UniRestore: Unified Perceptual and Task-Oriented Image Restoration Model Using Diffusion Prior

I-Hsiang Chen [1*] Wei-Ting Chen [1,2,4*] Yu-Wei Liu [1] Yuan-Chun Chiang [1]
Sy-Yen Kuo [1,3] Ming-Hsuan Yang [4,5]

[1]National Taiwan University [2]Microsoft [3]Chang Gung University
[4]UC Merced [5]Google Research

CVPR 2025 Highlight

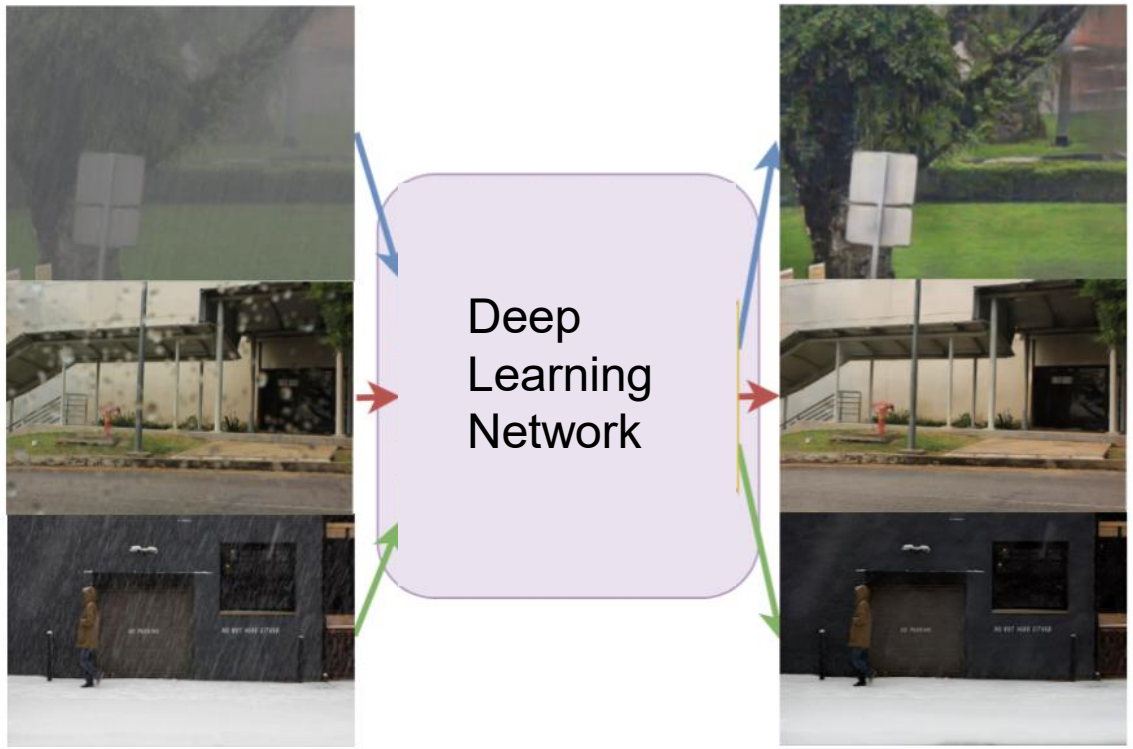
Presenter: Yufei Zhang

2025.12.21

Background : Perceptual Image Restoration

北京大学 PEKING UNIVERSITY

Image Restoration
- Perceptual
- Task-oriented



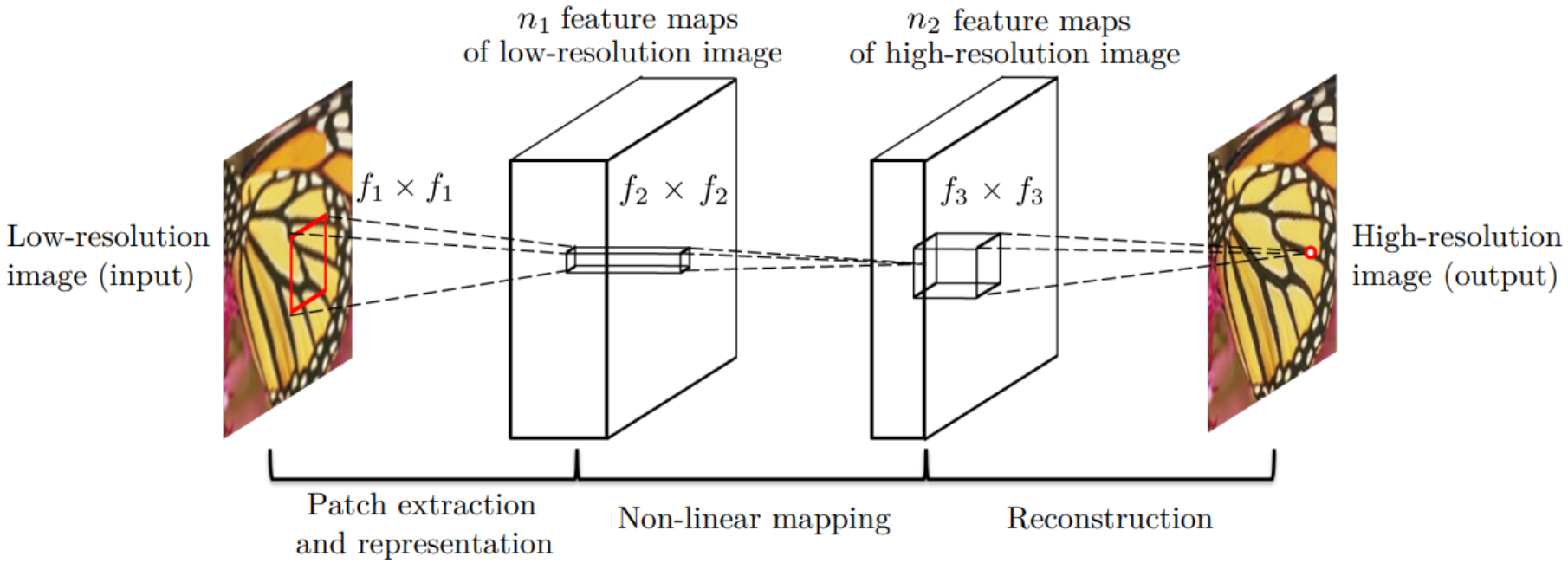
visual quality of images as perceived by humans

Background : Perceptual Image Restoration

北京大学 PEKING UNIVERSITY

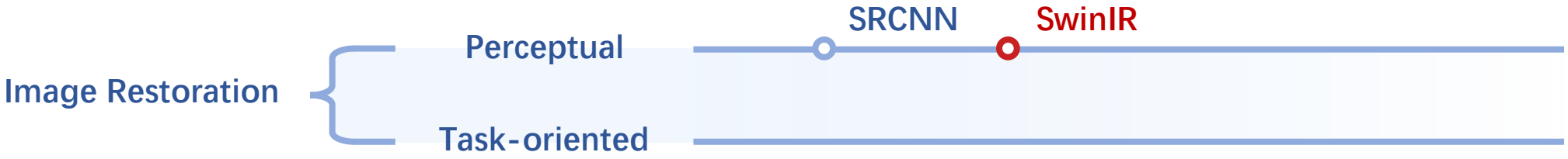
Image Restoration
- Perceptual
  - SRCNN
- Task-oriented

SRCNN

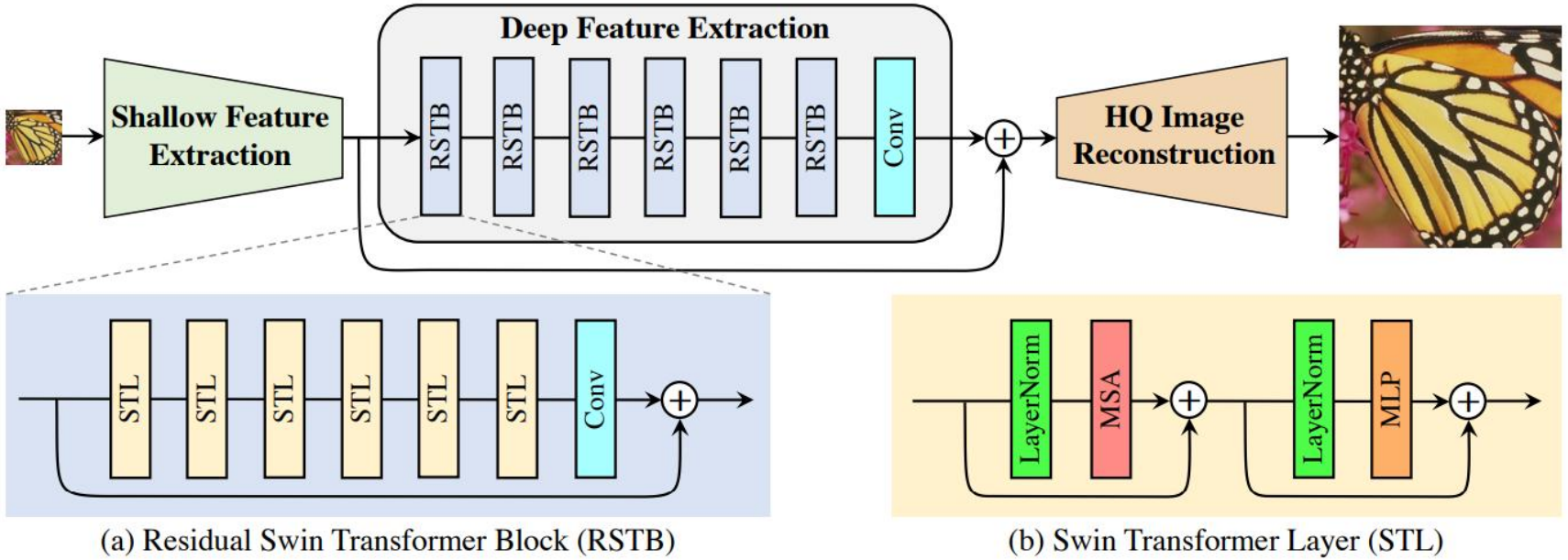


Chao Dong, et al. Image Super-Resolution Using Deep Convolutional Networks, CVPR, 2014

Background : Perceptual Image Restoration



SwinIR



(a) Residual Swin Transformer Block (RSTB)

(b) Swin Transformer Layer (STL)

PS. Multihead Self-Attention

Jingyun Liang, et al. SwinIR: Image Restoration Using Swin Transformer, CVPR, 2021

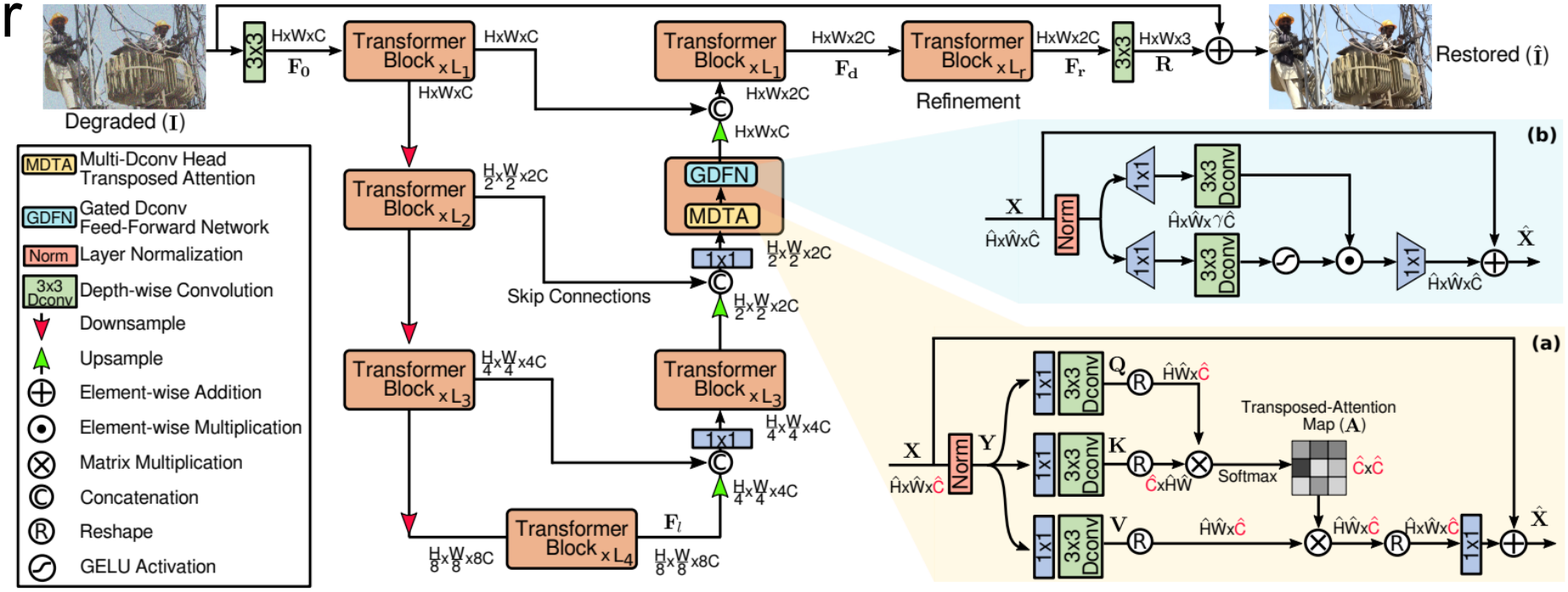
Background : Perceptual Image Restoration

北京大学 PEKING UNIVERSITY

Image Restoration

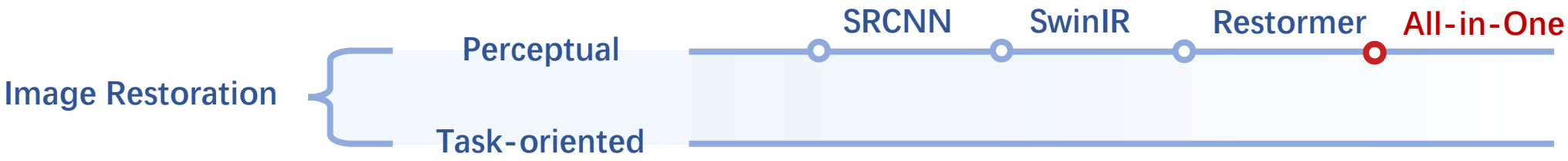
- Perceptual: SRCNN → SwinIR → Restormer
- Task-oriented

Restormer

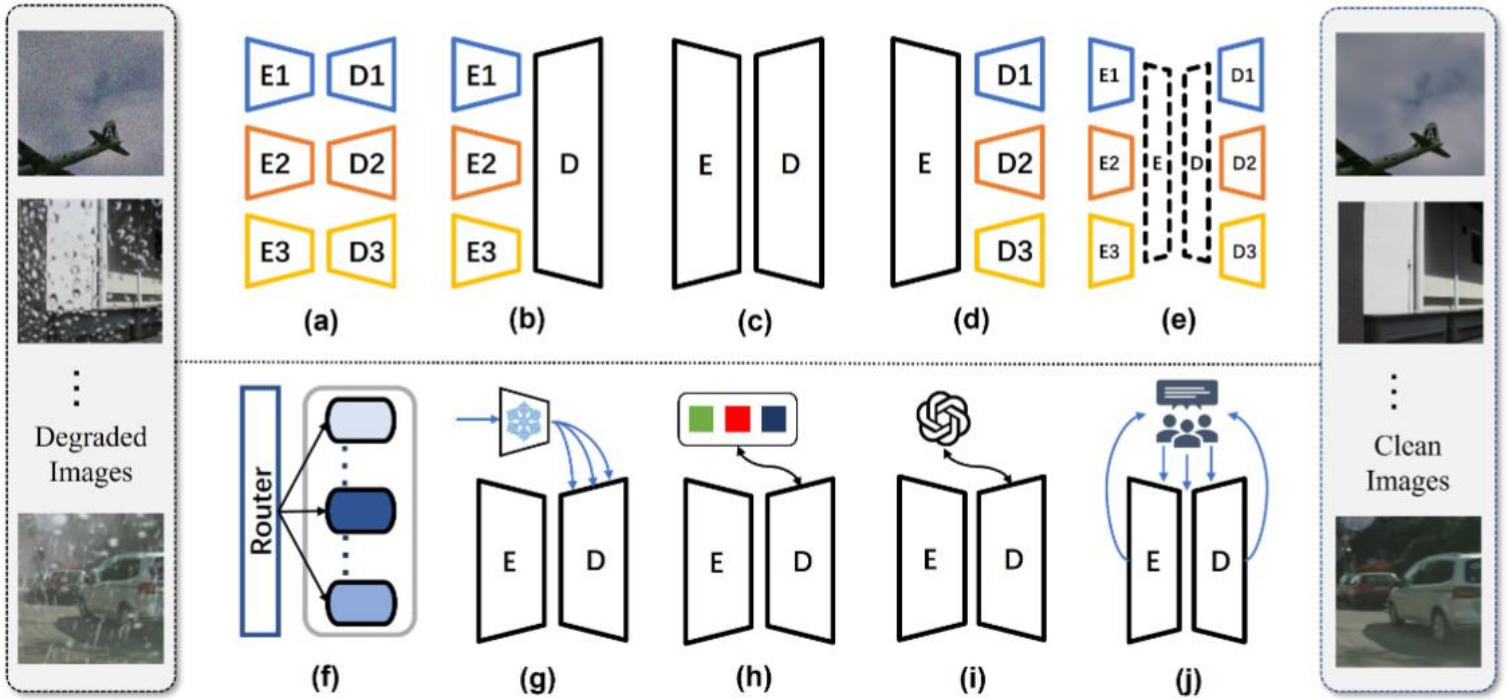


NxNxC → CxCxN (N=HxW)

Background : Perceptual Image Restoration

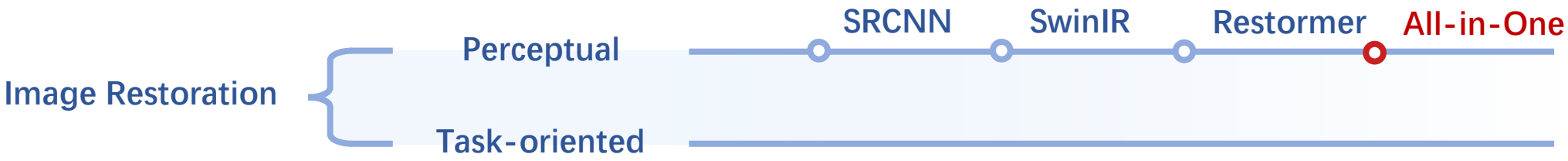


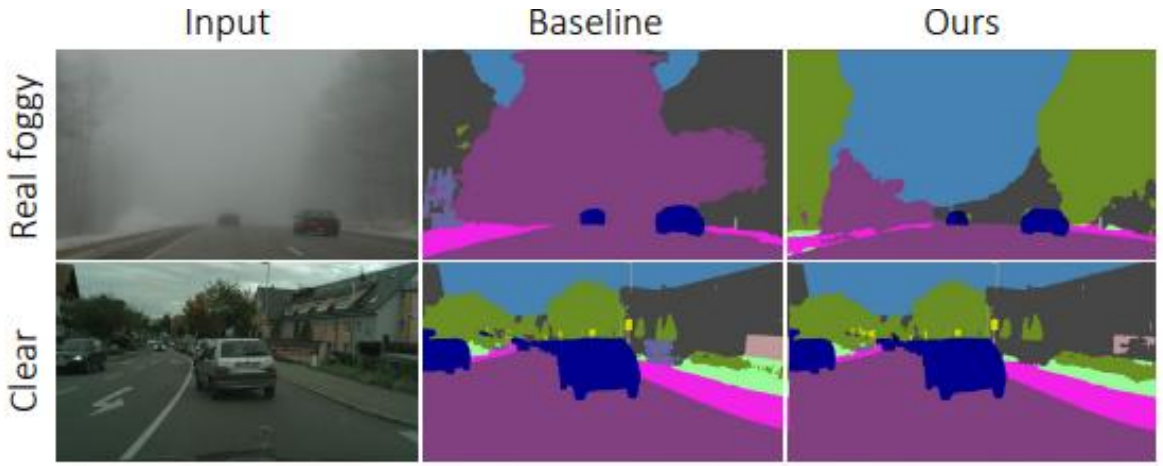
All-in-One



Junjun Jiang, et al. A Survey on All-in-One Image Restoration: Taxonomy, Evaluation and Future Trends, IEEE TPAMI, 2020

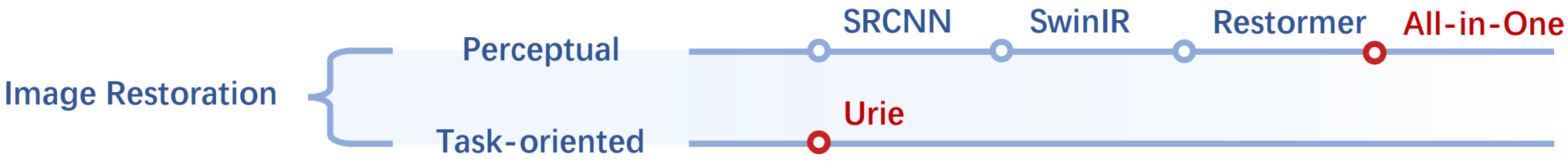
Background : Task-oriented Image Restoration





downstream task performance

Background : Task-oriented Image Restoration



Urie

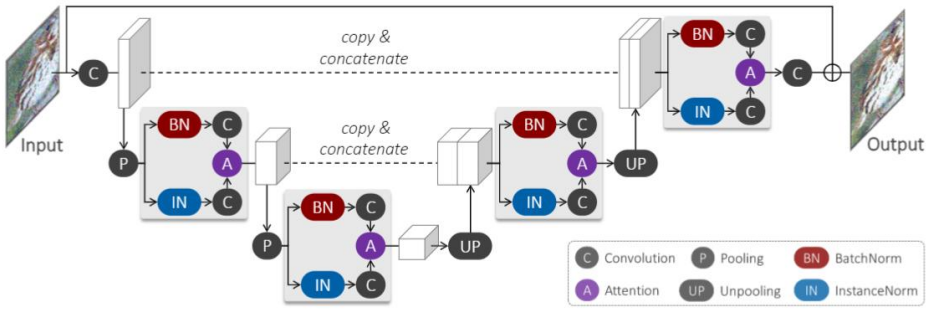


Fig. 1. Overall architecture of URIE. The Selective Enhancement Modules (SEM) are indicated by gray rectangles. Details of these modules are illustrated in Fig. 2.

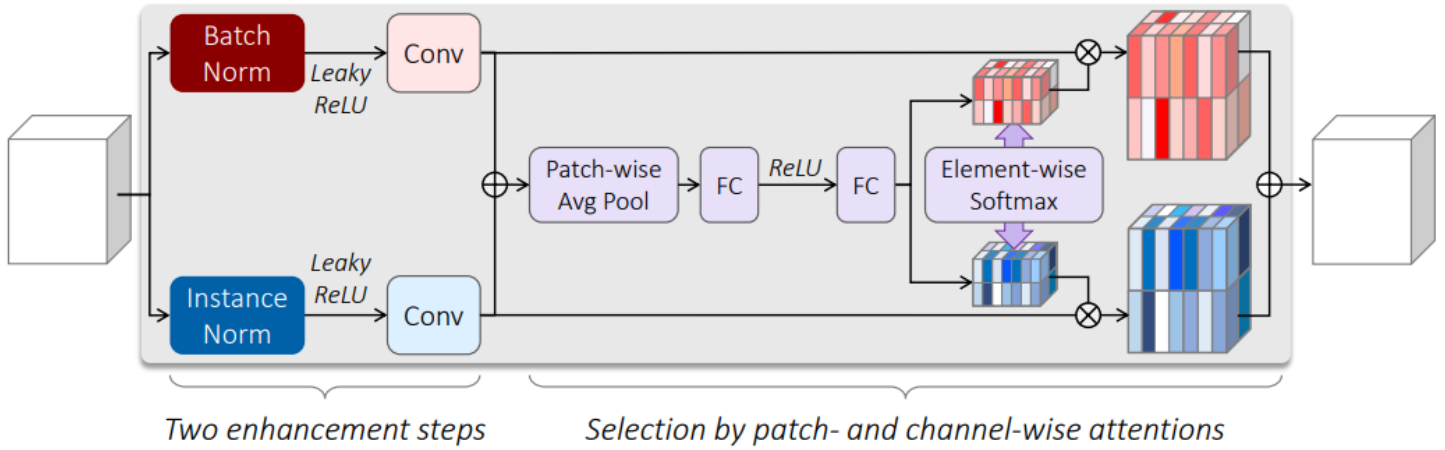
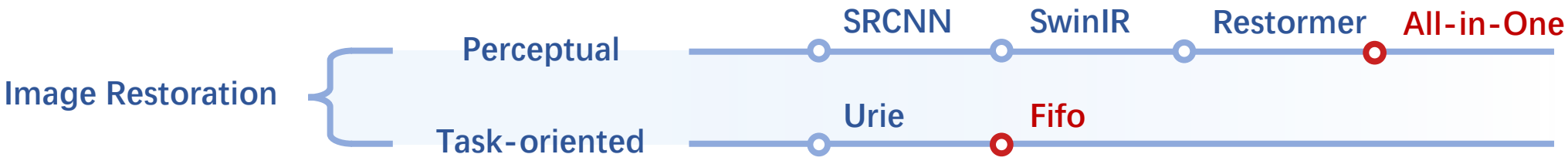


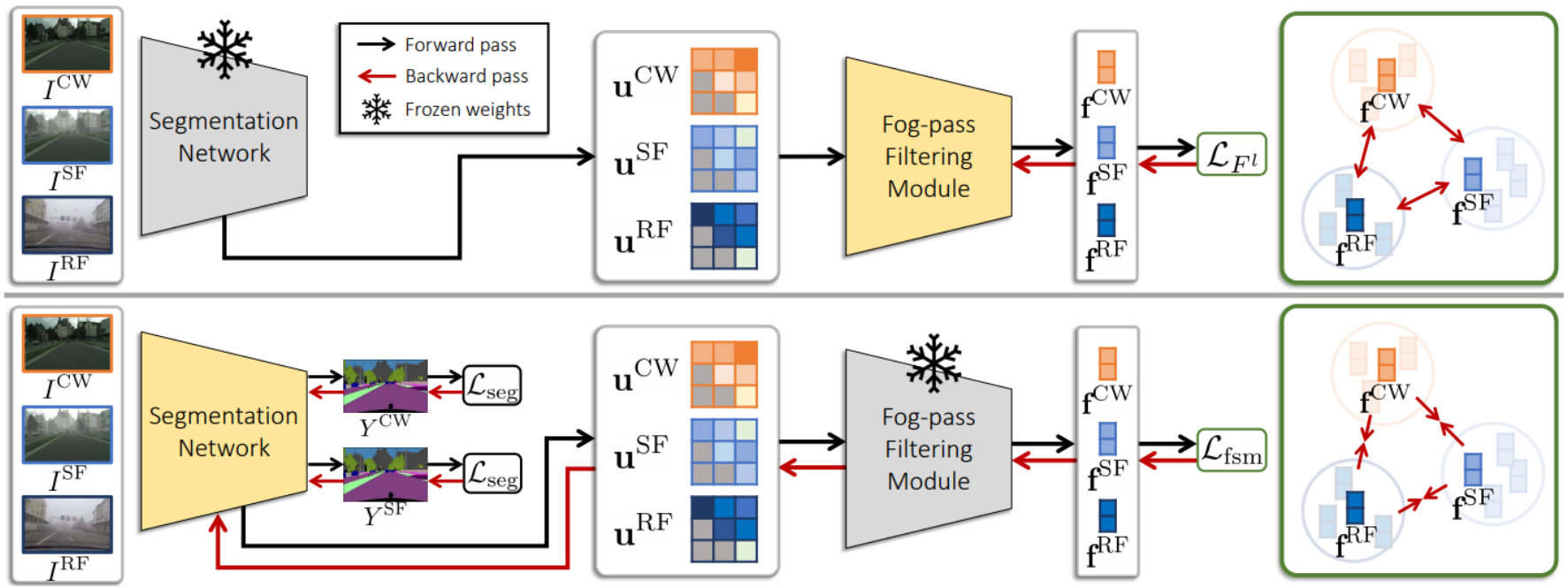
Fig. 2. Details of SEM. $\oplus$ and $\otimes$ indicate element-wise summation and multiplication between feature maps, respectively.

Taeyoung Son, et al. Urie: Universal image enhancement for visual recognition in the wild, ECCV, 2020

Background : Task-oriented Image Restoration

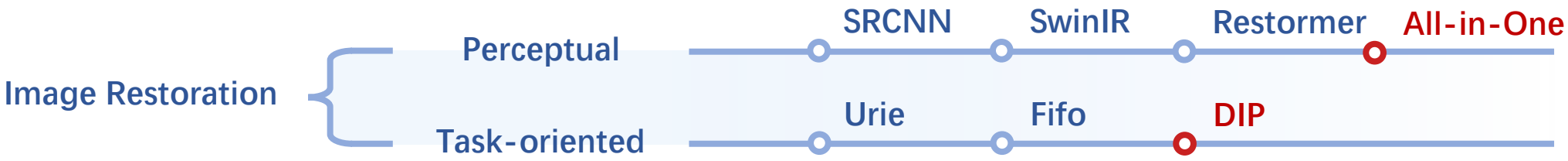


Fifo

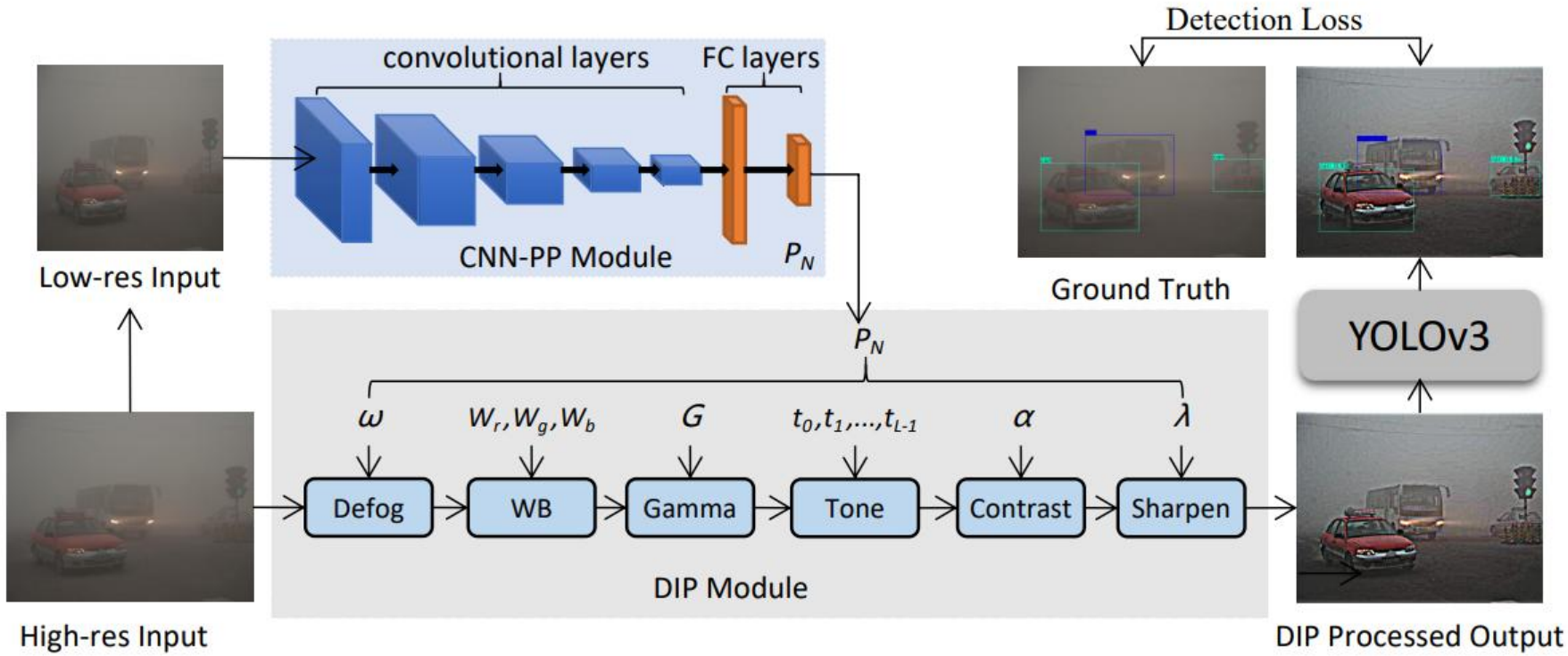


PS. clear weather (CW), synthetic fog (SF), and real fog (RF).

Background : Task-oriented Image Restoration



DIP



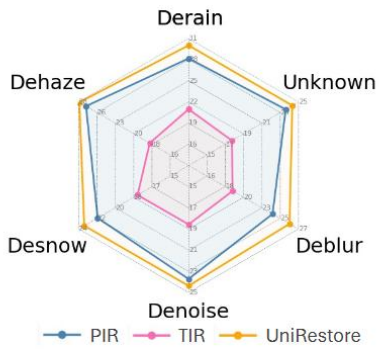
Wenyu Liu, et al. Image-adaptive yolo for object detection in adverse weather conditions, AAAI, 2022

Problems

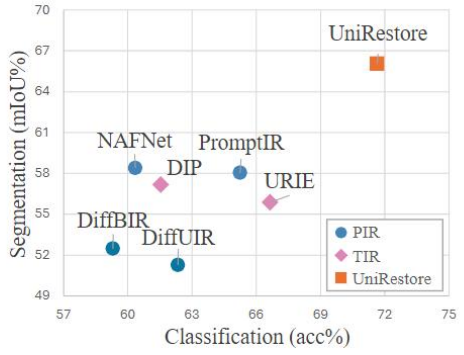
Perceptual Image Restoration (PIR)
Enhances image clarity and visual quality

Task-oriented Image Restoration (TIR)
Optimizes image for downstream tasks

LQ URIE HQ



(a) Perceptual IR

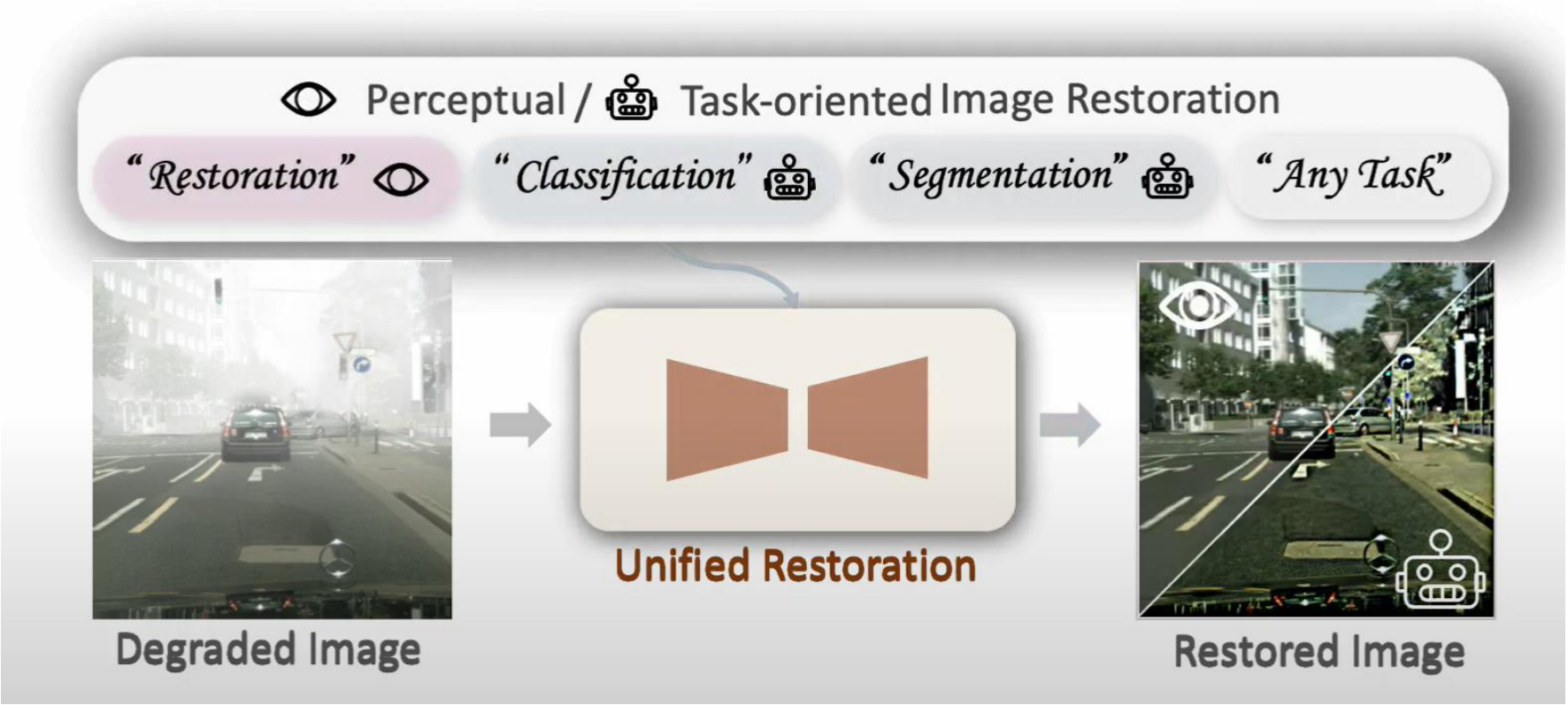


(b) Task-oriented IR

LQ PromptIR HQ

Motivation : UniRestore

How can we unify perceptual and task-oriented objectives in a single model?

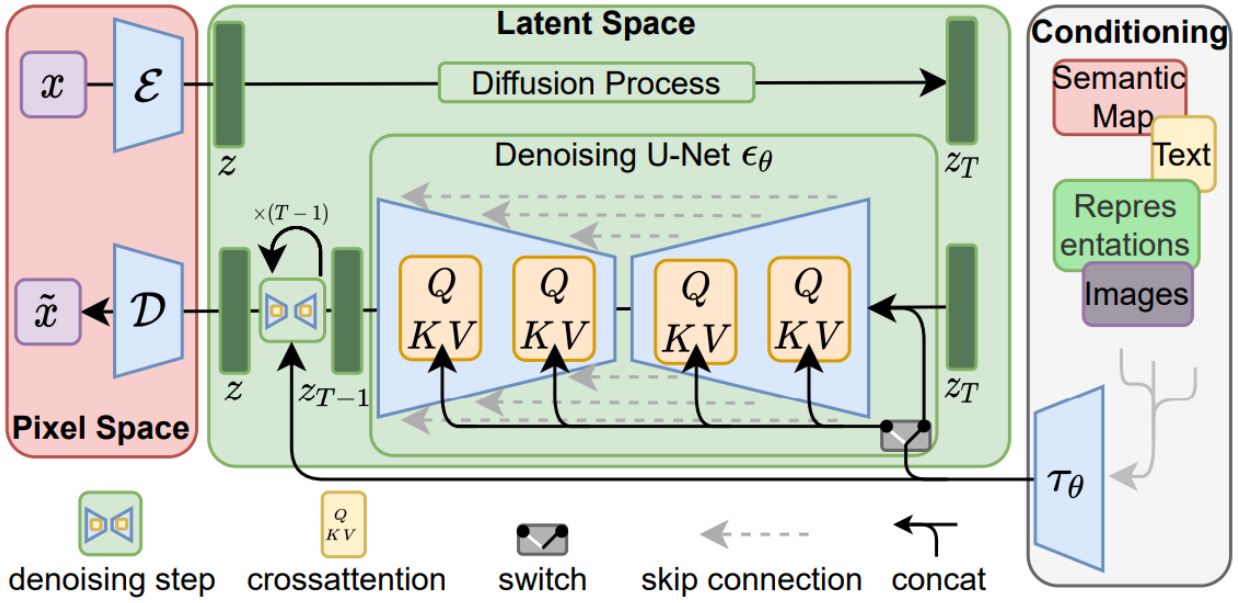


Diffusion Prior

Enhance feature for downstream task

Background : Diffusion Prior

Stable Diffusion



$$\mathbf{Z}_t = \sqrt{\overline{\alpha}_t}\mathbf{Z}_0 + \sqrt{1-\overline{\alpha}_t}\epsilon,\ \epsilon \in \mathcal{N}(0,\mathbf{I})$$

$$\begin{cases} \mathbf{Q} = \mathbf{W}_Q\phi(\mathbf{Z}_t);\ \mathbf{K} = \mathbf{W}_K\tau(\mathbf{y});\ \mathbf{V} = \mathbf{W}_V\tau(\mathbf{y}) \\ Attention(\mathbf{Q},\mathbf{K},\mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})\cdot\mathbf{V}, \end{cases}$$

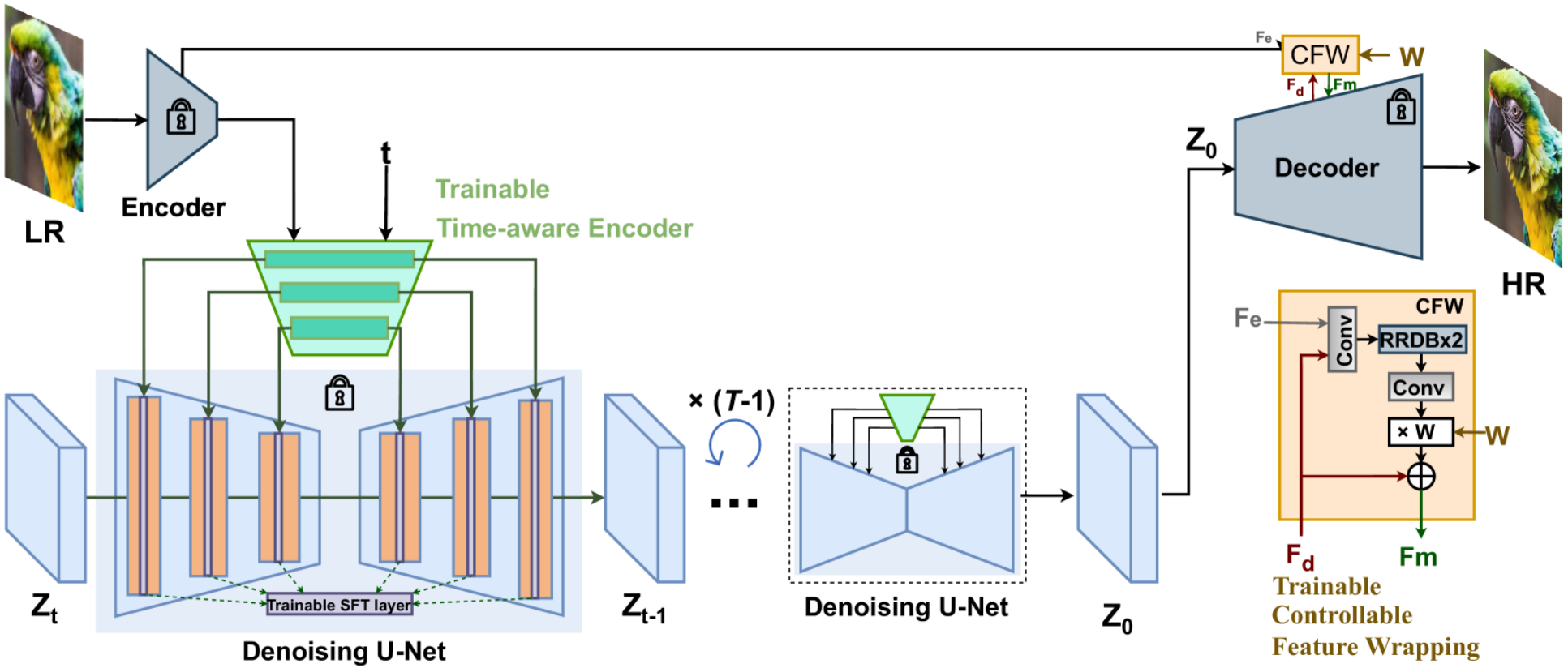
$$\mathcal{L} = \mathbb{E}_{\mathbf{Z}_t,\mathbf{C},\epsilon,t}(||\epsilon - \epsilon_\theta(\mathbf{Z}_t,\mathbf{C})||_2^2),$$

Robin Rombach, et al. High-resolution image synthesis with latent diffusion models, CVPR, 2022

Background : Diffusion Prior

北京大学 PEKING UNIVERSITY

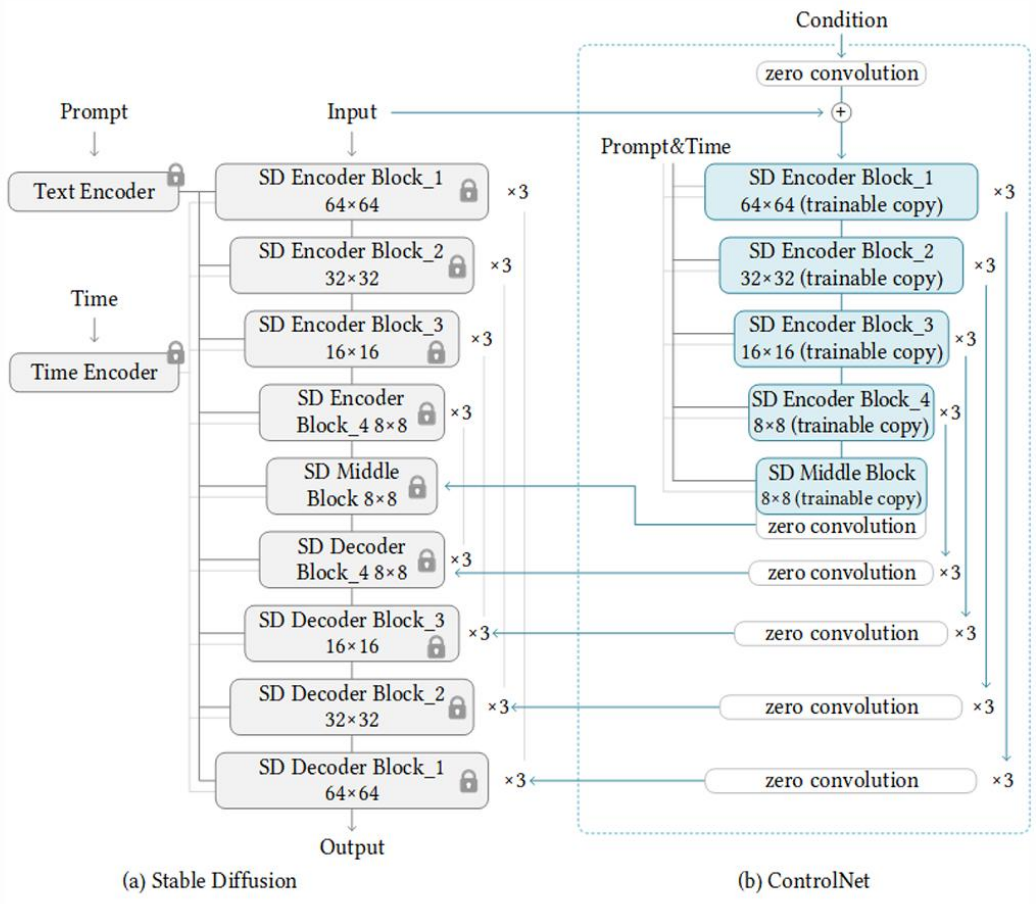
StableSR

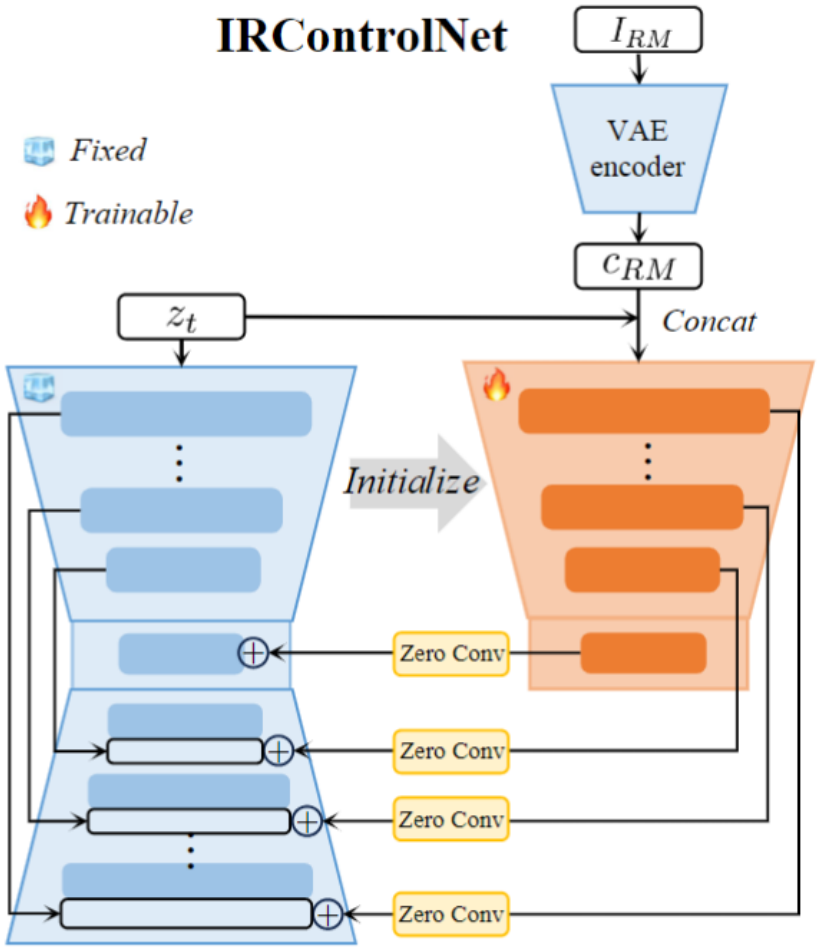


Jianyi Wang, et al. Exploiting Diffusion Prior for Real-World Image Super-Resolution, IJCV, 2024

Background : Diffusion Prior

DiffBIR





Background : Diffusion Prior

北京大学 PEKING UNIVERSITY

DiffBIR

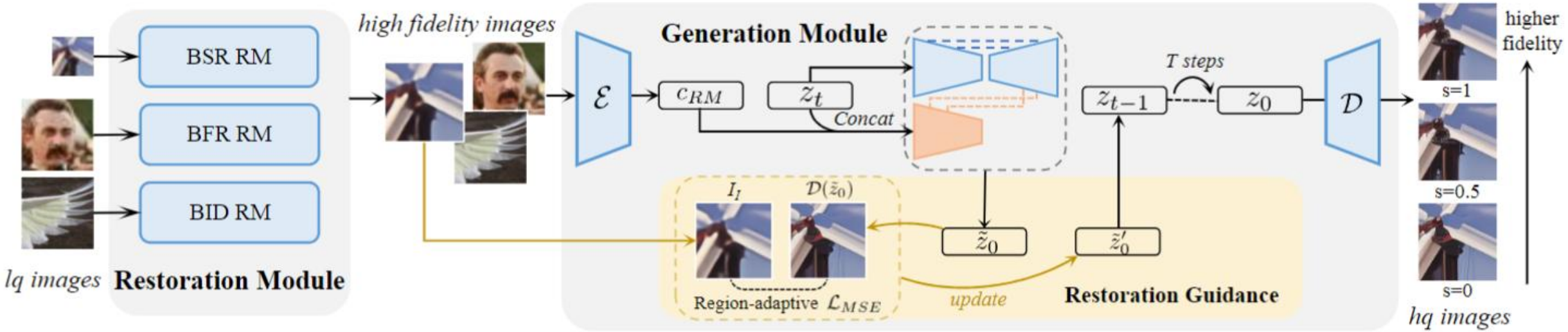


Figure 3. The two-stage pipeline of DiffBIR. 1) Restoration Module (RM) for degradation removal; 2) Generation Module (GM) for realistic image reconstruction with optional region-adaptive restoration guidance for a trade-off between *quality* and *fidelity*.

Xinqi Lin, et al. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior, ECCV, 2024

Overview

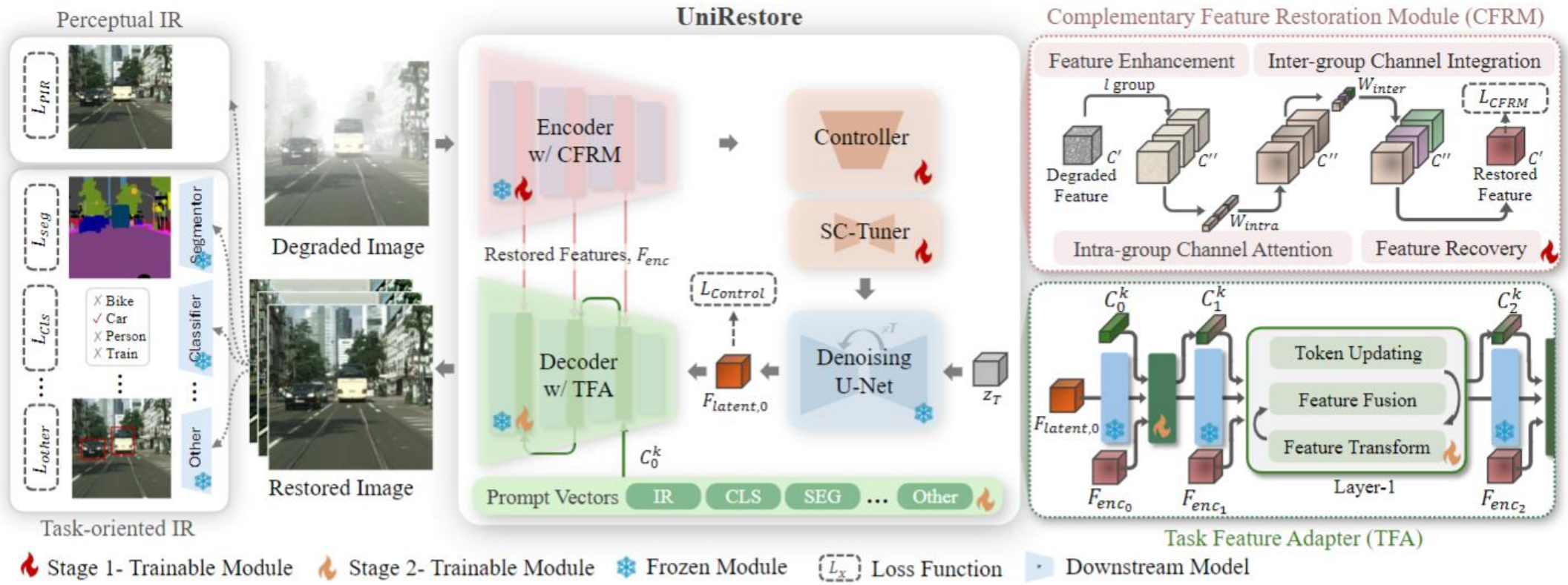
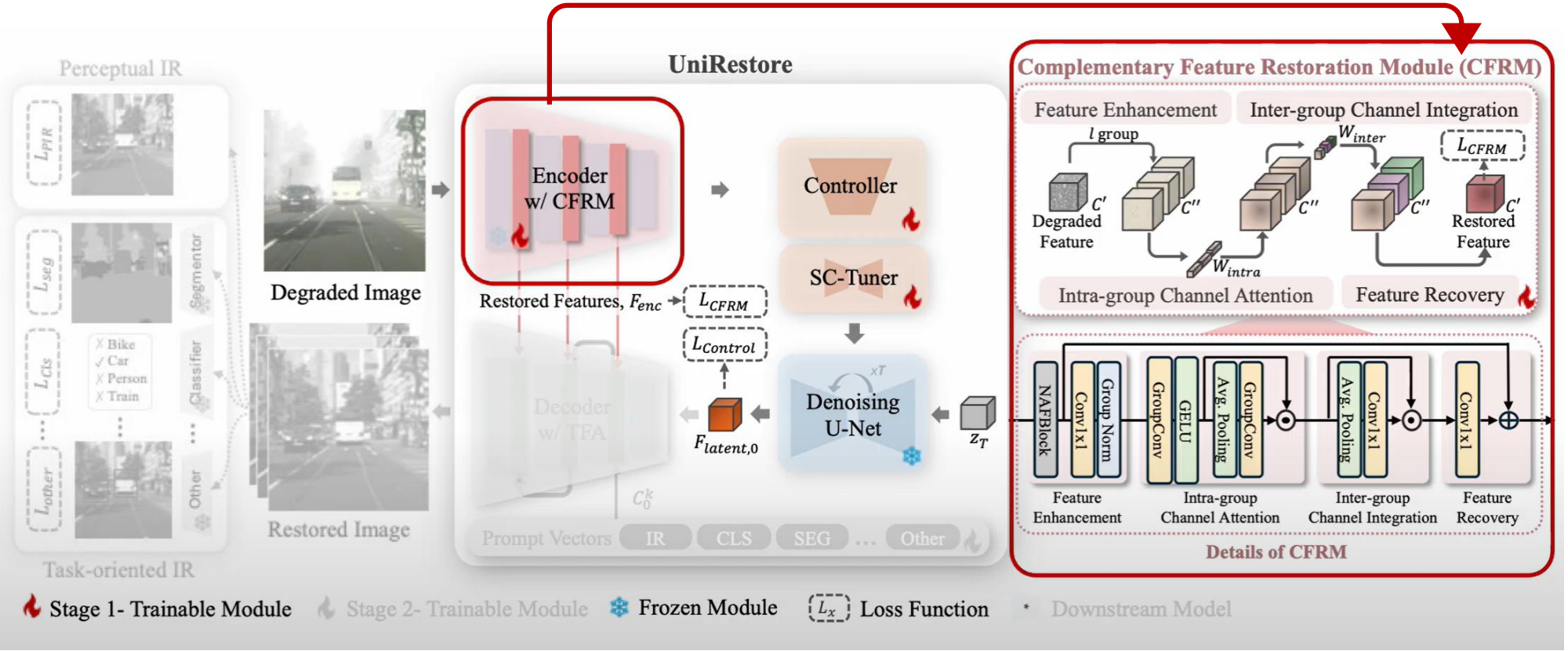


Figure 2. **Overview of UniRestore.** UniRestore augments the diffusion model by incorporating CFRMs and TFAs within the pre-trained autoencoder. The training process is divided into two stages: In the first stage, CFRM, Controller, and SC-Tuner are trained to restore clear encoder and latent features. In the second stage, the TFA is trained to adapt the restored encoder features and latent features for various downstream tasks, using task-specific prompts at the decoder to control the output restoration.

Method

Stage1: Complementary Feature Restoration Module



1) restore and enhance features within the encoder
2) provide complementary inputs to the decoder

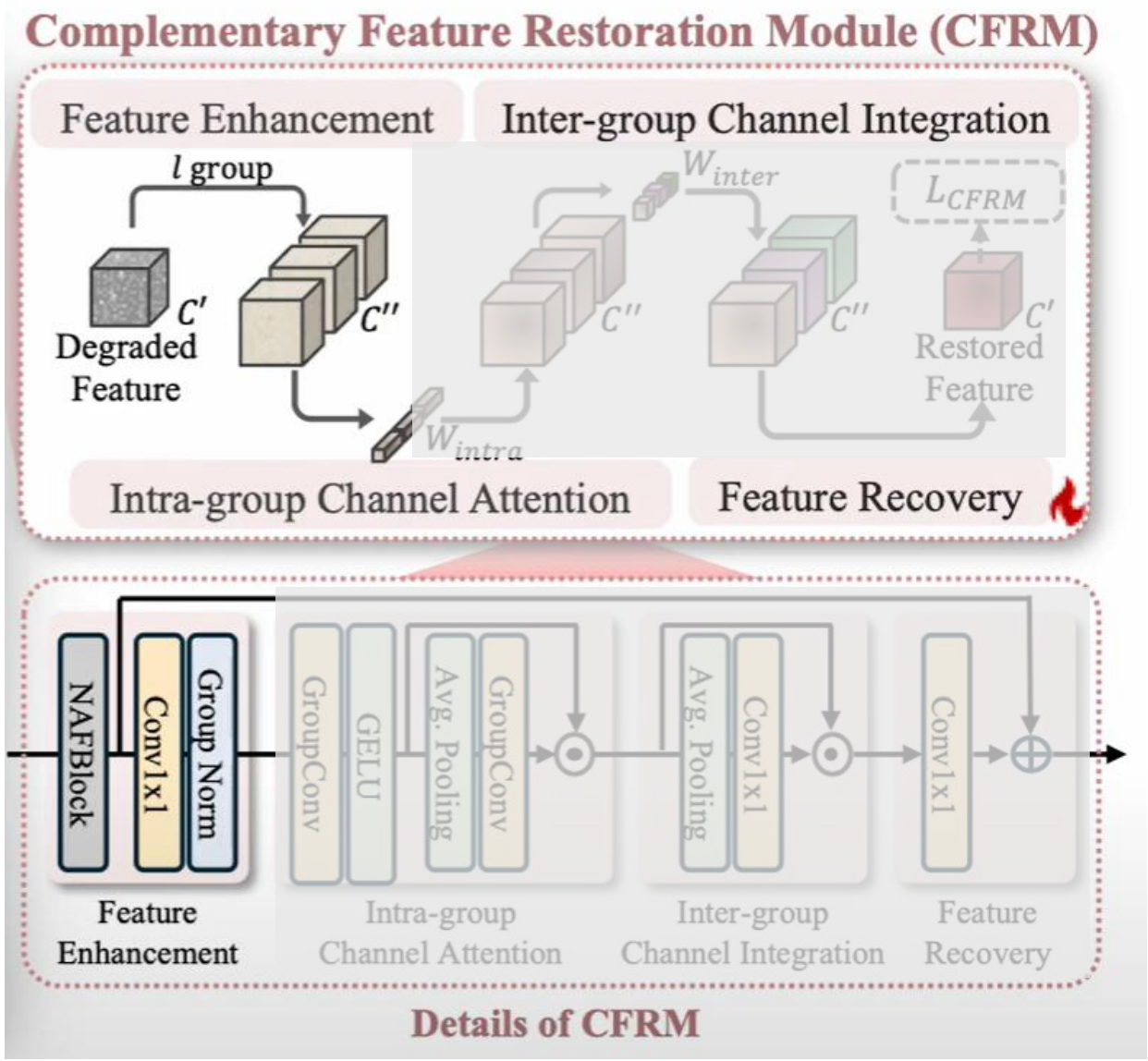
Method

北京大学 PEKING UNIVERSITY

Stage1: CFRM

1. **Feature Enhancement**
2. Intra-group Channel Attention
3. Inter-group Channel Integration
4. Feature Recovery

- NAFBlock provides initial enhancements.
- The extended channel increases expressive capability.
- Grouping prepares for subsequent intra-group/inter-group attention.



Method

NAFBlock

GELU: $GELU(x) = x\Phi(x) \approx 0.5x(1 + tanh[\sqrt{2/\pi}(x + 0.044715x^3)])$

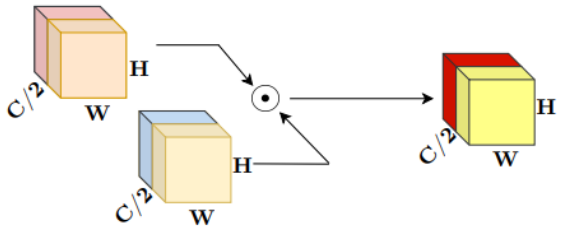
GLU: $Gate(\mathbf{X}, f, g, \sigma) = f(\mathbf{X}) \odot \sigma(g(\mathbf{X}))$,

$\downarrow$

$Gate(\mathbf{X}) = f(\mathbf{X}) \odot g(\mathbf{X})$

$\downarrow$

Simple Gate: $SimpleGate(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \odot \mathbf{Y}$



(c)Simple Gate

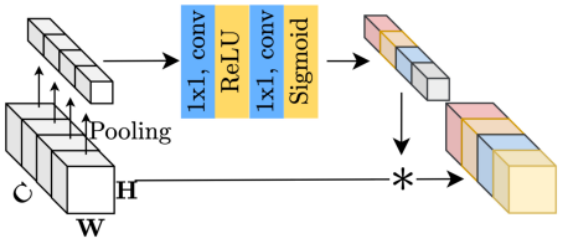
$CA(\mathbf{X}) = \mathbf{X} * \sigma(W_2 max(0, W_1 pool(\mathbf{X})))$

$\downarrow$

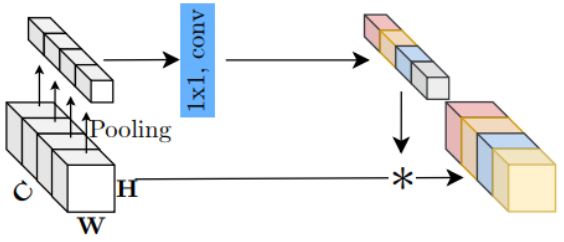
$CA(\mathbf{X}) = \mathbf{X} * \Psi(\mathbf{X})$

$\downarrow$

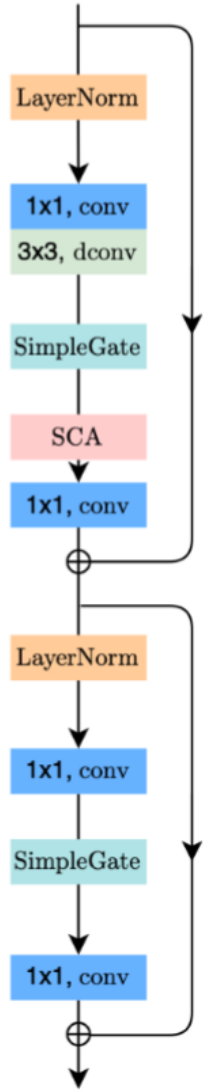
$SCA(\mathbf{X}) = \mathbf{X} * Wpool(\mathbf{X})$



(a)Channel Attention



(b)Simplified Channel Attention



(d)NAFNet's Block(ours)

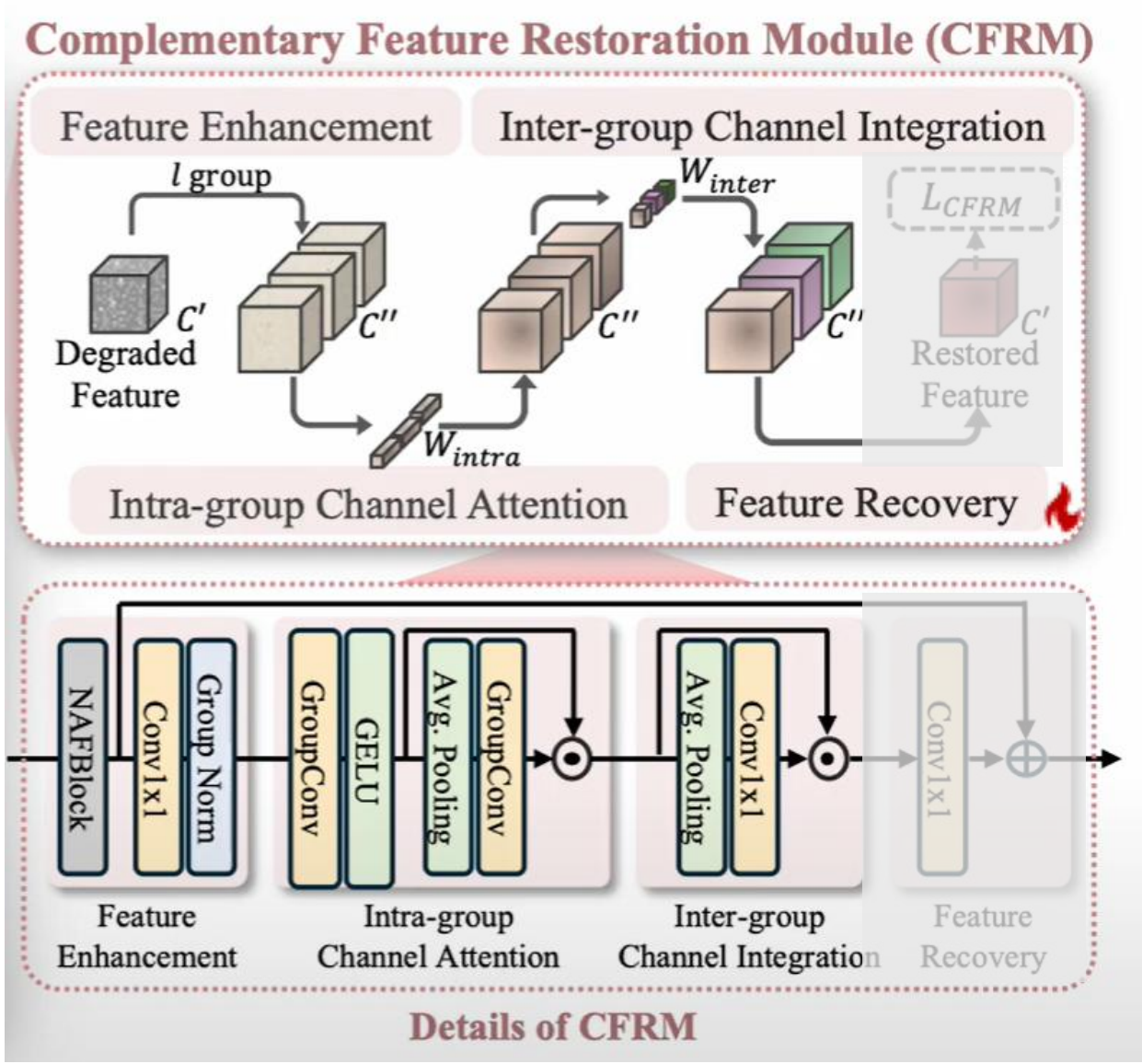
Liangyu Chen, et al. Simple Baselines for Image Restoration, ECCV, 2022

Method

Stage1: CFRM

1. Feature Enhancement
2. **Intra-group Channel Attention**
3. **Inter-group Channel Integration**
4. Feature Recovery

- Intra-group attention focuses on local features.
- inter-group integration provides global context.

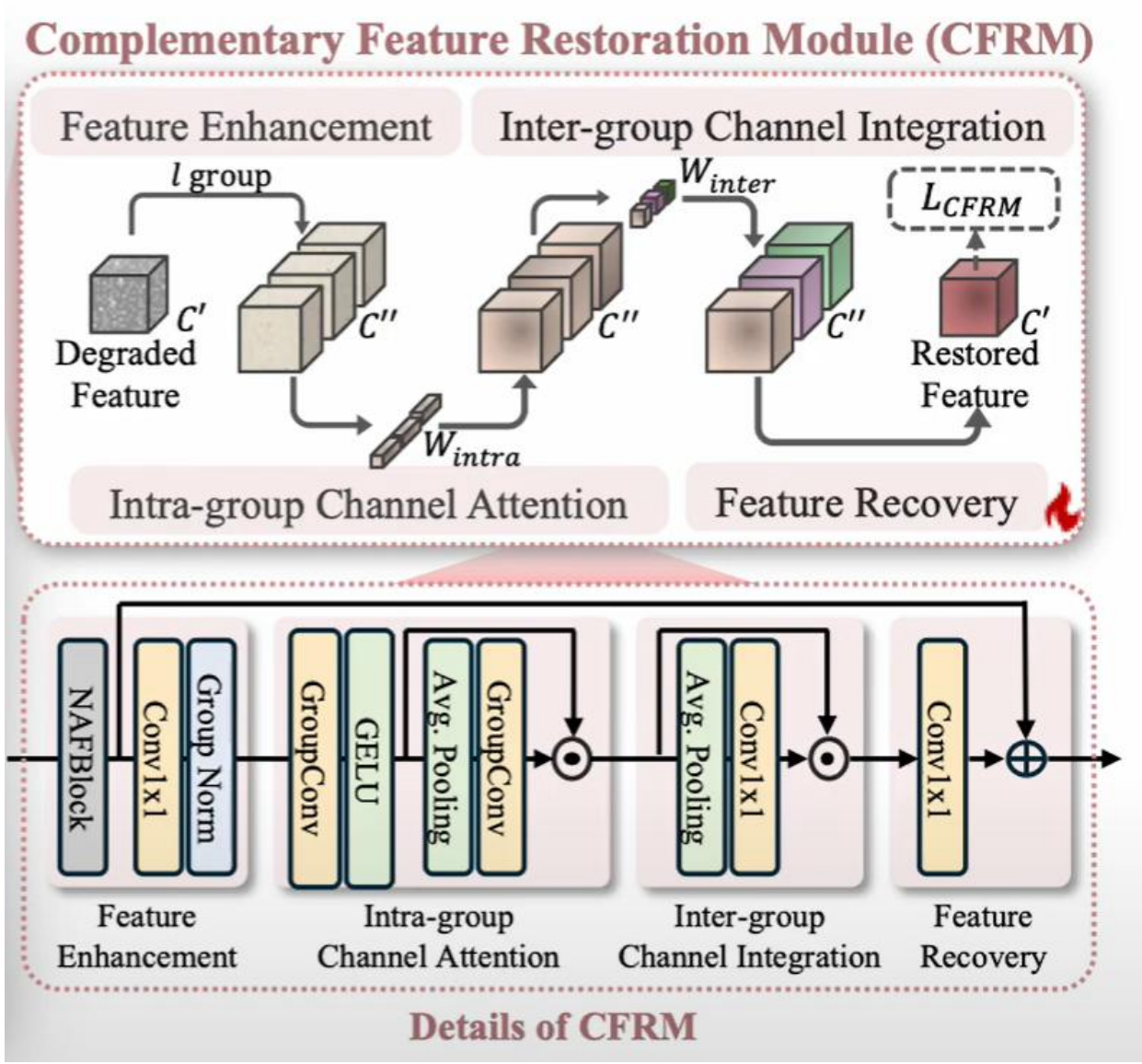


Method

Stage1: CFRM

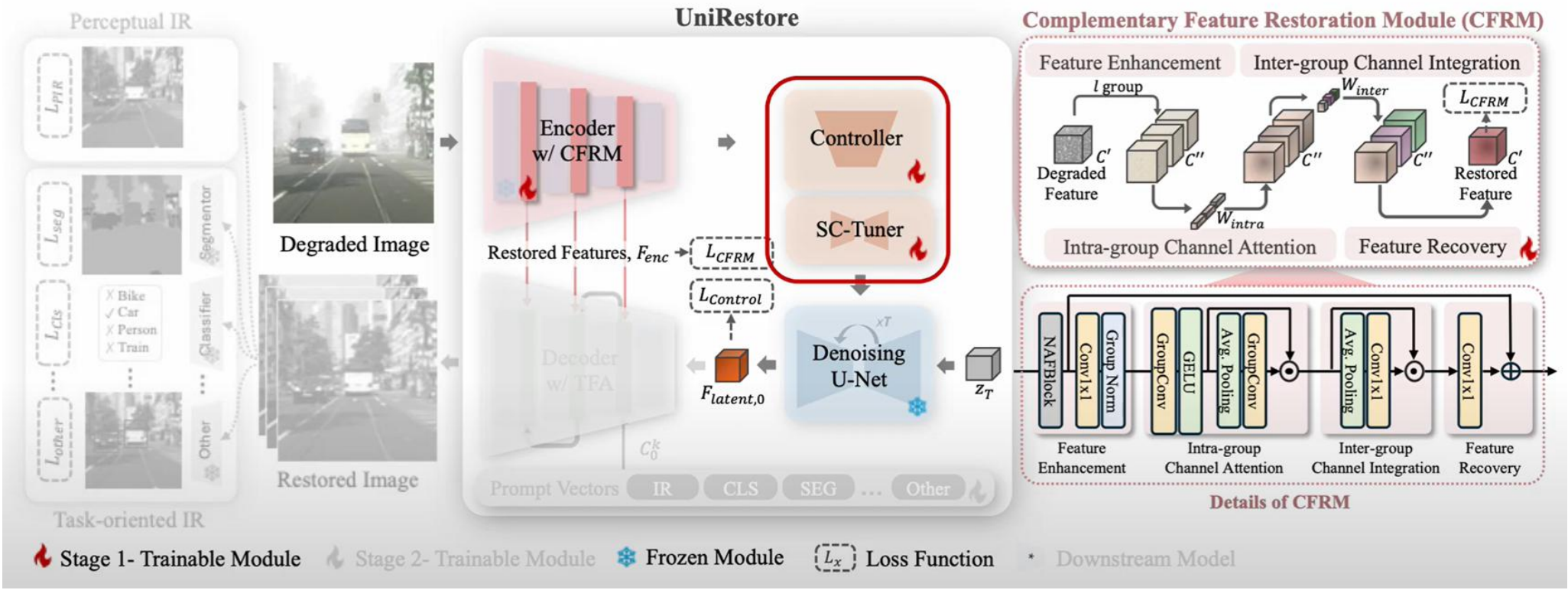
1. Feature Enhancement
2. Intra-group Channel Attention
3. Inter-group Channel Integration
4. **Feature Recovery**

- Refined features
- Weighted information complements the original information.



Method

Stage1.5: Controller & SC-Tuner

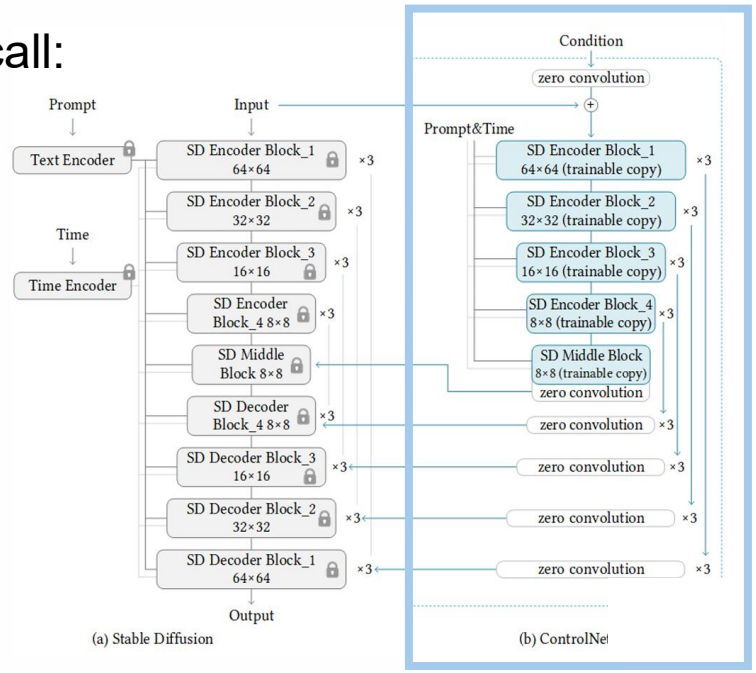


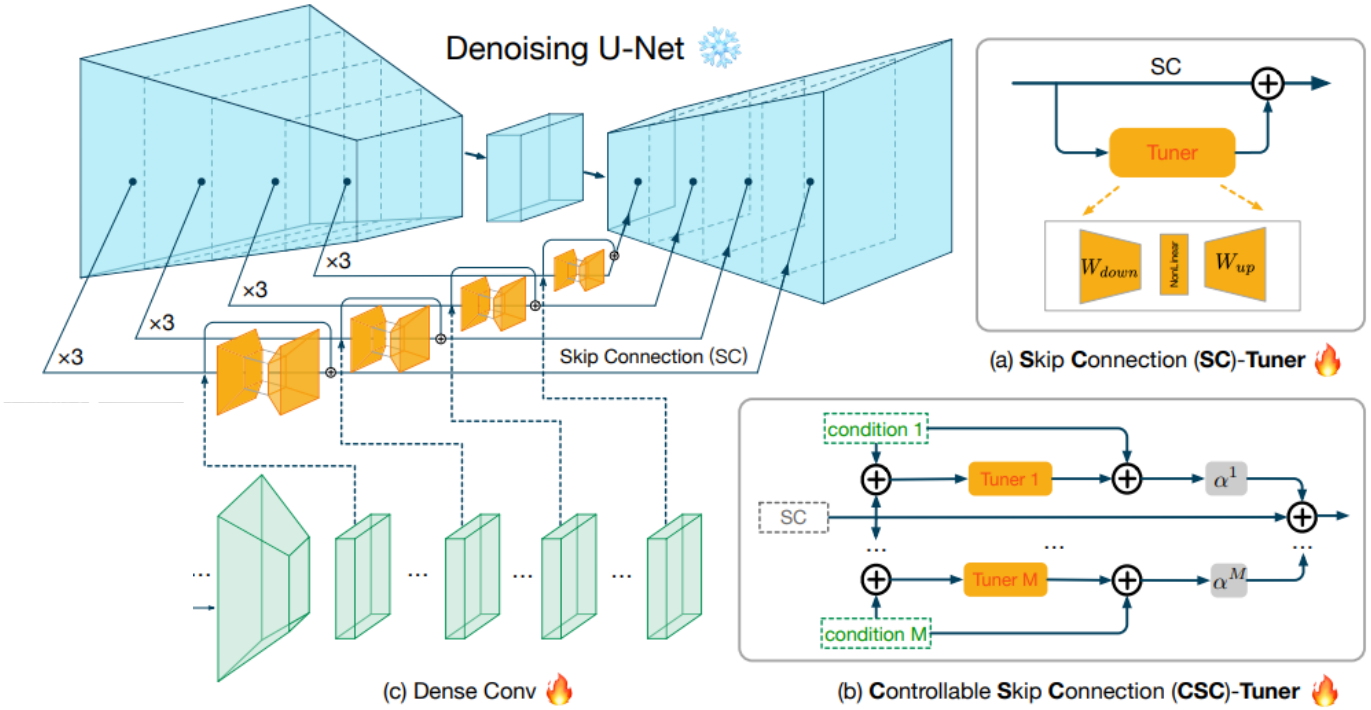
Latent features are then fed into the Controller, equipped with an SC-Tuner.

Method

Stage1.5: Controller & SC-Tuner

Recall:



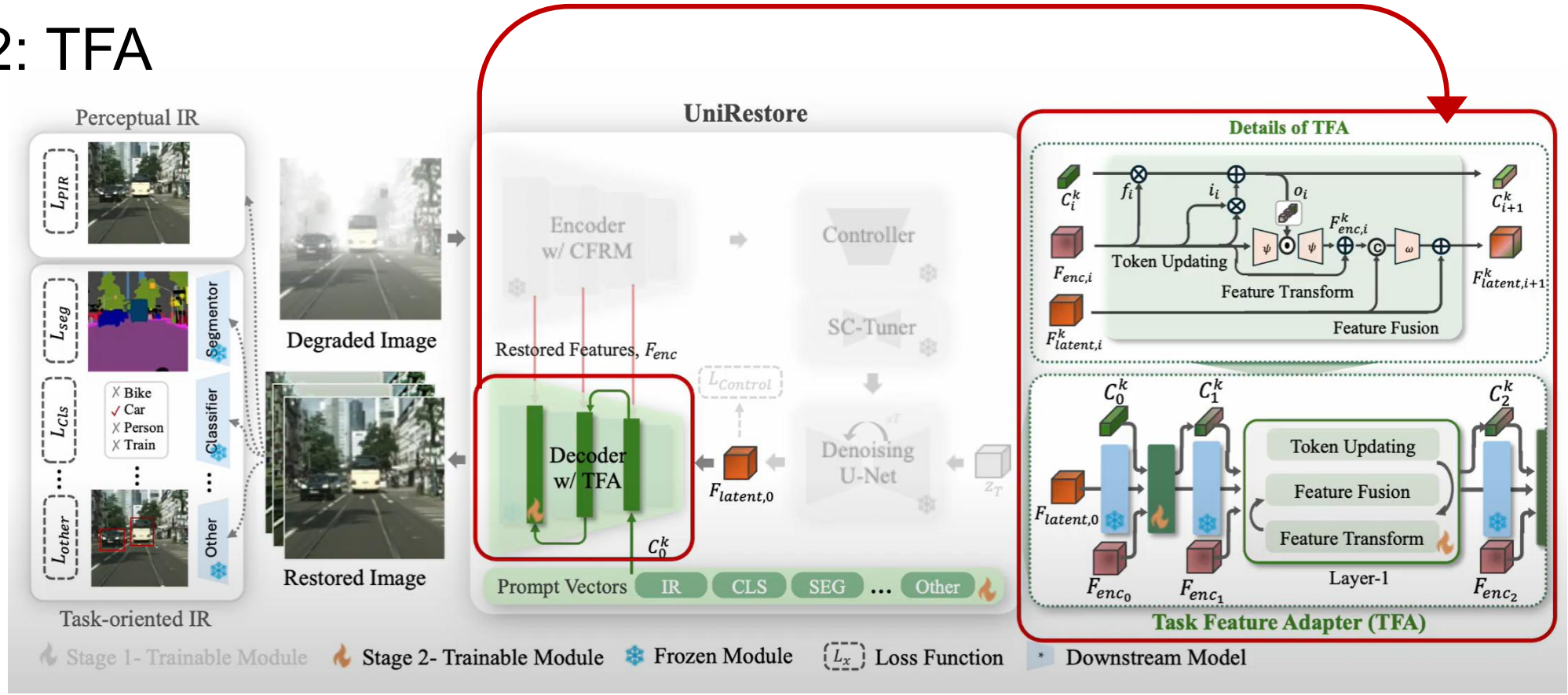


CFRM → Controller

Zeyinzi Jiang , et al. SCEdit: Efficient and Controllable Image Diffusion Generation via Skip Connection Editing, CVPR Highlight, 2024

Method

Stage2: TFA



Draw inspiration from prompt tuning and LSTM.

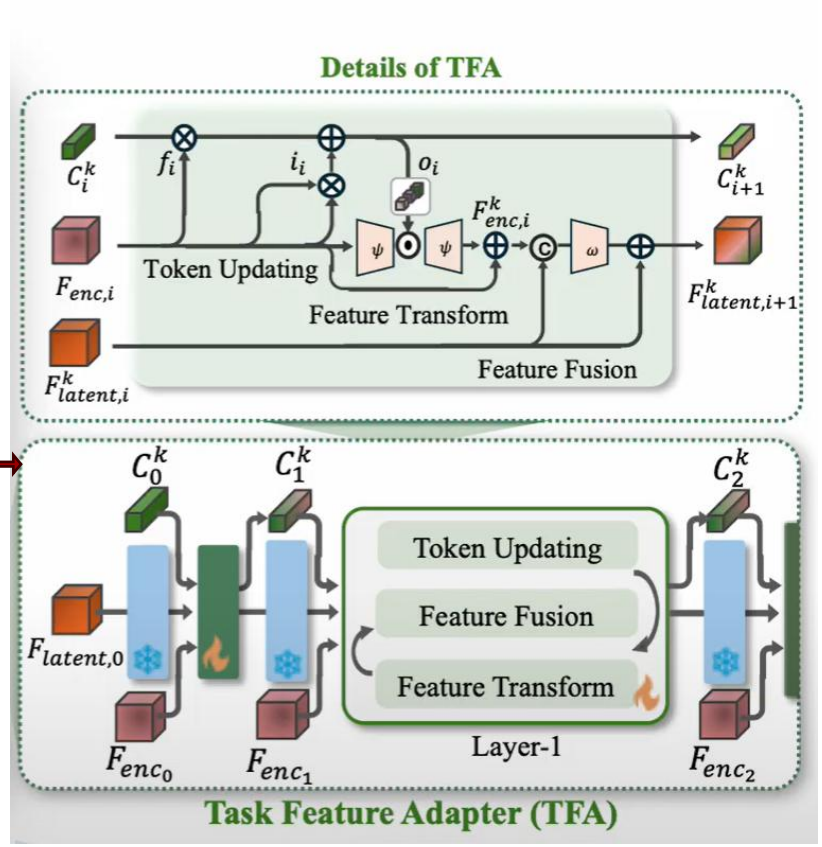
Initialize prompt vector: C_0^k

CFRM features in each encoder layer $F_{\text{enc},i}$

Decoder's output features in each layer i: $F_{\text{latent},i}^k$

Method

Stage2: TFA



Initialize prompt vector: C_0^k

CFRM features in each ecoder layer i: $F_{\text{enc},i}$

Decoder's output features in each layer i: $F_{\text{latent},i}^k$

Forget Gate: $$f_i = \sigma\left(\vartheta\left(F_{\text{enc},i}, \theta_{f,i}\right)\right)$$

Input Gate: $$i_i = \sigma\left(\vartheta\left(F_{\text{enc},i}, \theta_{i,i}\right)\right)$$

Token Update: $$C_{i+1}^k = f_i \otimes C_i^k + i_i \otimes \tanh\left(\vartheta\left(F_{\text{enc},i}, \theta_{c,i}\right)\right)$$

Output Gate: $$o_i = \tanh(\xi(C_{i+1}^k, \theta_{o,i}))$$

Feature Transform &Fusion: $$F_{\text{enc},i}^k = \psi(F_{\text{enc},i}, o_i, \theta_{t,i}) + F_{\text{enc},i}$$

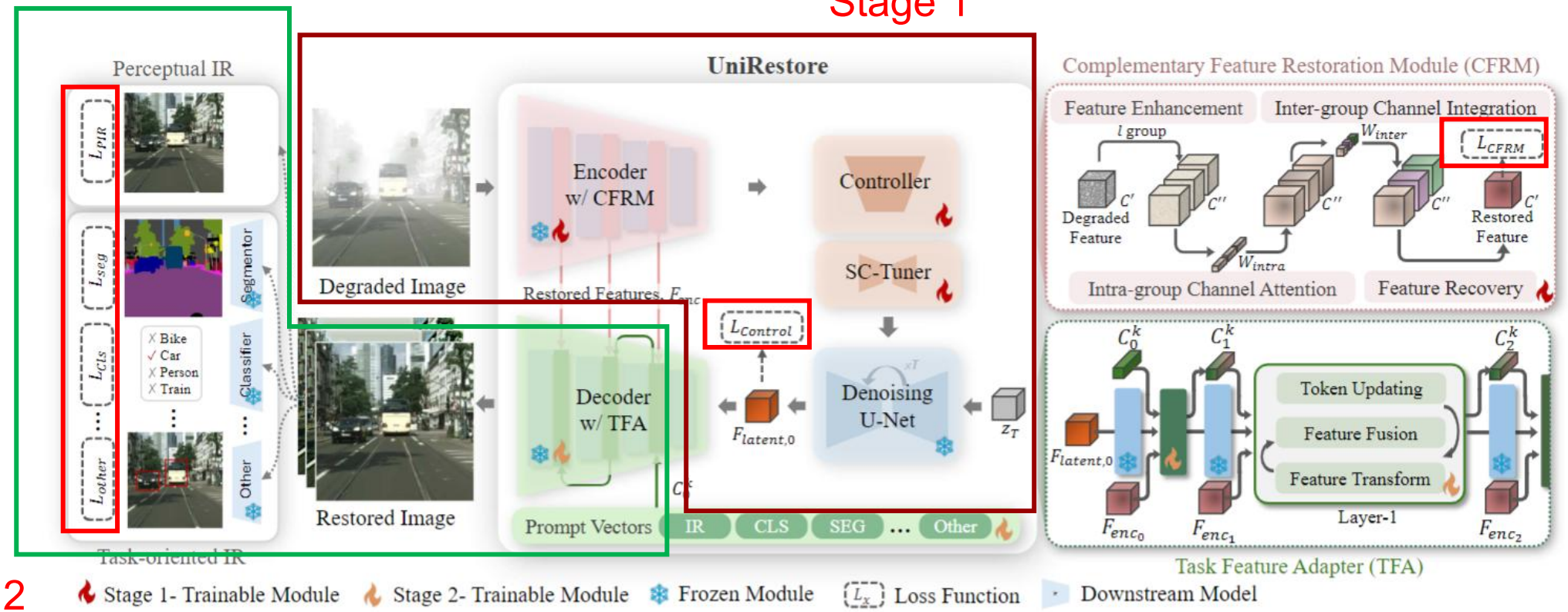
$$F_{\text{latent},i+1}^k = \omega((F_{\text{enc},i}^k, F_{\text{latent},i}^k), \theta_{l,i}) + F_{\text{latent},i}^k$$

Method

北京大学 PEKING UNIVERSITY

Training Pipeline – Two Stage

Stage 1



Stage 2

Stage 1- Trainable Module Stage 2- Trainable Module Frozen Module L_x Loss Function Downstream Model

$$\mathcal{L}_{\text{CFRM}} = \sum_{i=1}^{M} \lambda_i (f_i^{\text{Clear}} - f_i^{\text{Restored}})$$

$$\mathcal{L}_{\text{Control}} = \|z_0 - \hat{z}_0^t\|_2^2$$

$$\mathcal{L}_{\text{Stage 2}} = \sum_{i=1}^{N} \beta_{\text{Task}}^i \mathcal{L}_{\text{Task}}^i$$

Experiments

TIR methods：DIP 、URIE

PIR methods：NAFNet 、PromptIR

Diffusion-based approaches：DiffBIR 、DiffUIR

Two settings

#1 **Original objective**:

Trained only for their intended purpose, denoted as “method”.

#2 **UniRestore**:

Trained on the PIR training set and then fine-tuned on multiple downstream tasks, as “method*”.

Experiments

PIR Training Datasets:

- a blend of the DIV2K , Flickr2K , and OST datasets

PIR Testing Datasets:

- seen: test set of DIV2K
- unseen: Rain100L, RESIDE, UHDSnow, GoPro,
  'Noise' comprising Urban100, BSD68, CBSD68, Kodak,
  McMaster, Set12

PIR Evaluation:

- PSNR & SSIM

Experiments

北京大学 PEKING UNIVERSITY

PIR Performance

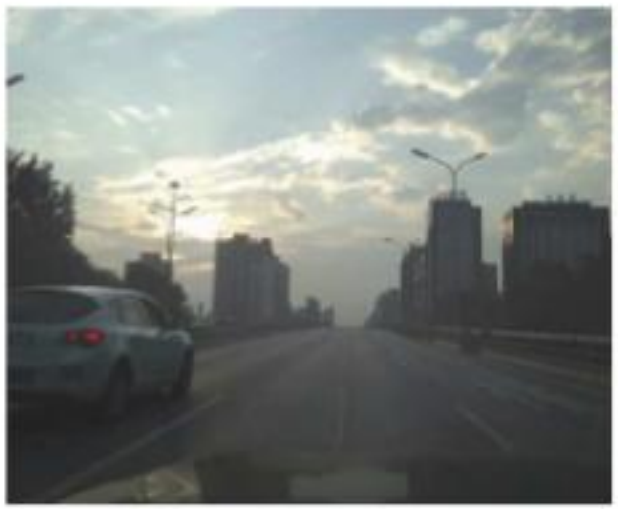
Methods	*Seen Dataset* DIV2K [1]		*Unseen Datasets* Rain100L [67]		RESIDE [26]		UHDSnow [59]		Noise [20, 37]		GoPro [38]		*Average*	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DIP [34]	18.47	0.5810	22.65	0.7884	21.30	0.7819	19.03	0.8089	15.41	0.2494	23.08	0.8041	17.13	0.5734
DIP* [34]	18.62	0.5516	23.16	0.8097	19.83	0.7586	16.77	0.7830	14.51	0.2328	21.05	0.7624	16.28	0.5569
URIE [52]	17.72	0.5202	20.97	0.7293	18.30	0.7449	18.11	0.7626	18.57	0.5180	19.21	0.5683	18.81	0.6406
URIE* [52]	17.98	0.5967	19.97	0.6993	20.37	0.7694	16.18	0.7526	17.41	0.3624	18.57	0.4624	18.41	0.6071
NAFNet [2]	22.23	0.7905	24.57	0.8178	25.13	0.8632	20.71	0.8672	23.22	0.6951	22.18	0.8042	23.01	0.8063
NAFNet* [2]	19.81	0.7005	20.51	0.7314	21.24	0.8178	18.39	0.7958	20.38	0.6019	19.79	0.7293	20.02	0.7295
PromptIR [42]	23.90	0.8321	28.17	0.9034	27.26	0.8957	22.10	0.8877	23.72	0.7269	23.93	0.8221	24.85	0.8447
PromptIR* [42]	21.94	0.7421	24.76	0.8134	24.16	0.8317	19.13	0.8265	19.68	0.6283	20.18	0.7657	21.64	0.7680
DiffBIR [33]	22.76	0.8053	27.25	0.8695	26.97	0.8770	20.84	0.8785	23.67	0.7661	23.49	0.8076	24.16	0.8340
DiffBIR* [33]	18.32	0.6847	23.48	0.8143	23.13	0.8068	18.29	0.8167	21.59	0.6419	20.13	0.7413	20.82	0.7510
DiffUIR [77]	23.79	0.8397	28.25	0.9154	27.12	0.8820	20.74	0.8753	24.27	0.7481	23.93	0.8241	24.68	0.8474
DiffUIR* [77]	21.47	0.7742	25.44	0.8276	23.58	0.8174	18.62	0.8318	22.76	0.6691	21.71	0.7649	22.26	0.7808
UniRestore	**24.32**	**0.8434**	**30.02**	**0.9237**	**27.91**	**0.9043**	**23.44**	**0.8943**	**24.37**	**0.7811**	**25.94**	**0.8541**	**26.00**	**0.8668**

Table 1. **Performance comparison of existing methods on one seen and five unseen PIR datasets.**

TIR PIR Diffusion-based

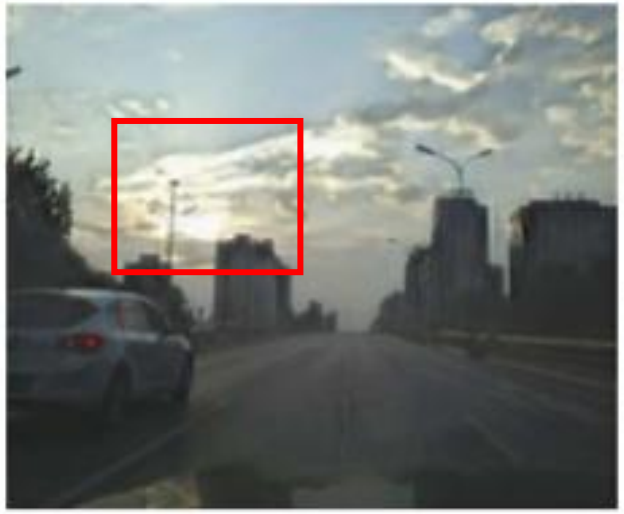
Experiments

PIR Performance

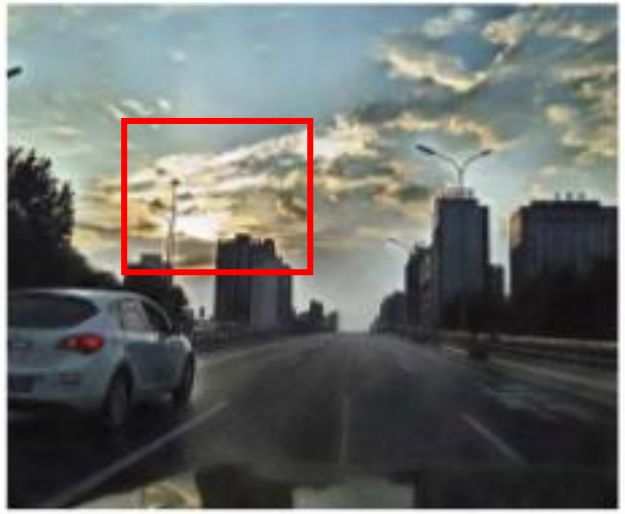


LQ

HQ

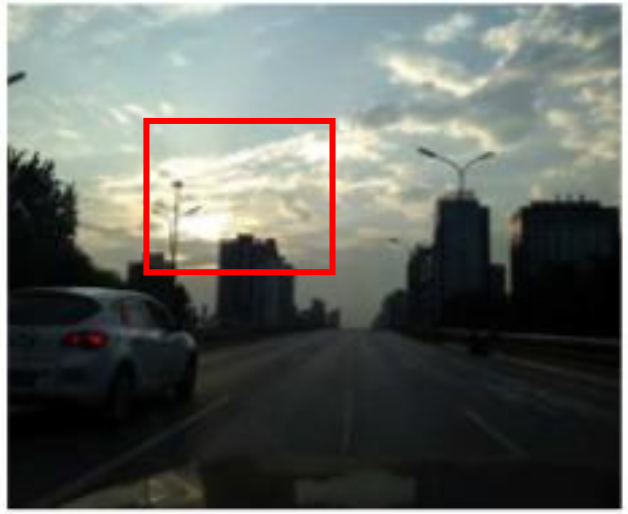


URIE



PromptIR

DiffUIR



UniRestore

TIR

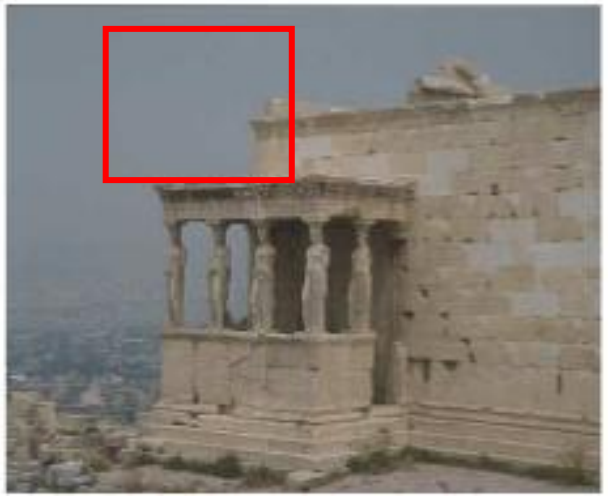
PIR

Diffusion-based

Experiments

PIR Performance

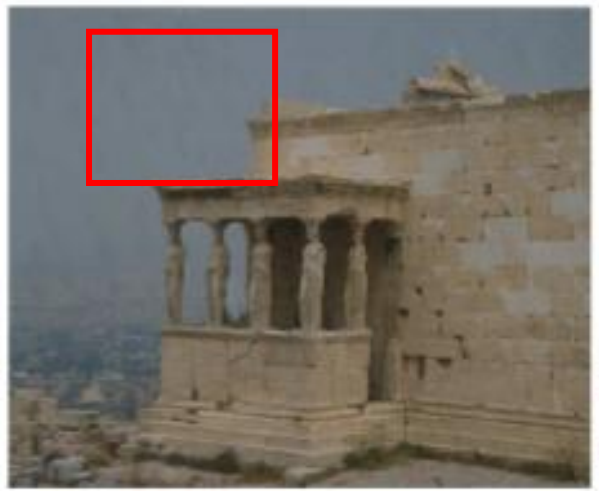
LQ



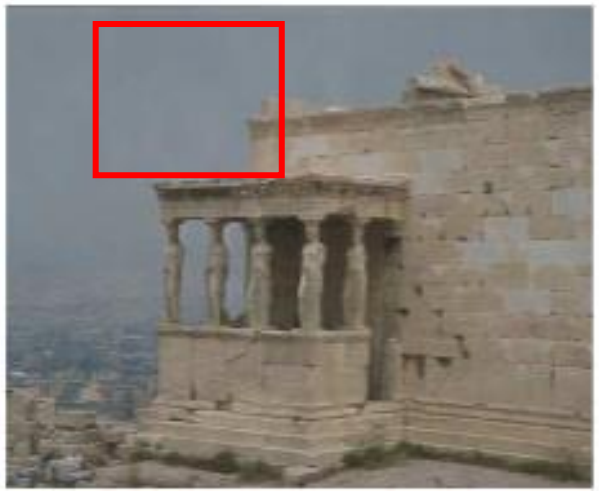
HQ

URIE

PromptIR



DiffUIR



UniRestore

TIR PIR Diffusion-based

Experiments

北京大学 PEKING UNIVERSITY

TIR Training Datasets：

- image classification (randomly select 80,000 images from ImageNet)
- semantic segmentation (Cityscapes datasets)
- synthesized with 15 types of degradation

TIR Testing Datasets：

- seen： test set of ImageNet, Cityscapes
- unseen： CUB, ACDC

TIR Evaluation：

- image classification (ACC)
- semantic segmentation (mIoU)

Experiments

北京大学 PEKING UNIVERSITY

TIR Performance

TIR for image classification.

Inputs	Seen Dataset ImageNet [12]		Unseen Dataset CUB [56]	
	ResNet-50 [15] ↑	ViT-B [14] ↑	ResNet-50 [15] ↑	ViT-B [14] ↑
LQ	*51.75*	*67.65*	*33.69*	*44.83*
DIP [34]	61.55	72.05	47.91	54.10
DIP* [34]	59.80	70.35	45.99	52.48
URIE [52]	66.65	73.95	49.64	57.24
URIE* [52]	65.20	72.15	46.89	54.93
NAFNet [2]	60.35	70.80	46.47	53.82
NAFNet* [2]	57.65	68.25	43.17	51.88
PromptIR [42]	65.25	73.90	49.52	58.04
PromptIR* [42]	64.05	73.00	48.52	57.39
DiffBIR [33]	59.30	68.05	41.68	52.38
DiffBIR* [33]	57.55	66.85	40.65	51.34
DiffUIR [77]	62.35	72.10	46.75	57.28
DiffUIR* [77]	61.15	71.60	45.44	56.31
UniRestore	**71.65**	**77.05**	**53.70**	**60.79**
HQ	*72.80*	*78.70*	*58.22*	*64.41*

TIR for semantic segmentation.

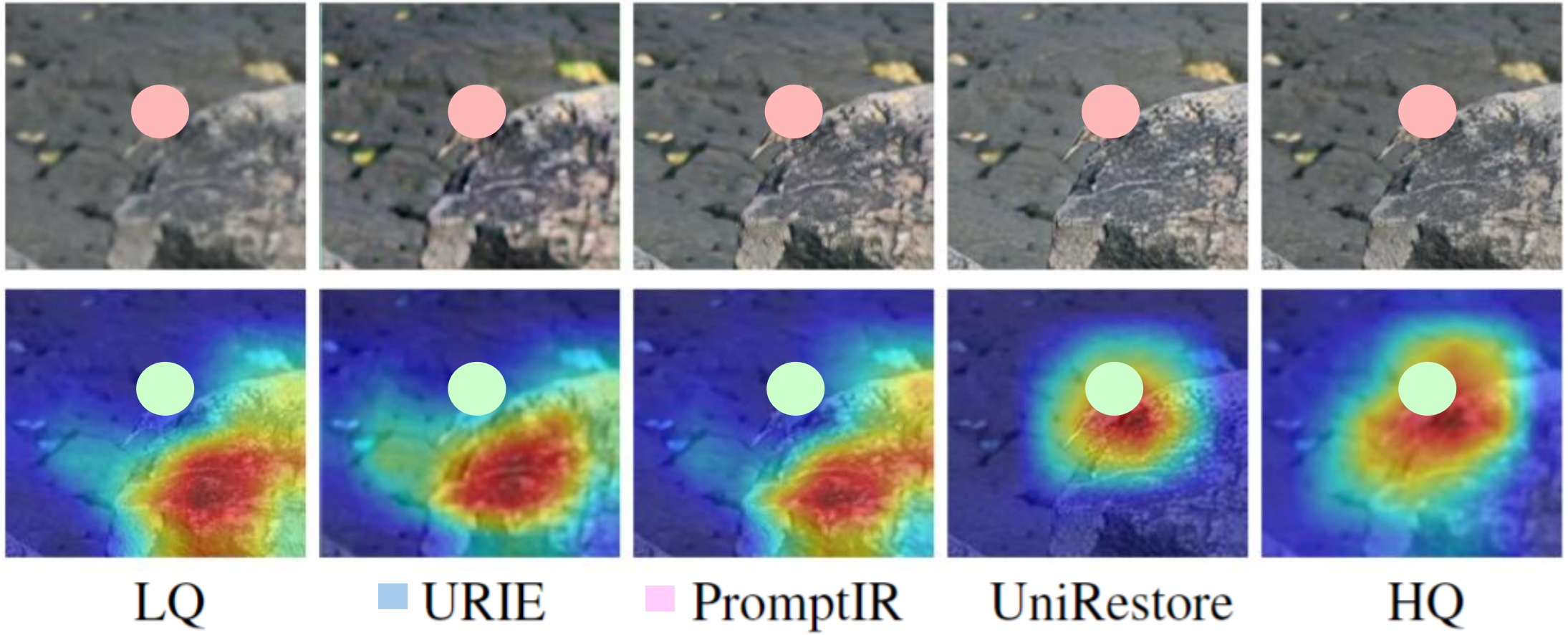
Inputs	Seen Dataset Cityscapes [47]		Seen Dataset FoggyCityscapes [47]	Unseen Dataset ACDC [48]
	DeepLabv3+ [3]	RefineNet-lw [39]	RefineNet-lw [39]	RefineNet-lw [39]
LQ	*40.36*	*40.75*	*65.20*	*28.30*
DIP [34]	57.17	57.67	67.81	38.19
DIP* [34]	51.81	50.35	67.16	32.98
URIE [52]	55.88	51.45	65.93	37.90
URIE* [52]	50.56	48.23	65.93	32.71
NAFNet [2]	58.41	58.19	66.06	37.59
NAFNet* [2]	51.91	53.29	65.40	36.03
PromptIR [42]	58.05	57.54	66.76	37.86
PromptIR* [42]	54.67	52.25	63.44	35.51
DiffBIR [33]	52.49	53.68	66.29	36.28
DiffBIR* [33]	48.90	48.56	63.26	33.12
DiffUIR [77]	51.28	51.46	66.24	35.78
DiffUIR* [77]	47.92	45.01	62.82	34.83
UniRestore	**66.05**	**65.73**	**70.77**	**39.27**
HQ	*75.64*	*75.66*	*75.66*	-

TIR PIR Diffusion-based

Experiments

北京大学 PEKING UNIVERSITY

TIR Performance : Classification

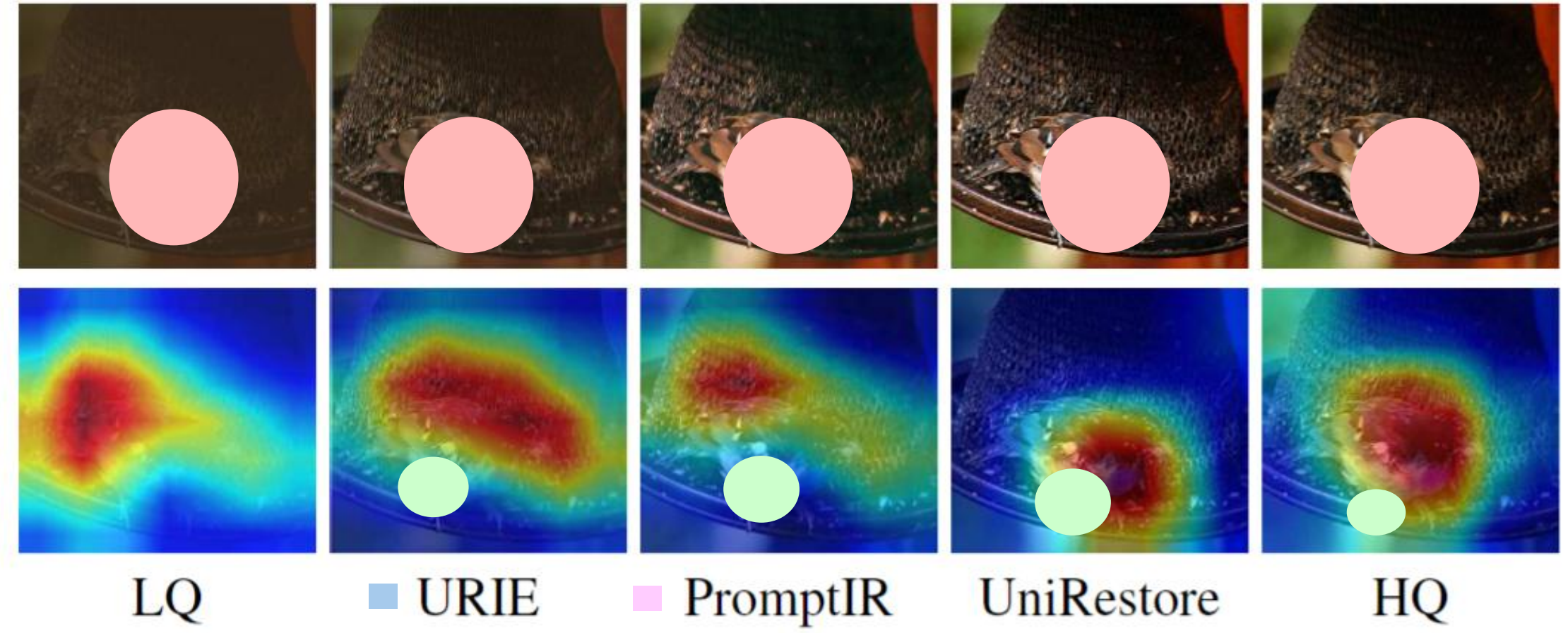


TIR PIR Diffusion-based

Experiments

北京大学 PEKING UNIVERSITY

TIR Performance : Classification

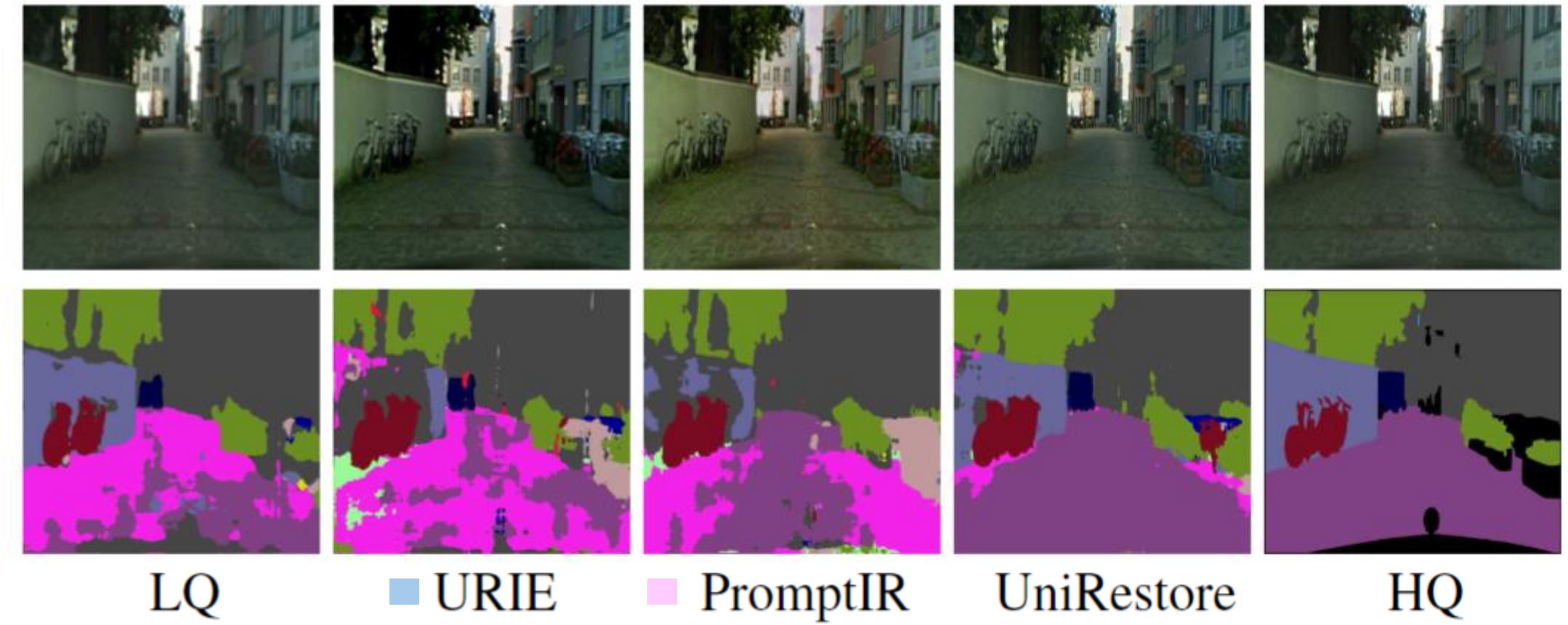


TIR PIR Diffusion-based

Experiments

北京大学 PEKING UNIVERSITY

TIR Performance : Segmentation

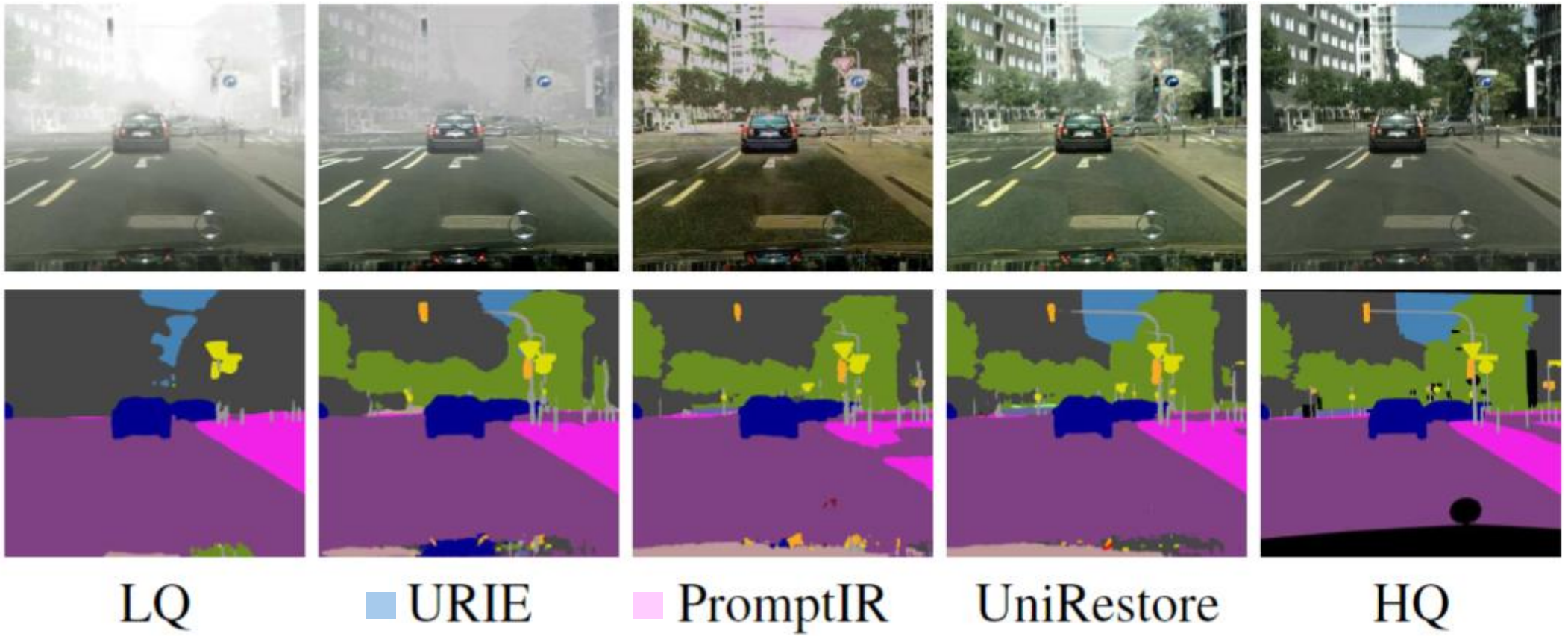


TIR PIR Diffusion-based

Experiments

北京大学 PEKING UNIVERSITY

TIR Performance : Segmentation



TIR PIR Diffusion-based

Ablation Study

北京大学 PEKING UNIVERSITY

Effectiveness of Proposed Modules

Methods	PIR PSNR ↑	Cls ACC ↑	Seg mIoU ↑
Baseline	19.35	57.65	46.76
UniRestore w/o CFRM	21.43	63.10	55.48
UniRestore w/o TFA	22.16	64.25	58.13
UniRestore	**24.32**	**71.65**	**66.05**

Input UniRestore w/o Diffusion UniRestore

Ablation Study

北京大学 PEKING UNIVERSITY

Investigation of TFA

Methods	# of Tuned Parameters	PIR PSNR ↑	Cls ACC ↑	Seg mIoU ↑
Multiple Adapters	65.17M	23.06	68.95	64.64
Multiple TFAs	63.03M	**25.48**	71.20	65.78
UniRestore-SP	21.01M	23.91	70.05	64.99
UniRestore	21.03M	24.32	**71.65**	**66.05**

(i) **Multiple Adapters:** concatenates the output of the denoising U-Net with the restored features from CFRM and processes them through the same number of convolutional blocks as in TFA
(ii) **Multiple TFAs:** optimizes each task with its own TFA
(iii) **UniRestore-SP:** employs a single TFA with a single prompt for all tasks
(iv) **UniRestore:** utilizes one TFA with specific prompts for each task

Ablation Study

北京大学 PEKING UNIVERSITY

Extendability Evaluation

- an additional downstream task—object detection
- based on the model trained for PIR, image classification, and semantic segmentation
- update only with a new learnable prompt, optimizing it using the object detection loss

Method	LQ	DIP [34]	PromptIR [42]	**UniRestore**
mAP ↑	45.63	54.29	50.61	**58.06**

- existing methods that require retraining models on complete task datasets

Takeaway

北京大学 PEKING UNIVERSITY

- UniRestore, an approach capable of addressing PIR and TIR simultaneously.
- Adapting diffusion features for diverse applications.
- Complementary feature restoration module that restores features within the encoder. (inter & intra group)
- Task feature adapter that dynamically and efficiently combines these restored features with diffusion features for downstream tasks. (LSTM)

Thanks for Listening!