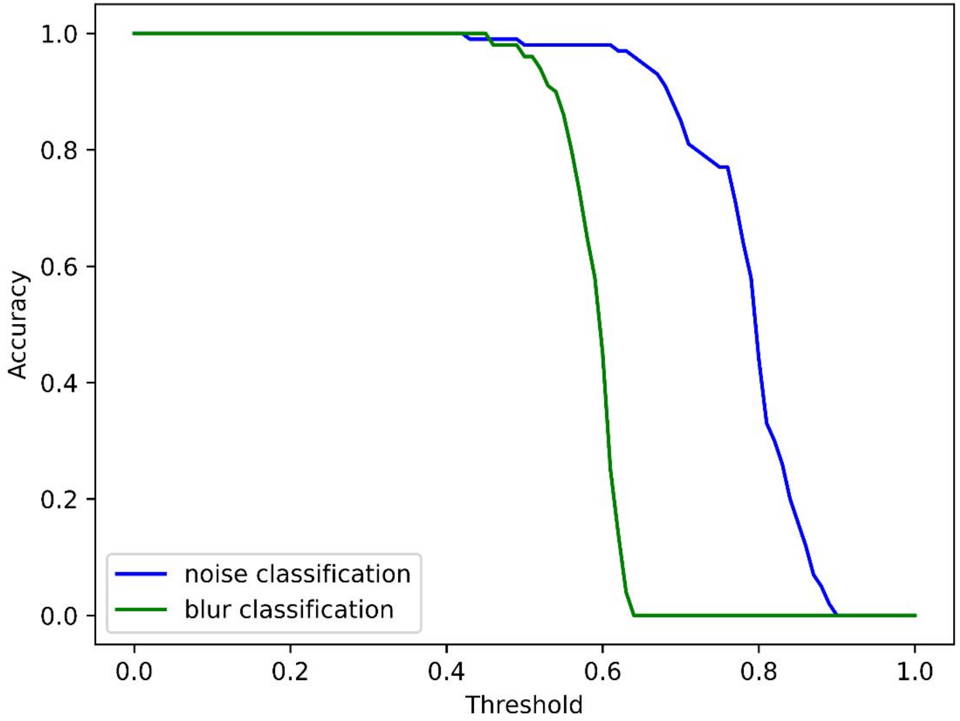
Experiments — Degradation Classification

Predict the degradation of input images without training in the supervision of degradation labels

we calculate the overlap score (OS) to measure the intersection of the two sets of filters:

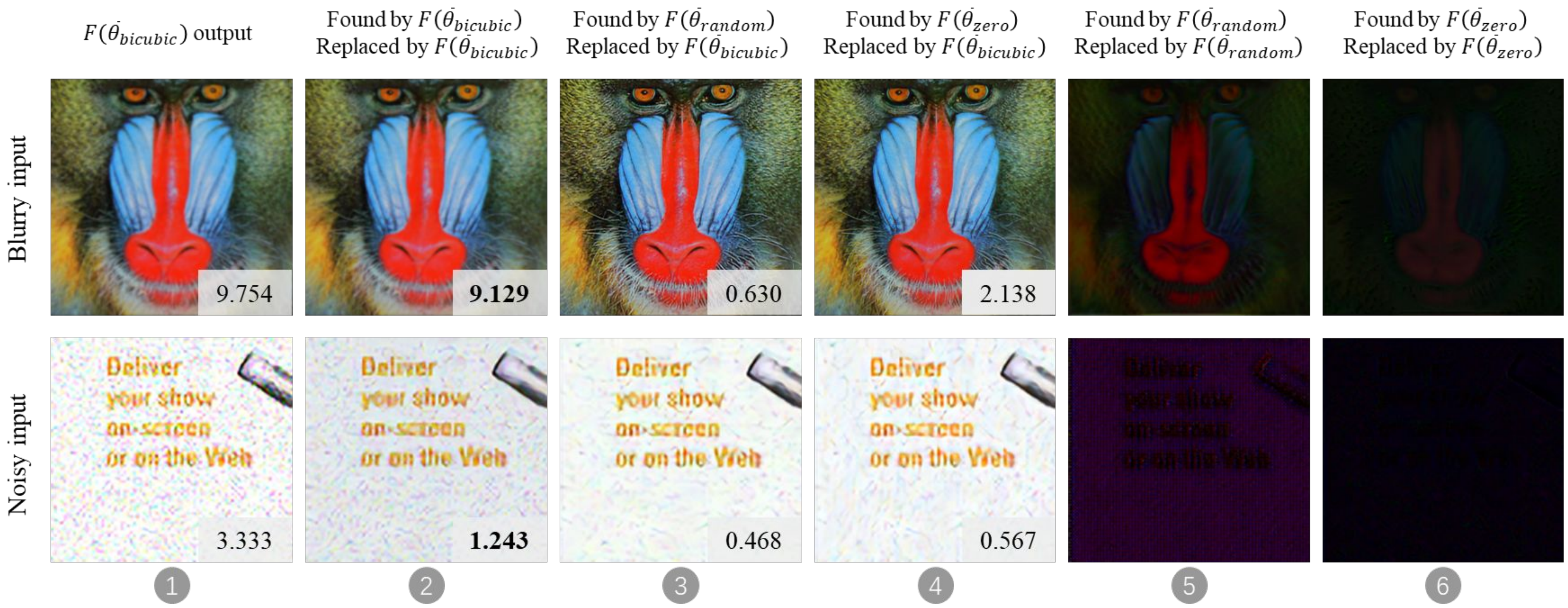
$$\mathrm{OS}(x, \mathcal{D}) = \frac{|\{\mathtt{filter}^{\mathcal{D}}\} \cap \{\mathtt{filter}^{x}\}|}{|\{\mathtt{filter}^{x}\}|}$$

$\{filter^D\}$: the set of discovered filters for degradation D
$\{filter^x\}$: the set of discovered filters for input x



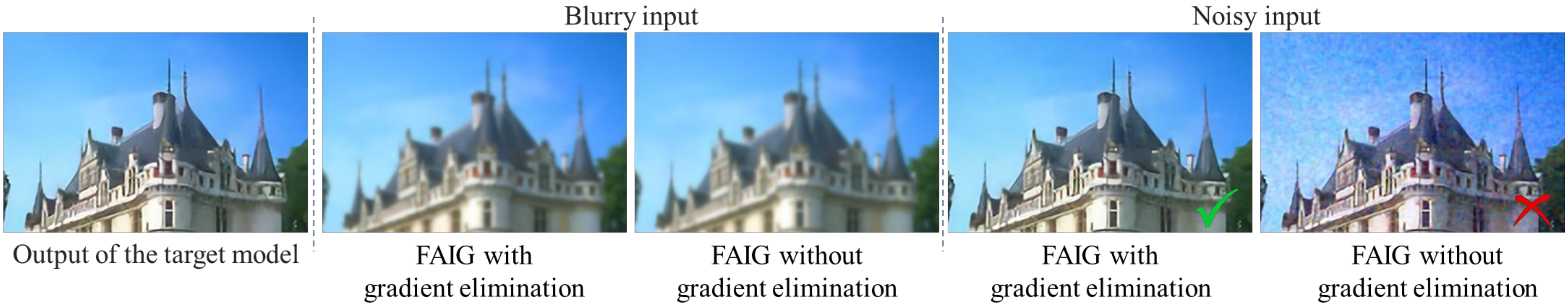
By setting the thresholds: T^{noise} and T^{blur} to 0.6 and 0.5, the prediction accuracy can reach **98%** and **96%**.

Ablations — Impact of Different Baseline



The proposed fine-tuning strategy for baseline models is more effective in finding discriminative filters for specific degradations.

Ablations — Importance of Gradient



$$\mathtt{FAIG}_i^{\mathcal{D}}(\theta) = \frac{1}{\|\mathcal{X}|}\Big(\underbrace{\sum_{x\in\mathcal{X}} |\mathtt{FAIG}_i(\theta, x^{\mathcal{D}})|}_{\text{attribution for degradtion}\mathcal{D}} - \underbrace{\sum_{x\in\mathcal{X}} |\mathtt{FAIG}_i(\theta, x^{\sim\mathcal{D}})|}_{\text{attribution for other degradtions}}\Big)$$

FAIG with gradient elimination can find neurons that are more discriminative for a specific degradation.

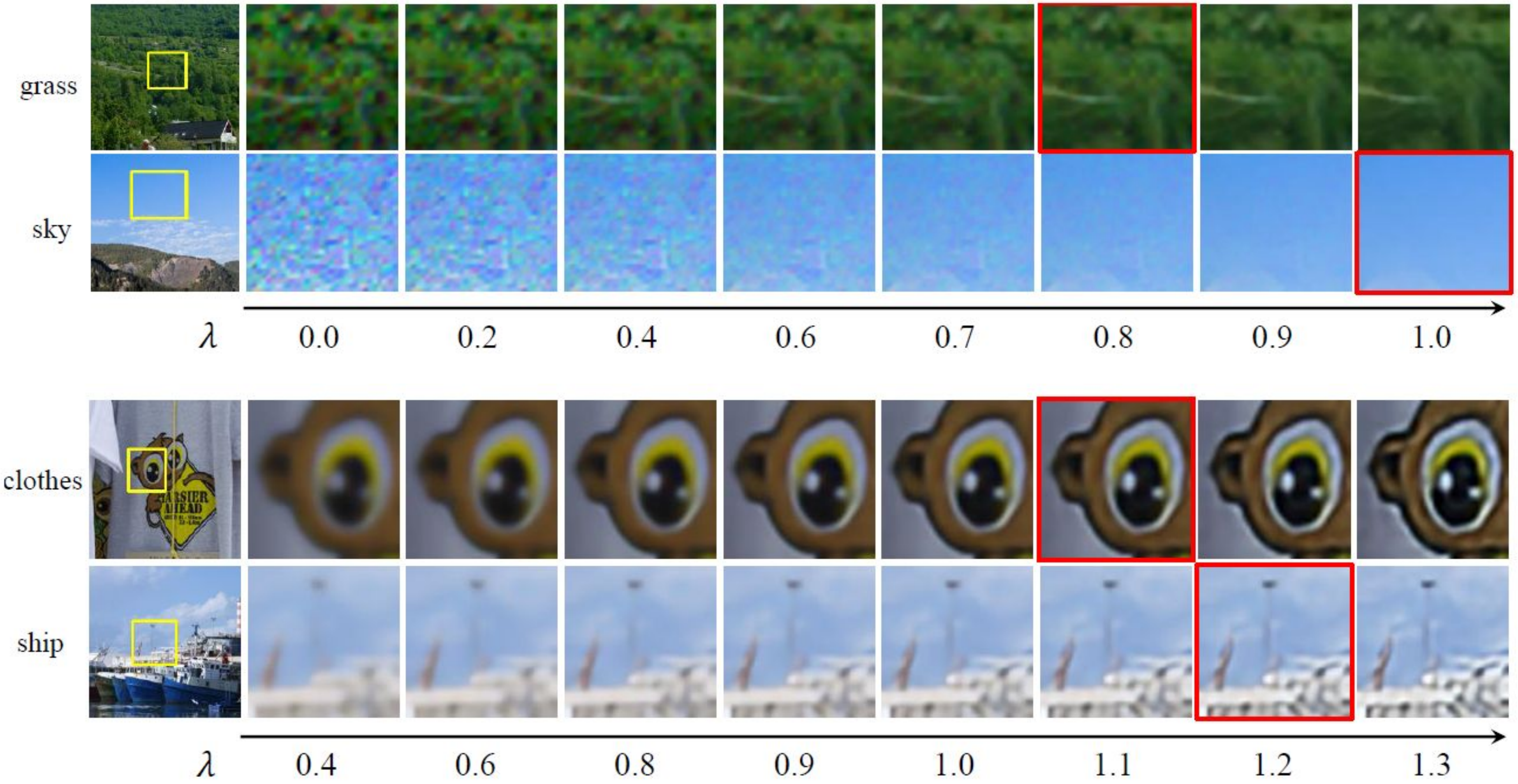
Applications -- Controllable restoration

We can interpolate the corresponding parameters (at the same location) of discovered filters between the baseline model $F(\bar{\theta})$ and the target model $F(\theta)$

$$\theta_{\text{interp}} = (1-\lambda)\bar{\theta} + \lambda\theta$$

Denoising

Deblur



Conclusions

- Provide an understanding of the mechanism of blind SR networks
  - Two-branch networks V.S. A unified one-branch networks.
- Propose a new diagnostic tool – Filter Attribution Integral Gradient (FAIG)
  - Find discriminative filters for specific degradations in blind SR
- Exploiting the interpretability of blind SR would bring great significance for future works in
  - designing more efficient architectures;
  - diagnosing an SR network, such as determining the boundary of network restoration capacity and improving algorithm robustness.

Thanks for

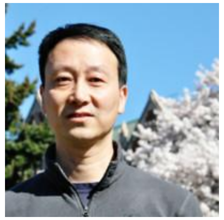
Finding *Discriminative Filters* for Specific Degradations in Blind Super-Resolution

Liangbin Xie

Xintao Wang

Chao Dong

Zhongang Qi



Ying Shan

Codes