Multi-scale Attention Network for Single Image Super-Resolution

Yan Wang, Yusen Li, Gang Wang, Xiaoguang Liu

CVPR SEATTLE, WA JUNE 17-21, 2024

Introduction

Motivation

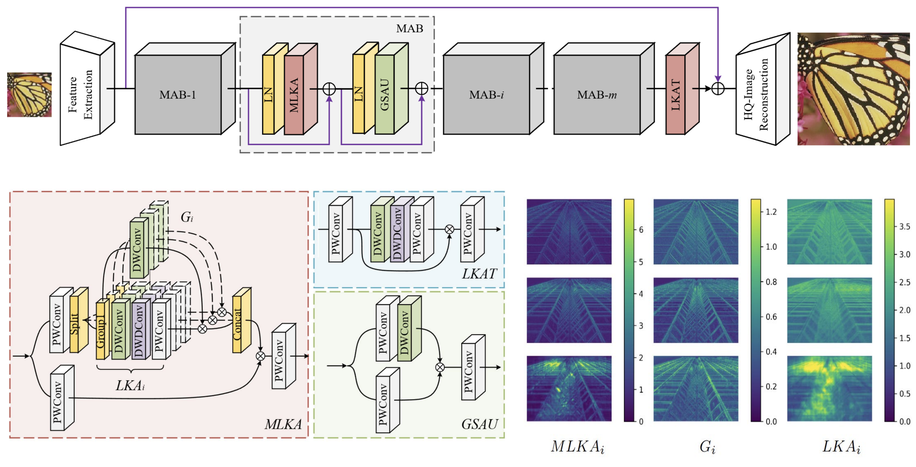
- Fixed-size LKA is inflexible to fully exploit the long-range correlations.
- Dilation convolution in LKA may cause blocking artifacts, leading to performance drop.
- ConvNet can perform well as Transformer in low-level vision.

Backbone	#Params	#FLOPs	Set5	Set14	B100	U100
ConvNeXt-S [35]	833K	47.3G	32.22	28.62	27.60	26.07
VAN-S [15]	818K	46.0G	32.27	28.70	27.63	26.18
SwinIR-light [30]	896K	49.6G	32.30	28.73	27.65	26.30
MAN-light	840K	47.1G	32.33	28.76	27.67	26.31

Contribution

- We propose **M**ulti-scale **L**arge **K**ernel **A**ttention (**MLKA**) to model long-range dependencies at various levels by combining large kernel with gate and multi-scale mechanisms
- We integrate gate mechanisms and spatial attention to construct a simplified feed-forward network called **GSAU**.
- Through simply stacking the proposed modules, we present **M**ulti-scale **A**ttention **N**etwork (**MAN**) family capable of achieving a trade-off between model complexity and performance in lightweight and performance-oriented SR tasks.

Method



MLKA

- **Large kernel attention:** LKA adaptively builds the long-range relationship by decomposing a $K \times K$ convolution into three convolutions:

$$LKA(\boldsymbol{X}) = f_{PW} f_{DWD} f_{DW}(\boldsymbol{X})$$

- **Multi-scale mechanism:** splits the input features into $\boldsymbol{n}$-pieces $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ of $\left\lfloor \frac{C}{n} \right\rfloor \times H \times W$, each piece feature uses a LKA to generate a homogeneous scale attention maps.
- **Gated aggregation:** to avoid the block effect, as well as to learn more local information, we leverage spatial gate to dynamically adapt LKA by:

$$MLKA_i(\boldsymbol{X}_i) = LKA_i(\boldsymbol{X}_i) \otimes G_i(\boldsymbol{X}_i).$$

Results

Method	Scale	#Params	#FLOPs	Set5 [2]		Set14 [55]		BSD100 [38]		Urban100 [18]		Manga109 [39]	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RCAN [58]	×2	15.4M	3.5T	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	39.44	0.9786
SAN [8]	×2	15.9M	3.1T	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792
HAN [41]	×2	63.6M	14.6T	38.27	0.9614	34.16	0.9217	32.41	0.9027	33.35	0.9385	39.46	0.9785
IGNN [59]	×2	49.5M	-	38.24	0.9613	34.07	0.9217	32.41	0.9025	33.23	0.9383	39.35	0.9786
NLSA [40]	×2	41.8M	9.6T	38.34	0.9618	34.08	0.9231	32.43	0.9027	33.42	0.9394	39.59	0.9789
DFSA+ [37]	×2	-	-	38.38	0.9620	34.33	0.9232	32.50	0.9036	33.66	0.9412	39.98	0.9798
MAN	×2	8.7M	1.7T	38.42	0.9622	34.40	0.9242	32.53	0.9043	33.73	0.9422	40.02	0.9801
MAN+	×2	8.7M	-	**38.44**	**0.9623**	**34.49**	**0.9248**	**32.55**	**0.9045**	**33.86**	**0.9430**	**40.13**	**0.9804**
SwinIR† [30]	×2	11.8M	2.3T	38.42	0.9623	34.46	0.9250	32.53	0.9041	33.81	0.9427	39.92	0.9797
RCAN [58]	×3	15.6M	1.6T	34.74	0.9299	30.65	0.8482	29.32	0.8111	29.09	0.8702	34.44	0.9499
SAN [8]	×3	15.9M	1.6T	34.75	0.9300	30.59	0.8476	29.33	0.8112	28.93	0.8671	34.30	0.9494
HAN [41]	×3	64.3M	6.5T	34.75	0.9299	30.67	0.8483	29.32	0.8110	29.10	0.8705	34.48	0.9500
IGNN [59]	×3	49.5M	-	34.72	0.9298	30.66	0.8484	29.31	0.8105	29.03	0.8696	34.39	0.9496
NLSA [40]	×3	44.7M	4.6T	34.85	0.9306	30.70	0.8485	29.34	0.8117	29.25	0.8726	34.57	0.9508
DFSA+ [37]	×3	-	-	34.92	0.9312	30.83	0.8507	29.42	0.8128	29.44	0.8761	35.07	0.9525
MAN	×3	8.7M	0.8T	34.91	0.9312	30.88	0.8514	29.43	0.8138	29.52	0.8782	35.06	0.9526
MAN+	×3	8.7M	-	**34.97**	**0.9315**	**30.91**	**0.8522**	**29.47**	**0.8144**	**29.65**	**0.8799**	**35.21**	**0.9533**
SwinIR† [30]	×3	11.9M	1.0T	34.97	0.9318	30.93	0.8534	29.46	0.8145	29.75	0.8826	35.12	0.9537
RCAN [58]	×4	15.6M	0.9T	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
SAN [8]	×4	15.9M	0.9T	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169
HAN [41]	×4	64.2M	3.8T	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094	31.42	0.9177
IGNN [59]	×4	49.5M	-	32.57	0.8998	28.85	0.7891	27.77	0.7434	26.84	0.8090	31.28	0.9182
NLSA [40]	×4	44.2M	3.0T	32.59	0.9000	28.87	0.7891	27.78	0.7444	26.96	0.8109	31.27	0.9184
DFSA+ [37]	×4	-	-	32.79	0.9019	29.06	0.7922	27.87	0.7458	27.17	0.8163	31.88	**0.9266**
MAN	×4	8.7M	0.4T	32.81	0.9024	29.07	0.7934	27.90	0.7477	27.26	0.8197	31.92	0.9230
MAN+	×4	8.7M	-	**32.87**	**0.9030**	**29.12**	**0.7941**	**27.93**	**0.7483**	**27.39**	**0.8223**	**32.13**	0.9248
SwinIR† [30]	×4	11.9M	0.6T	32.92	0.9044	29.09	0.7950	27.92	0.7489	27.45	0.8254	32.03	0.9260