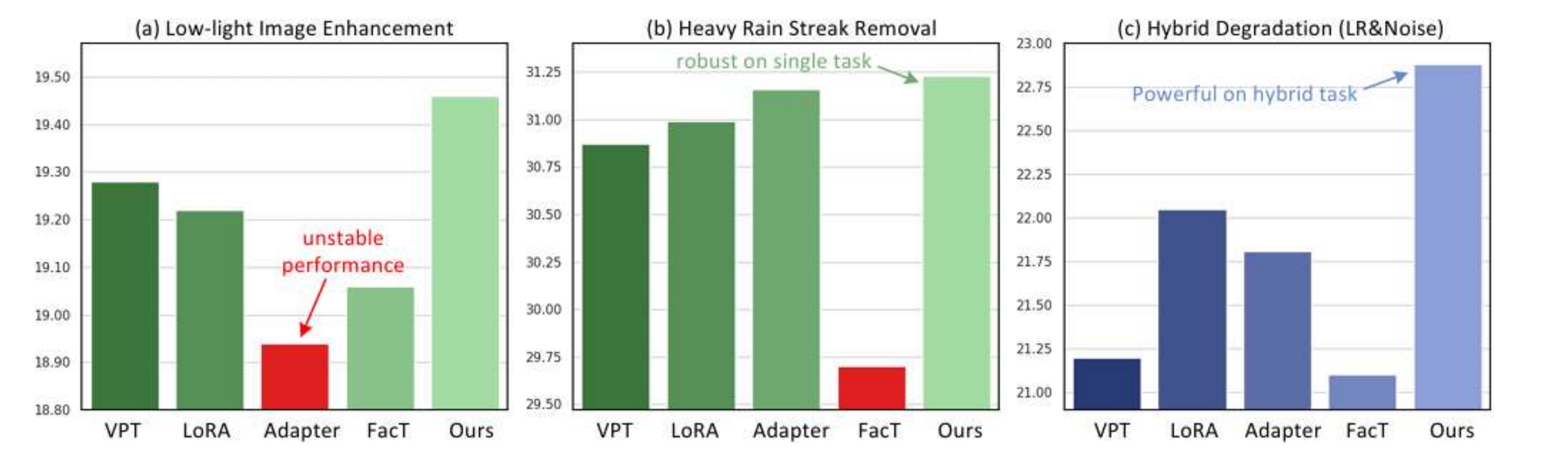
Parameter Efficient Adaptation for Image Restoration with Heterogeneous Mixture-of-Experts

Hang Guo[1], Tao Dai[2,*], Yuanchao Bai[3], Bin Chen[3], Xudong Ren[1], Zexuan Zhu[1], Shu-tao Xia[1,4]
[1]Tsinghua University; [2]Shenzhen University; [3]Harbin Institute of Technology; [4]Peng Cheng Laboratory
*Corresponding author

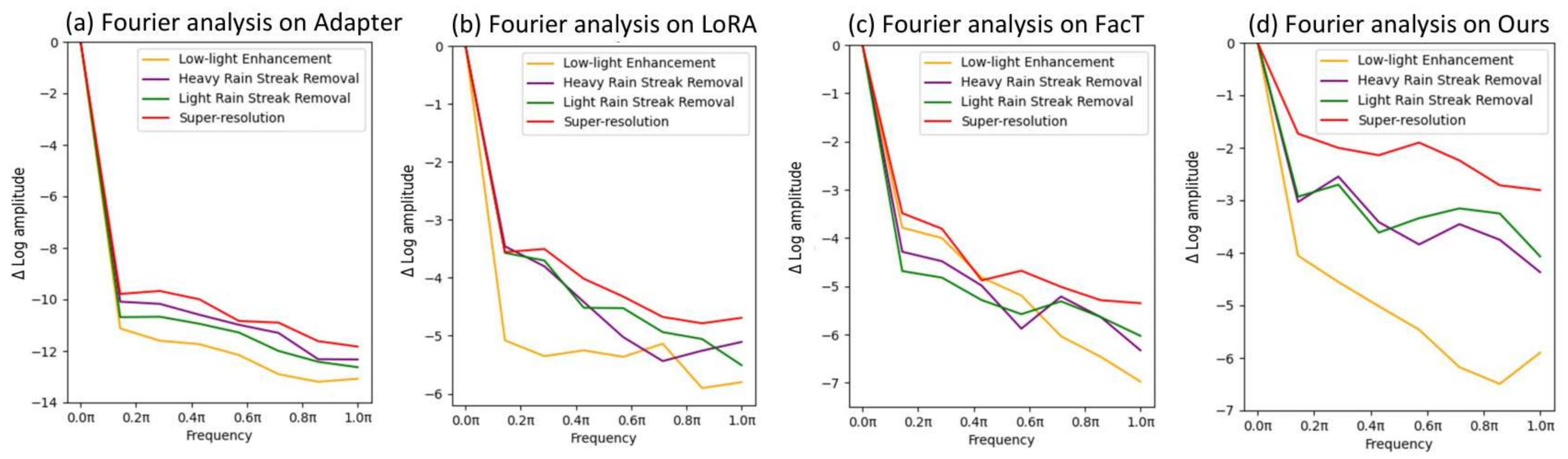
Observation



Directly applying current PETL methods to image restoration

- Unstable performance on single degradation
- Sub-optimal results on hybrid degradation

Motivation



Reasons behind observation

- Existing methods exhibit homogeneous frequency representations even when faced with different degradations, i,.e., they cannot figure out different degradations.
- This problem hinders them to learn different representations for different degradations, leading to the above phenomenon

Solution

We can use the structure of Mixture-of-Experts to learn distinct representations for different degradations!

Challenge 1:

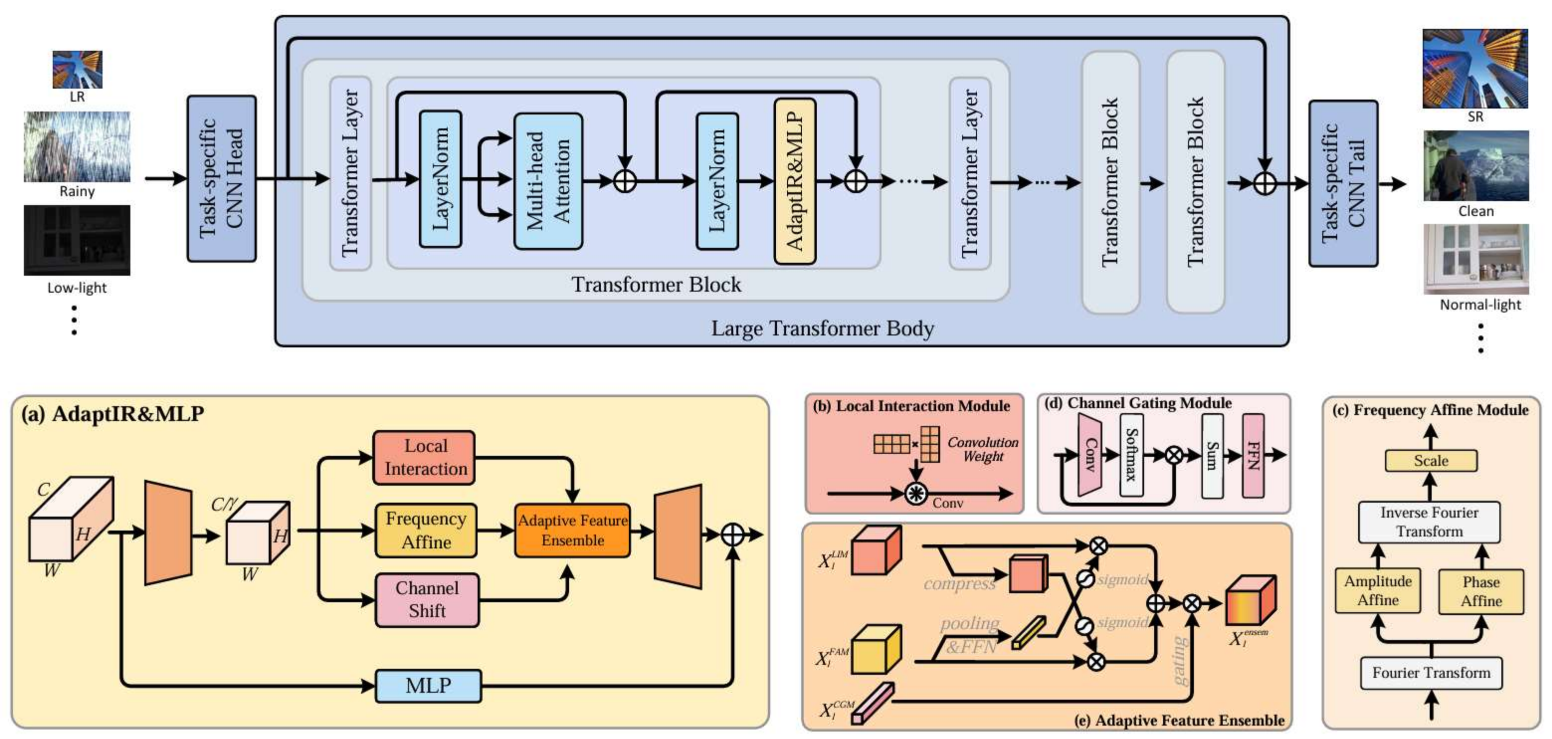
How to learning distinct representation? There is danger from the mode collapse to one representation

Challenge 2:

How to learning distinct representation under low parameter budgets?

Method

Overview



Designing orthogonal branches to force the learning of heterogeneous representations!

- Local Interaction Module
  local spatial modeling

$$W' = UV^\top$$
$$X_l^{LIM} = \mathrm{Reshape}(W') \circledast X_l^{intrin}$$

- Frequency Affine Module
  global spatial modeling

$$[Mag_l, Pha_l] = \mathrm{FFT}(X_l^{intrin}),$$
$$X_l^{FAM} = \mathrm{Conv}(\mathrm{iFFT}(\mathrm{to\_complex}(\phi_1(Mag_l), \phi_2(Pha_l)))),$$

- Channel Gating Module
  channel modeling

$$\mathcal{M}_l = \mathrm{Softmax}(\mathrm{Conv}(X_l^{intrin}))$$
$$X_l^{CGM} = \mathrm{FFN}(\sum_{h,w} \mathcal{M}_l \otimes X_l^{intrin})$$

- Adaptive Feature Ensemble

Advantages of AdaptIR

High efficiency: Tuning only 0.6% of pre-trained parameters within 8h!

High effectiveness: adapts to a various unseen tasks with comparable or even better performance!

Comparison to SoTA

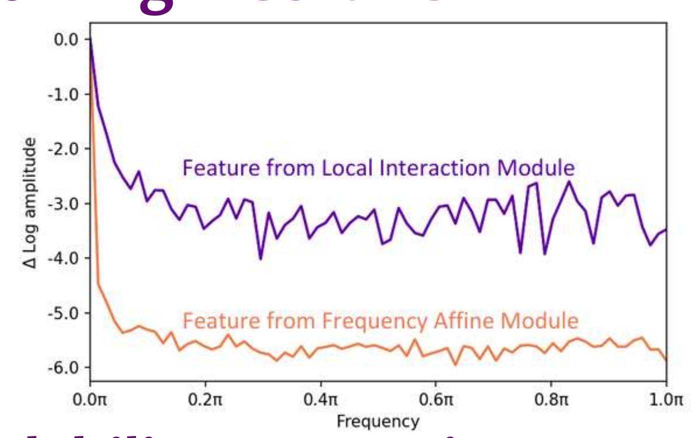
Restoration with Ideal Reference

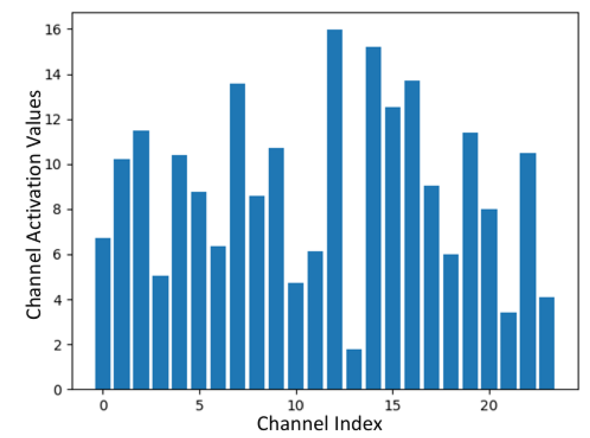
Method	Degradation	#param	Set5 PSNR	Set5 SSIM	Set14 PSNR	Set14 SSIM	BSDS100 PSNR	BSDS100 SSIM	Urban100 PSNR	Urban100 SSIM	Manga109 PSNR	Manga109 SSIM
Full-ft	LR4&Noise30	119M	27.24	0.7859	25.56	0.6686	25.02	0.6166	24.02	0.6967	26.31	0.8245
Pretrain	LR4&Noise30	-	19.74	0.3569	19.27	0.3114	19.09	0.2783	18.54	0.3254	19.75	0.3832
SSF [24]	LR4&Noise30	373K	25.41	0.6720	24.02	0.5761	24.06	0.5411	21.89	0.5514	23.33	0.6736
VPT [9]	LR4&Noise30	884K	24.11	0.5570	22.97	0.4722	22.91	0.4336	21.20	0.4527	22.61	0.5570
Adapter [8]	LR4&Noise30	691K	25.60	0.6862	24.16	0.5856	24.17	0.5498	22.05	0.5640	23.61	0.6904
LoRA [21]	LR4&Noise30	995K	25.19	0.6371	23.82	0.5405	23.82	0.5026	21.81	0.5193	23.30	0.6396
Adaptfor. [7]	LR4&Noise30	677K	26.10	0.7138	24.58	0.6095	24.44	0.5686	22.52	0.5976	24.38	0.7296
FacT [10]	LR4&Noise30	537K	25.70	0.6963	24.24	0.5944	24.25	0.5586	21.10	0.5727	23.63	0.6993
MoE	LR4&Noise30	667K	26.35	0.7335	24.80	0.6254	24.59	0.5835	22.77	0.6188	24.73	0.7517
Ours	LR4&Noise30	697K	26.48	0.7441	24.88	0.6345	24.67	0.6279	22.88	0.5932	24.96	0.7625
Full-ft	LR4&JPEG30	119M	27.21	0.7778	25.49	0.6563	25.08	0.6076	23.54	0.6687	25.48	0.7971
Pretrain	LR4&JPEG30	-	25.23	0.6702	24.12	0.5917	24.19	0.5627	21.74	0.5654	22.93	0.6732
SSF [24]	LR4&JPEG30	373K	26.26	0.7321	24.81	0.6285	24.71	0.5882	22.44	0.6085	23.92	0.7350
VPT [9]	LR4&JPEG30	884K	26.63	0.7497	25.14	0.6414	24.89	0.5974	22.96	0.6377	24.53	0.7591
Adapter [8]	LR4&JPEG30	691K	26.73	0.7554	25.22	0.6448	24.92	0.5999	23.09	0.6447	24.74	0.7677
LoRA [21]	LR4&JPEG30	995K	26.64	0.7501	25.17	0.6424	24.91	0.5983	23.02	0.6405	24.64	0.7619
Adaptfor. [7]	LR4&JPEG30	677K	26.74	0.7562	23.08	0.6441	25.22	0.6447	24.92	0.5996	24.72	0.7669
FacT [10]	LR4&JPEG30	537K	26.71	0.7557	25.22	0.6450	24.93	0.5998	23.08	0.6446	24.74	0.7681
MoE	LR4&JPEG30	667K	26.80	0.7590	25.26	0.6465	24.04	0.6009	23.14	0.6477	24.81	0.7708
Ours	LR4&JPEG30	697K	26.91	0.7646	25.34	0.6502	24.98	0.6032	23.25	0.6541	25.02	0.7791

Restoration in the Wild

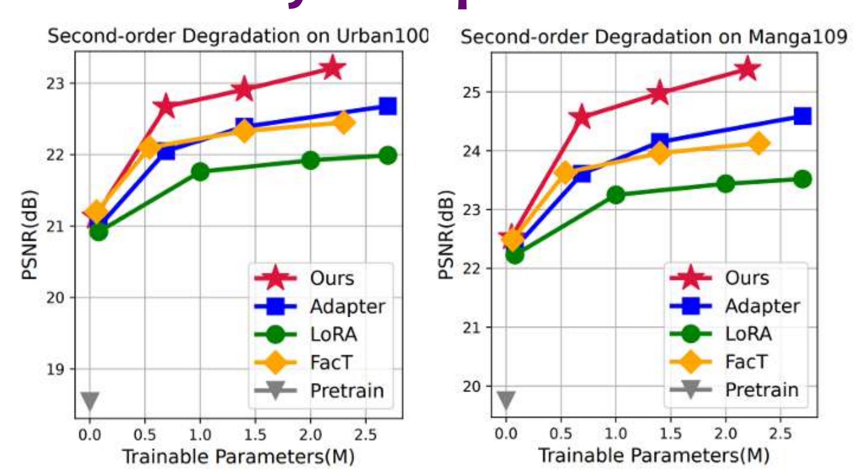
Method	#param	GPU memory	training time	light derain	denoise σ=25	denoise σ=30
AirNet [4]	8.7M	~11G	~48h	34.90/0.967	31.90/0.914	28.68/0.861
PromptIR [5]	97.1M	~128G	~48h	36.37/0.972	32.09/0.919	28.99/0.871
Ours	697K	~8G	~10h	41.27/0.988	32.64/0.926	29.16/0.875

Working Mechanism





Scalability Comparison



Additional Resources

paper

code