Sample-specific Masks for Visual Reprogramming-based Prompting

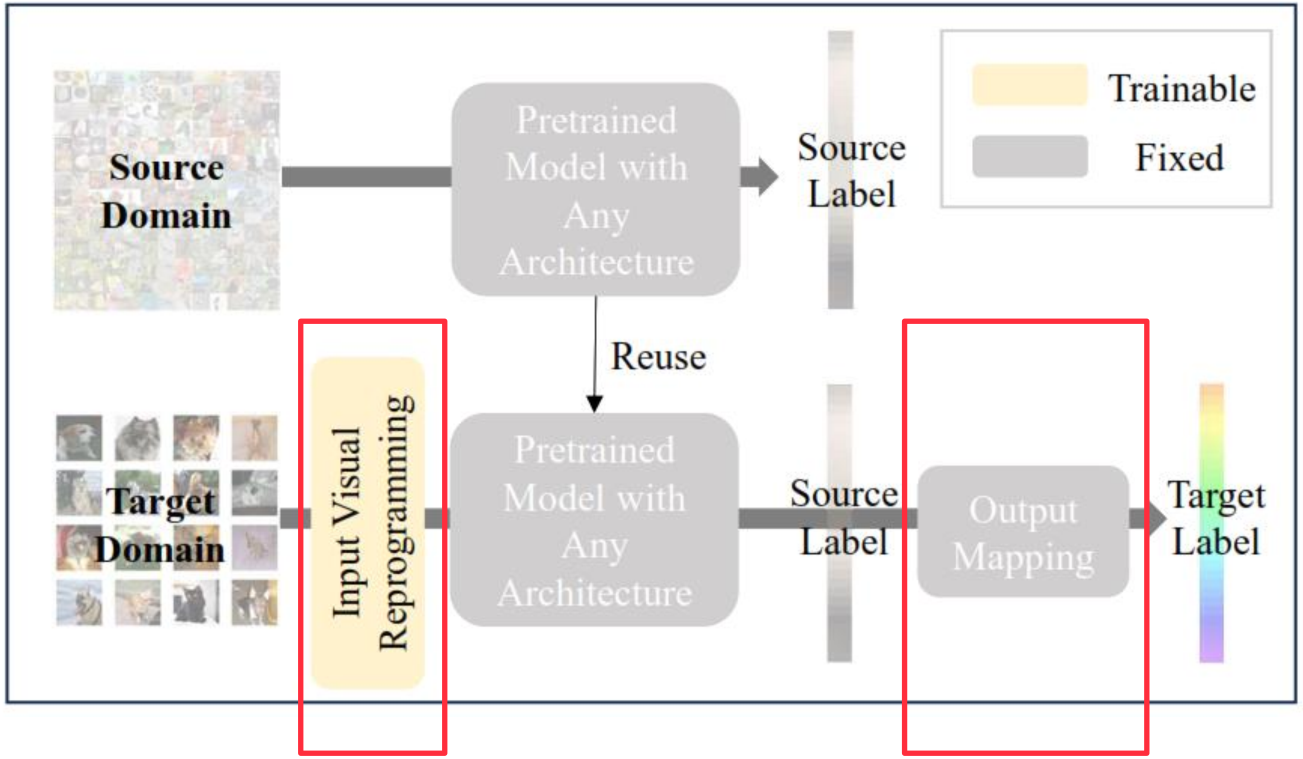
Chengyi Cai, Zesheng Ye, Lei Feng, Jianzhong Qi, Feng Liu*

THE UNIVERSITY OF MELBOURNE

ICML 2024

Background: Visual Reprogramming-based Prompting

➢ Reusing Pre-trained Models in Downstream Tasks



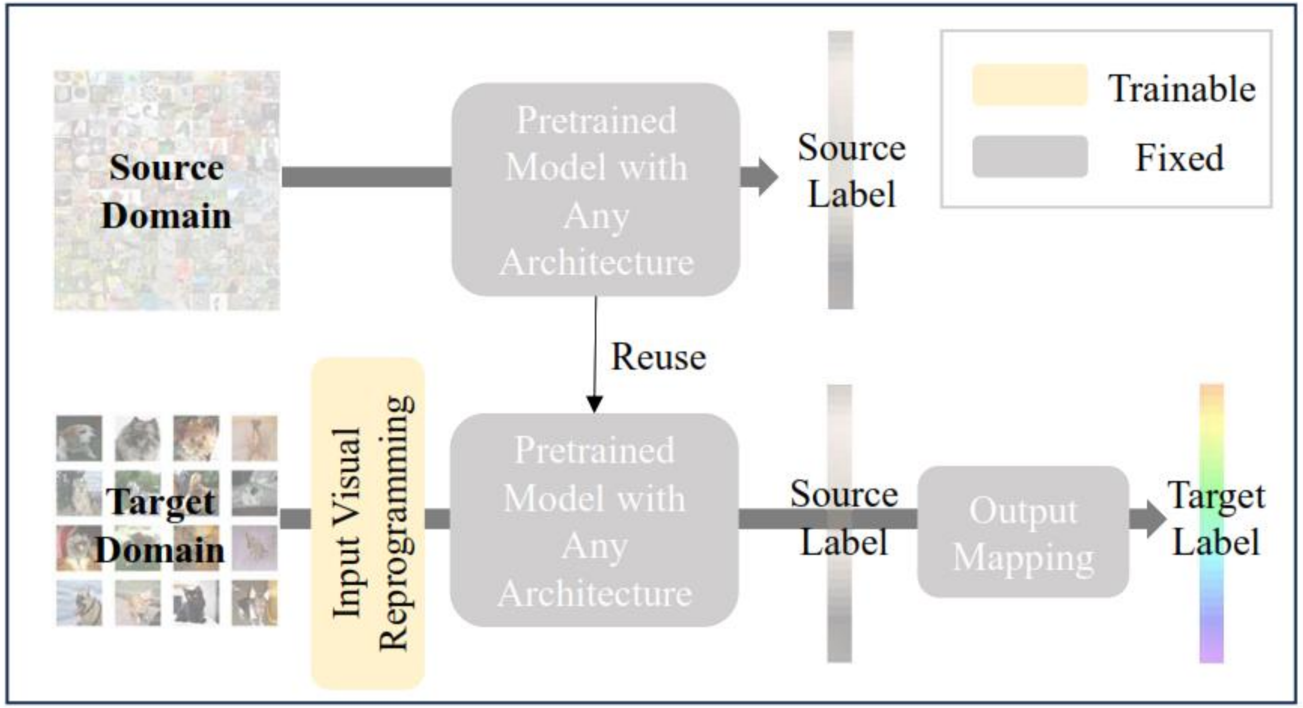
- Reusing Pre-trained Vision Models:

(1) Input Visual Reprogramming

(2) Output Mapping

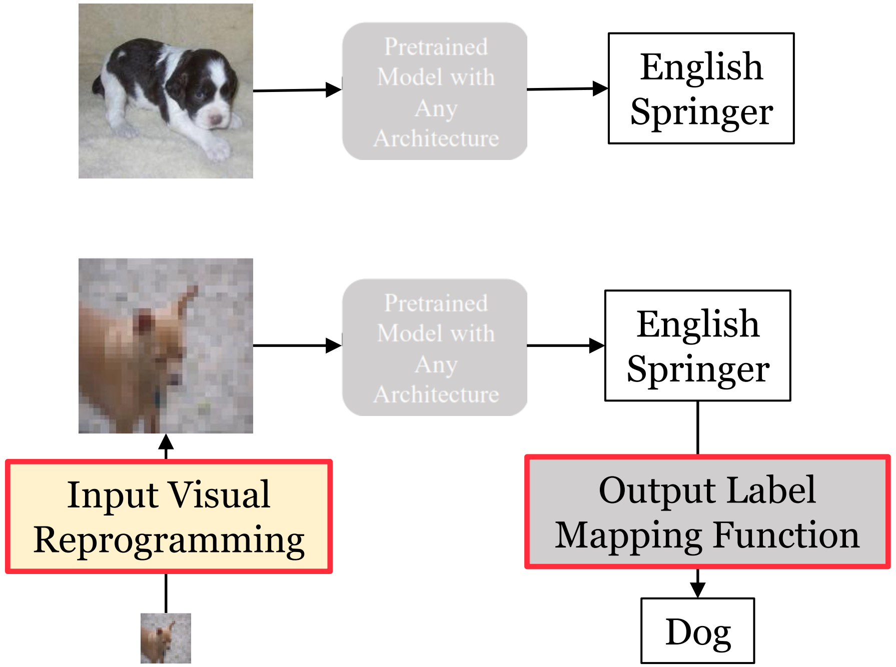
Background: Visual Reprogramming-based Prompting

➢ Reusing Pre-trained Models in Downstream Tasks



- Example

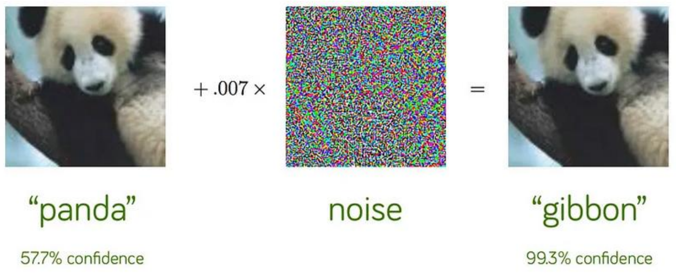
ImageNet-1k → CIFAR10



Background: Visual Reprogramming-based Prompting

➢ Visual (Adversarial) Reprogramming

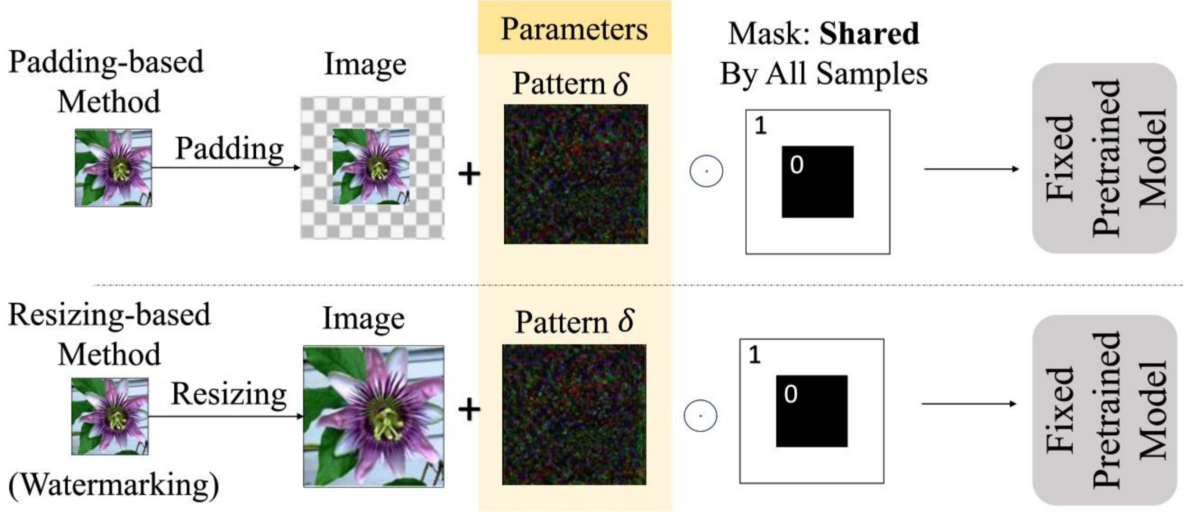
Origin of the Concept:
Adversarial Attacks



Goal: Hindering Pre-Trained Models

VS

Visual (Adversarial) Reprogramming



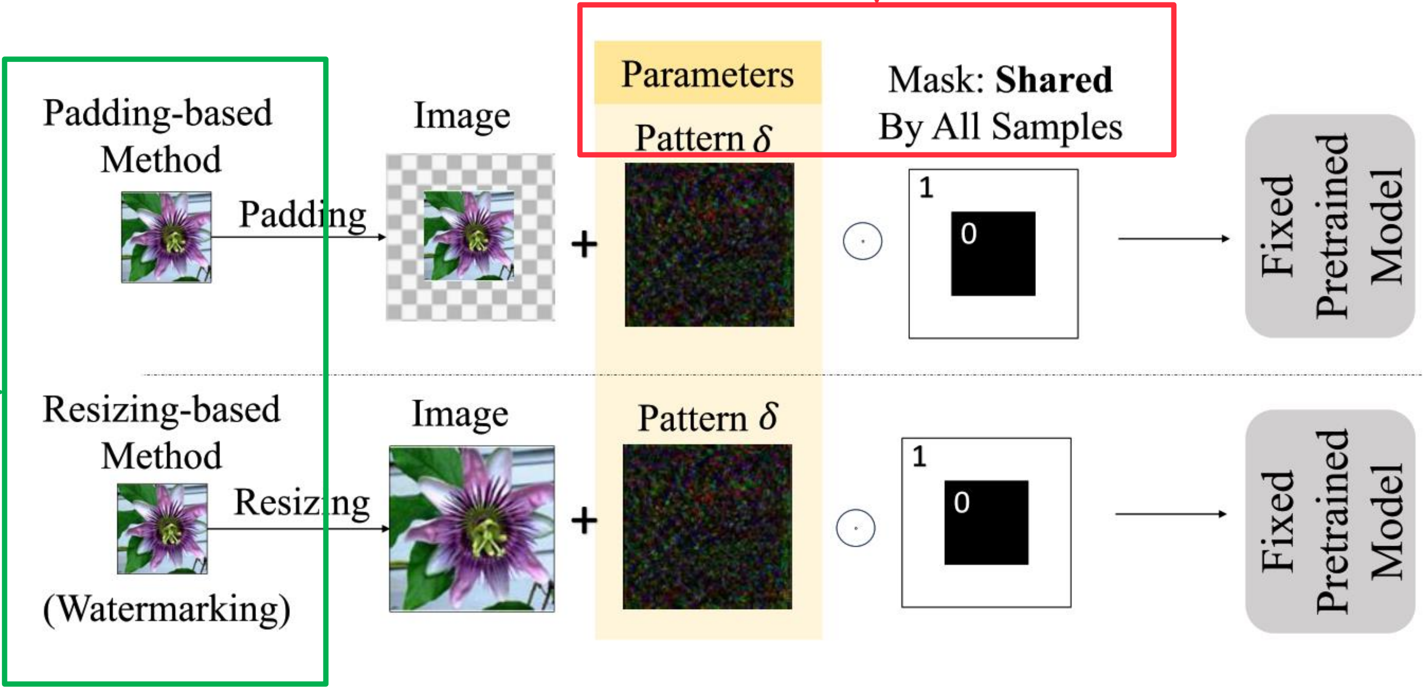
Goal: Reusing Pre-Trained Models

Background: Visual Reprogramming-based Prompting

➢ Visual (Adversarial) Reprogramming

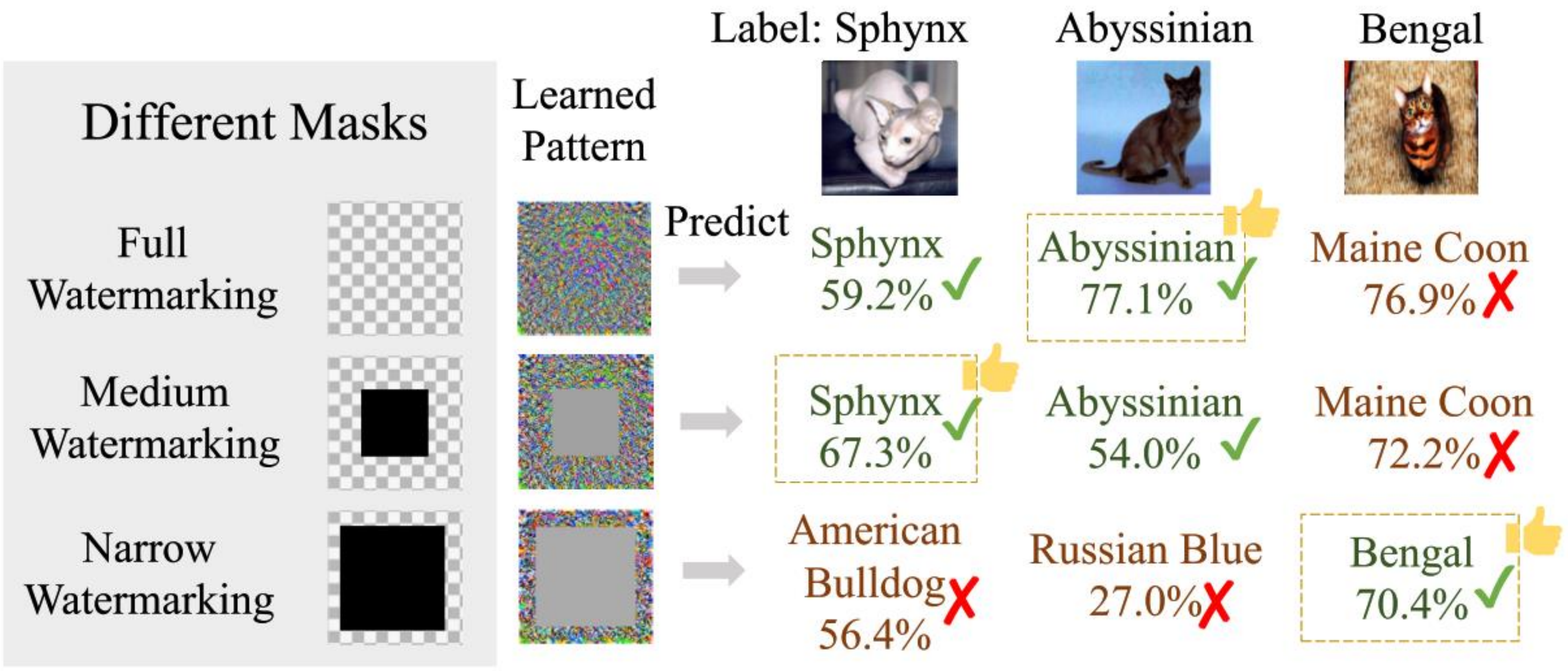
Two Main Components:
(1) Trainable Noise Patterns
(2) Shared Masks

Two Types of Methods:
(1) Padding-based
(2) Resizing-based (Watermarking)



Drawbacks of Shared Masks

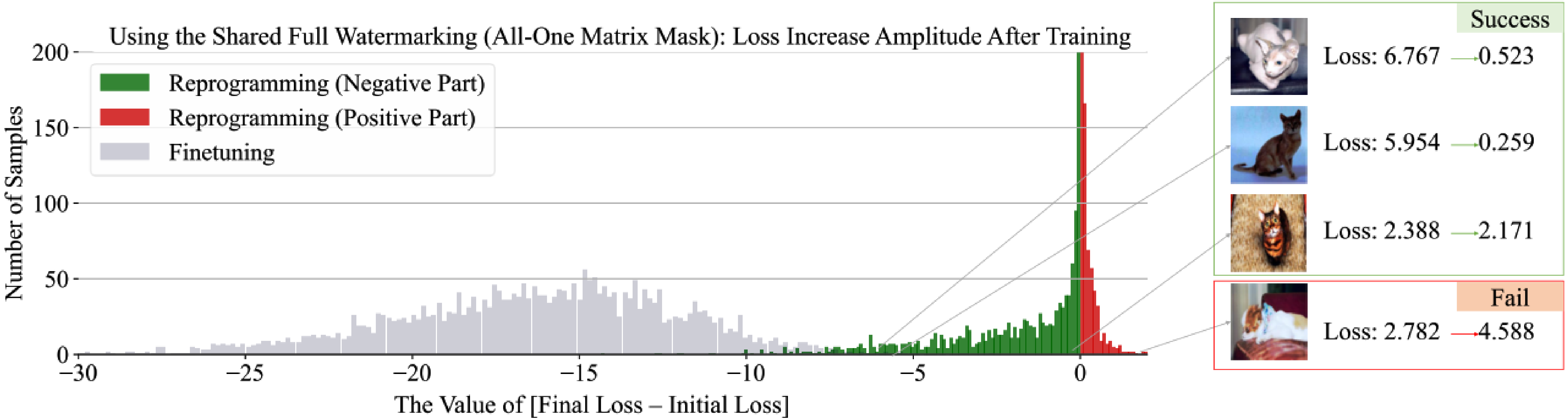
➢ **Drawback Over Individual Images**



Different masks are needed for individual images!

Drawbacks of Shared Masks

➢ Drawback in The Statistical View



The training loss for some samples even rises!

Sample-specific Multi-channel Masks (SMM)

ICML International Conference On Machine Learning

TMLR TRUSTWORTHY MACHINE LEARNING AND REASONING

POSTERA CRESCAM LAUDE THE UNIVERSITY OF MELBOURNE

➢ Problem Setting and Goa

$$\min_{\theta\in\Theta,\omega\in\Omega} \frac{1}{n}\sum_{i=1}^{n} \ell(f_{\mathrm{out}}(f_{\mathrm{P}}(f_{\mathrm{in}}(x_i^{\mathrm{T}}|\theta))|\mathcal{Y}_{\mathrm{sub}}^{\mathrm{P}},\omega), y_i^{\mathrm{T}})$$

Input VR – trainable parameters: $f_{\mathrm{in}}(.|\theta) \ : \ \mathcal{X}^{\mathrm{T}} \mapsto \mathcal{X}^{\mathrm{P}}$

Output Label Mapping – non-parametric function: $f_{\mathrm{out}}(.|\mathcal{Y}_{\mathrm{sub}}^{\mathrm{P}},\omega) \quad \mathcal{Y}_{\mathrm{sub}}^{\mathrm{P}} \mapsto \mathcal{Y}^{\mathrm{T}}$

➢ Methods

Resizing Function

Shared Masks

Reprogramming Pattern

A Shared Mask: $\mathcal{F}^{\mathrm{shr}}(f_{\mathrm{P}}') = \{f|f(x) = f_{\mathrm{P}}'(\boxed{r(x)} + \boxed{M} \odot \boxed{\delta}), \forall x \in \mathcal{X}\}$

Sample-specific Patterns: $\mathcal{F}^{\mathrm{sp}}(f_{\mathrm{P}}') = \{f|f(x) = f_{\mathrm{P}}'(r(x) + \boxed{f_{\mathrm{mask}}}(r(x))), \forall x \in \mathcal{X}\}$

Sample-specific Masks

Our SMM: $\mathcal{F}^{\mathrm{smm}}(f_{\mathrm{P}}') = \{f|f(x) = f_{\mathrm{P}}'(r(x) + \boxed{f_{\mathrm{mask}}}(r(x)) \odot \delta), \forall x \in \mathcal{X}\}$

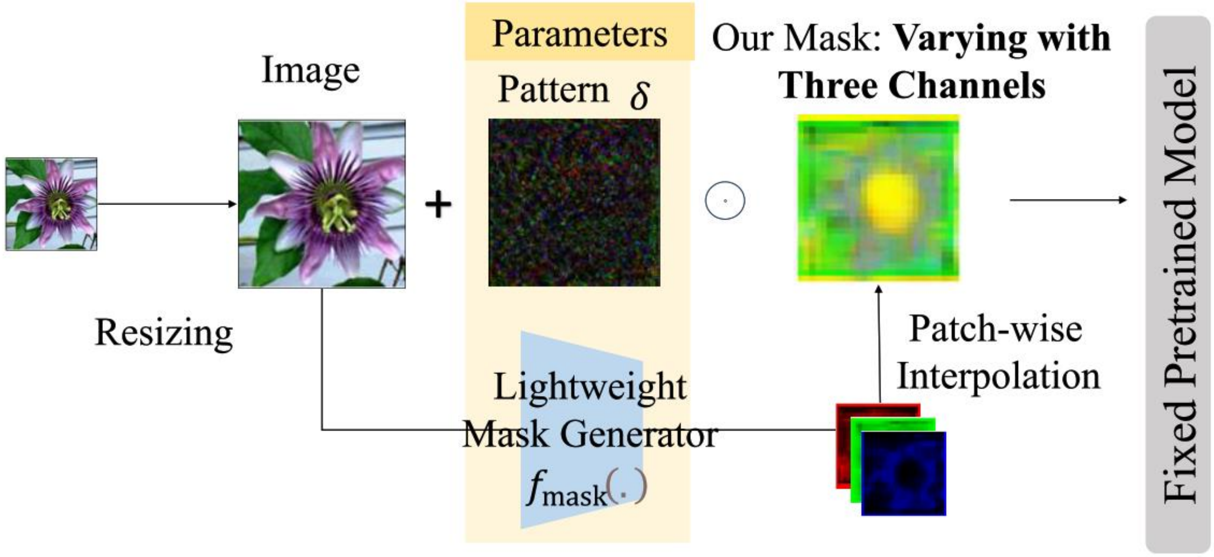
➢ Theory

Approximation Error $\mathrm{Err}_{\mathcal{D}_{\mathrm{T}}}^{\mathrm{apx}}(\mathcal{F}^{\mathrm{sp}}(f_P')) \geq \mathrm{Err}_{\mathcal{D}_{\mathrm{T}}}^{\mathrm{apx}}(\mathcal{F}^{\mathrm{smm}}(f_P')) \quad \mathrm{Err}_{\mathcal{D}_{\mathrm{T}}}^{\mathrm{apx}}(\mathcal{F}^{\mathrm{shr}}(f_P')) \geq \mathrm{Err}_{\mathcal{D}_{\mathrm{T}}}^{\mathrm{apx}}(\mathcal{F}^{\mathrm{smm}}(f_P'))$ ➞ **Lower**

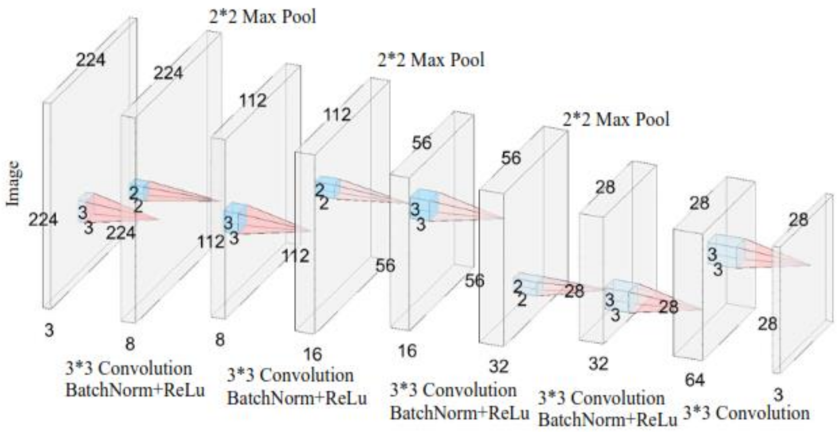
Estimation Error Introducing less than 0.2% extra parameters
Not increasing the risk of over-fitting in experiments ➞ **Negligible**

Sample-specific Multi-channel Masks (SMM)

- **Framework and Modules**



- Module 1: Lightweight Mask Generator



- Module 2: Patch-wise Interpolation
  → Interpolating by Copying

Sample-specific Multi-channel Masks (SMM)

➢ Experimental Results

Pre-trained	ResNet-18 (ImageNet-1K)					ResNet-50 (ImageNet-1K)				
Methods	Pad	Narrow	Medium	Full	Ours	Pad	Narrow	Medium	Full	Ours
CIFAR10	65.5 ±0.1	68.6 ±2.8	68.8 ±1.1	68.9 ±0.4	**72.8** ±0.7	76.6±0.3	77.4±0.5	77.8±0.2	79.3±0.3	**81.4**±0.6
CIFAR100	24.8±0.1	36.9±0.6	34.9±0.2	33.8±0.2	**39.4**±0.6	38.9±0.3	42.5±0.2	43.8±0.2	47.2±0.1	**49.0**±0.2
SVHN	75.2±0.2	58.5±1.1	71.1±1.0	78.3±0.3	**84.4**±2.0	75.8±0.4	59.1±1.3	71.5±0.8	79.5±0.5	**82.6**±2.0
GTSRB	52.0±1.2	46.1±1.5	56.4±1.0	76.8±0.9	**80.4**±1.2	52.5±1.4	38.9±1.3	52.6±1.3	76.5±1.3	**78.2**±1.1
Flowers102	27.9±0.7	22.1±0.1	22.6±0.5	23.2±0.5	**38.7**±0.7	24.6±0.6	19.9±0.6	20.9±0.6	22.6±0.1	**35.9**±0.5
DTD	**35.3**±0.9	33.1±1.3	31.7±0.5	29.0±0.7	33.6±0.4	40.5±0.5	37.8±0.7	38.4±0.2	34.7±1.3	**41.1**±1.1
UCF101	23.9±0.5	27.2±0.9	26.1±0.3	24.4±0.9	**28.7**±0.8	34.6±0.2	38.4±0.2	37.2±0.2	35.2±0.2	**38.9**±0.5
Food101	14.8±0.2	14.0±0.1	14.4±0.3	13.2±0.1	**17.5**±0.1	17.0±0.3	18.3±0.2	18.3±0.2	16.7±0.2	**19.8**±0.0
SUN397	13.0±0.2	15.3±0.1	14.2±0.1	13.4±0.2	**16.0**±0.3	20.3±0.2	22.0±0.1	21.5±0.1	21.1±0.1	**22.9**±0.0
EuroSAT	85.2±0.6	82.8±0.4	83.8±0.5	84.3±0.5	**92.2**±0.2	83.6±0.7	83.7±0.4	85.8±0.1	86.9±0.3	**92.0**±0.6
OxfordPets	65.4±0.7	73.7±0.2	71.4±0.2	70.0±0.6	**74.1**±0.4	76.2±0.6	76.4±0.3	75.6±0.3	73.4±0.3	**78.1**±0.2
Average	43.91	43.48	45.04	46.85	**52.53**	49.15	46.76	49.39	52.10	**56.35**

Pre-trained	ViT-B32 (ImageNet-1K)				
Method	Pad	Narrow	Medium	Full	Ours
CIFAR10	62.4	96.6	96.5	95.8	**97.4**
CIFAR100	31.6	74.4	75.3	75.0	**82.6**
SVHN	80.2	85.0	87.4	87.8	**89.7**
GTSRB	62.3	57.8	68.6	75.5	**80.5**
Flowers102	57.3	55.3	56.6	55.9	**79.1**
DTD	43.7	37.3	38.5	37.7	**45.6**
UCF101	33.6	44.5	**44.8**	40.9	42.6
Food101	37.4	47.3	48.6	49.4	**64.8**
SUN397	21.8	29.0	29.4	28.8	**36.7**
EuroSAT	**95.9**	90.9	90.9	89.1	93.5
OxfordPets	57.6	82.5	81.0	75.3	**83.8**
Average	53.1	63.7	65.2	64.7	**72.4**

- Applying SMM yields **higher accuracy** across commonly-used downstream datasets
- **Compatible** with different pre-trained model architectures

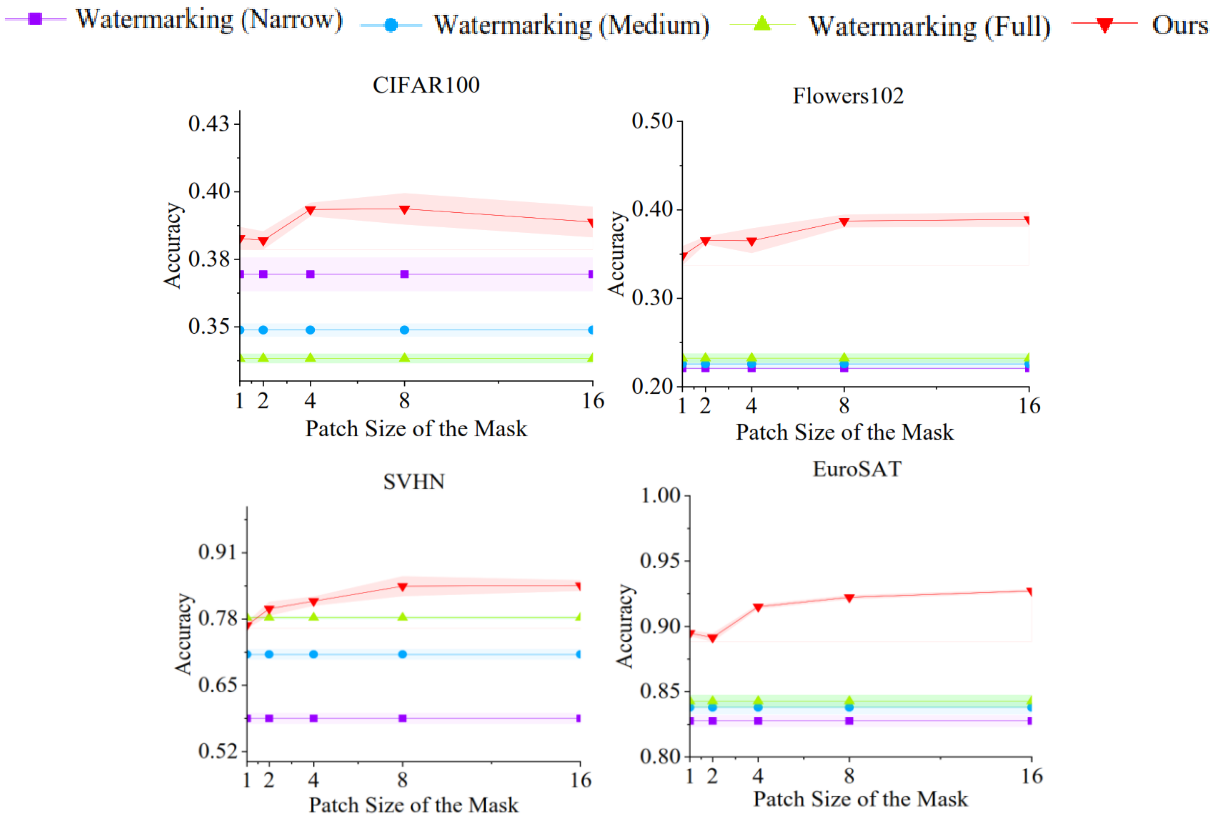
Sample-specific Multi-channel Masks (SMM)

➢ Impact of Masking

$$r(x) + f_{\text{mask}}(r(x)) \odot \delta$$

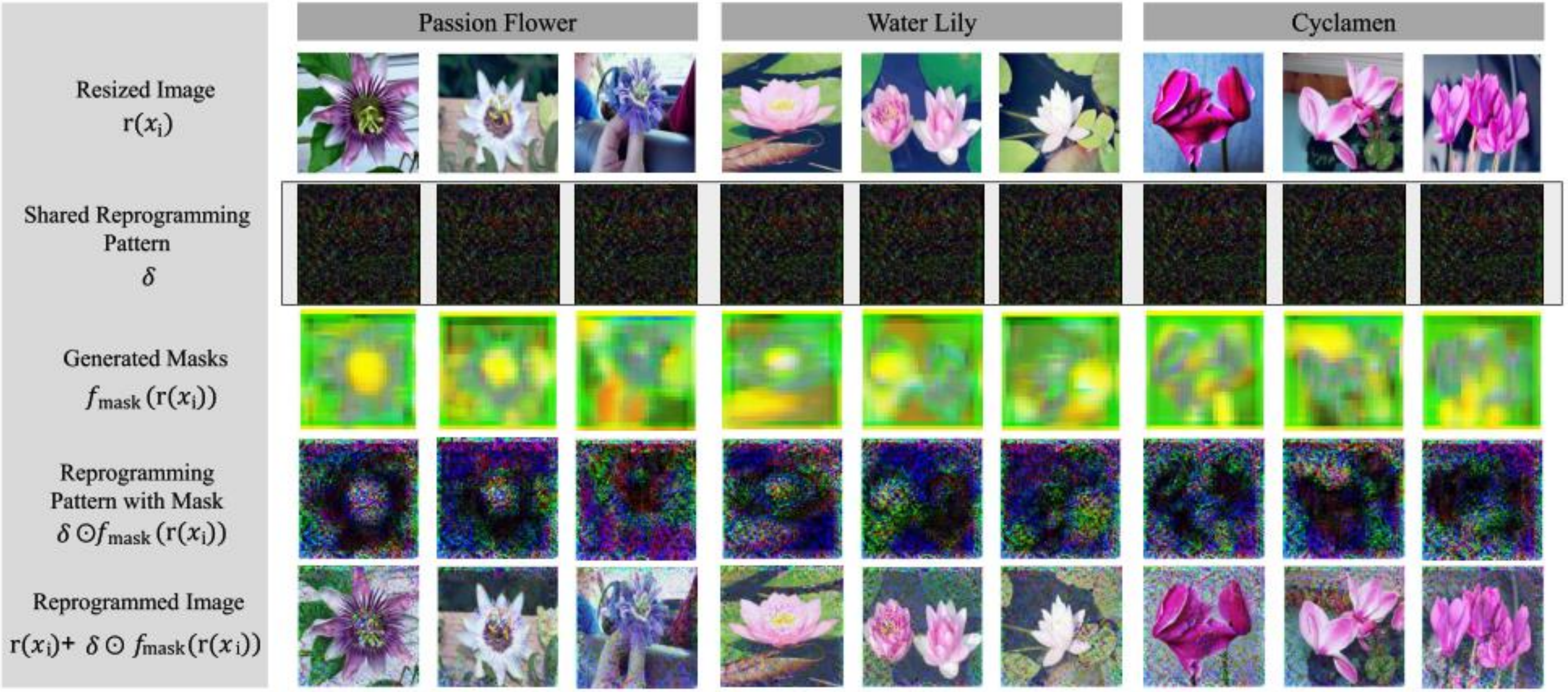
	ONLY δ	ONLY f_{mask}	SINGLE-CHANNEL $f^{\text{s}}_{\text{mask}}$	OURS
CIFAR10	68.9±0.4	59.0±1.6	72.6±2.6	**72.8**±0.7
CIFAR100	33.8±0.2	32.1±0.3	38.0±0.6	**39.4**±0.6
SVHN	78.3±0.3	51.1±3.1	78.4±0.2	**84.4**±2.0
GTSRB	76.8±0.9	55.7±1.2	70.7±0.8	**80.4**±1.2
FLOWERS102	23.2±0.5	32.2±0.4	30.2±0.4	**38.7**±0.7
DTD	29.0±0.7	27.2±0.5	32.7±0.5	**33.6**±0.4
UCF101	24.4±0.9	25.7±0.3	28.0±0.3	**28.7**±0.8
FOOD101	13.2±0.1	13.3±0.1	15.8±0.1	**17.5**±0.1
SUN397	13.4±0.2	10.5±0.1	15.9±0.1	**16.0**±0.3
EUROSAT	84.3±0.5	89.2±0.9	90.6±0.5	**92.2**±0.2
OXFORDPETS	70.0±0.6	72.5±0.3	73.8±0.6	**74.1**±0.4
AVERAGE	46.85	42.59	49.70	**52.53**

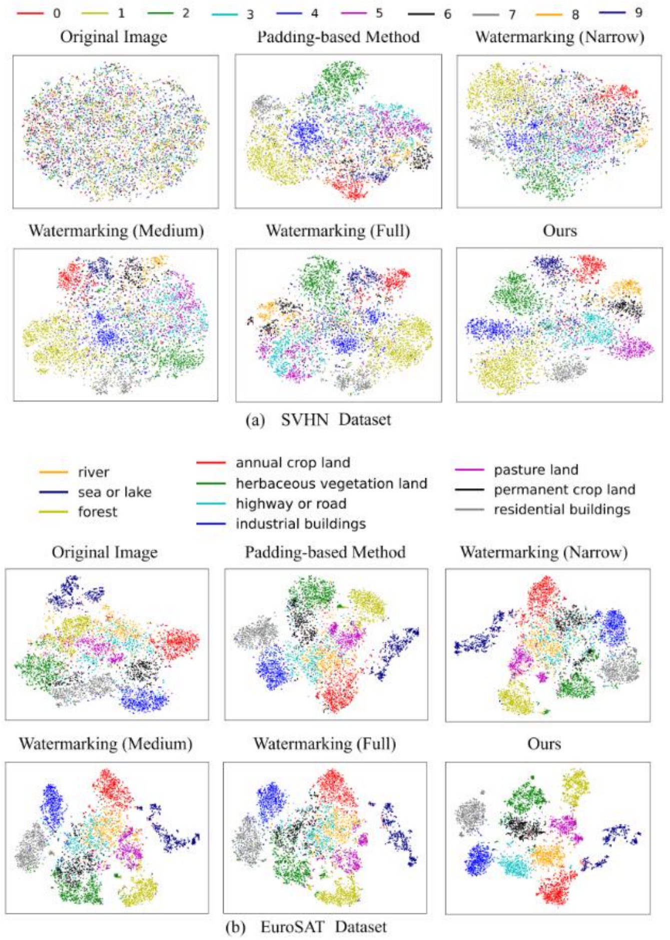
➢ Impact of Patch Size



Sample-specific Multi-channel Masks (SMM)

➢ Visualization Results





- Successfully resolves incorrectly clustered classes in the output feature space
- Able to retain the important parts of the image and remove the interference

Thanks For Listening