

# Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning

Xuehui Wang<sup>1,3</sup>, Qing Wang<sup>1</sup>, Yuzhi Zhao<sup>2</sup>, Junchi Yan<sup>3</sup>, Lei Fan<sup>4</sup>, Long Chen<sup>1\*</sup>

<sup>1</sup>Sun Yat-sen University, China

<sup>2</sup>City University of Hong Kong, Hong Kong, China

<sup>3</sup>Shanghai Jiao Tong University, China

<sup>4</sup>Northwestern University, USA

# Background



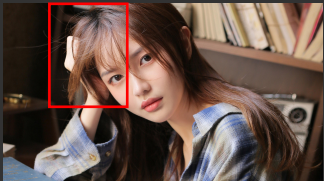
## What is Super Resolution (SR)?

*Ideal target: Upsample an image and keep its high fidelity.*

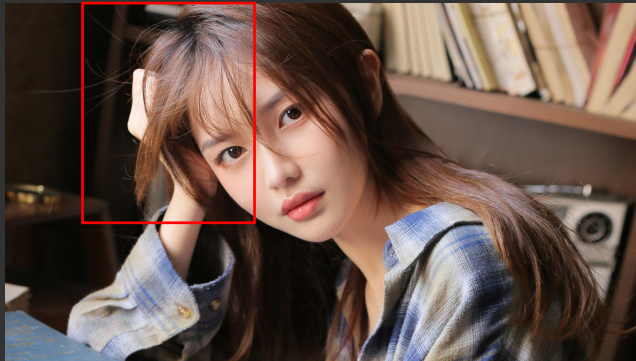
2000x1000

1000x500

500x250



origin



x2



x4

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning



# Background

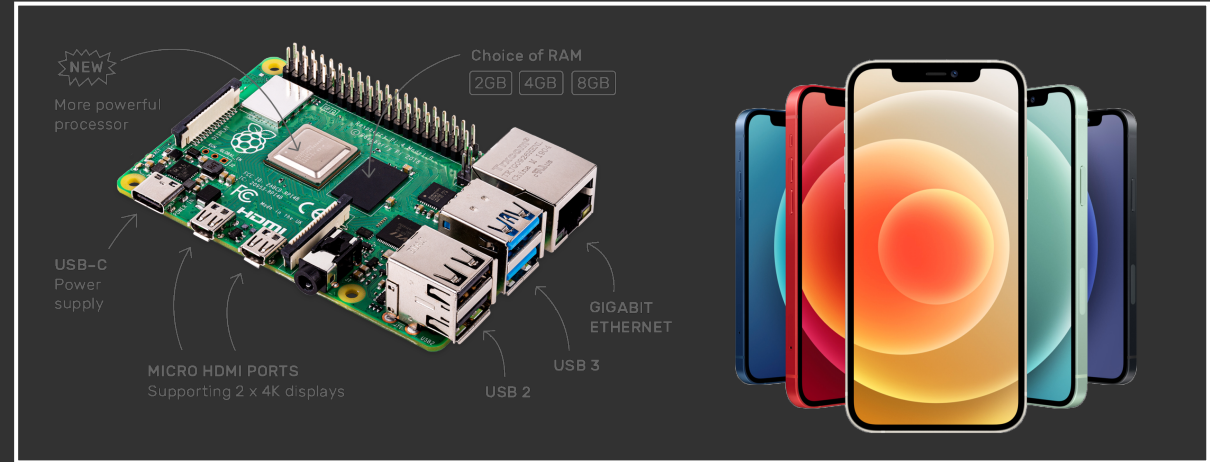
What is the problem for current SR?



apply



inference slowly



Impossible for Embedded devices

*How about directly design a  
lightweight SR method?*

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning



# Motivation



- ✓ Reconstruct the high frequency details from the low resolution images *hierarchically*

10 layers



20 layers



30 layers



40 layers



Different layers maybe learn the capacity of recovering *different frequency information*

*How to combine them in a limited resources situation is a direction.*

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning

# Motivation



We have the following observations:

- *Reducing the depth or the width of the network, Performing upsample at the end of the network, Adopting small kernels are useful measures.*  
**Over-fitting for some specific patterns, loss generalization ability.**
  - *Diverse compacts for different frequency information.*  
**Should be decided by network rather than us.**
- 

We decide to do these:

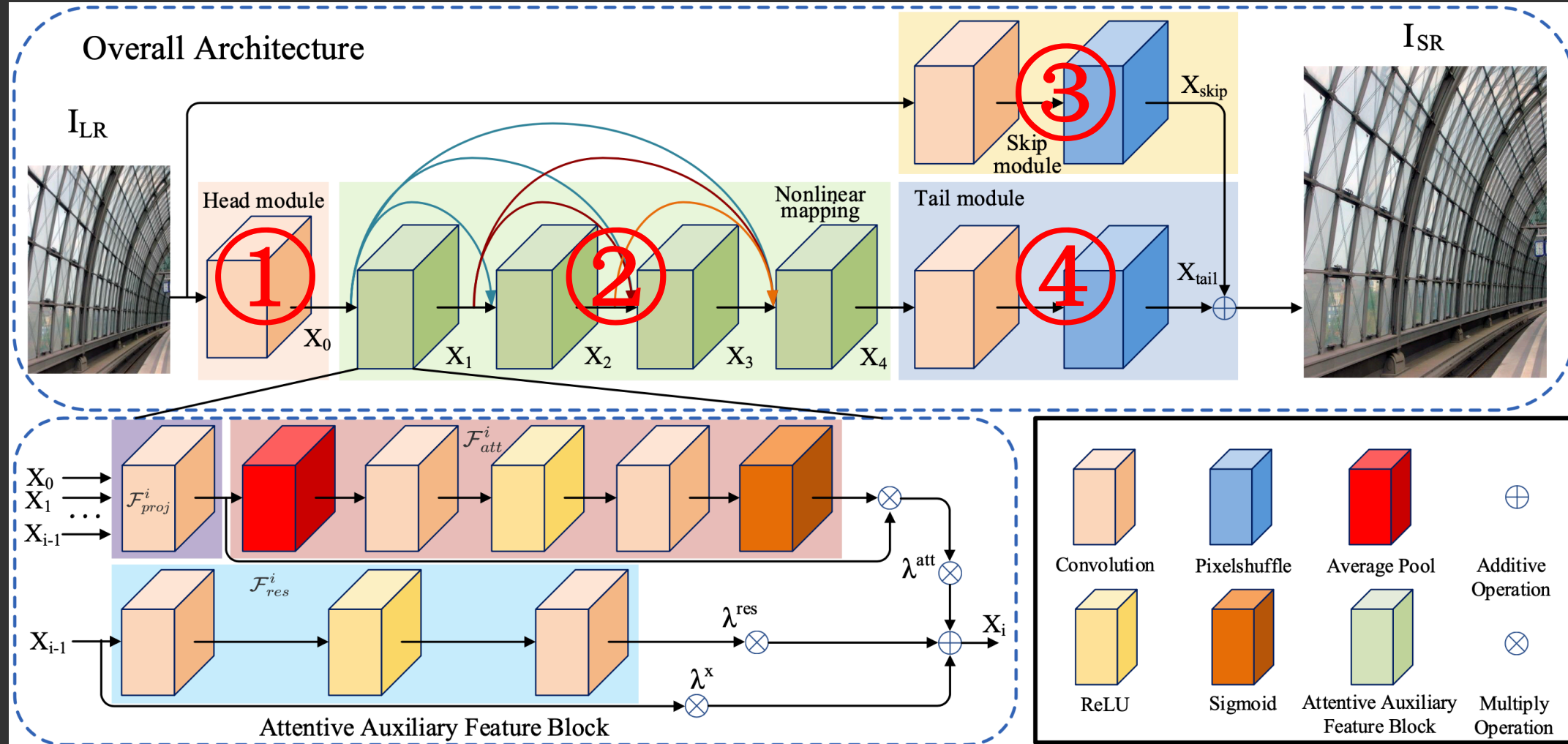
- *Explore to reuse auxiliary features in a attention-based mode.*
- *Build a same-space attention operation.*
- *Combine the effect of former blocks with a residual structure.*



# Methodology



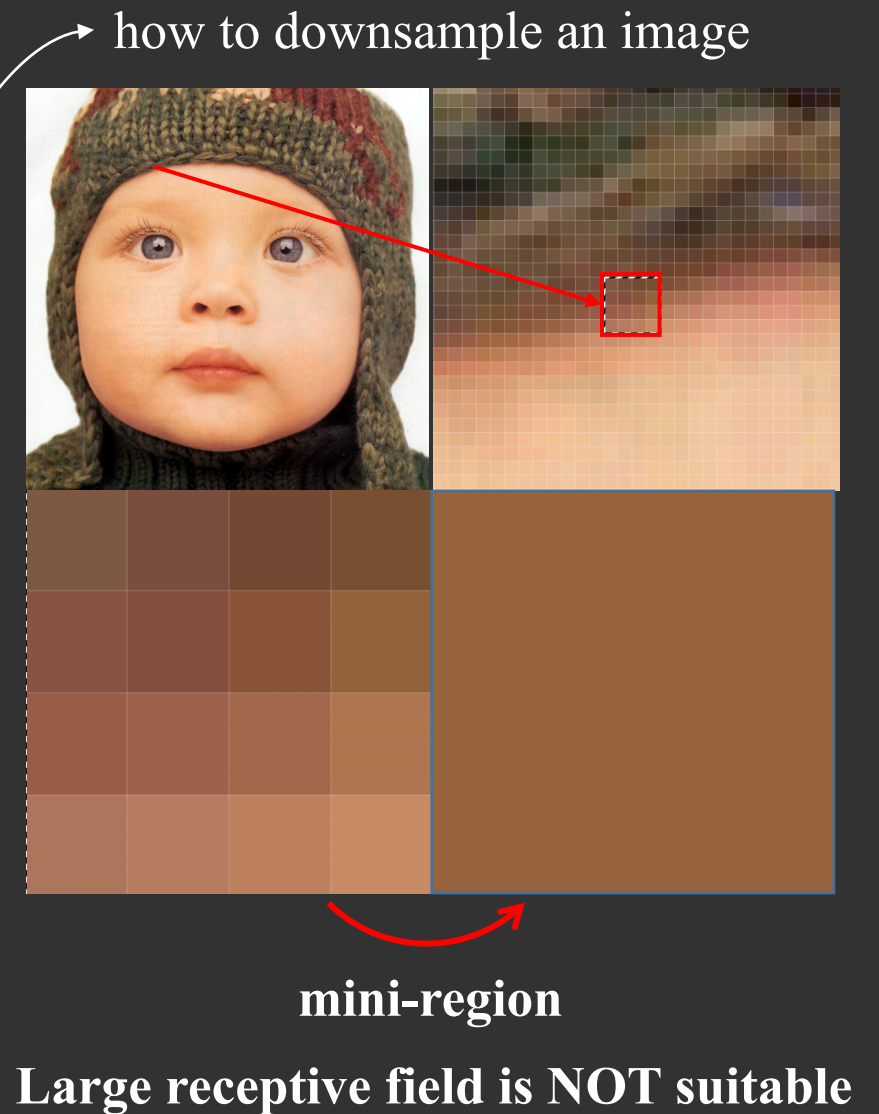
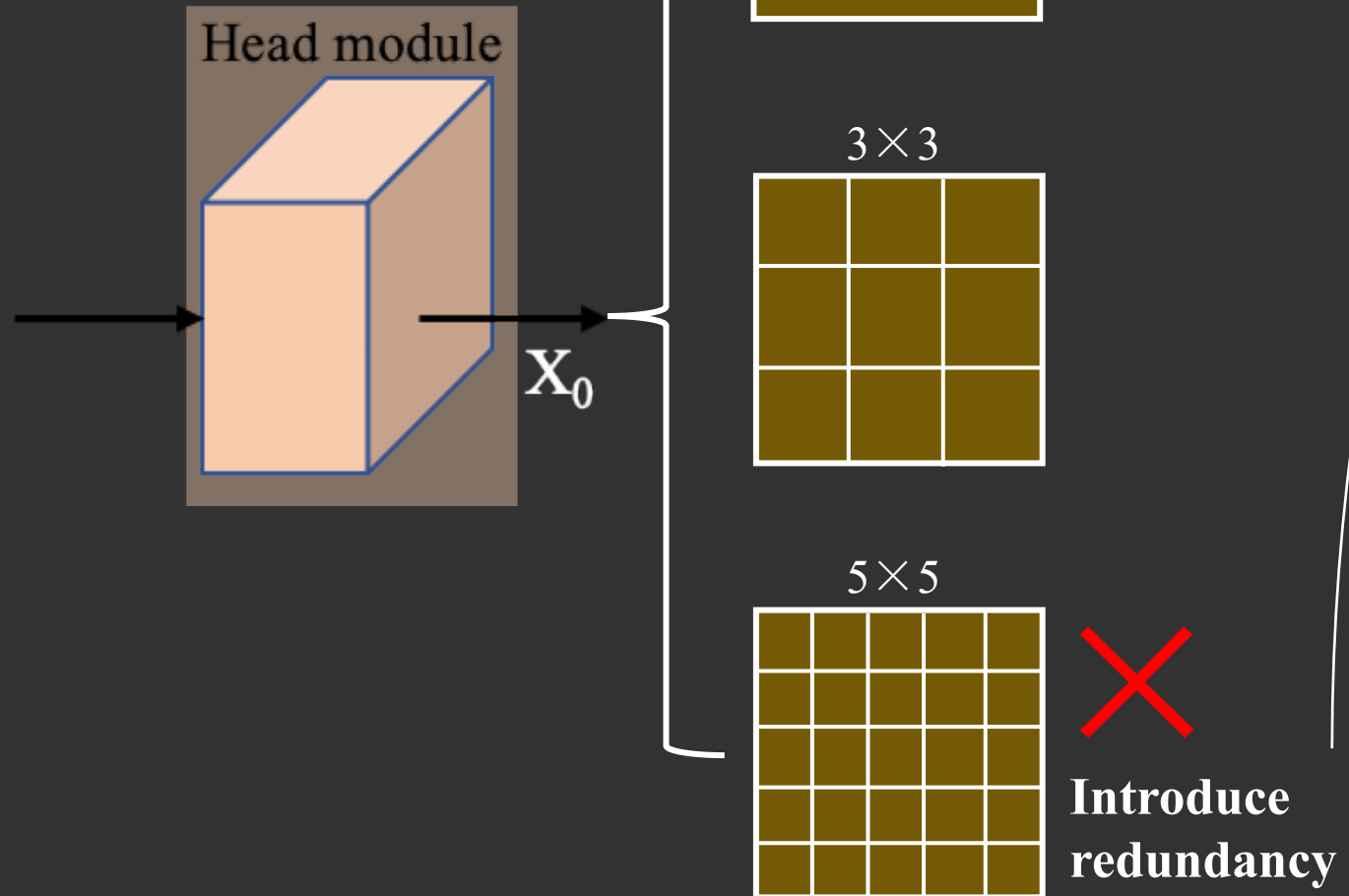
## 4 PARTS



Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning

# Methodology

## Head Module



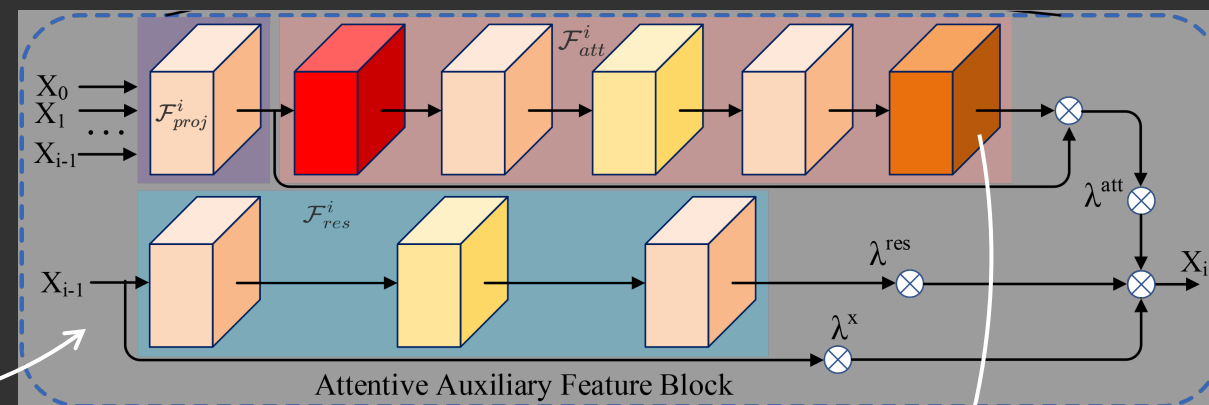
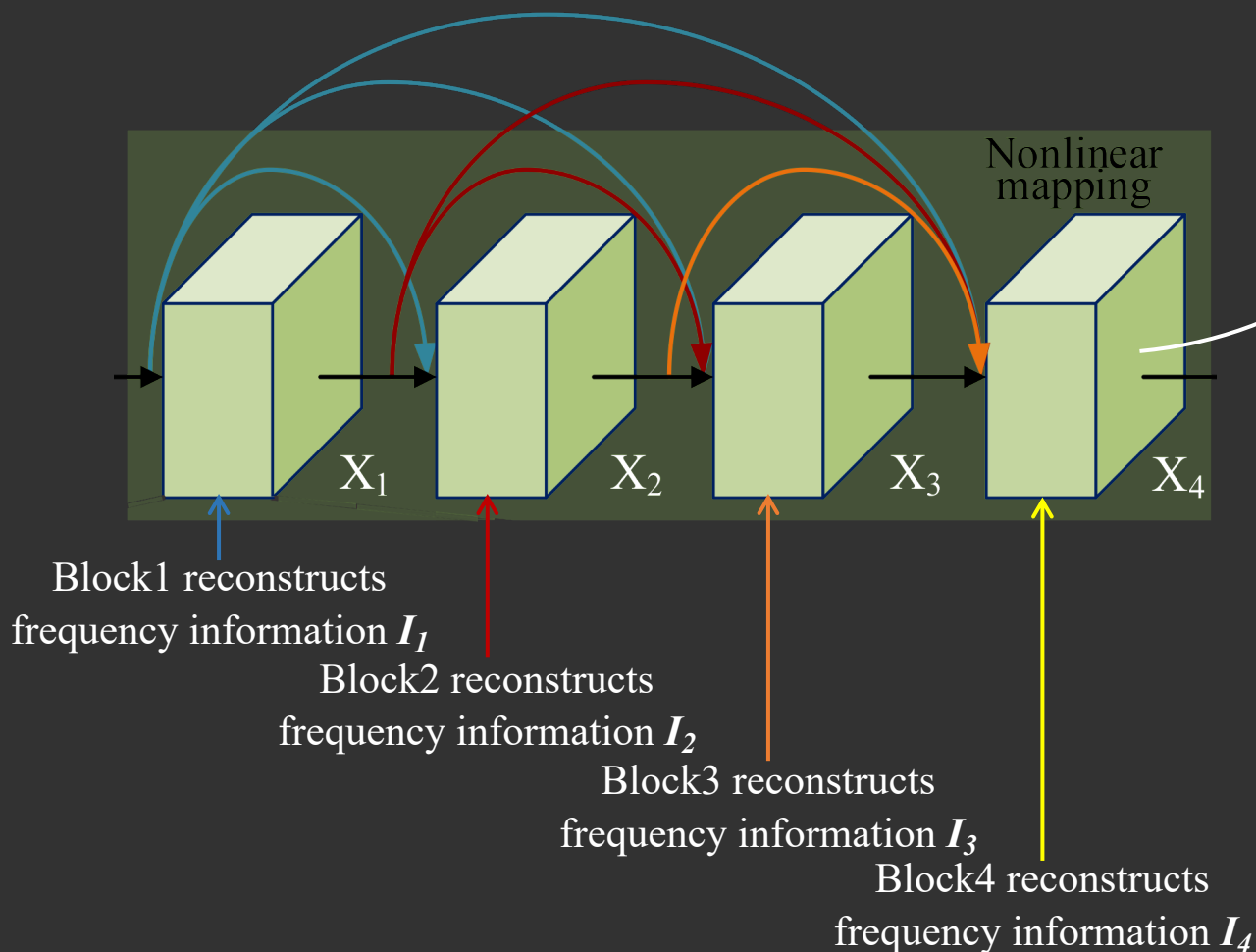
# Methodology



香港城市大學  
City University of Hong Kong



## Nonlinear Mapping Module



Weight for each channel in different blocks

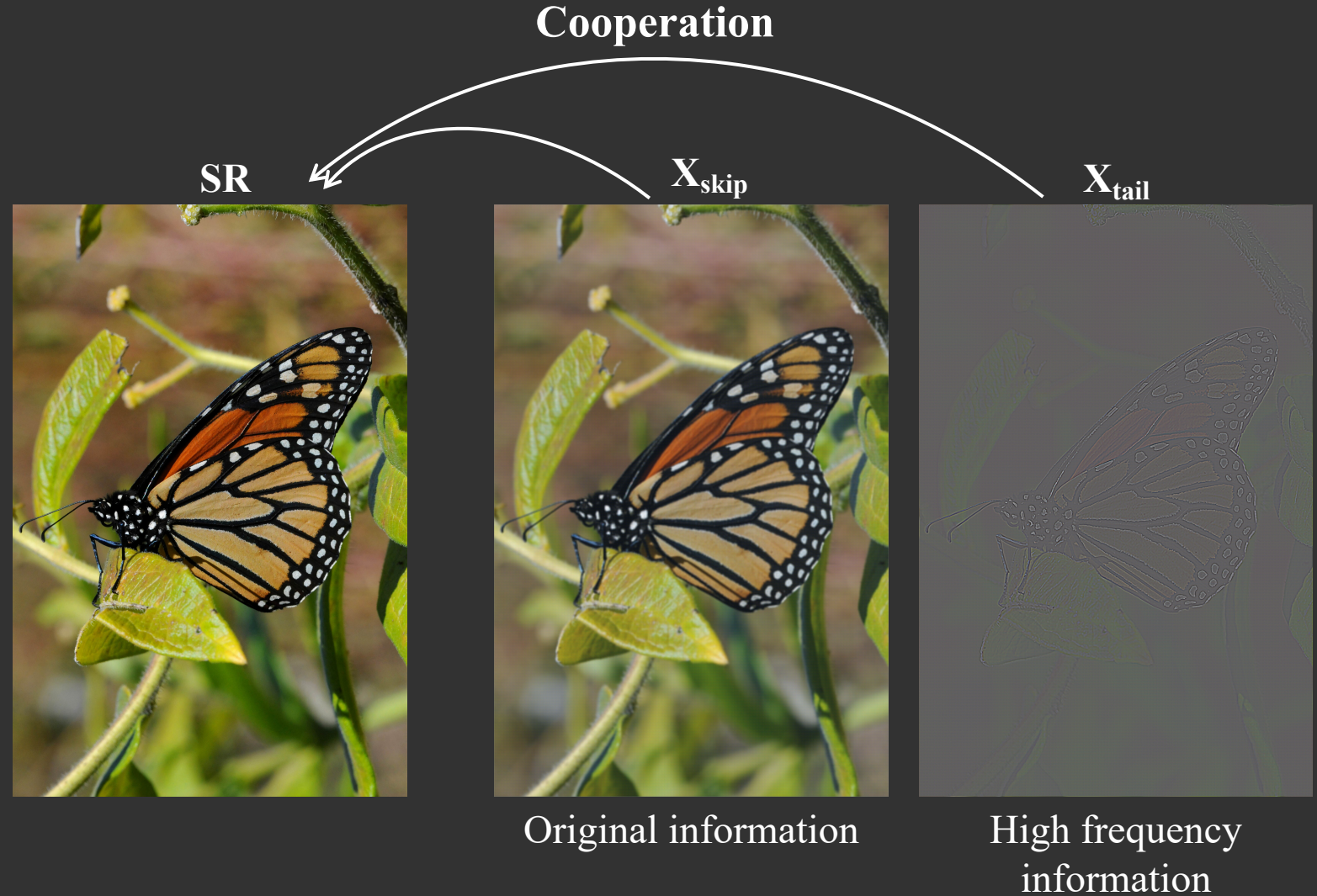
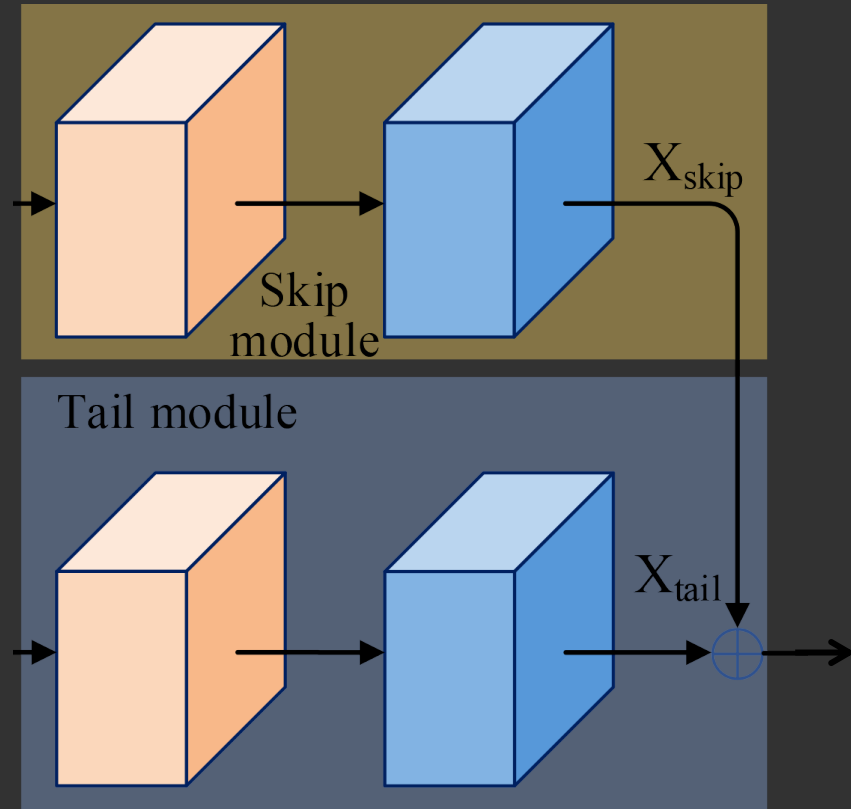




# Methodology



## Tail Module | Skip Module

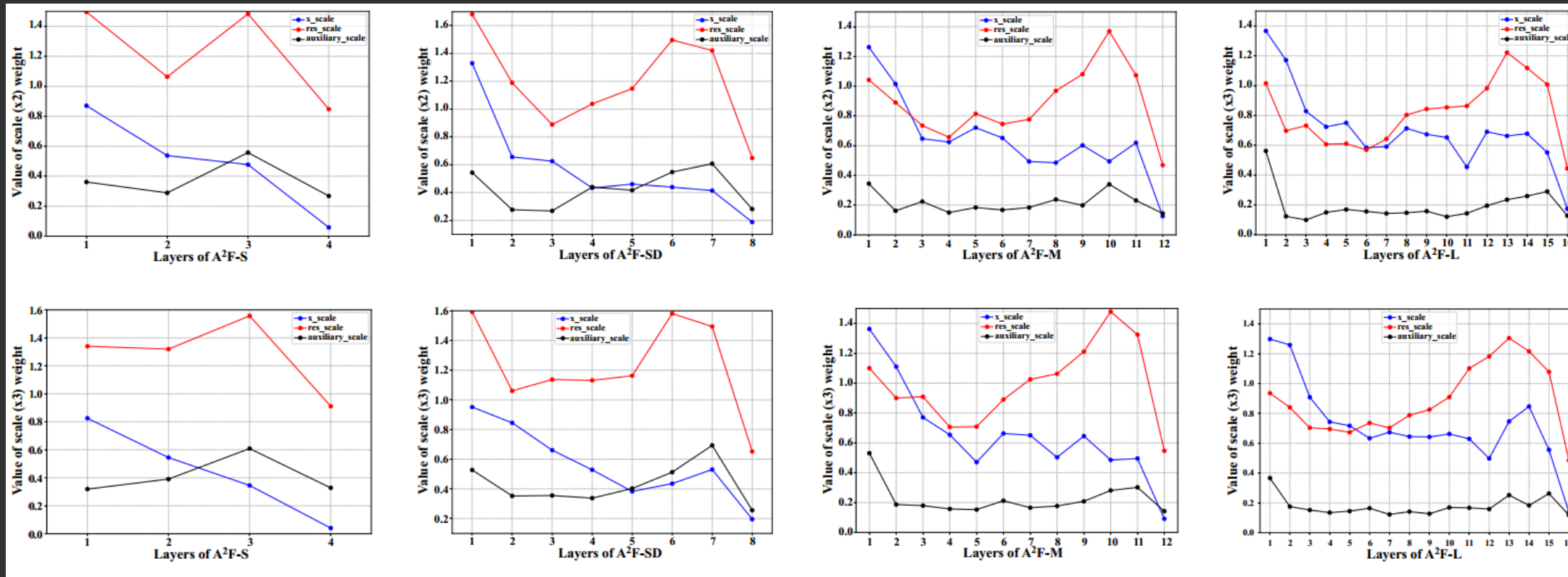


# Experiments



## Ablation Study

- Effect of Auxiliary Features
- Effect of Projection Unit and Channel Attention
- Kernel Selection in Head Module



The weight of three components for different layer in A<sup>2</sup>F-S/A<sup>2</sup>F-SD/A<sup>2</sup>F-M/A<sup>2</sup>F-L

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning

# Experiments



## Ablation Study

- Effect of Auxiliary Features
- Effect of Projection Unit and Channel Attention
- Kernel Selection in Head Module

Model	PU	CA	Param	MutiAdds	Set5	Set14	B100	Urban100	Manga109
BASELINE			1190K	273.9G	38.04	33.69	32.20	32.20	38.66
BASELINE-MP			1338K	308.0G	38.09	33.70	32.21	32.25	38.69
A <sup>2</sup> F-L-NOCA	✓		1329K	306.0G	38.08	33.75	32.23	32.39	38.79
A <sup>2</sup> F-L-NOCA-MP	✓		1368K	315.1G	38.09	33.77	32.23	32.35	38.79
A <sup>2</sup> F-L	✓	✓	1363K	306.1G	<b>38.09</b>	<b>33.78</b>	<b>32.23</b>	<b>32.46</b>	<b>38.95</b>

Results of ablation study on the projection unit and the channel attention

Convolutional Kernel Selection						
Kernel	Parameters	Set5	Set14	B100	Urban100	Manga109
1 × 1	319.2K	32.00	28.46	27.46	25.78	30.13
<b>3 × 3</b>	<b>319.6K</b>	<b>32.06</b>	<b>28.47</b>	<b>27.48</b>	<b>25.80</b>	<b>30.16</b>
5 × 5	320.4K	32.00	28.45	27.48	25.80	38.13
7 × 7	321.6K	31.99	28.44	27.48	25.78	30.10

Results of ablation study on different kernel size

**Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning**

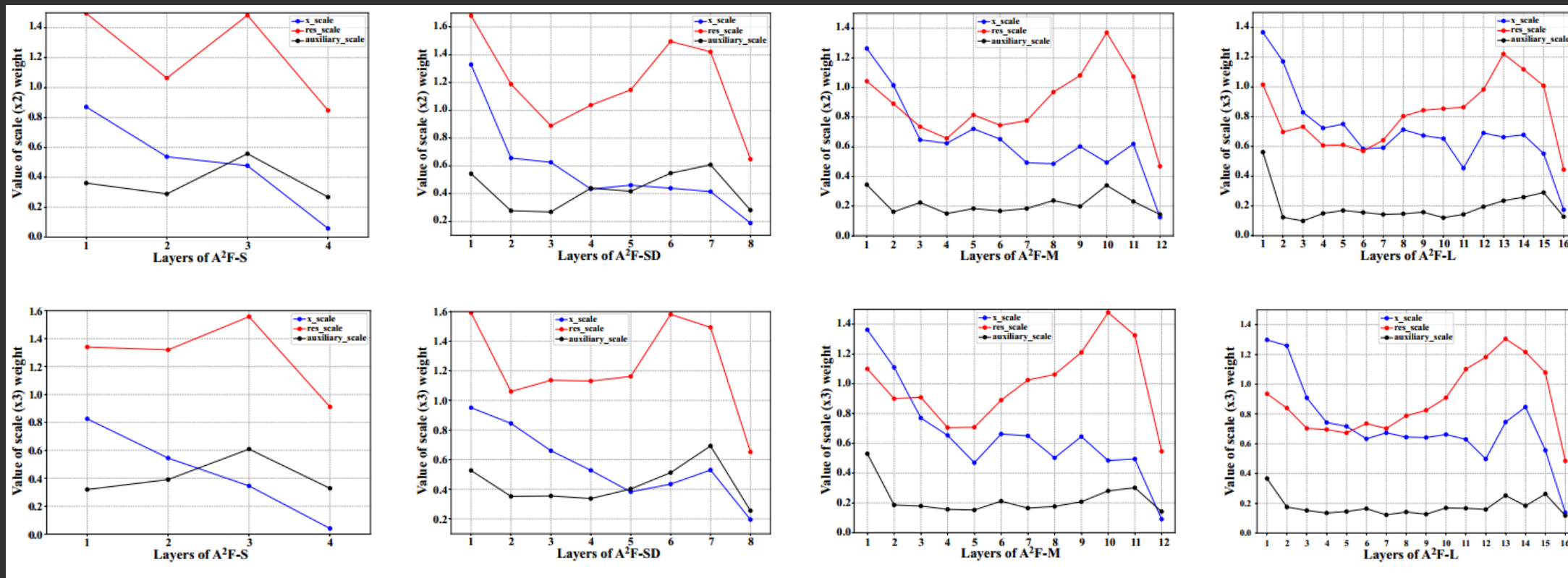


# Experiments



## Ablation Study What we know?

- ✓ Auxiliary features play a certain role for output feature in one block.
- ✓ Auxiliary features are more and more important in smaller model.



The weight of three components for different layer in A<sup>2</sup>F-S/A<sup>2</sup>F-SD/A<sup>2</sup>F-M/A<sup>2</sup>F-L

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning

# Experiments



## Ablation Study What we know?

- ✓ The projection unit and attention operation both explore the using of auxiliary features.
- ✓ Verify the reasonability of the head module's kernel size.

Model	PU	CA	Param	MutiAdds	Set5	Set14	B100	Urban100	Manga109
BASELINE			1190K	273.9G	38.04	33.69	32.20	32.20	38.66
BASELINE-MP			1338K	308.0G	38.09	33.70	32.21	32.25	38.69
A <sup>2</sup> F-L-NOCA	✓		1329K	306.0G	38.08	33.75	32.23	32.39	38.79
A <sup>2</sup> F-L-NOCA-MP	✓		1368K	315.1G	38.09	33.77	32.23	32.35	38.79
A <sup>2</sup> F-L	✓	✓	1363K	306.1G	<b>38.09</b>	<b>33.78</b>	<b>32.23</b>	<b>32.46</b>	<b>38.95</b>

Results of ablation study on the projection unit and the channel attention

Convolutional Kernel Selection						
Kernel	Parameters	Set5	Set14	B100	Urban100	Manga109
1 × 1	319.2K	32.00	28.46	27.46	25.78	30.13
<b>3 × 3</b>	<b>319.6K</b>	<b>32.06</b>	<b>28.47</b>	<b>27.48</b>	<b>25.80</b>	<b>30.16</b>
5 × 5	320.4K	32.00	28.45	27.48	25.80	38.13
7 × 7	321.6K	31.99	28.44	27.48	25.78	30.10

Results of ablation study on different kernel size



# Experiments



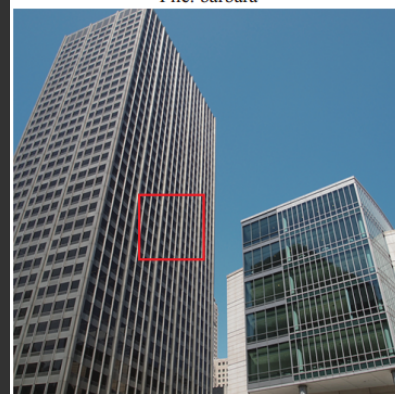
## Comparison with SOTAs

➤ PSNR, SSIM, LPIPS, Running time, Gflops, visualization results.

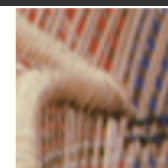
Scale	Size Scope	Model	Param	MutiAdds	Set5	Set14	B100	Urban100	Manga109
x2	< 5 × 10 <sup>2</sup> K	FSRCNN	12K	6G	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694
		SRCNN	57K	52.7G	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661
		DRRN	297K	6797G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.92/0.9760
		A <sup>2</sup> F-SD(ours)	313k	71.2G	37.91/0.9602	33.45/0.9164	32.08/0.8986	31.79/0.9246	38.52/0.9767
		A <sup>2</sup> F-S(ours)	320k	71.7G	37.79/0.9597	33.32/0.9152	31.99/0.8972	31.44/0.9211	38.11/0.9757
		FALSR-B	326K	74.7G	37.61/0.9585	33.29/0.9143	31.97/0.8967	31.28/0.9191	-
		AWSRN-SD	348K	79.6G	37.86/0.9600	33.41/0.9161	32.07/0.8984	31.67/0.9237	38.20/0.9762
		AWSRN-S	397K	91.2G	37.75/0.9596	33.31/0.9151	32.00/0.8974	31.39/0.9207	37.90/0.9755
		FALSR-C	408K	93.7G	37.66/0.9586	33.26/0.9140	31.96/0.8965	31.24/0.9187	-
	CARN-M	412K	91.2G	37.53/0.9583	33.26/0.9141	31.92/0.8960	31.23/0.9193	-	
	SRFBN-S	483K	-	37.78/0.9597	33.35/0.9156	32.00/0.8970	31.41/0.9207	38.06/0.9757	
	< 10 <sup>3</sup> K	IDN	552K	-	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	-
		VDSR	665K	612.6G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
		MemNet	677K	2662.4G	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	-
		LapSRN	813K	29.9G	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
		SelNet	974K	225.7G	37.89/0.9598	33.61/0.9160	32.08/0.8984	-	-
A <sup>2</sup> F-M(ours)		999k	224.2G	38.04/0.9607	33.67/0.9184	32.18/0.8996	32.27/0.9294	38.87/0.9774	
< 2 × 10 <sup>3</sup> K		FALSR-A	1021K	234.7G	37.82/0.9595	33.55/0.9168	32.12/0.8987	31.93/0.9256	-
		MoreMNAS-A	1039K	238.6G	37.63/0.9584	33.23/0.9138	31.95/0.8961	31.24/0.9187	-
		AWSRN-M	1063K	244.1G	38.04/0.9605	33.66/0.9181	32.21/0.9000	32.23/0.9294	38.66/0.9772
	A <sup>2</sup> F-L(ours)	1363k	306.1G	38.09/0.9607	33.78/0.9192	32.23/0.9002	32.46/0.9313	38.95/0.9772	
	AWSRN	1397K	320.5G	38.11/0.9608	33.78/0.9189	32.26/0.9006	32.49/0.9316	38.87/0.9776	
	SRMDNF	1513K	347.7G	37.79/0.9600	33.32/0.9150	32.05/0.8980	31.33/0.9200	-	
	CARN	1592K	222.8G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	-	
	DRCN	1774K	1797.4G	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723	
	MSRN	5930K	1365.4G	38.08/0.9607	33.70/0.9186	32.23/0.9002	32.29/0.9303	38.69/0.9772	
x4	< 10 <sup>3</sup> K	FSRCNN	12K	4.6G	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
		SRCNN	57K	52.7G	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
		DRRN	297K	6797G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8960
		A <sup>2</sup> F-SD(ours)	320k	18.2G	32.06/0.8928	28.47/0.7790	27.48/0.7373	25.80/0.7767	30.16/0.9038
		A <sup>2</sup> F-S(ours)	331k	18.6G	31.87/0.8900	28.36/0.7760	27.41/0.7305	25.58/0.7685	29.77/0.8987
		CARN-M	412K	32.5G	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.62/0.7694	-
		AWSRN-SD	444K	25.4G	31.98/0.8921	28.46/0.7786	27.48/0.7368	25.74/0.7746	30.09/0.9024
		SRFBN-S	483K	132.5G	31.98/0.8923	28.45/0.7779	27.44/0.7313	25.71/0.7719	29.91/0.9008
		IDN	552K	-	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	-
	< 2 × 10 <sup>3</sup> K	AWSRN-S	588K	37.7G	31.77/0.8893	28.35/0.7761	27.41/0.7304	25.56/0.7678	29.74/0.8982
		VDSR	665K	612.6G	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
		MemNet	677K	2662.4G	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	-
		LapSRN	813K	149.4G	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8845
		A <sup>2</sup> F-M(ours)	1010k	56.7G	32.28/0.8955	28.62/0.7828	27.58/0.7364	26.17/0.7892	30.57/0.9100
		AWSRN-M	1254K	72G	32.21/0.8954	28.65/0.7832	27.60/0.7368	26.15/0.7884	30.56/0.9093
		A <sup>2</sup> F-L(ours)	1374K	77.2G	32.32/0.8964	28.67/0.7839	27.62/0.7379	26.32/0.7931	30.72/0.9115
SelNet		1417K	83.1G	32.00/0.8931	28.49/0.7783	27.44/0.7325	-	-	
SRMDNF		1555K	89.3G	31.96/0.8930	28.35/0.7770	27.49/0.7340	25.68/0.7730	-	
< 10 <sup>4</sup> K	AWSRN	1587K	91.1G	32.27/0.8960	28.69/0.7843	27.64/0.7385	26.29/0.7930	30.72/0.9109	
	CARN	1592K	90.9G	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	-	
	DRCN	1774K	1797.4G	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816	
	SRDenseNet	2015K	389.9K	32.02/0.8934	28.35/0.7770	27.53/0.7337	26.05/0.7819	-	
	MSRN	6078K	349.8G	32.26/0.8960	28.63/0.7836	27.61/0.7380	26.22/0.7911	30.57/0.9103	



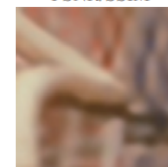
Dataset: Set14  
File: barbara



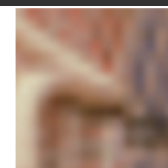
Dataset: Urban100  
File: img\_096



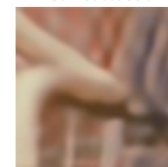
Ground Truth  
Params/Multi-adds  
PSNR/SSIM



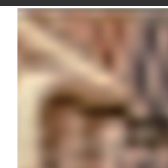
DRRN  
297K/6797G  
25.744/0.7314



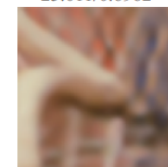
Bicubic  
Params/Multi-adds  
PSNR/SSIM



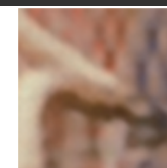
LapSRN  
813K/149.4G  
25.798/0.7426



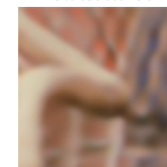
SRCNN  
57K/52.7G  
25.011/0.6962



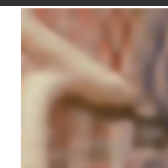
CARN-M  
412K/32.5G  
25.892/0.7463



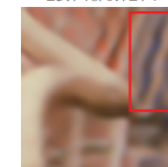
FSRCNN  
12K/4.6G  
25.709/0.7295



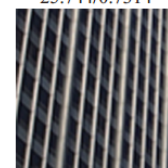
AWSRN-SD  
444K/25.4G  
25.874/0.7463



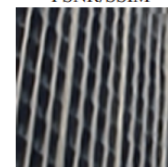
VDSR  
665K/612.6G  
25.748/0.7274



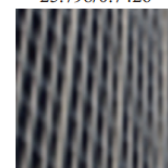
**A<sup>2</sup>F-SD(ours)**  
**320K/18.2G**  
**25.962/0.7491**



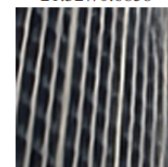
Ground Truth  
Params/Multi-adds  
PSNR/SSIM



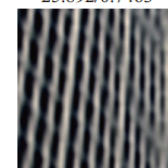
DRRN  
297K/6797G  
23.984/0.8250



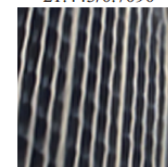
Bicubic  
Params/Multi-adds  
PSNR/SSIM



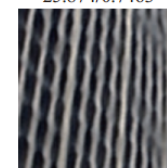
LapSRN  
813K/149.4G  
23.585/0.8193



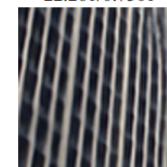
SRCNN  
57K/52.7G  
21.443/0.7090



CARN-M  
412K/32.5G  
24.129/0.8372



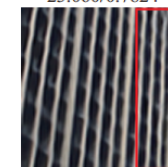
FSRCNN  
12K/4.6G  
22.261/0.7501



AWSRN-SD  
444K/25.4G  
23.852/0.8376



VDSR  
665K/612.6G  
23.060/0.7824



**A<sup>2</sup>F-SD(ours)**  
**320K/18.2G**  
**24.498/0.8491**

PSNR/SSIM on five datasets (x2, x4)

Visualization results on images sampled from Set14 and Urban100

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning



# Experiments



## Comparison with SOTAs

- PSNR, SSIM, LPIPS, Running time, Gflops, visualization results.

Methods	Params	GFLOPs	Set5	Set14	B100	Urban100	Manga109
AWSRN [28]	1587K	1.620G	0.1747	0.2853	0.3692	0.2198	0.1058
AWSRN-SD [28]	444K	-	0.1779	0.2917	0.3838	0.2468	0.1168
CARN [32]	1592K	1.620G	0.1761	0.2893	0.3799	0.2363	-
CARN-M [32]	412K	0.445G	0.1777	0.2938	0.3850	0.2524	-
SRFBN-S [31]	483K	0.323G	0.1776	0.2938	0.3861	0.2554	0.1396
IMDN [51]	715K	0.729G	0.1743	0.2901	0.3740	0.2350	0.1330
<b>A<sup>2</sup>F-SD</b>	320K	0.321G	0.1731	0.2870	0.3761	0.2375	0.1112
<b>A<sup>2</sup>F-L</b>	1374K	1.370G	0.1733	0.2846	0.3698	0.2194	0.1056

LPIPS/Gflops on five datasets  
for scale x4. Gflops are calculated  
based on the input size of 32x32

Model	Params	Multi-Adds	Running time(s)	PSNR
RCAN [7]	15590K	919.9G	0.8746	26.82
EDSR [3]	43090K	2896.3G	0.3564	26.64
D-DBPN [21]	10430K	685.7G	0.4174	26.38
SRFBN [31]	3631K	1128.7G	0.4291	26.60
SRFBN-S [31]	483K	132.5G	0.0956	25.71
VDSR [2]	665K	612.6G	0.1165	25.18
CARN-M [32]	412K	32.5G	0.0326	25.62
<b>A<sup>2</sup>F-SD</b>	<b>320K</b>	<b>18.2G</b>	<b>0.0145</b>	<b>25.80</b>
<b>A<sup>2</sup>F-L</b>	<b>1374K</b>	<b>77.2G</b>	<b>0.0324</b>	<b>26.32</b>

Running time comparison  
for scale x4 on Urban100

Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning



# Thank you!