Scale-Recurrent Multi-Residual Dense Network for Image Super-resolution

Kuldeep Purohit, Srimanta Mandal, and A. N. Rajagopalan (Team: REC-SR)

IPCV Lab, Department of Electrical Engg., Indian Institute of Technology Madras, India

kuldeeppurohit3@gmail.com, in.srimanta.mandal@ieee.org , raju@ee.iitm.ac.in

ECCV 2018 European Conference on Computer Vision

8 – 14 September 2018 I Munich, Germany

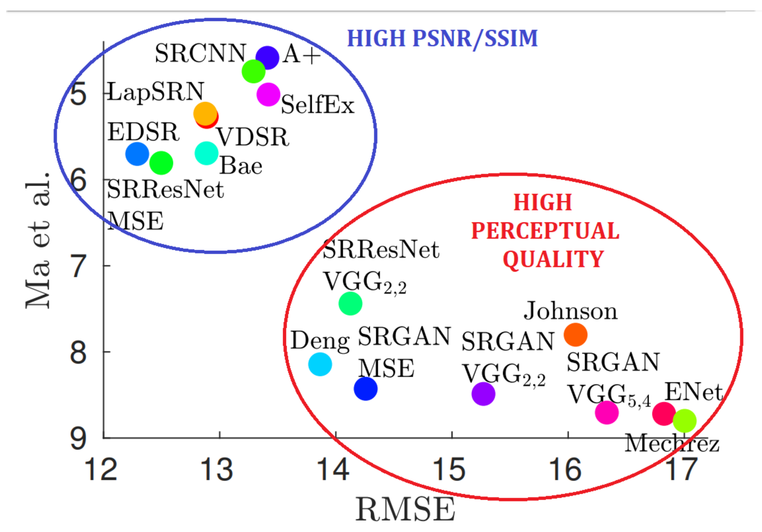
PIRM2018

1 Objective

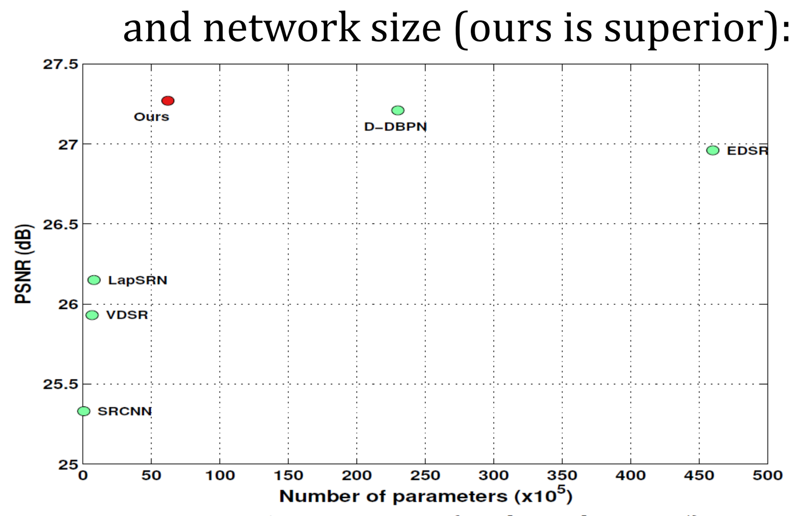
- **Single-image super-resolution (SISR):** inferring a high-resolution (HR) image from a single low-resolution input.
- **Convolutional neural networks (CNNs)** have achieved state-of-the-art SR results.
- CNN based methods struggle with the trade-off between the number of parameters and SR performance, specially for higher scale factors like 4 and 8.
- **PSNR/SSIM** have recently been shown [1] to correlate poorly with the perceptual quality of images.
- Main objective is to produce perceptually better results with a parametrically efficient network.

2 Related Works

Two streams of approaches:



Existing methods provide a poor trade-off between SR performance and network size (ours is superior):

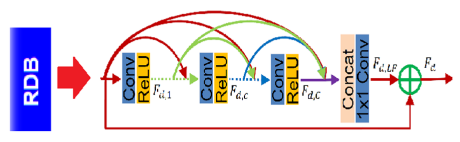


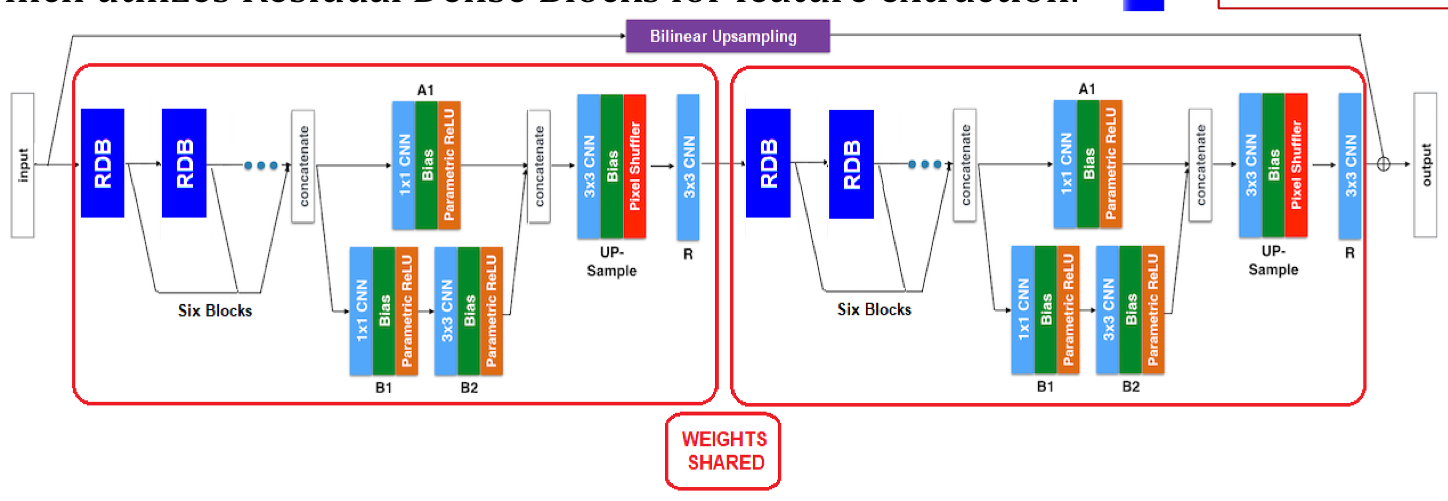
Approaches for 8× SR (Evaluated on Set5)

3 Our Scale-Recurrent Architecture

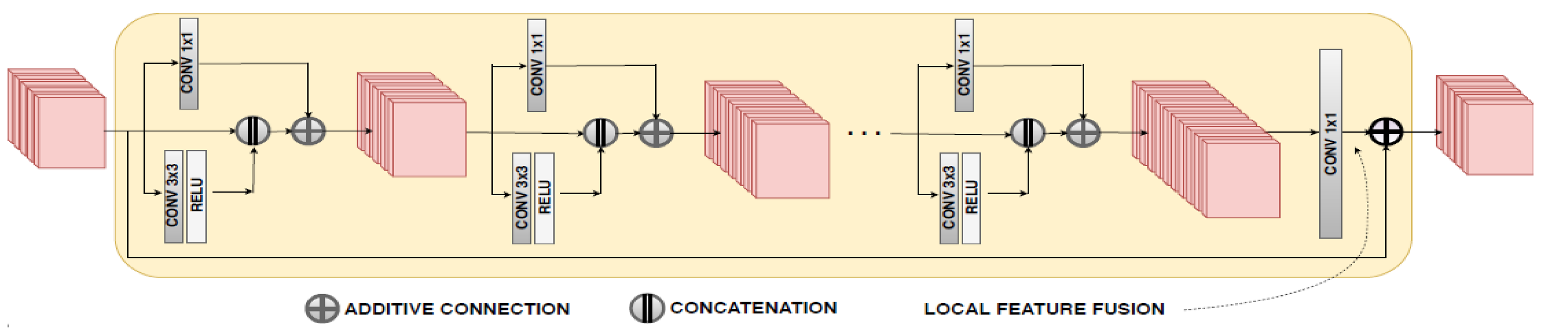
- **Super-Resolution (SR) via Scale –Recurrent Network:**

SR-RDN fully convolutional scale-recurrent neural network, which utilizes Residual Dense Blocks for feature extraction.





Multi-Residual Dense Block



MRDB contains a mix of residual and dense connection at each layer to promote better gradient flow and deeper feature extraction with fewer parameters.

4 Perception Distortion Trade-off

- Conventional loss functions such as MSE does not yield perceptually better results.
- Feature-space losses (Dosovitskiy and Brox, Johnson et al.) or adversarial losses (Goodfellow et al.) can be used to mitigate this issue.
- To improve results perceptually, we fine-tune our network with deep feature loss function alongside adversarial loss (conditional GAN). Different combinations of weights on these losses produce results with different qualities.
- **Loss configuration**

  Perceptual loss in feature space. Rather than computing errors in image space, they are computed in VGG's feature space. This loss is less sensitive to small local variations in favor of capturing higher-level statistics.

  Adversarial training GANs have proven to be a powerful mechanism to produce realistically looking images. In our setting, our super-resolution network acts as the generator network, and we use a discriminator with structure similar to [4], which takes up-sampled LR image also as an input.

$$Loss = \lambda_1 * L1\ loss + \lambda_2 * VGG\ loss + \lambda_3 * GAN\ loss$$

For region1: SRRDN: $\lambda_1 = 0.1, \lambda_2 = 5, \lambda_3 = 0$

For region3: SRRDN-GAN: $\lambda_1 = 0, \lambda_2 = 5, \lambda_3 = 0.15$

For region2: Soft-thresholding operation between the outputs of two networks.

$$I = I_{SRRDN} + S_\lambda (I_{SRRDN} - I_{SRRDN-GAN})$$

S_λ is a pixel-wise soft-thresholding operation that depends on λ (=0.2) which controls the amount of information to be combined from the two images[6].

- **EVALUATION:**

Perceptual Metric: $P(I) = \frac{1}{2}((10 - M(I) + N(I))$, where $M(I)$ and $N(I)$ represent Ma.et.al [2] and NIQE [3], respectively

5 Objective Quality Comparisons

Table 1. Quantitative results with BI degradation model. Bold indicates the best performance, red color second best and blue color indicates the third best performance.

Method	Scale	Set5		Set14		B100		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	×4	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
FSRCNN	×4	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280	27.90	0.8610
VDSR	×4	31.35	0.8830	28.02	0.7680	27.29	0.0726	25.18	0.7540	28.83	0.8870
LapSRN	×4	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560	29.09	0.8900
MemNet	×4	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	29.42	0.8942
EDSR	×4	32.46	0.8968	28.80	**0.7876**	27.71	**0.7420**	**26.64**	**0.8033**	**31.02**	0.9148
SRMDNF	×4	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731	30.09	0.9024
D-DBPN	×4	32.47	0.8980	**28.82**	0.7860	**27.72**	0.7400	26.38	0.7946	30.91	0.9137
RDN	×4	32.47	**0.8990**	28.81	0.7871	**27.72**	0.7419	26.61	20.8028	31.00	**0.9151**
MRDN (ours)	×4	**32.48**	0.8983	28.77	0.7855	27.71	0.7396	26.45	20.7956	30.92	0.9137
Bicubic	×8	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160	21.47	0.6500
SRCNN	×8	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440	22.46	0.6950
FSRCNN	×8	20.13	0.5520	19.75	0.4820	24.21	0.5680	21.32	0.5380	22.39	0.6730
SCN	×8	25.59	0.7071	24.02	0.6028	24.30	0.5698	21.52	0.5571	22.68	0.6963
VDSR	×8	25.93	0.7240	24.26	0.6140	24.49	0.5830	21.70	0.5710	23.16	0.7250
LapSRN	×8	26.15	0.7380	24.35	0.6200	24.54	0.5860	21.81	0.5810	23.39	0.7350
MemNet	×8	26.16	0.7414	24.38	0.6199	24.58	0.5842	21.89	0.5825	23.56	0.7387
MSLapSRN	×8	26.34	0.7558	24.57	0.6273	24.65	0.5895	22.06	0.5963	23.90	0.7564
EDSR	×8	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221	24.69	0.7841
D-DBPN	×8	27.21	0.7840	25.13	0.6480	24.88	0.6010	22.73	0.6312	**25.14**	**0.7987**
MRDN (ours)	×8	**27.271**	**0.786**	**25.154**	**0.6511**	**24.952**	**0.602**	**22.820**	**0.634**	24.993	0.795

QUALITATIVE RESULTS

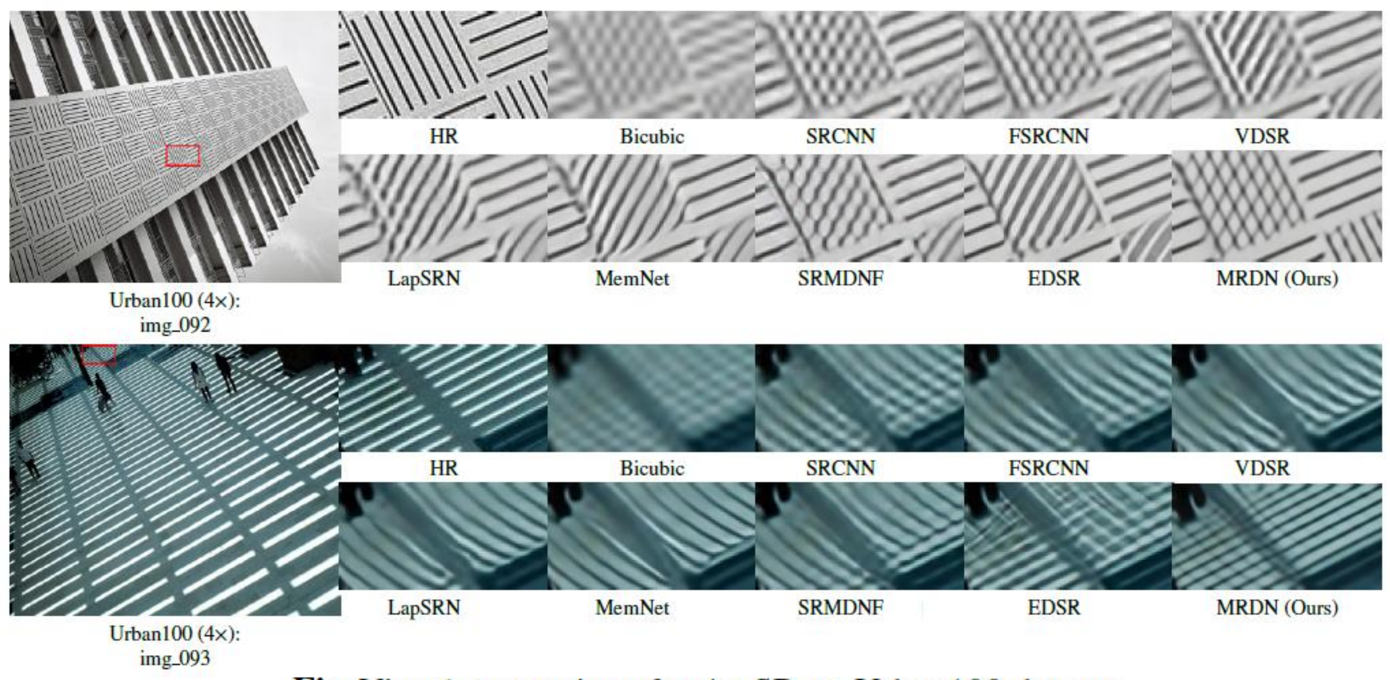
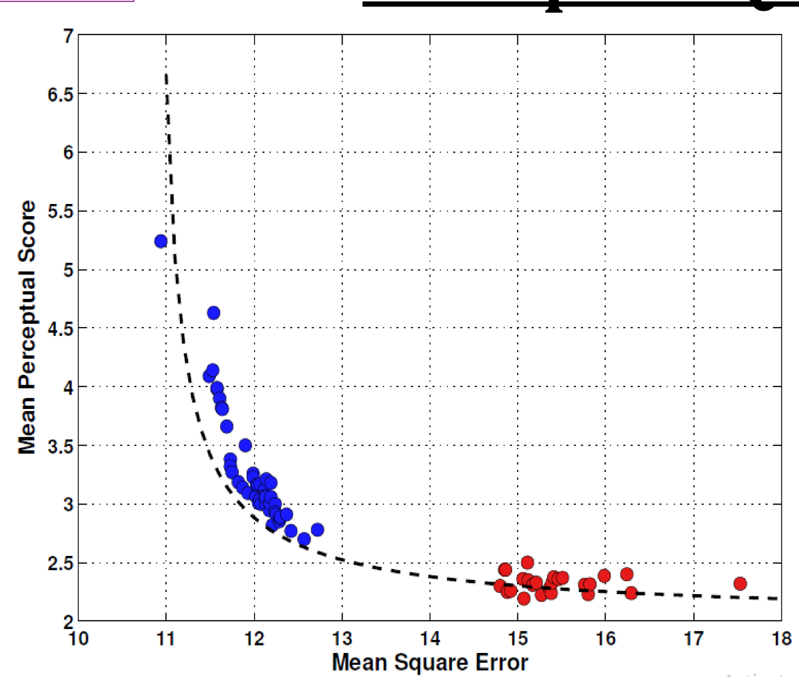


Fig. Visual comparison for 4× SR on Urban100 dataset.

Fig. Visual comparisons with existing approaches for super-resolution by a factor of 8

6 Perceptual Quality Assessment



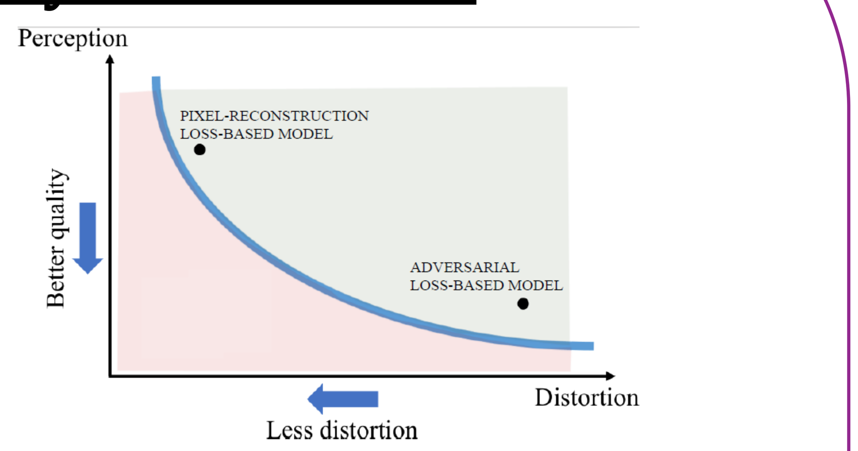


Table 2: (Region3) PSNR and P-score. Red colour indicates the best values.

Methods	Set5		Set14		BSD100		Urban100	
	PSNR	P-Score	PSNR	P-Score	PSNR	P-Score	PSNR	P-Score
SRGAN	29.40	3.61	26.02	2.91	25.16	2.59	22.79	3.45
ENET-PAT	28.56	2.93	25.75	3.01	25.38	2.93	23.68	3.47
MRDN-GAN	29.97	3.51	26.30	2.81	25.52	2.38	24.33	3.51

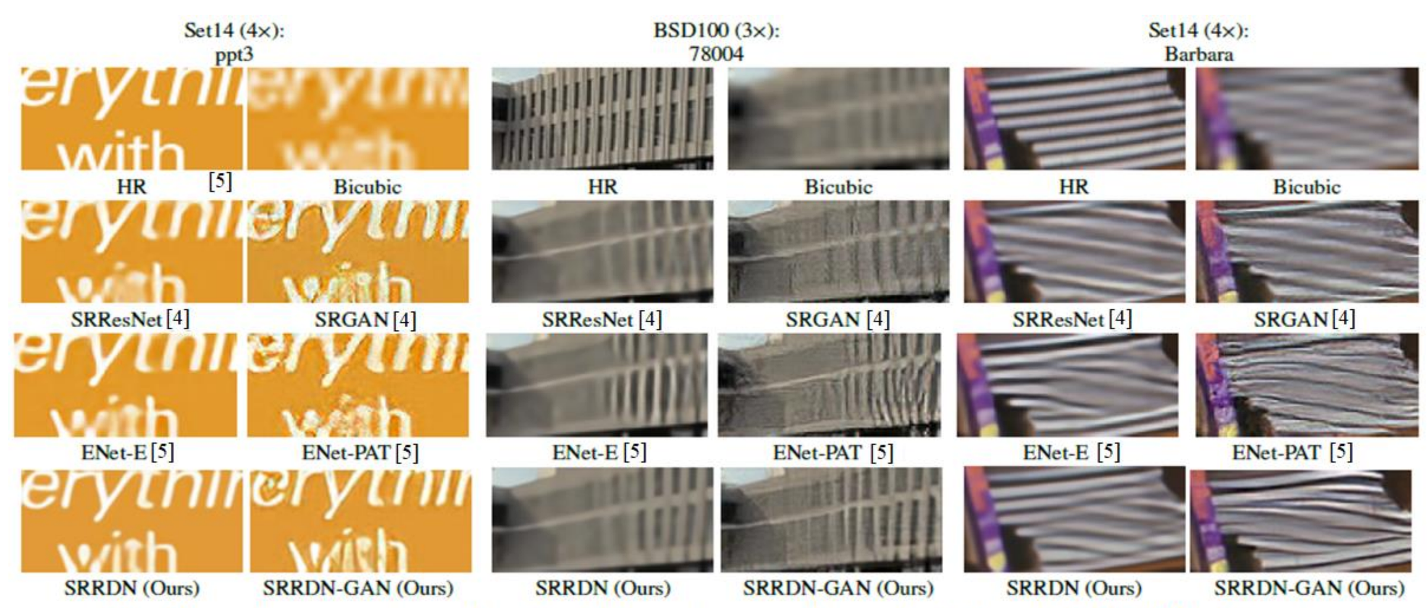


Fig. Visual comparison for 4× SR on images from Set14 and BSD100 datasets.

References

[1] Blau, Yochai, and Tomer Michaeli. "The perception-distortion tradeoff.", *CVPR* 2018.

[2] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang. Learning a no-reference quality metric for single-image super-resolution. CVIU, 2017

[3] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," in *IEEE Signal Processing Letters*, March 2013.

[4] Ledig, Christian, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." *CVPR* 2017.

[5] Sajjadi, Mehdi SM, Bernhard Schölkopf, and Michael Hirsch. "Enhancenet: Single image super- resolution through automated texture synthesis." *ICCV 2017*

[6] Deng, X.: Enhancing image quality via style transfer for single image super-resolution. IEEE Signal Processing Letters, 2018.