南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

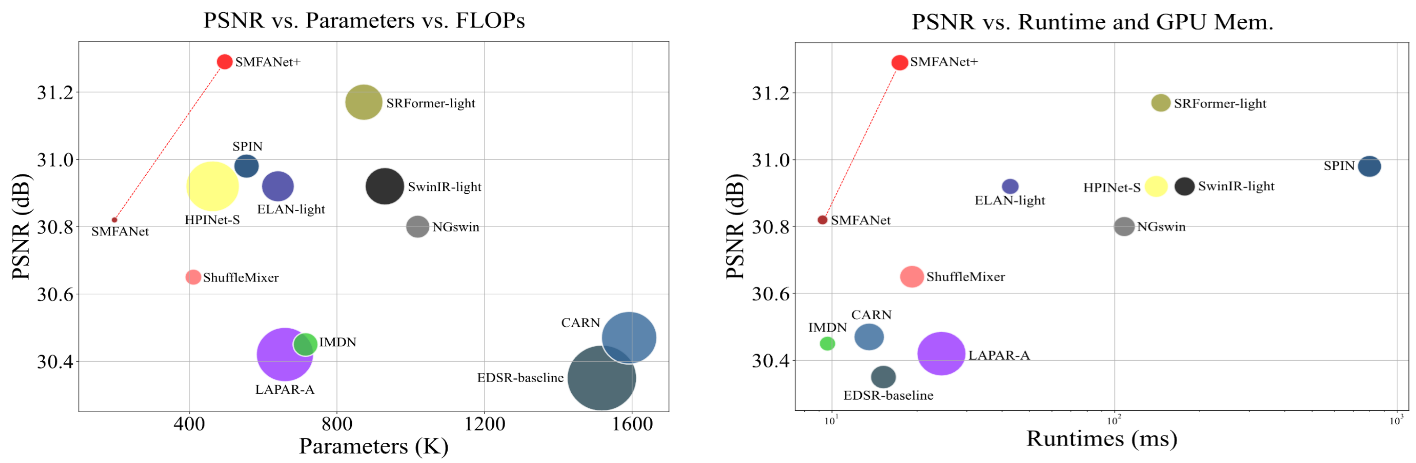
SMFANet: A Lightweight Self-Modulation Feature Aggregation Network for Efficient Image Super-Resolution

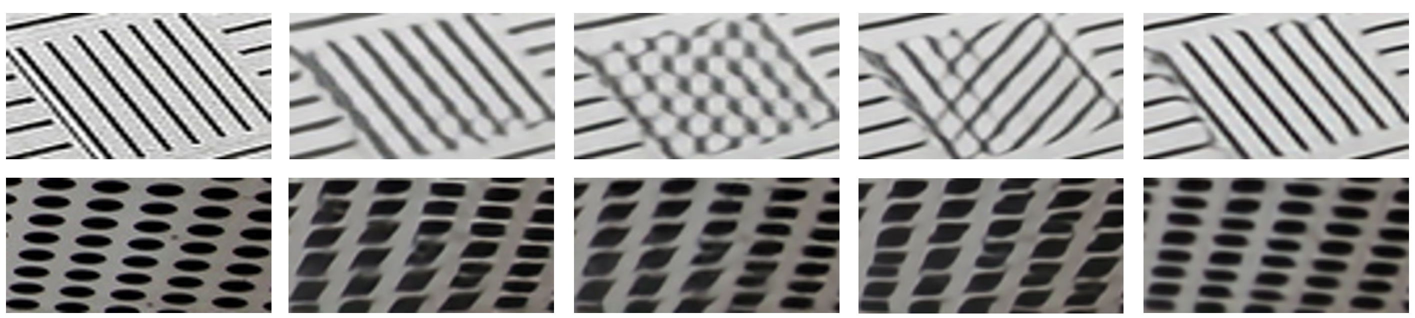
Mingjun Zheng*, Long Sun*, Jiangxin Dong, and Jinshan Pan

ECCV

Code & pre-trained models are available

Motivation

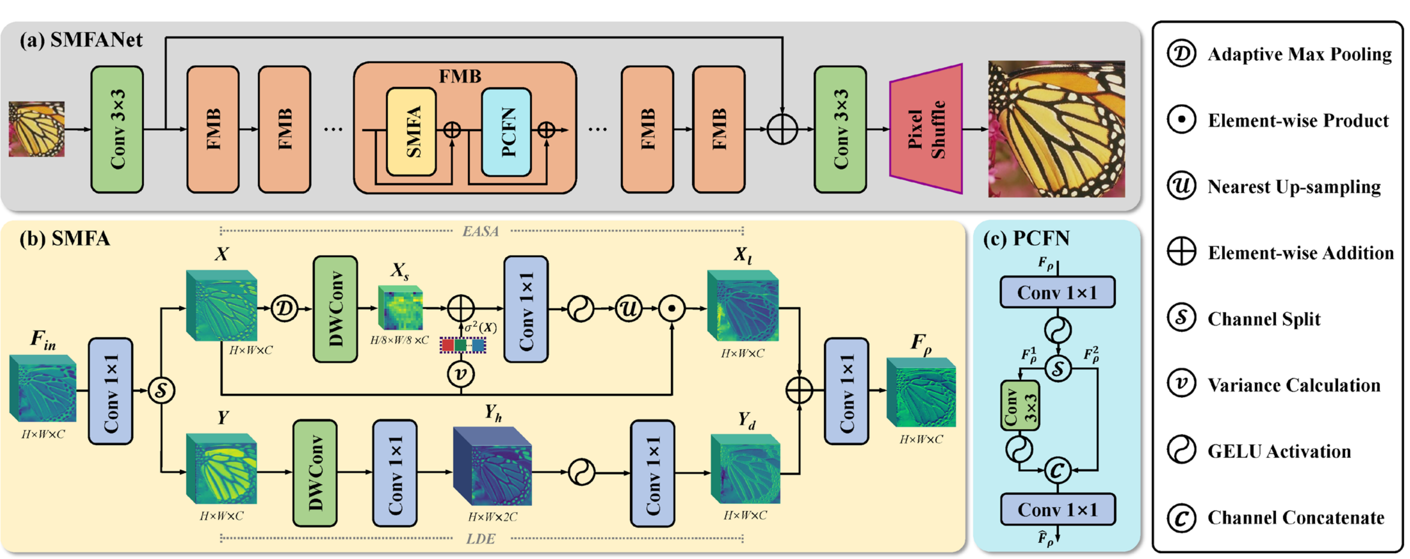




(a) GT (b) SwinIR-light (c) NGSwin (d) SRFormer-light (e) SMFANet+

- Previous ViT-based SR models require much more computational resources
- SA mechanism limits their capacity for capturing local details

Proposed Method



- We develop an efficient SMFA module to extract representative features, where the EASA branch for exploring non-local information and the LDE branch for capturing local features.
- We present a lightweight PCFN to further refine the features derived from the SMFA in spatial and channel dimensions.

Experimental Results

1. Memory and running time comparisons on ×4 SR

Type	Methods	#GPU Mem. [M]	#Avg. Time [ms]	Type	Methods	#GPU Mem. [M]	#Avg. Time [ms]
		Quality Metrics				Quality Metrics	
CNN	CARN	702.07	13.54	ViT	SwinIR-light	342.44	177.06
	EDSR-baseline	507.13	15.21		ELAN-light	241.34	42.75
	IMDN	204.27	9.64		HPINet-S	445.92	140.45
	LAPAR-A	1811.46	24.40		NGswin	372.85	108.25
	ShuffleMixer	474.79	19.21		SPIN	441.52	798.44
	SAFMN	65.26	8.44		SRFormer-light	320.95	145.99
	SMFANet	93.20	9.26		SMFANet+	247.51	17.39

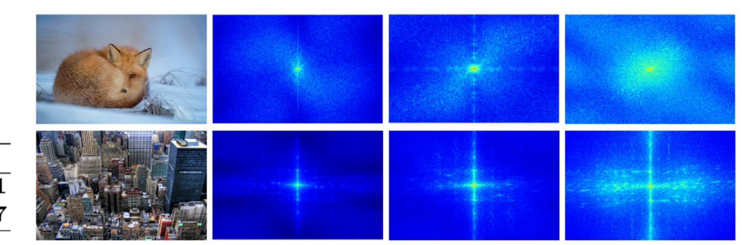
2. Comparison with ViT-based lightweight SR methods

Scale	Methods	#Params (K)	#FLOPs (G)	Set5	Set14	B100	Urban100	Manga109
×4	ESRT	752	298	32.19/0.8947	28.69/0.7833	27.69/0.7379	26.39/0.7962	30.75/0.9100
	SwinIR-light	930	65	32.44/0.8976	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151
	ELAN-light	640	54	32.43/0.8975	28.78/0.7858	27.69/0.7406	26.54/0.7982	30.92/0.9150
	HPINet-S	463	88	32.47/0.8987	28.80/0.7872	27.69/0.7416	26.59/0.8016	30.92/0.9143
	NGswin	1019	40	32.33/0.8963	28.78/0.7859	27.66/0.7396	26.45/0.7963	30.80/0.9128
	SPIN	555	42	32.48/0.8983	28.80/0.7862	27.70/0.7415	26.55/0.7998	30.98/0.9156
	SRFormer-light	873	63	32.51/0.8988	28.82/0.7872	27.73/0.7422	26.67/0.8032	31.17/0.9165
	SMFANet+ (DIV2K)	496	28	32.43/0.8979	28.77/0.7849	27.70/0.7400	26.45/0.7943	31.06/0.9138
	SMFANet+ (DF2K)	496	28	32.51/0.8985	28.87/0.7872	27.74/0.7412	26.56/0.7976	31.29/0.9163

Analysis

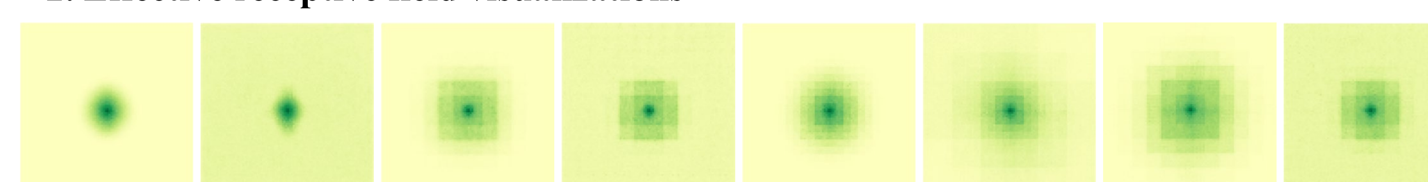
1. Effectiveness of the SMFA

Ablation	w/o SMFA	w/o EASA	w/o LDE	Baseline
Urban100	25.54/0.7646	25.94/0.7783	25.92/0.7772	**26.19/0.7861**
Manga109	29.76/0.8974	30.39/0.9061	30.22/0.9032	**30.72/0.9097**



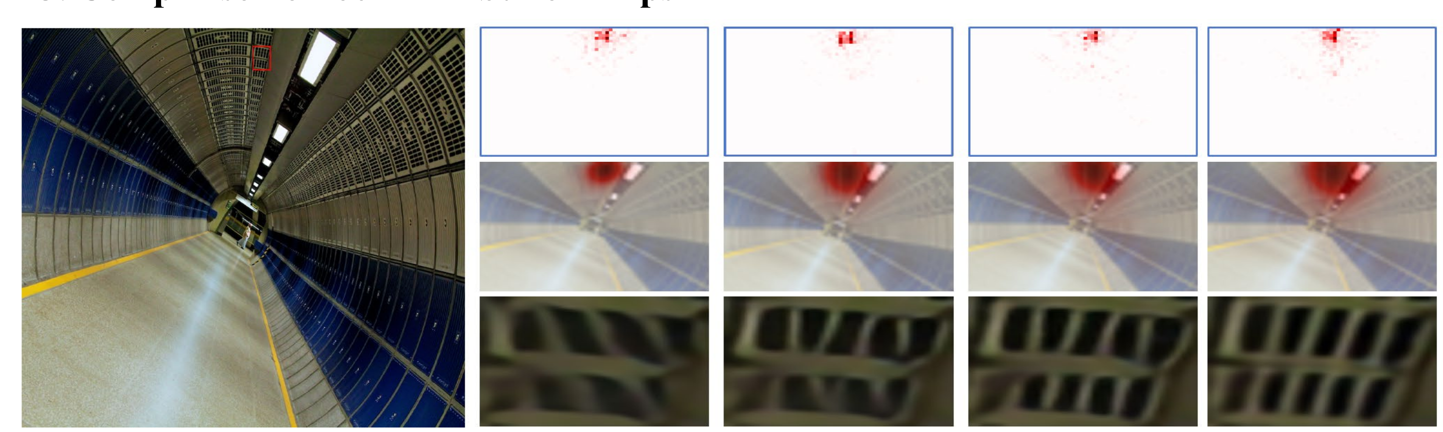
(a) HR image (b) Input F_{in} (c) X_l from EASA (d) Y_d from LDE

2. Effective receptive field visualizations



(a) EDSR-base (b) IMDN (c) SAFMN (d) SMFANet (e) SwinIR-light (f) NGSwin (g) SRFormer-light (h) SMFANet+

3. Comparison of local attribution maps



img078 from Urban100

DI: 4.0422 (a) SwinIR-light
DI: 9.7141 (b) ELAN-light
DI: 10.8982 (c)HPINet-S
DI: 14.3715 (d) SMFANet+