The 31st International Joint Conference on Artificial Intelligence
and the 25th European Conference on Artificial Intelligence

Lightweight Bimodal Network for Single-Image Super-Resolution via Symmetric CNN and Recursive Transformer

Guangwei Gao[1,3†], Zhengxue Wang[1†], Juncheng Li[2], Wenjie Li[1], Yi Yu[3], Tieyong Zeng[2]

[1]Nanjing University of Posts and Telecommunications [2] The Chinese University of Hong Kong
[3] National Institute of Informatics

Reporter: Zhengxue Wang

https://github.com/IVIPLab/LBNet

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

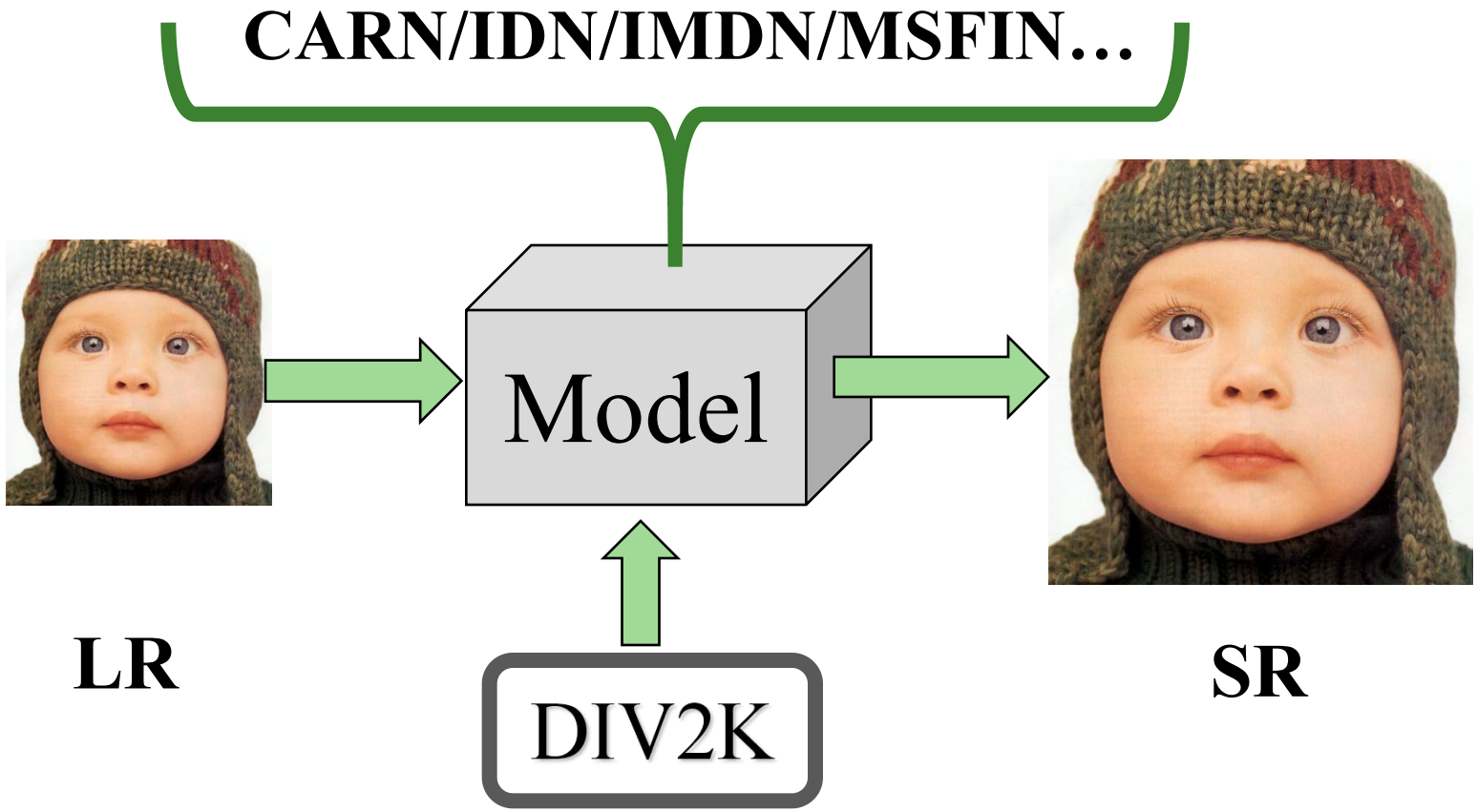
Background & Related Works

- Image Super-resolution
  - The purpose of single-image super-resolution (SISR) is **to reconstruct a high-resolution (HR) image from its degraded low-resolution (LR) counterpart**.

- Lightweight Image Super-resolution
  - **Lightweight** SR models are widely concerned for saving memory resources and computing resources.
  - In the case of **fewer parameters and computation**, a better performance is obtained.

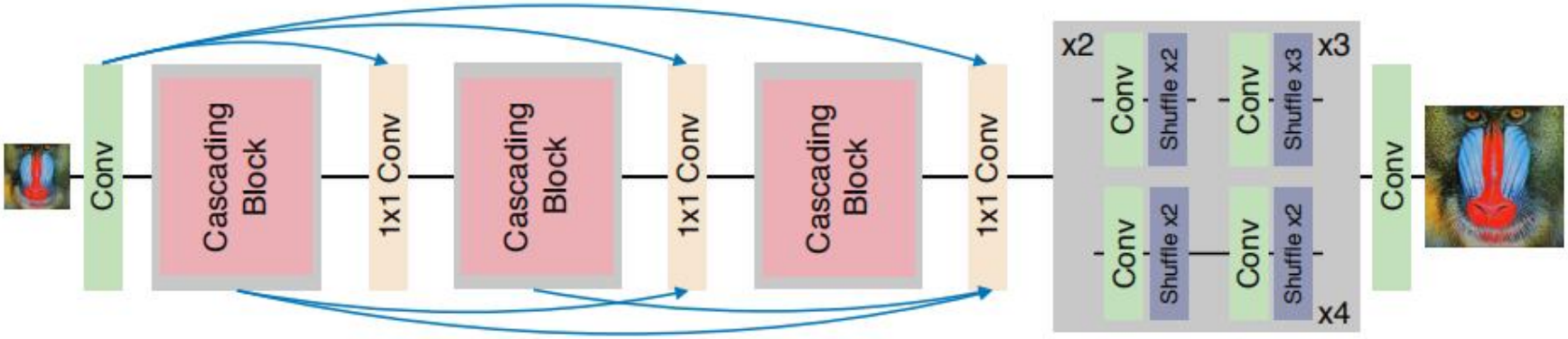
Background & Related Works

- How to reconstruct SR images?



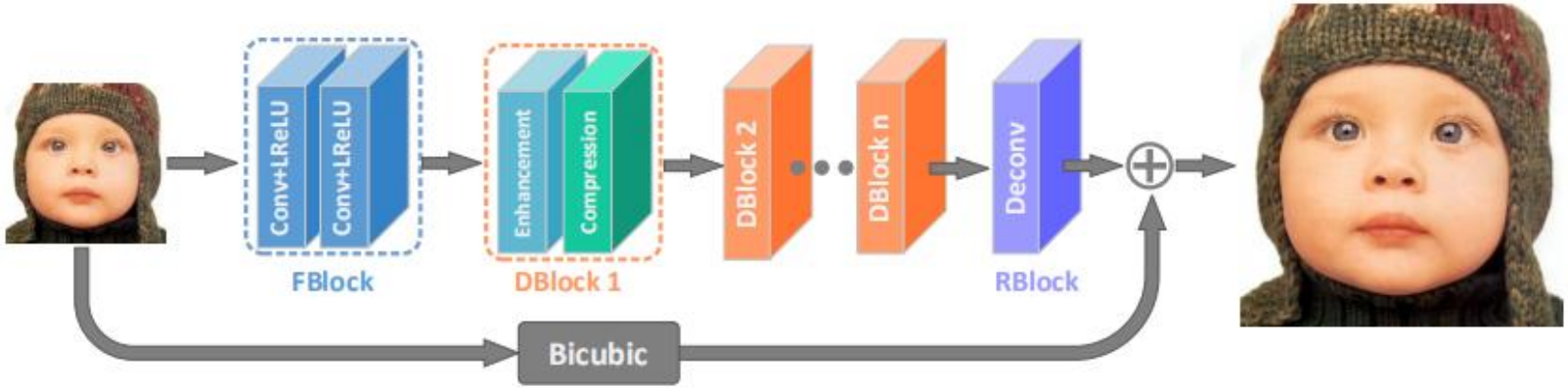
Background & Related Works

- Ahn's CARN



N. Ahn, et al. Fast, accurate, and lightweight super-resolution with cascading residual network, in *ECCV,* 2018, pp. 252–268.

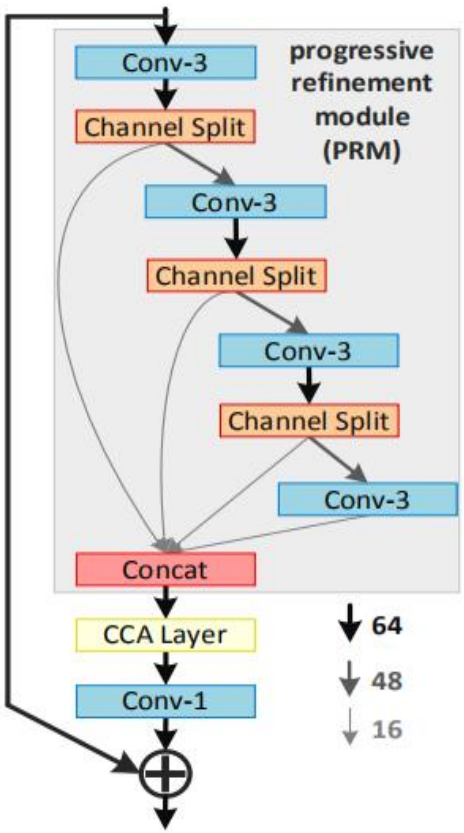
- Hui's IDN



Z. Hui, et al. Fast and accurate single image super-resolution via information distillation network, in *CVPR*, 2018, pp. 723–731.

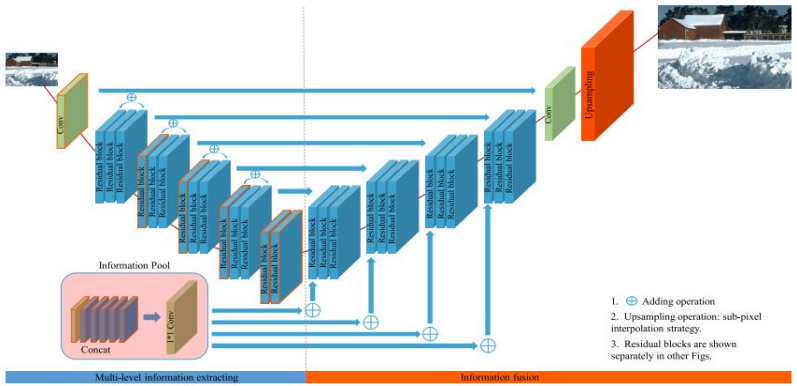
Background & Related Works

- Hui's



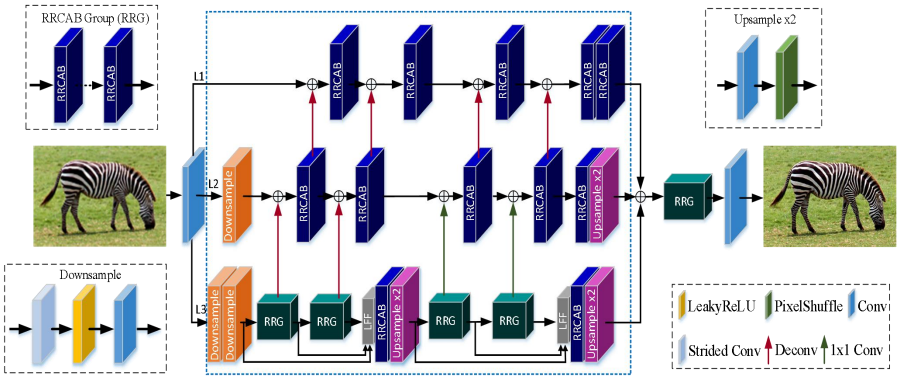
Z. Hui, et al. Lightweight image super-resolution with information multi-distillation network, in *ACM MM*, 2019, pp. 2024–2032.

- Li's s-LWSR



B. Li, et al. s-lwsr: Super lightweight super-resolution network, *IEEE TIP*, vol. 29, pp. 8368–8380, Aug. 2020.

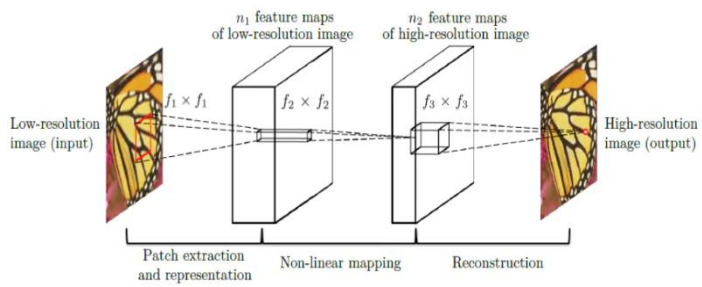
- Wang's MSFIN



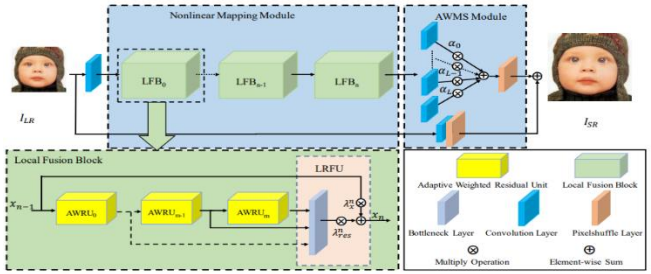
B. Li, et al. s-lwsr: Super lightweight super-resolution network, *IEEE TIP*, vol. 29, pp. 8368–8380, Aug. 2020.

Background & Related Works

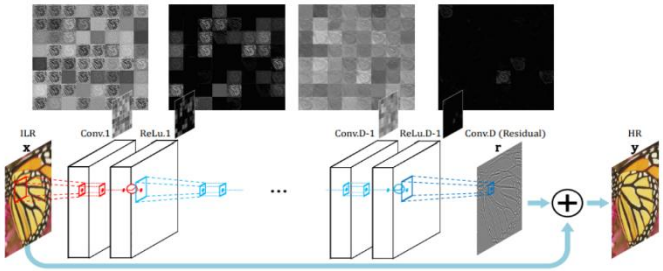
- Other Lightweight Methods



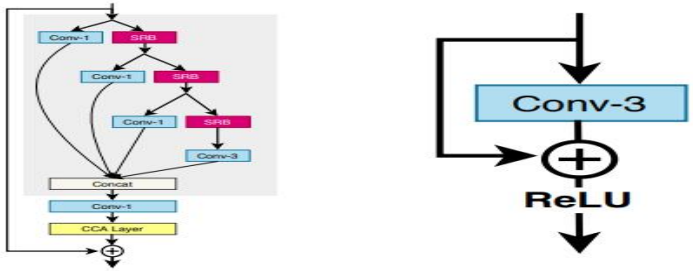
SRCNN



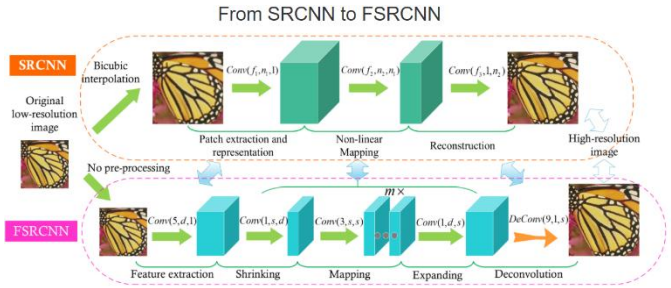
AWSRN



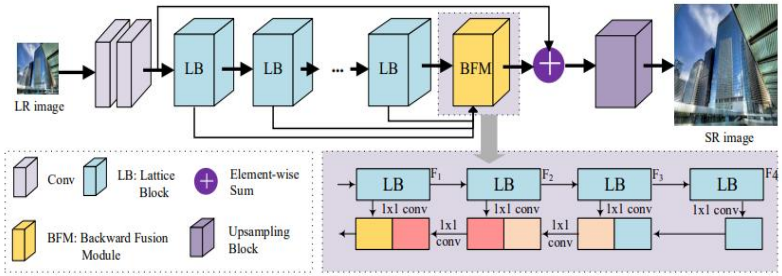
VDSR



RFDN



FSRCNN



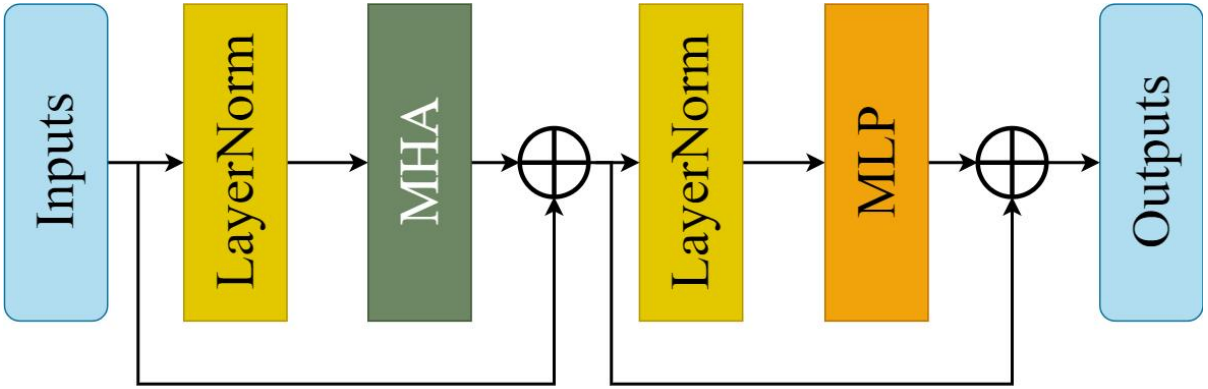
LatticeNet

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

Motivation

- Transformer Modules



$$F_{mid} = F_{in} + f_{MHA}(f_{norm}(F_{in}))$$

$$F_{out} = F_1 + f_{MLP}(f_{norm}(F_{mid}))$$

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V.$$

- Transformers can capture and model long-term image dependencies.

Motivation

Target

- We aim to to explore an efficient lightweight SISR model with low complexity, low model size and low execution time.

Contributions

- To better apply Transformer to lightweight SISR tasks, we propose a Recursive Transformer to learn the long-term dependence of images. Transformer will bring a lot of parameter consumption, and the recursive mechanism helps to use it to fully learn dependency information without increasing additional parameter consumption. This is the first attempt of the recursive mechanism in Transformer, which can refine the texture details by global information with few parameters and GPU memory consumption.
- To reduce the computational consumption caused by the repeated features extracted by CNN, the Feature Refinement Dual-Attention Block (FRDAB) are specially designed for feature extraction. Furthermore, a local feature fusion module (LFFM) is proposed for feature fusion.

Motivation

- Contributions

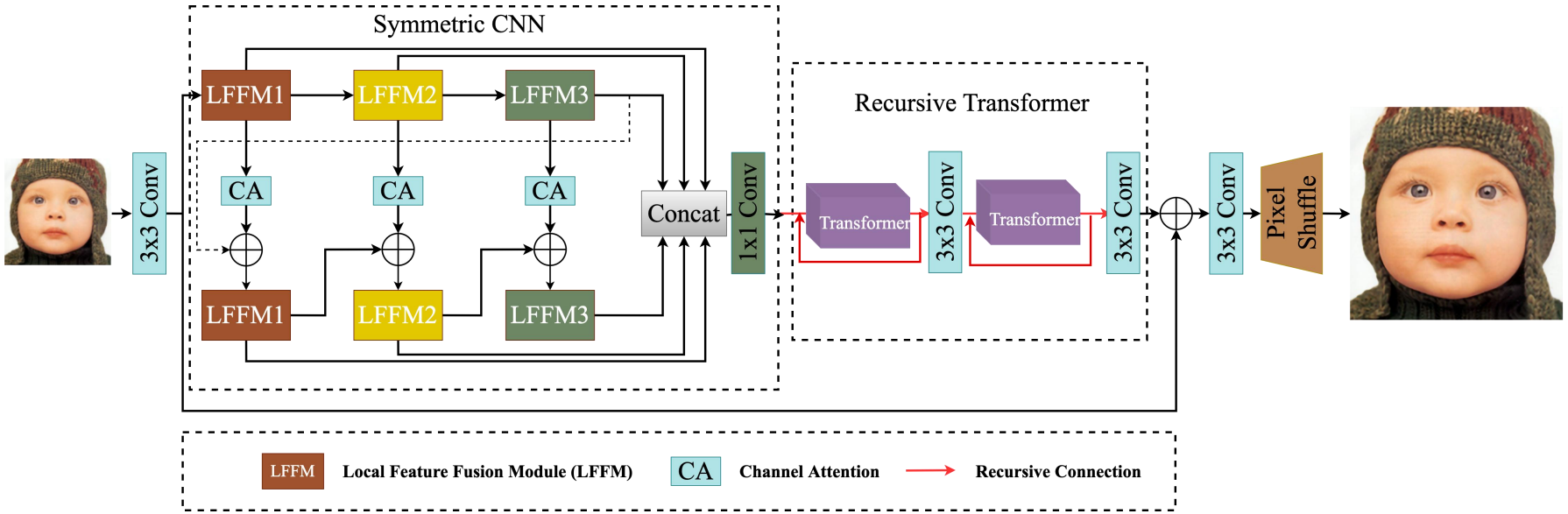
- Considering that as the depth of the CNN network increases, the complexity and parameters of the model will increase. The symmetric CNN network increases the feature extraction and representation capabilities of the network through the top CNN network and the bottom shared parameter branch network, and does not bring additional parameter consumption.
- To make full use of the local features extracted by the symmetric CNN network and the global dependency information learned by the recursive Transformer network, We propose a novel Lightweight Bimodal Network (LBNet) for SISR. LBNet elegantly integrates CNN and Transformer, enabling it to achieve a better balance between the performance, size and execution time of the model.

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

LBNet

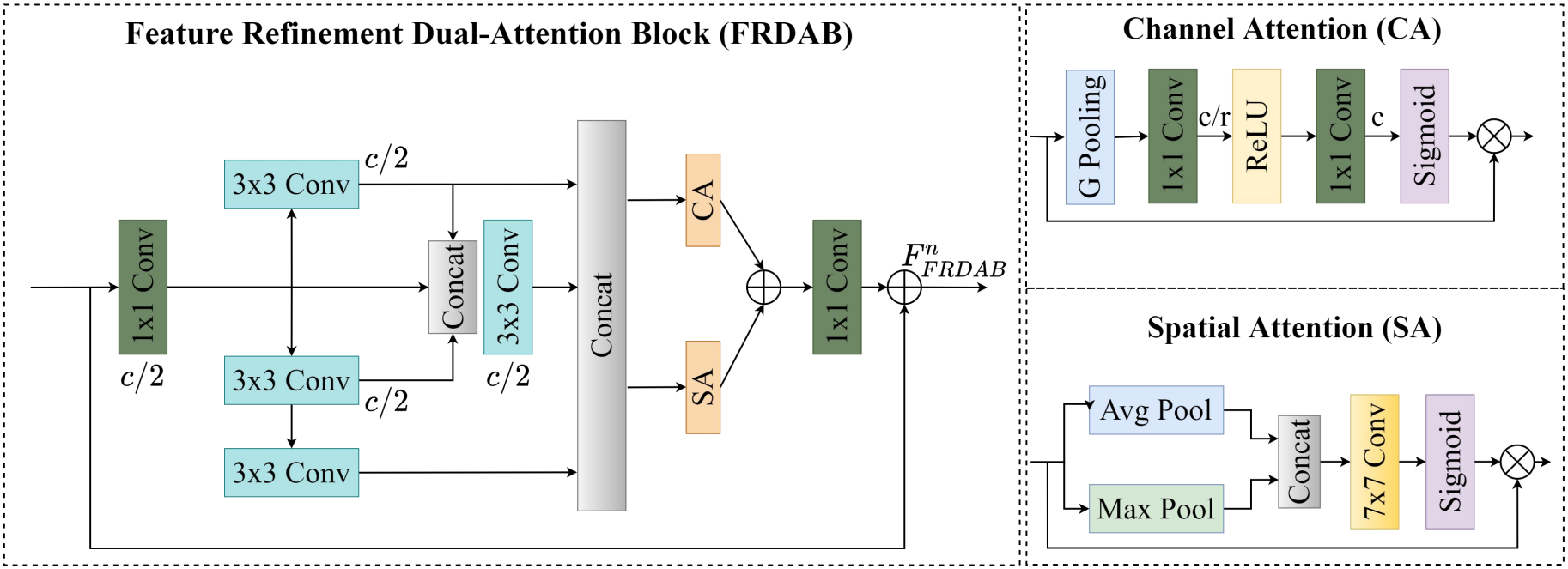
- Network Architecture



- Symmetric CNN is specially designed for local feature extraction, which mainly consists of some paired parameter sharing LFFM and CA modules.
- Recursive Transformer introduced the recursive mechanism to allow the Transformer to be fully trained without greatly increasing model parameters.

LBNet-FRDAB

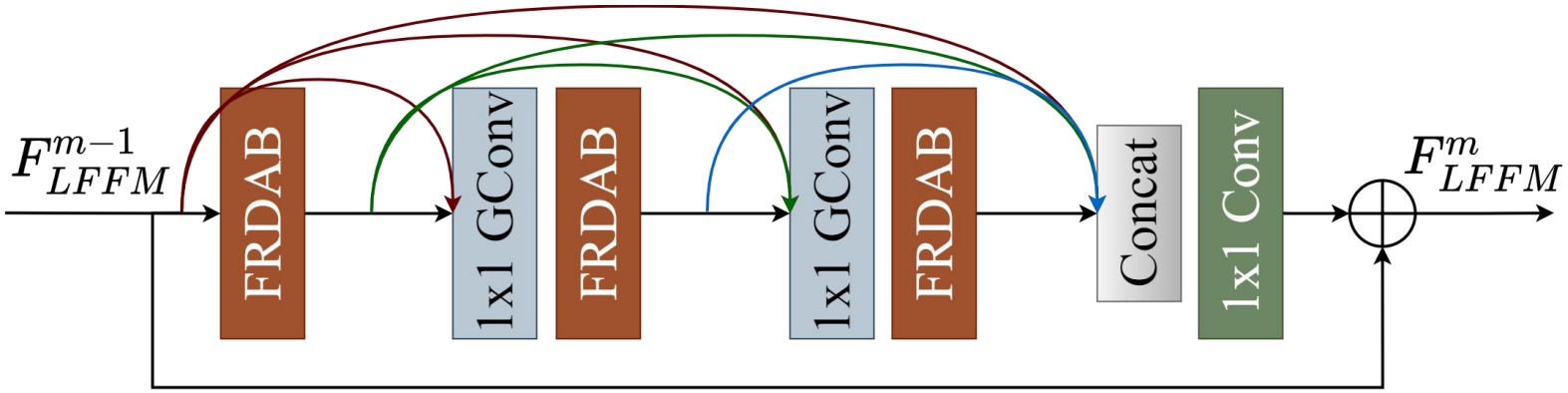
- Feature Refinement Dual-Attention Block



- FRDAB is a dual-attention block, which specially designed to reduce the computational cost of repetitive features extracted by the CNN network.
- The dual attention mechanism is used to channel and spatially reweight the extracted features.

LBNet-LFFM

- Local Feature Fusion Module



- Residual connections and dense connections are applied in LFFM to fully utilize the features extracted by FRDAB.
- Group convolutions are used for dimensionality reduction to reduce model parameters and computational cost.

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

Experiments & Discussion

- Quantitative comparisons

Method	Scale	Params	Mult-Adds	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic		-	-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
SRCNN [Dong *et al.*, 2014]		57K	52.7G	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.60/0.9663
FSRCNN [Dong *et al.*, 2016]		12K	6.0G	37.05/0.9560	32.66/0.9090	31.53/0.8920	29.88/0.9020	36.67/0.9710
VDSR [Kim *et al.*, 2016a]		665K	612.6G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
DRCN [Kim *et al.*, 2016b]		1774K	17974.3G	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
LapSRN [Lai *et al.*, 2018]		813K	29.9G	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
DRRN [Tai *et al.*, 2017]		297K	6796.9G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
IDN [Hui *et al.*, 2018]		553K	124.6G	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749
CARN [Ahn *et al.*, 2018]		1592K	222.8G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
CBPN [Zhu and Zhao, 2019]		1036K	240.7G	37.90/0.9590	33.60/0.9171	32.17/0.8989	32.14/0.9279	-
CBPN-S [Zhu and Zhao, 2019]	×2	430K	101.5G	37.69/0.9583	33.36/0.9147	32.02/0.8972	31.55/0.9217	-
IMDN [Hui *et al.*, 2019]		694K	158.8G	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
AWSRN-M [Wang *et al.*, 2019]		1063K	244.1G	38.04/0.9605	33.66/0.9181	32.21/ 0.9000	32.23/0.9294	38.66/0.9772
MADNet [Lan *et al.*, 2020]		878K	187.1G	37.85/0.9600	33.38/0.9161	32.04/0.8979	31.62/0.9233	-
GLADSR [Zhang *et al.*, 2020]		812K	187.2G	37.99/0.9608	33.63/0.9179	32.16/0.8996	32.16/0.9283	-
LAINet [Xiao *et al.*, 2021]		237K	-	37.94/0.9604	33.52/0.9174	32.12/0.8991	31.67/0.9242	-
DCDN [Li *et al.*, 2021b]		756K	-	38.01 /0.9606	33.52/0.9166	32.17/0.8996	32.16/0.9283	38.70/0.9773
SMSR [Wang *et al.*, 2021]		985K	351.5G	38.00/0.9601	33.64/0.9179	32.17/0.8990	32.19/0.9284	38.76/0.9771
LAPAR-A [Li *et al.*, 2021a]		548K	171.0G	38.01/0.9605	33.62/0.9183	32.19/0.8999	32.10/0.9283	38.67/0.9772
ECBSR [Zhang *et al.*, 2021]		596K	137.3G	37.90/0.9615	33.34/0.9178	32.10/0.9018	31.71/0.9250	-
LBNet-T (Ours)		404K	49.0G	37.95/0.9602	33.53/0.9168	32.07/0.8983	31.91/0.9253	38.59/0.9768
LBNet (Ours)		731K	153.2G	38.05/0.9607	33.65/0.9177	32.16/0.8994	32.30/0.9291	38.88/0.9775

Experiments & Discussion

- Visual results of LBNet with other SR methods (x2)

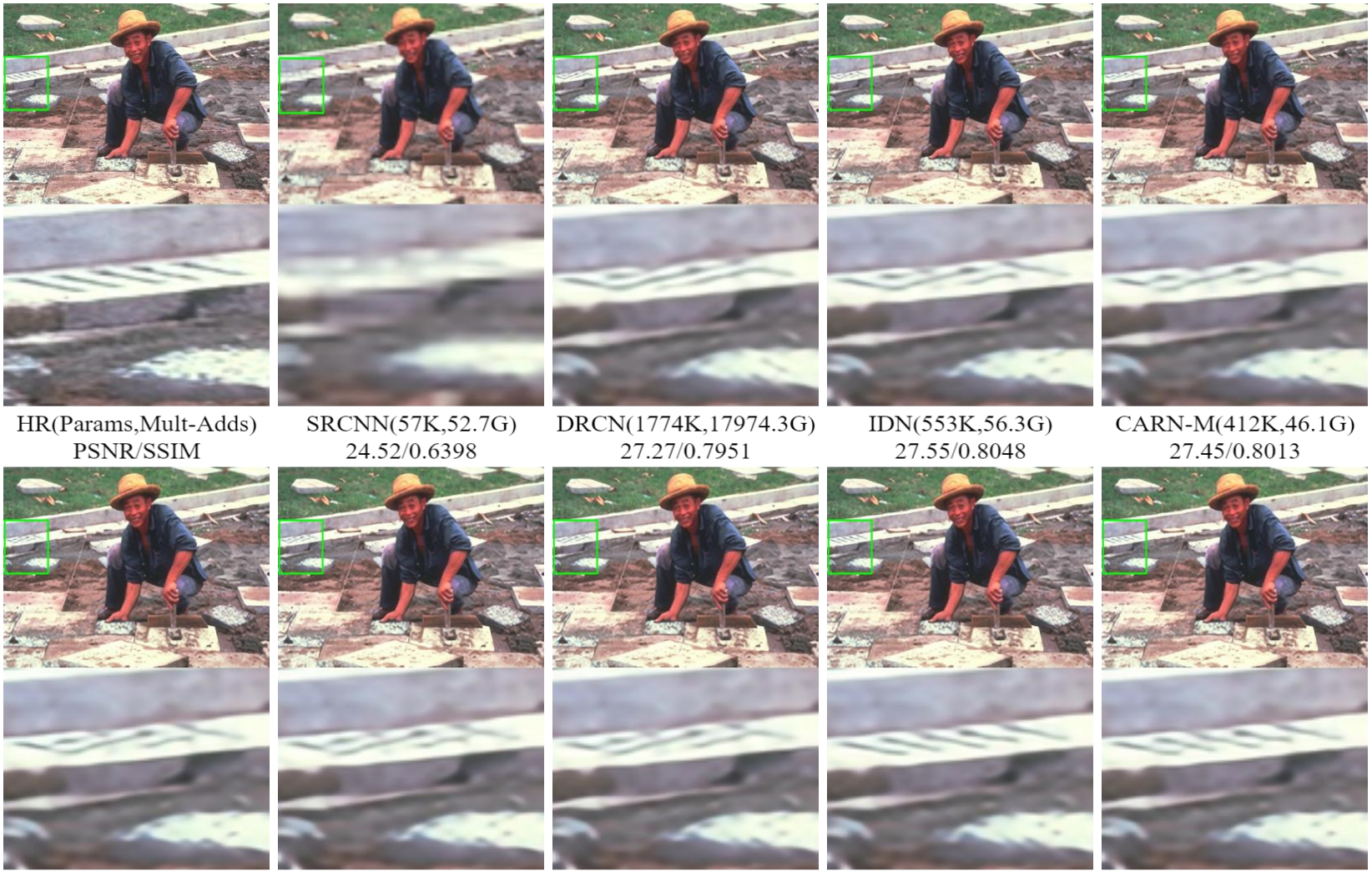
Experiments & Discussion

- Quantitative comparisons

Method	Scale	Params	Mult-Adds	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic		-	-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
SRCNN [Dong *et al.*, 2014]		57K	52.7G	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
FSRCNN [Dong *et al.*, 2016]		12K	5.0G	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	31.10/0.9210
VDSR [Kim *et al.*, 2016a]		665K	612.6G	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
DRCN [Kim *et al.*, 2016b]		1774K	17974.3G	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328
DRRN [Tai *et al.*, 2017]		297K	6796.9G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.74/0.9390
IDN [Hui *et al.*, 2018]		553K	56.3G	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.71/0.9381
CARN [Ahn *et al.*, 2018]		1592K	118.8G	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.43/0.9427
IMDN [Hui *et al.*, 2019]	×3	703K	71.5G	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
AWSRN-M [Wang *et al.*, 2019]		1143K	116.6G	34.42/0.9275	30.32/0.8419	29.13/0.8059	28.26/0.8545	33.64/0.9450
MADNet [Lan *et al.*, 2020]		930K	88.4G	34.16/0.9253	30.21/0.8398	28.98/0.8023	27.77/0.8439	-
GLADSR [Zhang *et al.*, 2020]		821K	88.2G	34.41/0.9272	30.37/0.8418	29.08/0.8050	28.24/0.8537	-
LAINet [Xiao *et al.*, 2021]		237K	-	34.26/0.9261	30.24/0.8404	29.04/0.8039	27.83/0.8453	-
DCDN [Li *et al.*, 2021b]		765K	-	34.41/0.9273	30.31/0.8417	29.08/0.8045	28.17/0.8520	33.54/0.9441
SMSR [Wang *et al.*, 2021]		993K	156.8G	34.40/0.9270	30.33/0.8412	29.10/0.8050	28.25/0.8536	33.68/0.9445
LAPAR-A [Li *et al.*, 2021a]		594K	114.0G	34.36/0.9267	30.34/0.8421	29.11/0.8054	28.15/0.8523	33.51/0.9441
EMASRN [Zhu *et al.*, 2021]		427K	-	34.36/0.9264	30.30/0.8411	29.05/0.8035	28.04/0.8493	33.43/0.9433
LBNet-T (Ours)		407K	22.0G	34.33/0.9264	30.25/0.8402	29.05/0.8042	28.06/0.8485	33.48/0.9433
LBNet (Ours)		736K	68.4G	34.47/0.9277	30.38/0.8417	29.13/0.8061	28.42/0.8559	33.82/0.9460

Experiments & Discussion

- Visual results of FDIWN with other SR methods (x3)



HR(Params,Mult-Adds) PSNR/SSIM	SRCNN(57K,52.7G) 24.52/0.6398	DRCN(1774K,17974.3G) 27.27/0.7951	IDN(553K,56.3G) 27.55/0.8048	CARN-M(412K,46.1G) 27.45/0.8013
CARN(1592K,118.8G) 27.63/0.8071	IMDN(703K,71.5G) 27.68/0.8075	MADNet(930K,88.4G) 27.57/0.8039	LBNet-T(407K,22.0G) 27.79/0.8105	LBNet(736K,68.4G) **27.83/0.8128**

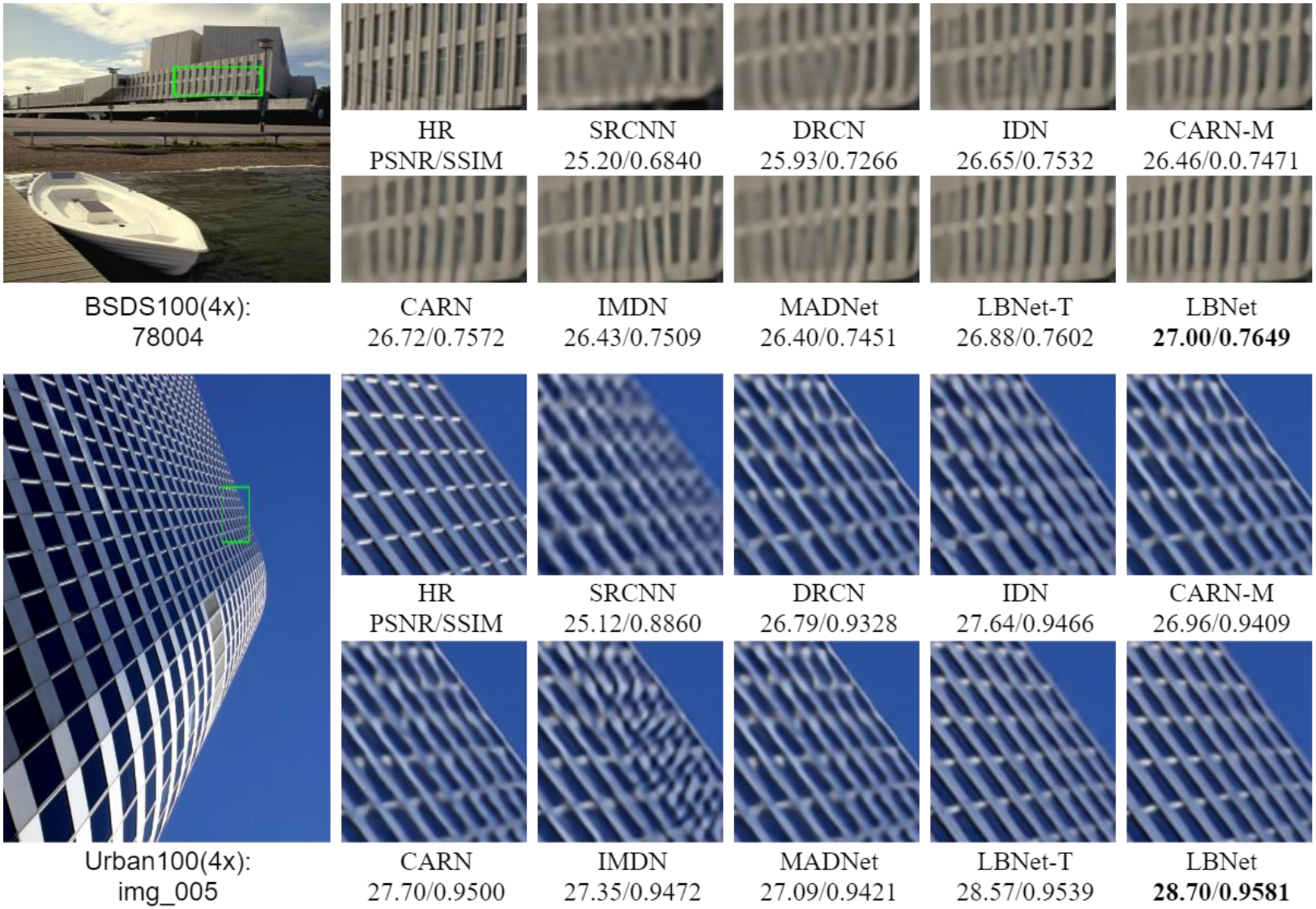
Experiments & Discussion

- Quantitative comparisons

Method	Scale	Params	Mult-Adds	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×4	-	-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
SRCNN [Dong *et al.*, 2014]		57K	52.7G	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
FSRCNN [Dong *et al.*, 2016]		12K	4.6G	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8610
VDSR [Kim *et al.*, 2016a]		665K	612.6G	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
DRCN [Kim *et al.*, 2016b]		1774K	17974.3G	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
LapSRN [Lai *et al.*, 2018]		813K	149.4G	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8900
DRRN [Tai *et al.*, 2017]		297K	6796.9G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8960
IDN [Hui *et al.*, 2018]		553K	32.3G	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8942
CARN [Ahn *et al.*, 2018]		1592K	90.9G	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.42/0.9070
CBPN [Zhu and Zhao, 2019]		1197K	97.9G	32.21/0.8944	28.63/0.7813	27.58/0.7356	26.14/0.7869	-
CBPN-S [Zhu and Zhao, 2019]		592K	63.1G	31.93/0.8908	28.50/0.7785	27.50/0.7324	25.85/0.7772	-
IMDN [Hui *et al.*, 2019]		715K	40.9G	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
AWSRN-M [Wang *et al.*, 2019]		1254K	72.0G	32.21/0.8954	28.65/0.7832	27.60/0.7368	26.15/0.7884	30.56/0.9093
MADNet [Lan *et al.*, 2020]		1002K	54.1G	31.95/0.8917	28.44/0.7780	27.47/0.7327	25.76/0.7746	-
GLADSR [Zhang *et al.*, 2020]		826K	52.6G	32.14/0.8940	28.62/0.7813	27.59/0.7361	26.12/0.7851	-
LAINet [Xiao *et al.*, 2021]		263K	-	32.12/0.8942	28.59/0.7810	27.55/0.7351	25.92/0.7805	-
DCDN [Li *et al.*, 2021b]		777K	-	32.21/0.8949	28.57/0.7807	27.55/0.7356	26.09/0.7855	30.41/0.9072
SMSR [Wang *et al.*, 2021]		1006K	89.1G	32.12/0.8932	28.55/0.7808	27.55/0.7351	26.11/0.7868	30.54/0.9085
LAPAR-A [Li *et al.*, 2021a]		659K	94.0G	32.15/0.8944	28.61/0.7818	27.61/0.7366	26.14/0.7871	30.42/0.9074
ECBSR [Zhang *et al.*, 2021]		603K	34.7G	31.92/0.8946	28.34/0.7817	27.48/0.7393	25.81/0.7773	-
EMASRN [Zhu *et al.*, 2021]		546K	-	32.17/0.8948	28.57/0.7809	27.55/0.7351	26.01/0.7838	30.41/0.9076
LBNet-T (Ours)		410K	12.6G	32.08/0.8933	28.54/0.7802	27.54/0.7358	26.00/0.7819	30.37/0.9059
LBNet (Ours)		742K	38.9G	32.29/0.8960	28.68/0.7832	27.62/0.7382	26.27/0.7906	30.76/0.9111

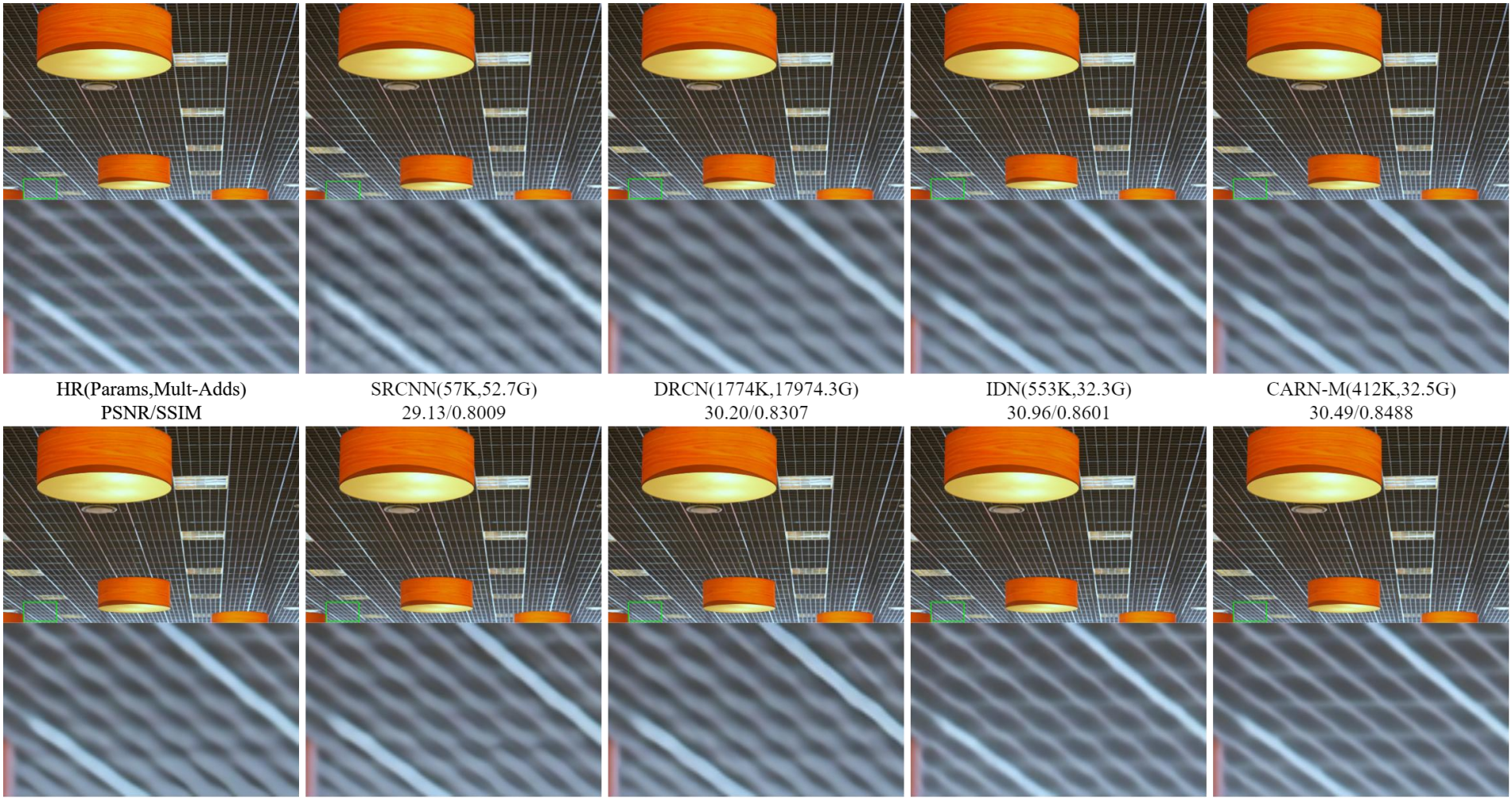
Experiments & Discussion

- Visual results of LBNet with other SR methods (x4)



Experiments & Discussion

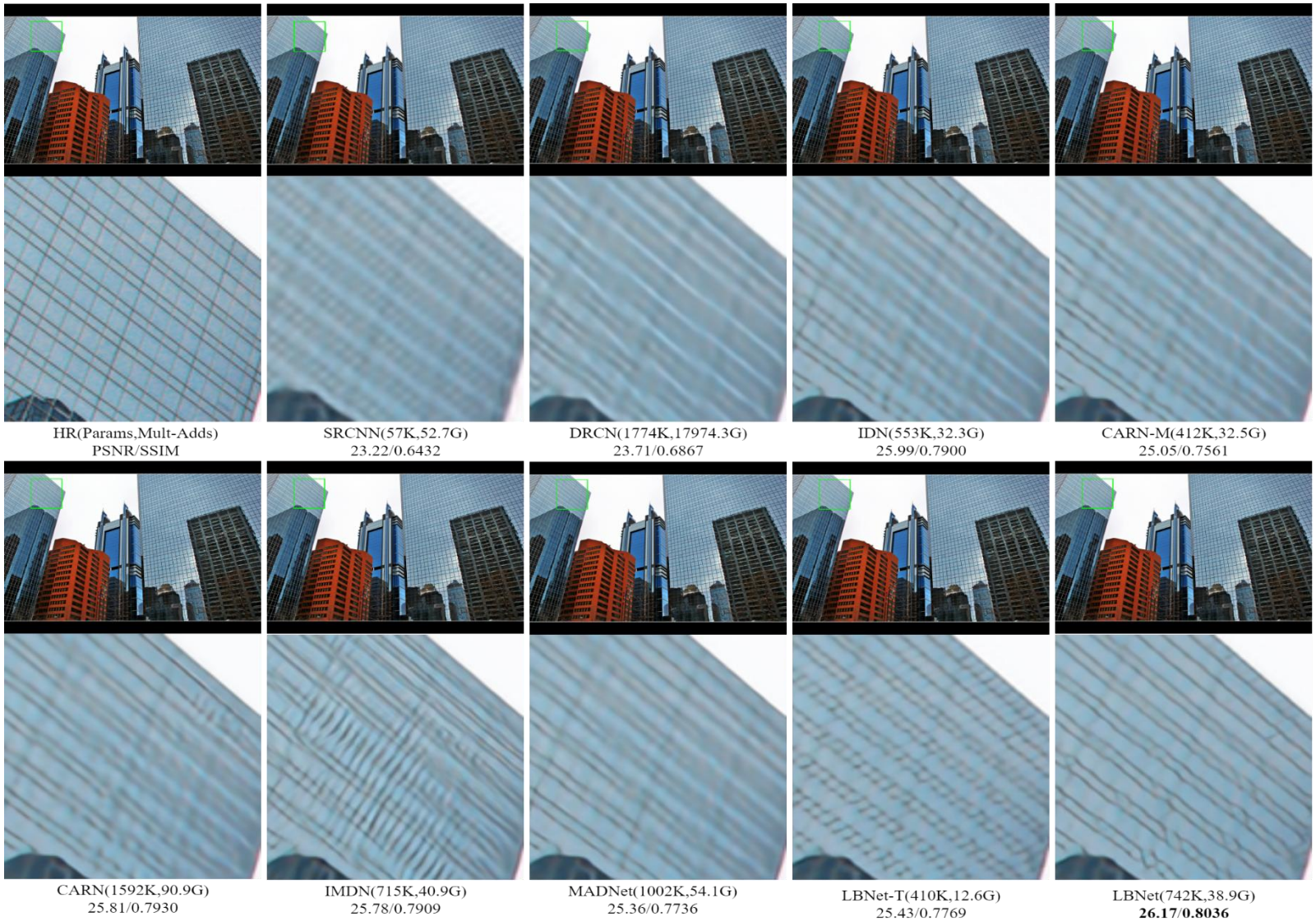
- Visual results of LBNet with other SR methods (x4)



HR(Params,Mult-Adds) PSNR/SSIM	SRCNN(57K,52.7G) 29.13/0.8009	DRCN(1774K,17974.3G) 30.20/0.8307	IDN(553K,32.3G) 30.96/0.8601	CARN-M(412K,32.5G) 30.49/0.8488
CARN(1592K,90.9G) 31.34/0.8648	IMDN(715K,40.9G) 31.50/0.8637	MADNet(1002K,54.1G) 30.71/0.8529	LBNet-T(410K,12.6G) 32.02/0.8719	LBNet(742K,38.9G) **32.60/0.8799**

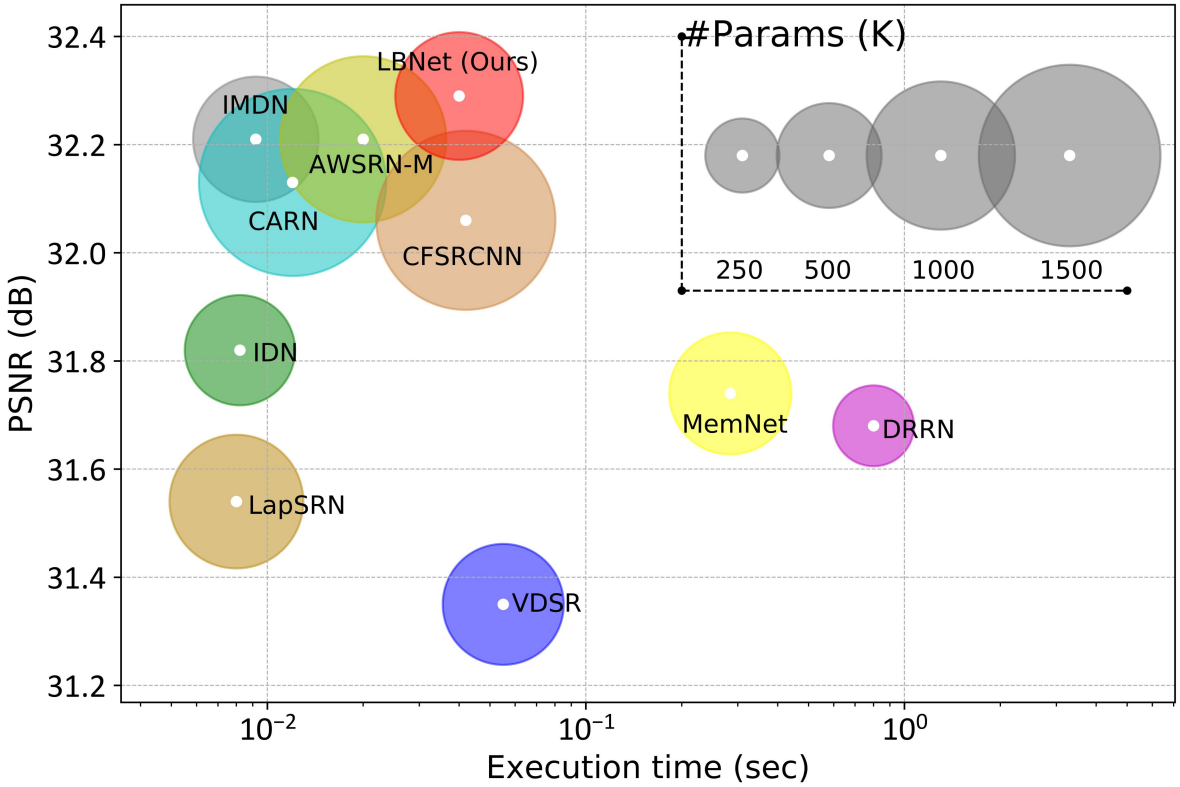
Experiments & Discussion

- Visual results of FDIWN with other SR methods (x4)



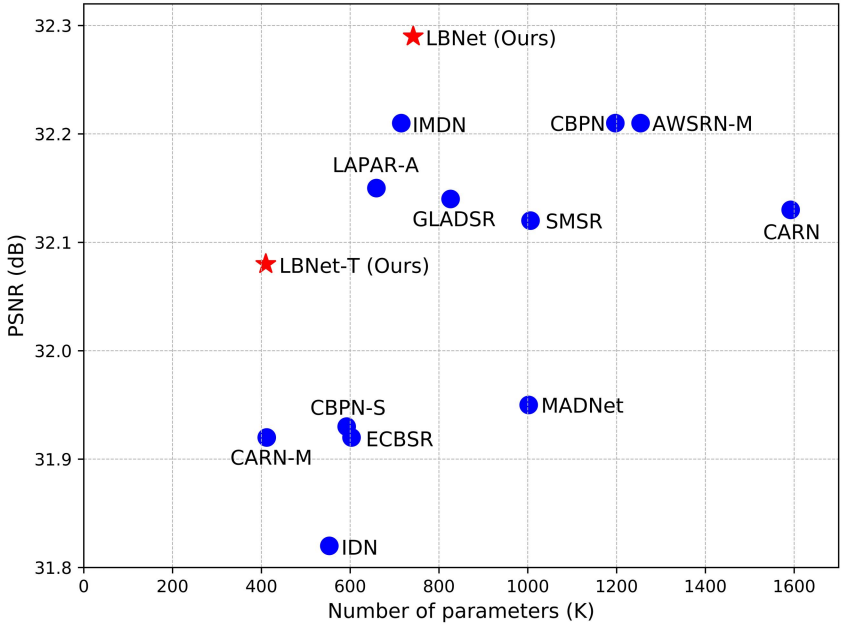
Experiments & Discussion

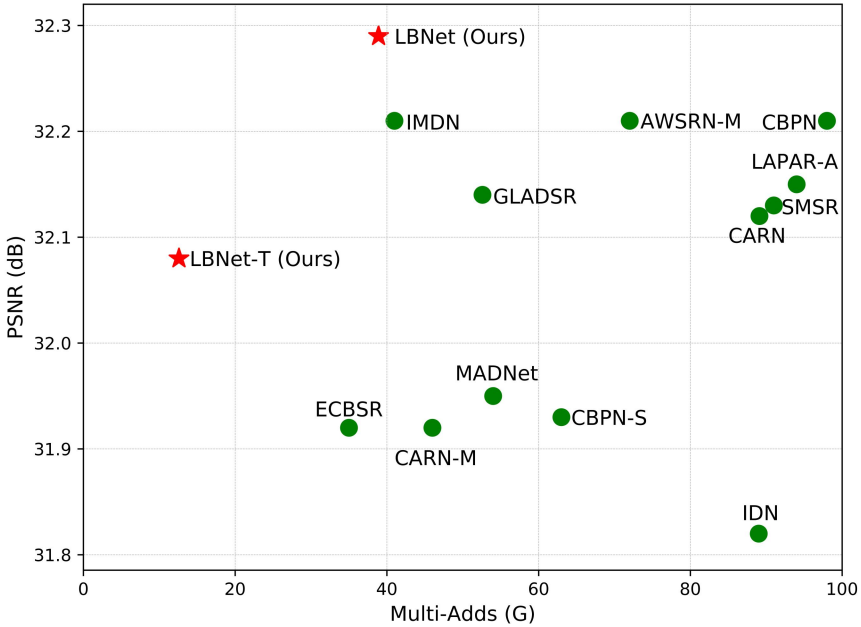
- Model complexity analysis



Experiments & Discussion

- Investigations of the model size and performance





Experiments & Discussion

- Symmetric CNN Investigations

Scale	FF	SA	CA	Params	Mult-Adds	PSNR/SSIM
×4	✓	✗	✗	96.1K	10.01G	25.28/0.7601
×4	✗	✓	✗	96.3K	10.03G	25.31/0.7614
×4	✗	✗	✓	96.5K	10.01G	**25.36/0.7622**

- FRDAB Investigations

Method	Params	Mult-Adds	PSNR/SSIM
LBNet+RCAB	228K	23.7G	29.94/0.9002
LBNet+IMDB	295K	31.3G	30.21/0.9043
LBNet+FRDAB (Ours)	365K	38.9G	**30.33/0.9059**

Scale	SA	CA	Params	Mult-Adds	PSNR/SSIM
×4	✗	✗	358.7K	38.80120G	30.18/0.9039
×4	✓	✗	359.6K	38.90281G	30.06/0.9025
×4	✗	✓	363.9k	38.80121G	30.30/0.9052
×4	✓	✓	364.8K	38.90282G	**30.33/0.9059**

Experiments & Discussion

- Recursive Transformer Investigations

Method	Params	Mult-Adds	Running time	PSNR/SSIM
w/o RT	365K	38.9028G	0.0168s	32.07/0.8929
with RT	742K	38.9032G	0.0274s	**32.23/0.8949**

Method	Params	Mult-Adds	Running time	PSNR/SSIM
TM-0	741.7K	38.9032G	0.0274s	32.23/0.8949
TM-1	741.7K	38.9036G	0.0356s	32.27/0.8958
TM-2	**741.7K**	**38.9039G**	**0.0401s**	**32.29/0.8960**
TM-3	741.7K	38.9043G	0.0516s	32.30/0.8960

Method	Params	Multi-Adds	Set5	Set14	BSD100	Urban100	Manga109	Average
SwinIR	897K	49.6G	32.44/0.8976	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151	29.26/0.8274
ESRT	751K	67.7G	32.19/0.8947	28.69/0.7833	27.69/0.7379	26.39/0.7962	30.75/0.9100	29.14/0.8244
LBNet (Ours)	**742K**	**38.9G**	32.29/0.8960	28.68/0.7832	27.62/0.7382	26.27/0.7906	30.76/0.9111	29.12/0.8238

Outline

- Background & Related Works
- Motivation
- Lightweight Bimodal Network (LBNet)
- Experiments & Discussion
- Summary

Summary

- The Local Feature Fusion Module (LFFM) and Feature Refinement Dual-Attention Block (FRDAB) are specially designed for feature extraction and utilization.

- The well designed Recursive Transformer can learn the long-term dependence of images. It is the first attempt of the recursive mechanism in Transformer, which can refine the texture details by global information with few parameters and GPU memory consumption.

- The experiment show that our proposed LBNet achieved a good balance between model size, performance, and computational cost.

Thanks !