IEEE

ICIP abu dhabi 2024

IEEE International Conference on Image Processing

27 - 30 October 2024, Abu Dhabi, UAE

IEEE Signal Processing Society — CELEBRATING 75 YEARS

Lightweight Recurrent Neural Network for Image Super-Resolution

CCDS.AI

Mir Sazzat Hossain   A. K. M. Mahbubur Rahman   Md Ashraful Amin   Amin Ahsan Ali

Center for Computational and Data Sciences (CCDS), Independent University, Bangladesh (IUB)

IUB

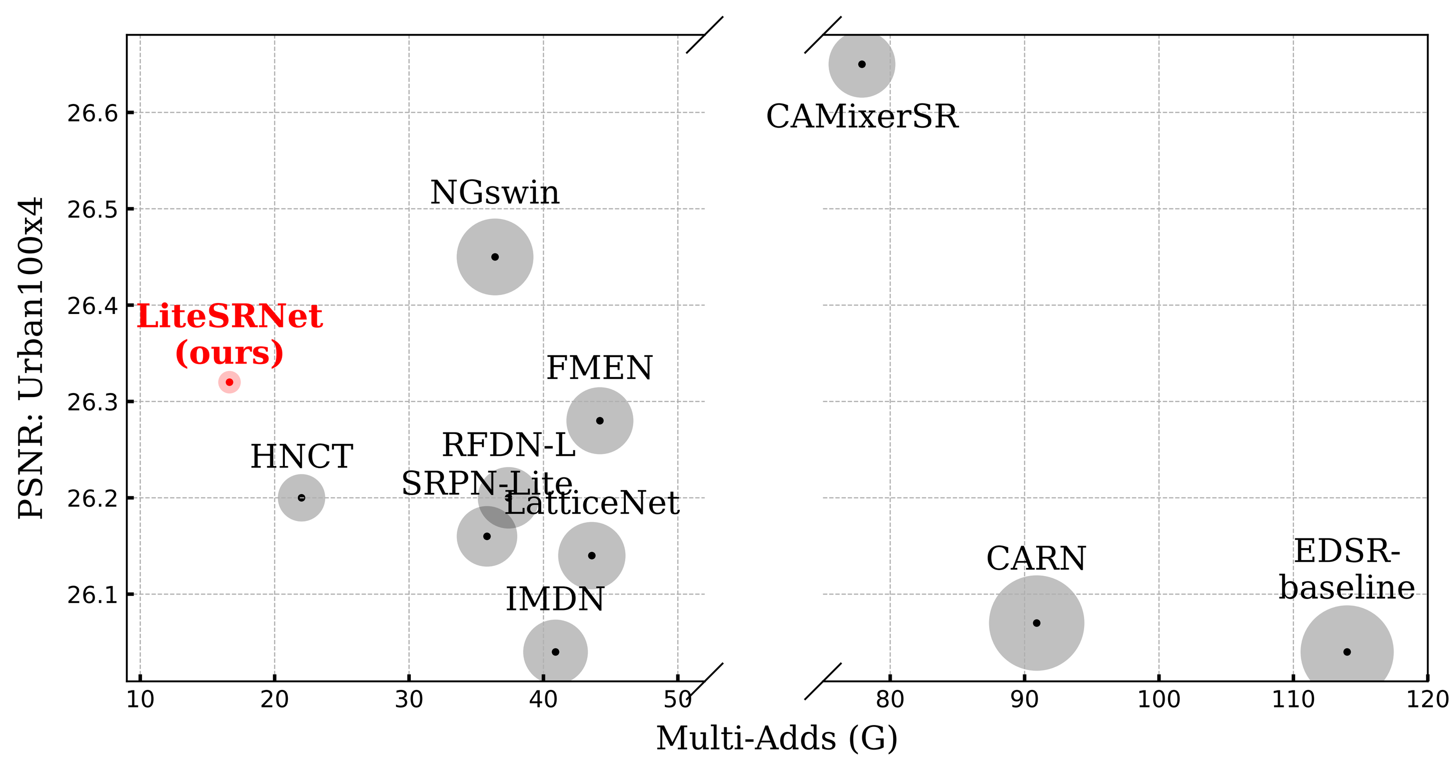

Figure 1. Comparing the trade-off between image quality and computational complexity for LiteSRNet with other SOTA models. The size of each circle is proportional to the num- ber of parameters in the model.HR Bicubic EDSR

Overview - Problem and Key Contributions

Problem Overview:

- Large-scale **Super-Resolution (SR)** models are computationally expensive.
- Hard to deploy on **resource-limited devices**.
- **Challenge:** How to achieve efficient super-resolution with fewer parameters?

Contributions:

- Developed a **lightweight RNN (LiteSRNet)** with less than 75k parameters.
- Achieved **comparable performance** to SOTA models with **10x fewer parameters**.
- **Computational efficiency:** Only 16.64 GFLOPs vs. SOTA models 53.8 GFLOPs.
- High **PSNR** and **SSIM** on Set5, Set14, BSD100, Urban100.

Methodology - LiteSRNet Architecture

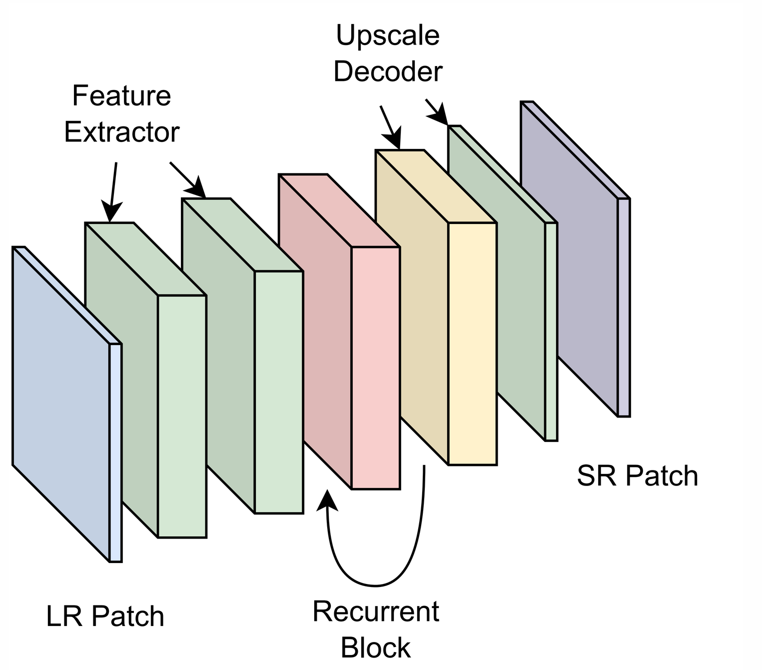

Figure 2. Recurrent Neural Network (RNN) based Single Image Super-resolution (SISR) model: LiteSRNet.

We extract features from 48×48 patches $P_{LR} \in \mathbb{R}^{(3\times H\times W)}$ cropped from original images, producing a 64-channel feature map for the recurrent block.

Algorithm 1 RNN-based SISR model (LiteSRNet)

Require: $F_2 \in \mathbb{R}^{(64\times H\times W)}, N, S_f$

Ensure: $F_{rec} \in \mathbb{R}^{(64\times H\times W)}$

$F \leftarrow \emptyset$
$F_{rec} \leftarrow F_2$
for $i = 1$ to N **do**
    $F_{rec} \leftarrow Conv2D(F_{rec})$
    $F_{rec} \leftarrow ReLU(F_{rec})$
    for $j = 1$ to $\lfloor \frac{S_f}{2} \rfloor$ **do**
        $F_{dec} \leftarrow TransposeConv2D(F_{rec})$
        $F_{dec} \leftarrow ReLU(F_{dec})$
    end for
    $F_{dec} \leftarrow Conv2D(F_{dec})$
    $F \leftarrow F \cup F_{dec}$
end for
return $F[-1]$

Loss Function: The final loss combines two terms:

$$L = \text{MSE}(P_{SR}, P_{HR}) + \alpha \cdot \text{Perceptual}(P_{SR}, P_{HR})$$

- **MSE Loss**: Minimizes pixel-level differences between **super-resolved** and **high-resolution** images.
- **Perceptual Loss**: Ensures high-level feature matching using pre-trained VGG16.

Results - Performance Comparison

Method	Training Dataset	Scale	No. of Params	Set5		Set14		BSD100		Urban100	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR-baseline	DIV2K	×2	1,370k	37.99	0.9604	33.57	0.9175	32.16	0.8994	31.98	0.9272
RFDN-L	DIV2K		626k	38.08	0.9606	33.67	0.9190	32. 18	0.8996	32.24	0.9290
SRPN-Lite	DIV2K		609k	38.10	0.9608	33.70	0.9189	32.25	0.9005	32.26	0.9294
HNCT	DIV2K		357k	38.08	0.9608	33.65	0.9182	32.22	0.9001	32.22	0.9294
FMEN	DIV2K+F2K		748k	38.10	0.9609	33.75	0.9192	32.26	0.9007	32.41	0.9311
NGswin	DIV2K		998k	38.05	0.9610	33.79	0.9199	32.27	0.9008	32.53	0.9324
CAMixerSR	DIV2K+F2K		746K	38.28	0.9614	34.04	0.9218	32.37	0.9021	33.04	0.9364
LiteSRNet (Ours)	**DIV2K**		**67k**	**38.04**	**0.9605**	**33.70**	**0.9185**	**32.24**	**0.8996**	**32.40**	**0.9294**
EDSR-baseline	DIV2K	×3	1,555K	34.37	0.9270	30.28	0.8417	29.09	0.8052	28.15	0.8527
RFDN-L	DIV2K		633K	34.47	0.9280	30.35	0.8421	29.11	0.8053	28.32	0.8547
SRPN-Lite	DIV2K		615K	34.47	0.9276	30.38	0.8425	29.16	0.8061	28.22	0.8534
HNCT	DIV2K		363K	34.47	0.9275	30.44	0.8439	29.15	0.8067	28.28	0.8557
FMEN	DIV2K+F2K		757K	34.45	0.9275	30.40	0.8435	29.17	0.8063	28.33	0.8562
NGswin	DIV2K		1,007K	34.52	0.9282	30.53	0.8456	29.19	0.8078	28.52	0.8603
CAMixerSR	DIV2K+F2K		-	-	-	-	-	-	-	-	-
LiteSRNet (Ours)	**DIV2K**		**68k**	**34.44**	**0.9278**	**30.36**	**0.8421**	**29.11**	**0.8052**	**28.30**	**0.8545**
EDSR-baseline	DIV2K	×4	1,518K	32.09	0.8938	28.58	0.7813	27.57	0.7357	26.04	0.7849
RFDN-L	DIV2K		643K	32.28	0.8957	28.61	0.7818	27.58	0.7363	26.20	0.7883
SRPN-Lite	DIV2K		623K	32.24	0.8958	28.69	0.7836	27.63	0.7373	26.16	0.7875
HNCT	DIV2K		373K	32.31	0.8957	28.71	0.7834	27.63	0.7381	26.20	0.7896
FMEN	DIV2K+F2K		769K	32.24	0.8955	28.70	0.7839	27.63	0.7379	26.28	0.7908
NGswin	DIV2K		1,019K	32.33	0.8963	28.78	0.7859	27.66	0.7396	26.45	0.7963
CAMixerSR	DIV2K+F2K		765K	32.60	0.9003	28.91	0.7889	27.78	0.7434	26.80	0.8068
LiteSRNet (Ours)	**DIV2K**		**75k**	**32.20**	**0.8943**	**28.70**	**0.7836**	**27.63**	**0.7375**	**26.32**	**0.7885**

Table 1. Quantitative comparison of our proposed model with other SOTA models. DIV2K+F2k is the combination of DIV2K and Flickr2K [1]. DIV2K+291 is the combination of DIV2K and 291 [21, 22] images. The best results are highlighted in blue and the second-best results are highlighted in red.

Figure 3. t-SNE visualizations of representations from the fine-tuned encoders of (a) BYOL, (b) SimCLR, and (c) Supervised G-CNN. Blue points denote FRI class, and orange for FRII. Improved clustering in our models, indicated by Silhouette and Davies Bouldin scores.

Ablation Study - Effect of Recurrent Block Depth

Scale Factor	Model Depth	PSNR	SSIM
×2	13	37.11	0.9565
	16	37.88	0.9598
	19	38.04	0.9605
×3	13	33.90	0.9232
	16	34.08	0.9247
	19	34.44	0.9278
×4	13	30.00	0.8565
	16	31.34	0.8785
	19	32.20	0.8943

Table 2. Comparing image quality metrics for LiteSRNet with varied depths, evaluated on the Set5 dataset.

Scale Factor	Model Depth	Multi-Adds (G)	Memory Footprint (M)	Inference Time (s)
×2	13	20.02	2.10	0.31
	16	29.29	2.10	0.46
	19	38.57	2.10	0.59
×3	13	11.49	1.05	0.14
	16	16.90	1.05	0.22
	19	22.32	1.05	0.28
×4	13	7.76	0.71	0.15
	16	12.20	0.71	0.24
	19	16.64	0.71	0.33

Table 3. Comparing computational complexity and inference time for LiteSRNet with varied depths.

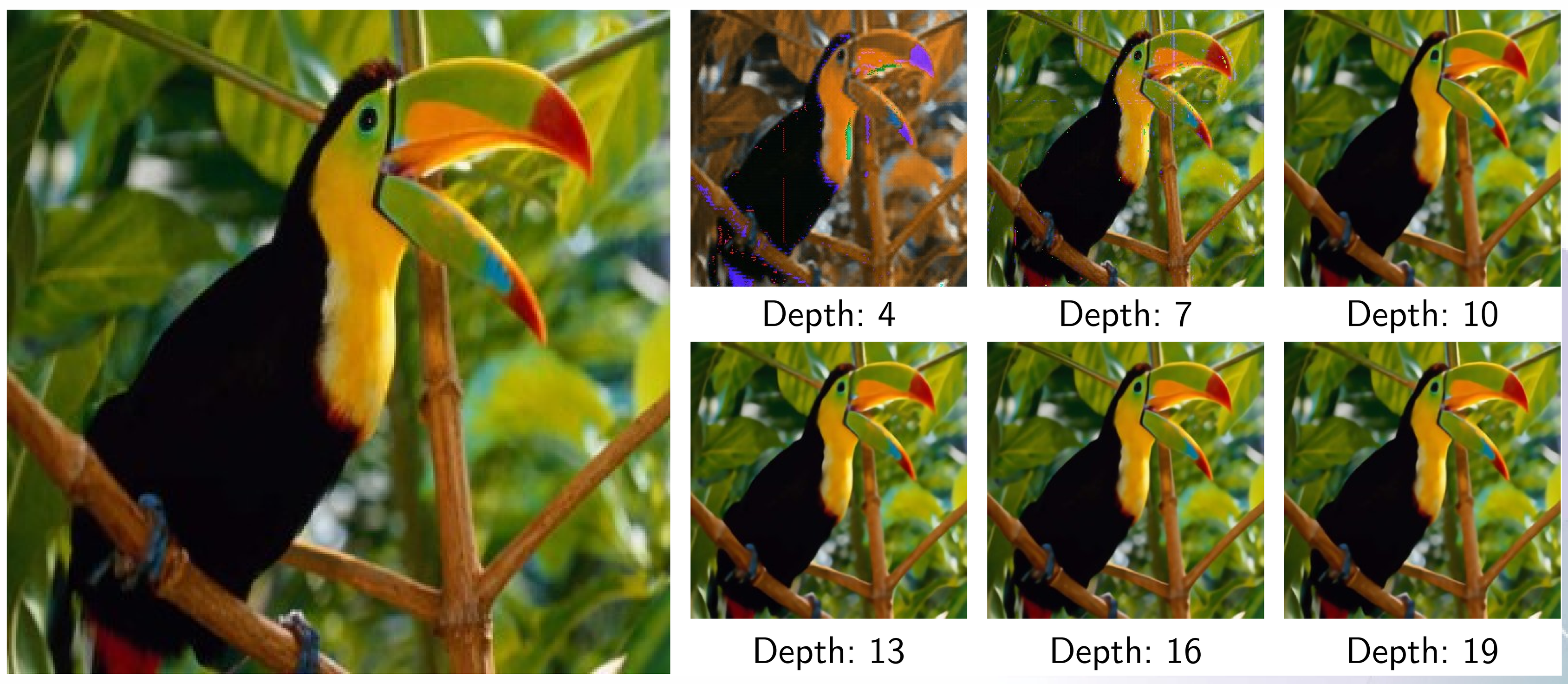

Figure 4. Visual comparison of outputs by our model and other SOTA models at ×4 upscaling. Our model consistently generates visually appealing images, comparable to others.

Acknowledgment

This project is supported by a grant from the Sponsored Research Unit of Independent University, Bangladesh (IUB).