Task-Aware Image Restoration using All-In-One Models

2025 Summer Seminar

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
Matti Zinke

Contents

- Background
- InstructIR: High-Quality Image Restoration Following Human Instructions [ECCV 2024]
- Complexity Experts are Task-Discriminitive Learners for Any Image Restoration [CVPR 2025]

Background

- What is Image Restoration?
  - Process that aims to recover clean or original version of an image
    - Remove blur, noise, weather
    - Upscale image, fill in missing pixels
- Why All-in-One models?
  - Previous methods focus either on specific degradations or have general framework but train models task-specific
    - Task-specific training is resource intensive
    - All-in-One restoration models proposed
      - Consider multiple degradations at once

- InstructIR: High-Quality Image Restoration Following Human Instructions [ECCV 2024]

Introduction

- All-in-one model that uses human language to restore images
- Text guidance can help guide blind restoration models better than image-based degradation classification
  - Thought is based on great potential shown by diffusion models using text prompts
  - Users usually have an idea on what is wrong in an image
    - Can use this information to guide the model

< Example model usage >

Method

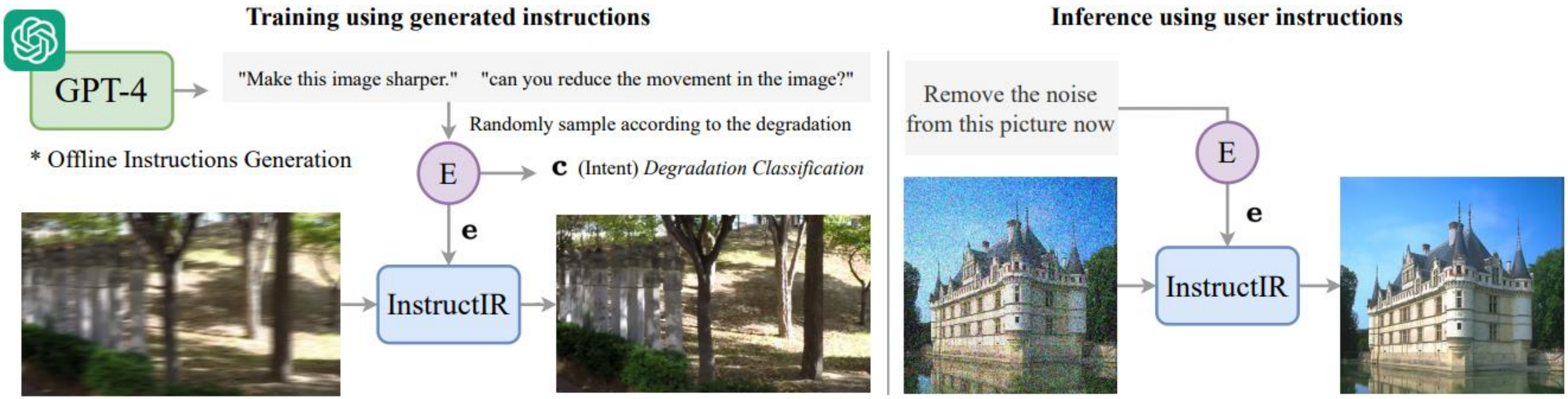
- Degradation prompts dataset
  - Human written instruction offer clear and expressive way to interact
    - Enables to clearly pinpoint degradations
    - Easier than inputting clean images and increases usability
  - Generated over 10.000 different prompts in total
    - Based on example prompts for each of the seven tasks
    - Have different levels of difficulty
      - Mimics different users like kids or experts
    - Filtered out ambiguous or unclear prompts
      - “Improve this image”, “Make the image clearer)

Degradation Prompts			
Denoising	Can you clean the dots from my image? Fix the grainy parts of this photo Remove the noise from my picture	Super-Res.	Make my photo bigger and better Add details to this image Increase the resolution of this photo
Deblurring	Can you reduce the movement in the image? My picture’s not sharp, fix it Deblur my picture, it’s too fuzzy	Low-light	The photo is too dark, improve exposure Increase the illumination in this shot My shot has very low dynamic range
Dehazing	Can you make this picture clearer? Help, my picture is all cloudy Remove the fog from my photo	Enhancement	Make it pop! Adjust the color balance for a natural look Apply a cinematic color grade to the photo
Deraining	I want my photo to be clear, not rainy Clear the rain from my picture Remove the raindrops from my photo	General	Fix my image please make the image look better

< Example degradation prompts >

Method

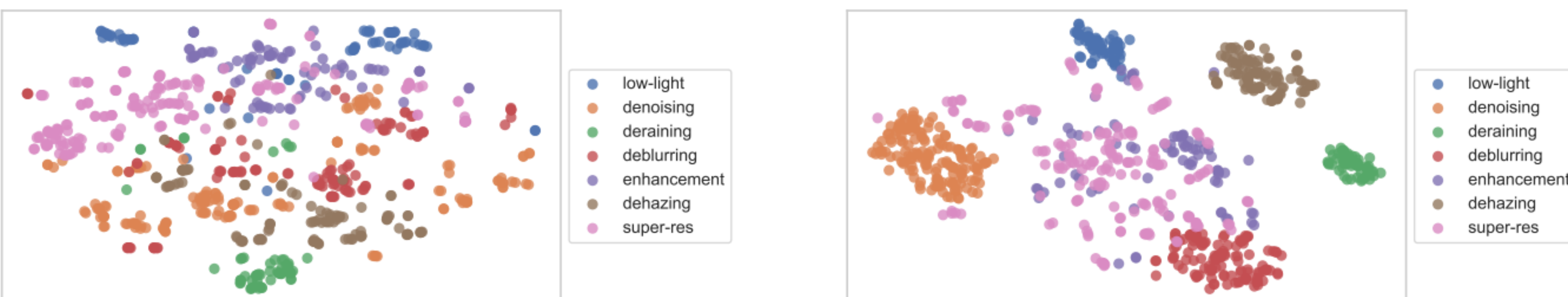
- Instruction-based filtering
  - Prompts are sampled randomly during training depending on input degradation
  - Text encoder maps prompt to fixed-size vector representation
    - Uses pure text-based BGE-micro-v2 sentence transformer
      - Used for speed and compactness, in comparison to CLIP
      - User prompts also contain little visual information, making CLIP unfitting
    - Fine-tune text encoder to adapt it for the restoration task



< Text prompt generation and usage >

Method

- Text encoder fine-tuning
  - Training full text encoder would lead to overfitting on training set
  - Instead freeze text encoder and train projection head
    - Untrained decoder is able to cluster the instructions to some extent
    - Clusters are clearly improved after training for most image enhancements
      - Enhancement and super-res clustered together due to similar prompting for them
  - Add classification head to improve training further
    - Classifies image degradation correctly to over 95%



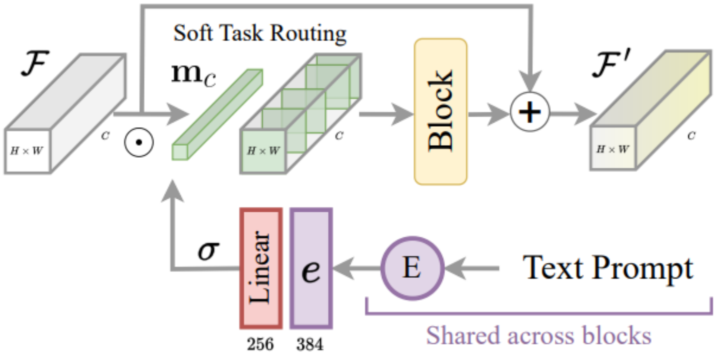
(a) t-SNE of embeddings *before* training *i.e.* frozen text encoder

(b) t-SNE of embeddings *after* training our learned projection

< t-SNE visualization of learned text embeddings >

Method

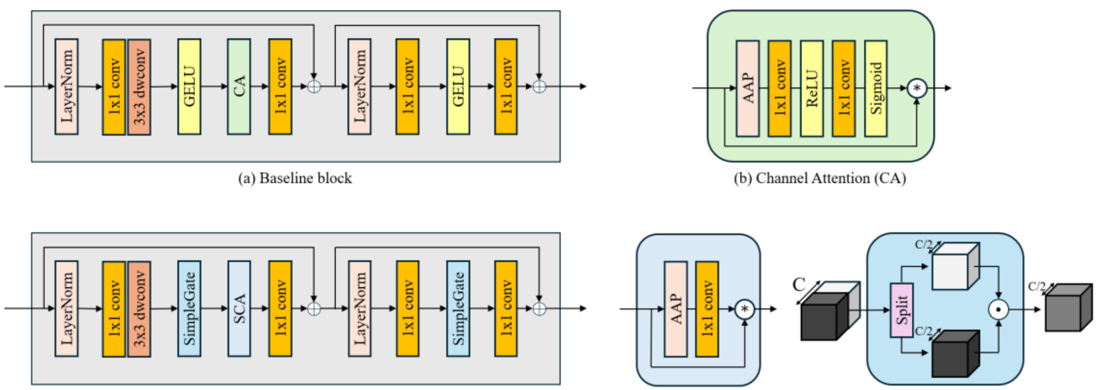
- Network architecture
  - Key aspect of InstructIR is integration of encoded instruction as mechanism of control for image model
    - Propose “Instruction Condition Block” (ICB) to enable task-specific transformations within model
    - Conventional task routing applies task-specific binary masks to channel features
      - Cannot use this technique as model does not know degradation a-priori
    - Mask allows model to select most relevant channels depending image information and instruction
    - Features with high weights contribute most to restoration, also enforces learning diverse filters



< ICB Block >

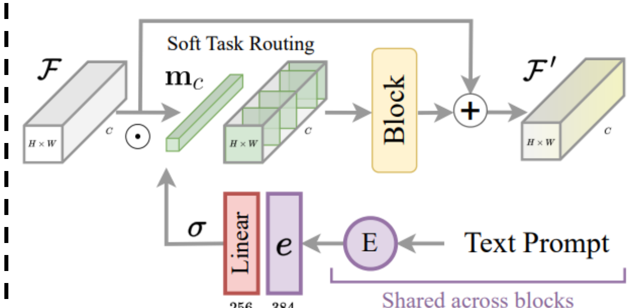
Method

- Network architecture
  - Model consists of image restoration model and adds text encoder
    - Use NAFNet as image restoration model
      - Follows U-Net architecture
    - NAFNet usually only learns one task at a time
      - InstructIR uses routing technique to learn multiple tasks at once
    - Model uses NAFBlock followed by ICBs to condition features
    - Text encoder used is BGE-micro-v2 sentence transformer



(a) Baseline block
(b) Channel Attention (CA)
(c) NAFNet block
(d) Simple Channel Attention (SCA)
(e) SimpleGate

< NAFNet Structure >



< ICB Block >

Experiment

- Quantitative results
  - All-in-one on 3 degradations

Methods	Dehazing SOTS [42]	Deraining Rain100L [21]	Denoising ablation study (BSD68 [53]) $\sigma = 15$	$\sigma = 25$	$\sigma = 50$	Average
BRDNet [73]	23.23/0.895	27.42/0.895	32.26/0.898	29.76/0.836	26.34/0.836	27.80/0.843
LPNet [25]	20.84/0.828	24.88/0.784	26.47/0.778	24.77/0.748	21.26/0.552	23.64/0.738
FDGAN [19]	24.71/0.924	29.89/0.933	30.25/0.910	28.81/0.868	26.43/0.776	28.02/0.883
MPRNet [97]	25.28/0.954	33.57/0.954	33.54/0.927	30.89/0.880	27.56/0.779	30.17/0.899
DL [21]	26.92/0.931	32.62/0.931	33.05/0.914	30.41/0.861	26.90/0.740	29.98/0.875
AirNet [43]	27.94/0.962	34.90/0.967	33.92/0.933	31.26/0.888	28.00/0.797	31.20/0.910
PromptIR [62]	**30.58/0.974**	36.37/0.972	33.98/0.933	31.31/0.888	28.06/0.799	32.06/0.913
InstructIR-3D	30.22/0.959	**37.98/0.978**	**34.15/0.933**	**31.52/0.890**	**28.30/0.804**	**32.43/0.913**
InstructIR-5D	27.10/0.956	36.84/0.973	34.00/0.931	31.40/0.887	28.15/0.798	31.50/0.909
InstructIR w/o text	26.84/0.948	34.02/0.960	33.70/0.929	30.94/0.882	27.78/0.780	30.65/0.900

< Comparison of all-in-one models for 3 restoration tasks (3D) >

Experiment

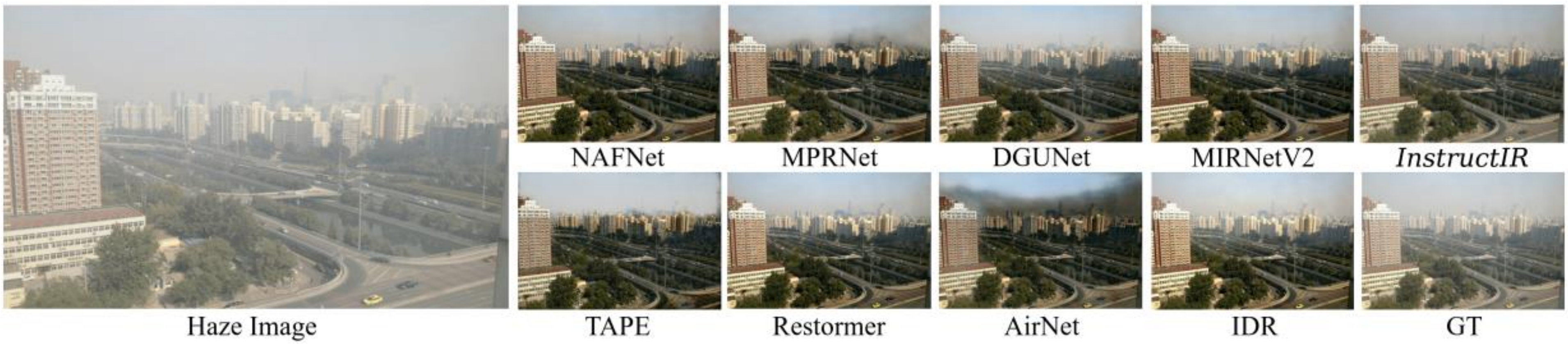
- Quantitative results
  - All-in-one on 5 degradations

Methods	Deraining Rain100L [90]		Dehazing SOTS [42]		Denoising BSD68 [53]		Deblurring GoPro [58]		Low-light Enh. LOL [84]		Average		Params
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	(M)
HINet [10]	35.67	0.969	24.74	0.937	31.00	0.881	26.12	0.788	19.47	0.800	27.40	0.875	88.67
DGUNet [57]	36.62	0.971	24.78	0.940	31.10	0.883	27.25	0.837	21.87	0.823	28.32	0.891	17.33
MIRNetV2 [96]	33.89	0.954	24.03	0.927	30.97	0.881	26.30	0.799	21.52	0.815	27.34	0.875	5.86
SwinIR [45]	30.78	0.923	21.50	0.891	30.59	0.868	24.52	0.773	17.81	0.723	25.04	0.835	0.91
Restormer [95]	34.81	0.962	24.09	0.927	31.49	0.884	27.22	0.829	20.41	0.806	27.60	0.881	26.13
NAFNet [9]	35.56	0.967	25.23	0.939	31.02	0.883	26.53	0.808	20.49	0.809	27.76	0.881	17.11
DL [21]	21.96	0.762	20.54	0.826	23.09	0.745	19.86	0.672	19.83	0.712	21.05	0.743	2.09
Transweather [76]	29.43	0.905	21.32	0.885	29.00	0.841	25.12	0.757	21.21	0.792	25.22	0.836	37.93
TAPE [46]	29.67	0.904	22.16	0.861	30.18	0.855	24.47	0.763	18.97	0.621	25.09	0.801	1.07
AirNet [43]	32.98	0.951	21.04	0.884	30.91	0.882	24.35	0.781	18.18	0.735	25.49	0.846	8.93
InstructIR w/o text	35.58	0.967	25.20	0.938	31.09	0.883	26.65	0.810	20.70	0.820	27.84	0.884	17.11
IDR [102]	35.63	0.965	25.24	0.943	31.60	0.887	27.87	0.846	21.34	0.826	<u>28.34</u>	<u>0.893</u>	15.34
InstructIR-**5D**	36.84	0.973	27.10	0.956	31.40	0.887	29.40	0.886	23.00	0.836	**29.55**	**0.907**	15.8
InstructIR-**6D**	36.80	0.973	27.00	0.951	31.39	0.888	29.73	0.892	22.83	0.836	**29.55**	**0.908**	15.8
InstructIR-**7D**	36.75	0.972	26.90	0.952	31.37	0.887	29.70	0.892	22.81	0.836	29.50	0.907	15.8

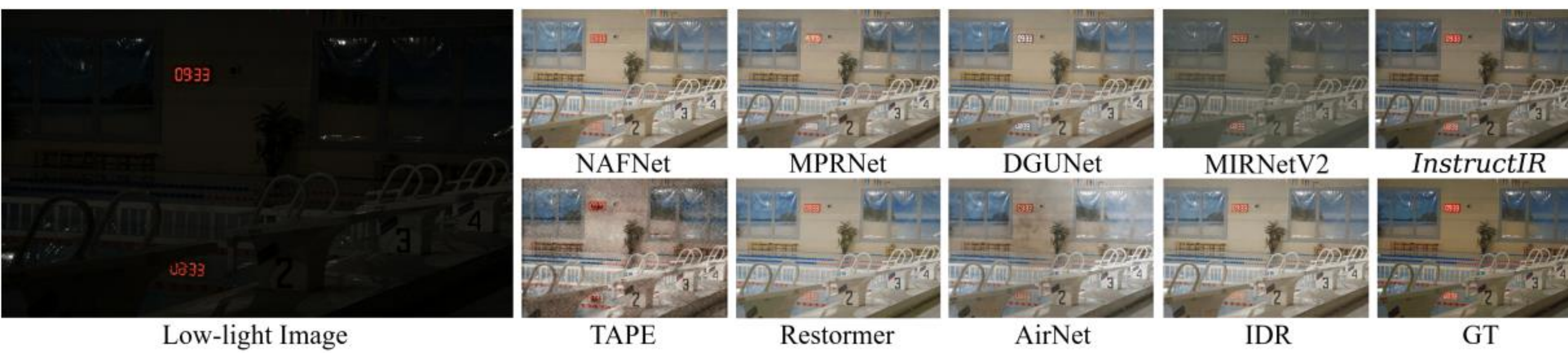
< Comparison of general image restoration and all-in-one models for 5 restoration tasks (5D) >

Experiment

- Qualitative results



< Qualitative results for the dehazing task >

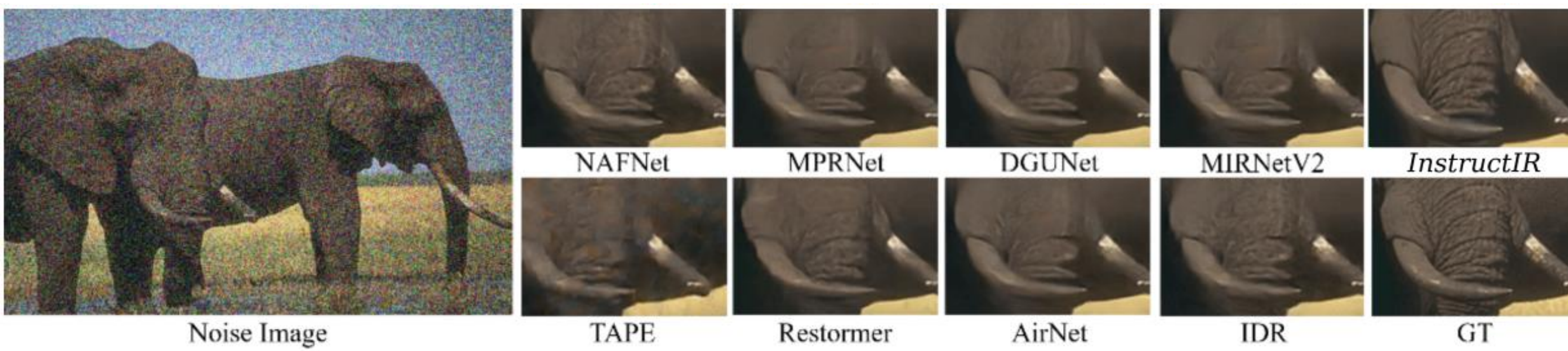


< Qualitative results for the low-light enhancement task >

Experiment

- Qualitative results

< Qualitative results for the deblurring task >



< Qualitative results for the denoising task >

Experiment

- Qualitative results

< Example of how the user interacts with InstructIR >

Experiment

- Ablation

< Impact of out-of-distribution instructions >

Language Level	Deraining	Denoising	Deblurring	LOL
Basic & Precise	36.84/0.973	31.40/0.887	29.47/0.887	23.00/0.836
Basic & Ambiguous	36.24/0.970	31.35/0.887	29.21/0.885	21.85/0.827
Real Users †	36.84/0.973	31.40/0.887	29.47/0.887	23.00/0.836

< Impact of prompt quality >

Summary

- Contributions
  - First approach that utilizes real human-written instructions to solve multi-task image restoration
  - Achieves all-in-one model which covers more tasks than previous works
  - Achieved SOTA performance on image denoising, deraining, deblurring, dehazing and low-light
- Limitations
  - Model struggles with images containing more than one degradation
    - Limits complex real-world images
    - Currently, all all-in-one models have this problem
    - Could be surpassed with more realistic training data
  - Cannot handle degradations that are out-of-distribution
    - Similar problem for all other related methods
  - Model trained on all 7 tasks at once has lower performance than 5 task model

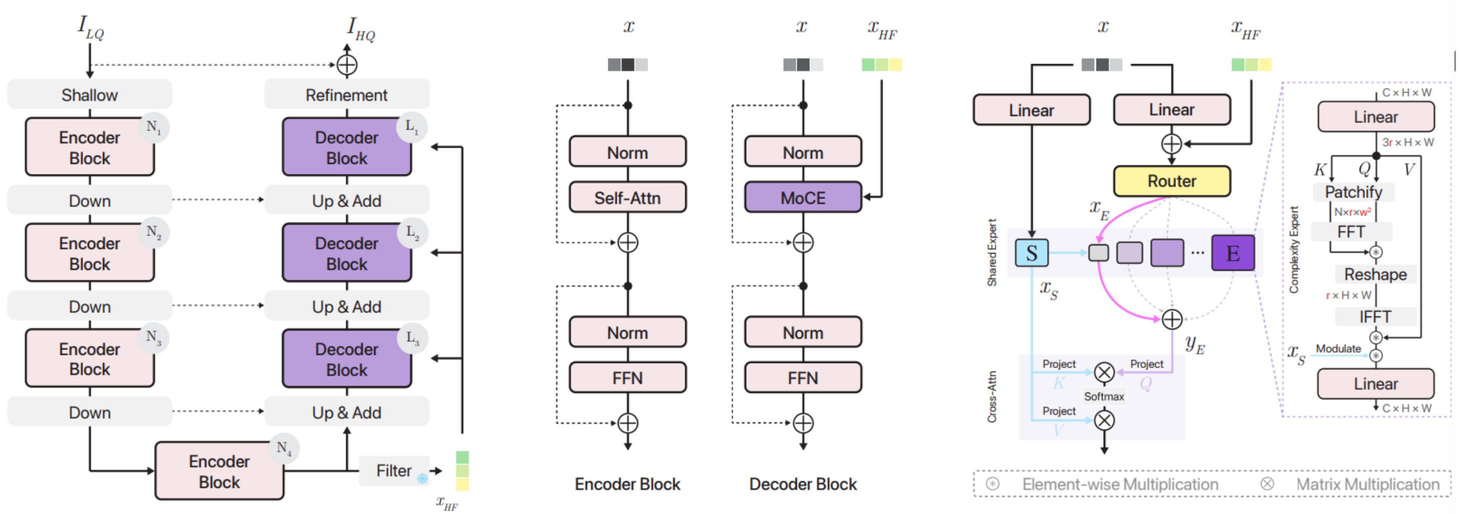
- Complexity Experts are Task-Discriminative Learners for Any Image Restoration [CVPR 2025]

Introduction

- Most all-in-one models use additional prompts (visual, language-based)
  - Often suffer from inefficiencies as parameters remain unused
- Suggest use of mixture-of-experts (MoE) for task-specific processing
  - Routing mechanism usually based on language or degradation priors
  - Leads to imbalanced optimization
    - Some experts generalize well, and others struggle with intended tasks limiting benefits
- Currently two big limitations for MoE models
  - Architecture of models are uniform
    - Fails to account for varying complexity requirements across different restoration tasks
      - E.g. motion blur demands localized processing with strong spatial awareness
      - Haze removal on the other hand requires broader contextual understanding
  - Second is challenge of appropriately routing tasks to experts being complicated
    - Due to unknown complexity of each degradation type a-priori

Method

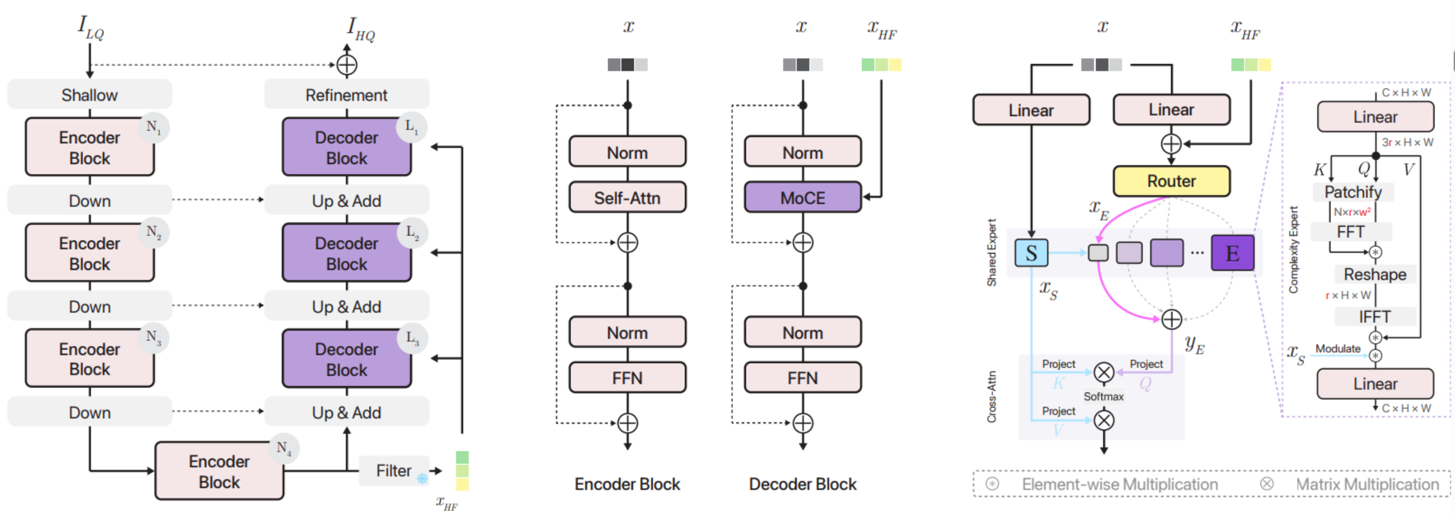
- Introduce Mixture-of-complexity-experts framework
  - Key innovation is designing expert blocks with increasing computational complexity and receptive fields
    - Allows model to adaptively match processing capacity with task requirements
    - Consists of n (n = 4) complexity experts E and a single shared expert S
      - Interaction made through two-level gating mechanism
      - Enables capturing of both degradation-specific features and inter-degradation relationships



(a) *Architecture overview* (b) *Transformer blocks.* (c) *Mixture-of-complexity-experts*

Method

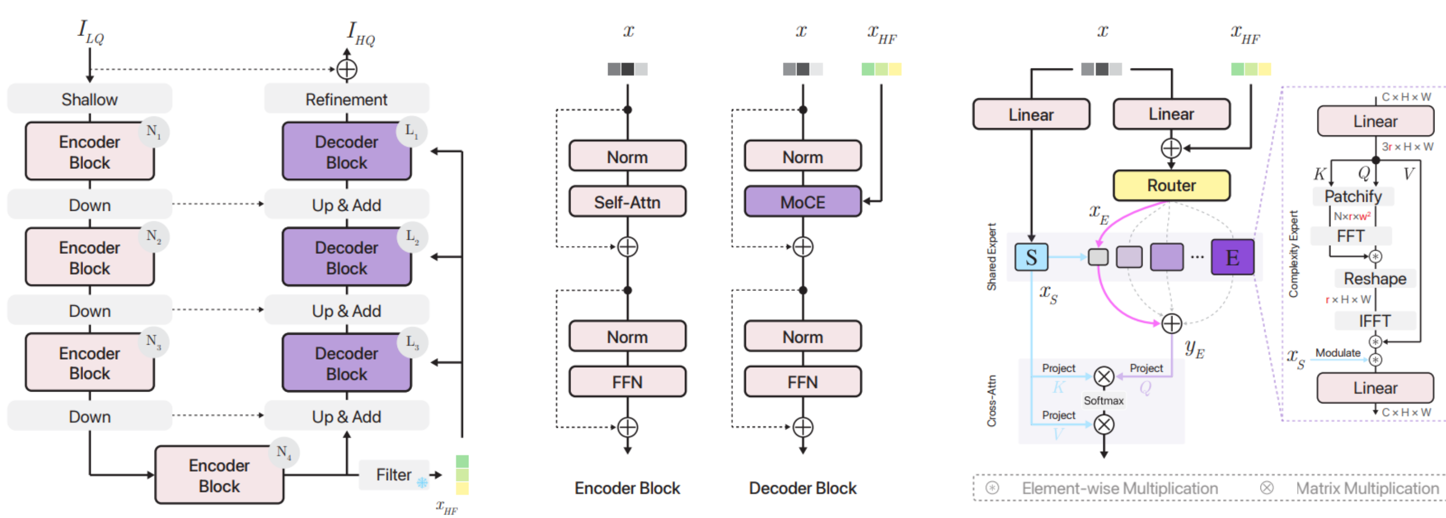
- Introduce Mixture-of-complexity-experts framework
  - Efficiency is important with increasing expert numbers and capacity
    - Nested expert structure with progressively reducing channel dimensionality r to control computational overhead
    - Simultaneously increase receptive field (window partition size w) within each expert to balance localized and global processing
    - Employs FFT-based approximation for efficient matrix multiplication
    - Shared expert employed by transposed self-attention module in channel dimension



(a) *Architecture overview* (b) *Transformer blocks.* (c) *Mixture-of-complexity-experts*

Method

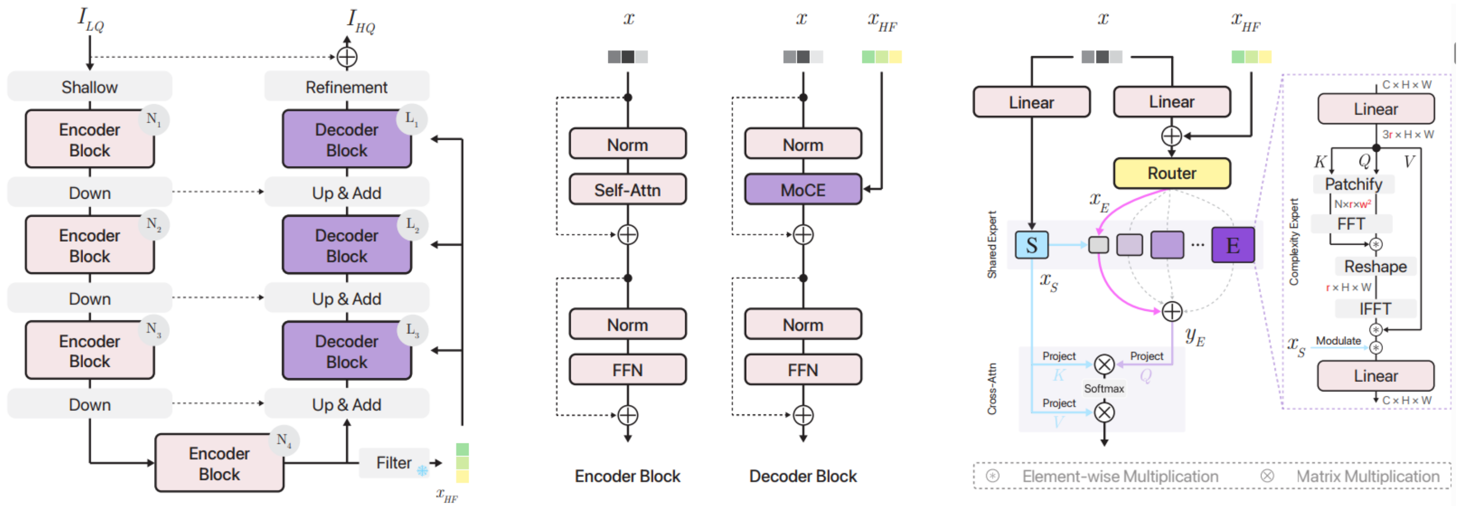
- Introduce Mixture-of-complexity-experts framework
  - Introduce complexity-aware allocation mechanism that preferentially directs tasks to lower-complexity experts
    - Addresses existing efficient routing challenge
    - Ensures effective task-specific allocation, directing inputs to experts with appropriate complexity levels
    - Implemented in each decoder block
      - Associate input features with corresponding specialized complexity experts E



(a) *Architecture overview* (b) *Transformer blocks.* (c) *Mixture-of-complexity-experts*

Method

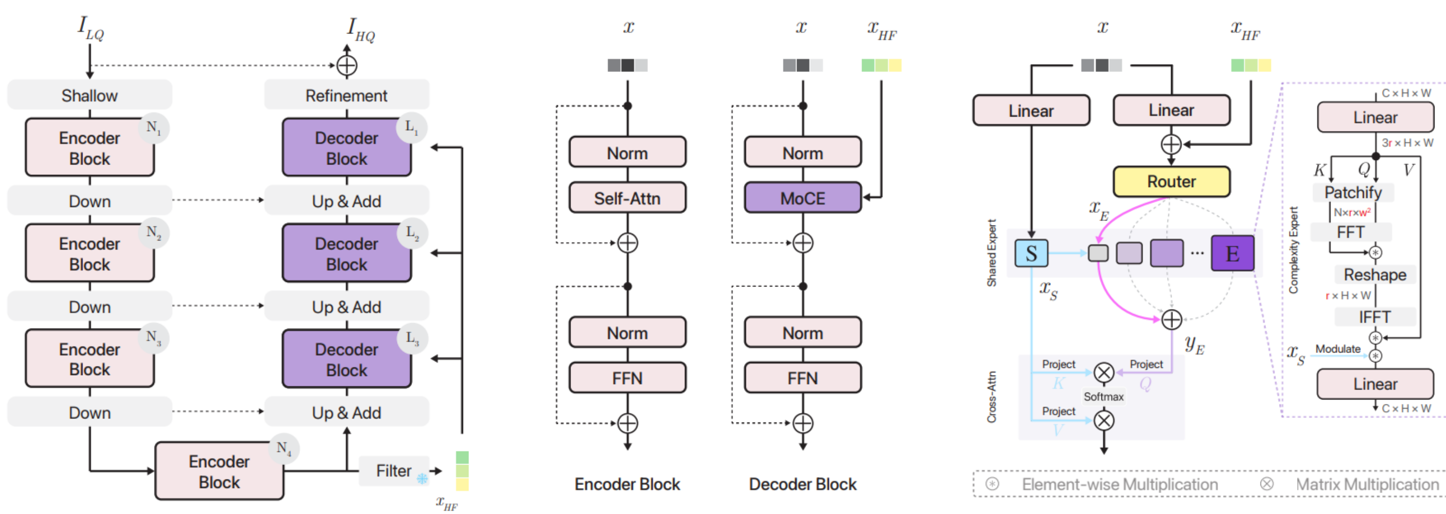
- Introduce Mixture-of-complexity-experts framework
  - Problem in image enhancement is scale-invariant tokenization of input images to ensure consistency across varying resolutions
  - Use image-level routing strategy where experts are selected for entire input image
    - Routing function allocates input samples based on required computation needed to corresponding complexity expert
    - Done by selecting top-k (k = 1) elements of softmax distribution
    - Use auxiliary loss to enhance matching of task to ideal expert



(a) *Architecture overview* (b) *Transformer blocks.* (c) *Mixture-of-complexity-experts*

Method

- Network architecture
  - U-shaped architecture with asymmetric encoder-decoder design
  - 3x3 convolution first extracts shallow features from degraded input
    - Then passed through 4 levels of encoding and decoding stages
  - Use transformer blocks with added MoCE layers in decoder
  - Additionally improve decoder's feature enrichment
    - Use high-frequency guidance via Sobel-filtered global feature vector



(a) *Architecture overview* (b) *Transformer blocks.* (c) *Mixture-of-complexity-experts*

Experiments

- Compare model on multiple settings
  - All-in-one on 3 degradations
    - Trained on dehazing, deraining and denoising degradations
    - Employ two models
      - MoCE-IR-S, a small model with 11M parameters
      - MoCE-IR, a heavier model with 25M parameters
  - All-in-one on 5 degradation
    - Trained on dehazing, deraining, denoising, deblurring and low-light degradations
    - Employ two models
      - MoCE-IR-S, a small model with 11M parameters
      - MoCE-IR, a heavier model with 25M parameters
  - Composited degradations
    - Model trained on a mix of multiple degradations at once
      - Up to three degradations at once (Low-light, haze, rain or snow)

Experiments

- Quantitative Results
  - All-in-one on 3 degradations

	Method	Params.	*Dehazing* SOTS		*Deraining* Rain100L		*Denoising* BSD68$_{\sigma=15}$		BSD68$_{\sigma=25}$		BSD68$_{\sigma=50}$		Average	
Light	BRDNet [48]	-	23.23	.895	27.42	.895	32.26	.898	29.76	.836	26.34	.693	27.80	.843
	LPNet [14]	-	20.84	828	24.88	.784	26.47	.778	24.77	.748	21.26	.552	23.64	.738
	FDGAN [11]	-	24.71	.929	29.89	.933	30.25	.910	28.81	.868	26.43	.776	28.02	.883
	DL [13]	2M	26.92	.931	32.62	.931	33.05	.914	30.41	.861	26.90	.740	29.98	.876
	AirNet [24]	9M	27.94	.962	34.90	.967	33.92	.933	31.26	.888	28.00	.797	31.20	.910
	MoCE-IR-S (*ours*)	11M	**30.94**	**.979**	**38.22**	**.983**	**34.08**	**.933**	**31.42**	**.888**	**28.16**	**.798**	**32.57**	**.916**
Heavy	MPRNet [64]	16M	25.28	.955	33.57	.954	33.54	.927	30.89	.880	27.56	.779	30.17	.899
	PromptIR [36]	36M	30.58	.974	36.37	.972	33.98	.933	31.31	.888	28.06	.799	32.06	.913
	Gridformer [51]	34M	30.37	.970	37.15	.972	33.93	.931	31.37	.887	28.11	.801	32.19	.912
	Art-PromptIR [54]	33M	30.83	.979	37.94	.982	34.06	.934	31.42	.891	28.14	.801	32.49	.917
	DA-CLIP* [30]	125M	29.46	.963	36.28	.968	30.02	.821	24.86	.585	22.29	.476	-	-
	UniProcessor* [12]	1002M	**31.66**	.979	38.17	.982	34.08	**.935**	31.42	**.891**	28.17	**.803**	32.70	**.918**
	MoCE-IR (*ours*)	25M	31.34	**.979**	**38.57**	**.984**	**34.11**	.932	**31.45**	.888	**28.18**	.800	**32.73**	.917

< Comparison of all-in-one models for 3 restoration tasks >

Experiments

- Quantitative Results

  - All-in-one on 5 degradations

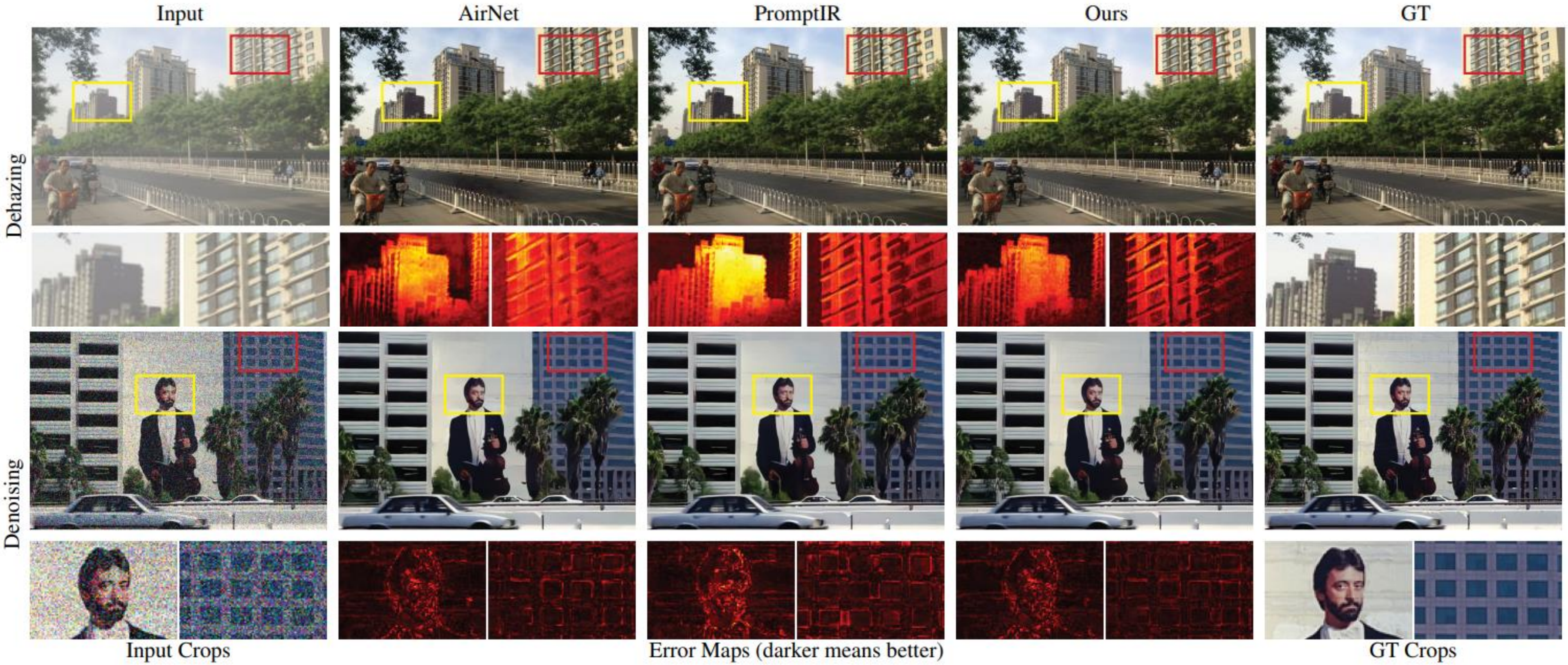
	Method	Params.	*Dehazing*		*Deraining*		*Denoising*		*Deblurring*		*Low-Light*		Average	
			SOTS		Rain100L		$\text{BSD68}_{\sigma=25}$		GoPro		LOLv1			
Light	SwinIR* [26]	1M	21.50	.891	30.78	.923	30.59	.868	24.52	.773	17.81	.723	25.04	.835
	DL [13]	2M	20.54	.826	21.96	.762	23.09	.745	19.86	.672	19.83	.712	21.05	.743
	TAPE [27]	1M	22.16	.861	29.67	.904	30.18	.855	24.47	.763	18.97	.621	25.09	.801
	AirNet [24]	9M	21.04	.884	32.98	.951	30.91	.882	24.35	.781	18.18	.735	25.49	.847
	MoCE-IR-S (*ours*)	11M	**31.33**	**.978**	**37.21**	**.978**	**31.25**	**.884**	**28.90**	**.877**	**21.68**	**.851**	**30.08**	**.913**
Heavy	NAFNet* [6]	17M	25.23	.939	35.56	.967	31.02	.883	26.53	.808	20.49	.809	27.76	.881
	DGUNet* [33]	17M	24.78	.940	36.62	.971	31.10	.883	27.25	.837	21.87	.823	28.32	.891
	Restormer* [65]	26M	24.09	.927	34.81	.962	31.49	.884	27.22	.829	20.41	.806	27.60	.881
	MambaIR [16]	27M	25.81	.944	36.55	.971	31.41	.884	28.61	.875	22.49	.832	28.97	.901
	Transweather [49]	38M	21.32	.885	29.43	.905	29.00	.841	25.12	.757	21.21	.792	25.22	.836
	IDR [66]	15M	25.24	.943	35.63	.965	**31.60**	.887	27.87	.846	21.34	.826	28.34	.893
	Gridformer [51]	34M	26.79	.951	36.61	.971	31.45	.885	29.22	.884	22.59	.831	29.33	.904
	InstructIR-5D [9]	17M	27.10	.956	36.84	.973	31.40	.873	29.40	.886	23.00	.836	29.55	.908
	MoCE-IR (*ours*)	25M	**30.48**	**.974**	**38.04**	**.982**	31.34	**.887**	**30.05**	**.899**	**23.00**	**.852**	**30.58**	**.919**

  - All-in-one on multiple degradations at once

Method	Params.	*CDD11-Single*								*CDD11-Double*										*CDD11-Triple*				Avg.	
		Low (L)		Haze (H)		Rain (R)		Snow (S)		L+H		L+R		L+S		H+R		H+S		L+H+R		L+H+S			
AirNet [24]	9M	24.83	.778	24.21	.951	26.55	.891	26.79	.919	23.23	.779	22.82	.710	23.29	.723	22.21	.868	23.29	.901	21.80	.708	22.24	.725	23.75	.814
PromptIR [36]	36M	26.32	.805	26.10	.969	31.56	.946	31.53	.960	24.49	.789	25.05	.771	24.51	.761	24.54	.924	23.70	.925	23.74	.752	23.33	.747	25.90	.850
WGWSNet [71]	26M	24.39	.774	27.90	.982	33.15	.964	34.43	.973	24.27	.800	25.06	.772	24.60	.765	27.23	.955	27.65	.960	23.90	.772	23.97	.771	26.96	.863
WeatherDiff [72]	83M	23.58	.763	21.99	.904	24.85	.885	24.80	.888	21.83	.756	22.69	.730	22.12	.707	21.25	.868	21.99	.868	21.23	.716	21.04	.698	22.49	.799
OneRestore [17]	6M	26.48	**.826**	32.52	.990	33.40	.964	34.31	.973	25.79	**.822**	25.58	.799	25.19	.789	**29.99**	.957	**30.21**	.964	24.78	.788	24.90	**.791**	28.47	.878
MoCE-IR-S (*ours*)	11M	**27.26**	.824	**32.66**	**.990**	**34.31**	**.970**	**35.91**	**.980**	**26.24**	.817	**26.25**	**.800**	**26.04**	**.793**	29.93	**.964**	30.19	**.970**	**25.41**	**.789**	**25.39**	.790	**29.05**	**.881**

Experiments

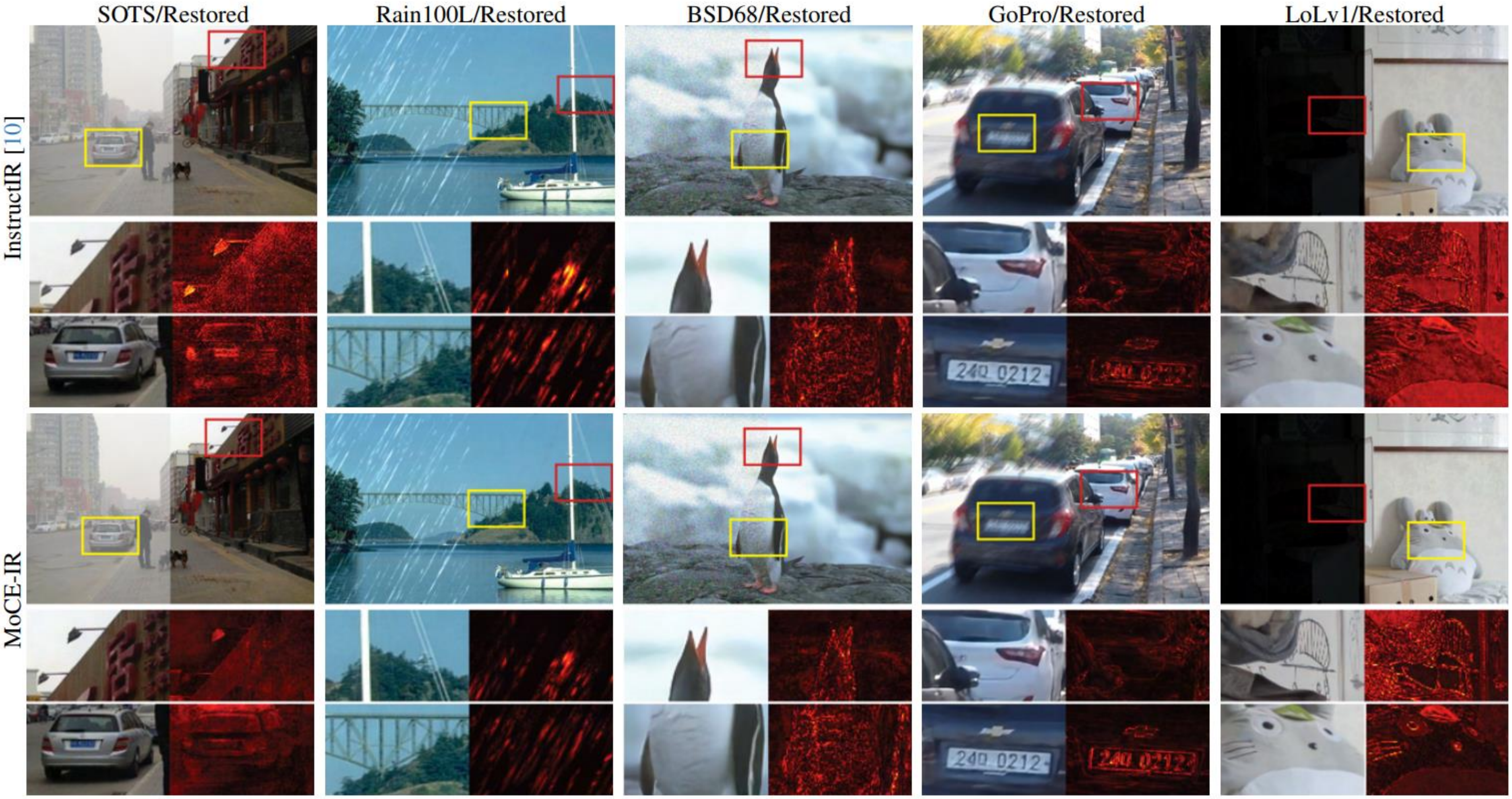
- Qualitative Results



< Visualization of dehazing and denoising task >

Experiments

- Qualitative Results



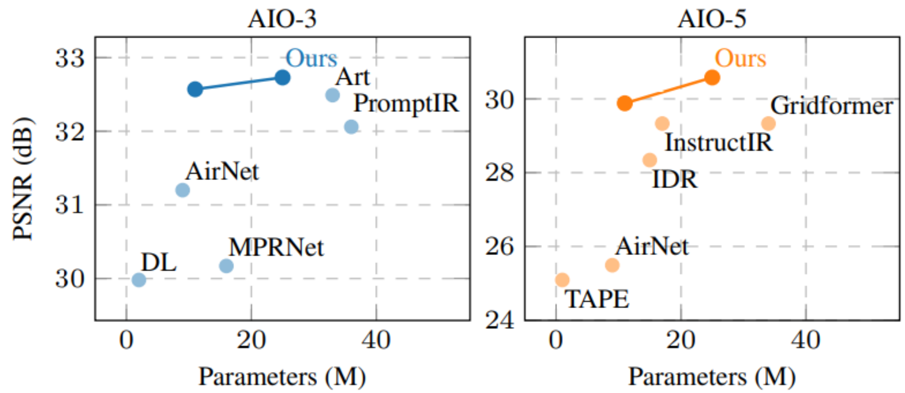
< Comparison of InstructIR and MoCE-IR on different image restoration tasks>

서강대학교 SOGANG UNIVERSITY

VDS LAB

Experiments

- Ablation



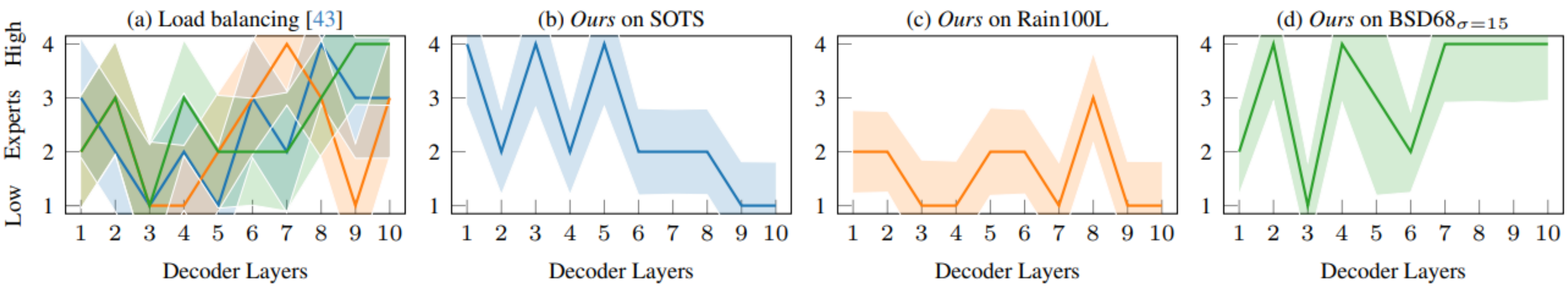
< Quality and computational efficiency >

Method	Params.	Memory	FLOPS	Runtime
AirNet [24]	**8.93M**	4829M	238G	42.17 ± 0.23
PromptIR [36]	35.59M	9830M	132G	41.28 ± 0.43
IDR [66]	15.34M	4905M	98G	-
MoCE-IR (*ours*)	25.35M	5887M	80.59 ± 5.21G	23.36 ± 2.47
MoCE-IR-S (*ours*)	11.47M	**4228M**	**36.93 ± 2.32G**	**22.15** ± 2.59

< Memory utilization >

Experiments

- Ablation



< Expert scaling / Routing analysis >

Degradation	Routing	Learned Choice	Manual Choice $\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$
Rain Noise	Load Balance [45]	Not Applicable	28.27 33.33	28.24 33.37	28.25 33.37	28.27 33.40
Rain Noise	Complexity Bias (ours)	$\mathcal{E}_1$ $\mathcal{E}_4$	**30.45** 33.92	30.21 33.93	29.73 33.93	24.65 **34.00**

< Expert generalization ability >

Summary

- Contributions
  - Selectively activates complexity experts based on input requirements
    - Unifies task-specific and holistic learning in single architecture
  - Develops complexity-aware routing mechanism
    - Balances restoration quality with computational efficiency by adaptive expert allocation
  - SOTA all-in-one image restoration model with improved efficiency
- Limitations
  - Current image-level routing imposes scalability constraints
  - Potential mismatch in synthetic-to-real adaptation
  - Speed and efficiency could be enhanced by using mixed-precision across experts

Thank you!