ESRGAN+ : Further Improving Enhanced Super-resolution Generative Adversarial Network

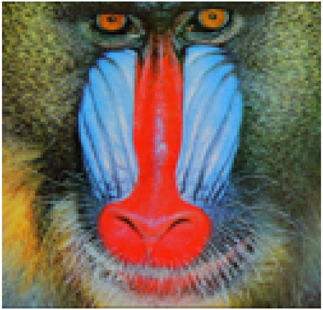
Nathanaël Carraz Rakotonirina
Andry Rasoanaivo

ICASSP 2020

Plan

Super-resolution

- Super-resolution (SR) is the task of generating a high-resolution (HR) image using low-resolution (LR) ones.
- It is referred as Single Image Super-Resolution (SISR) when only one LR image is used.
- It is applied in medical imaging, surveillance, security and helps improve other computer vision tasks.



(a) Low-resolution image of size 125×120

(b) Super-resolved image of size 500×480

Figure: Super-resolution with scale factor $4\times$.

Main categories

- Methods of super-resolution are mainly divided into three categories:
  - Interpolation-based methods
  - Reconstruction-based methods
  - Learning-based methods
- Learning-based approaches outperform the others.
- State-of-the-art methods use deep learning (convolutional neural networks).

Perception-distorsion

- Distorsion measures (like PSNR) used to evaluate image restoration algorithms do not reflect the perceptual quality of the images.
- Images with high PSNR values can be perceptually unsatisfying.
- Unlike PSNR-based methods, perceptual-based methods aim to improve the visual quality of the images.
- SRGAN (Super-Resolution Generative Adversarial Network) introduced a perceptual loss to generate photo-realistic images.
- ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) improves SRGAN by introducing a new architecture and by improving the perceptual loss.

Architecture

- An additional level of residual learning is added in the generator network.
- We call the resulting model ESRGAN+
- This allows the generator to have higher capacity.
- ESRGAN+ benefits from:
  - Feature exploration through the dense connections
  - Feature exploitation through the residuals

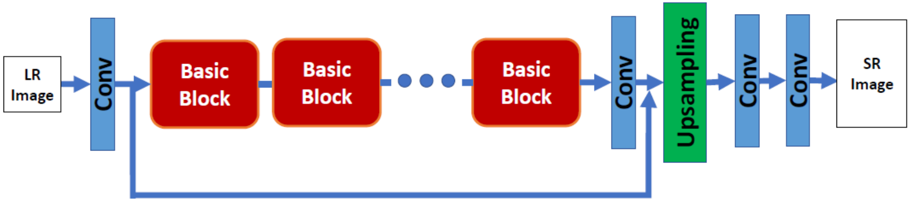


Figure: The basic architecture of SRResNet.

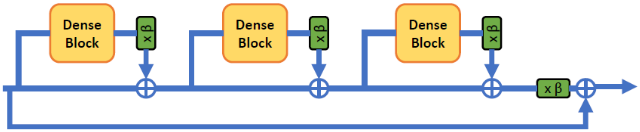
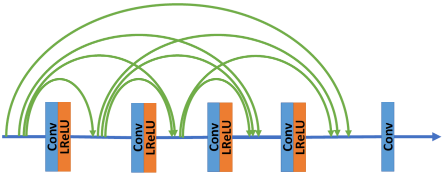
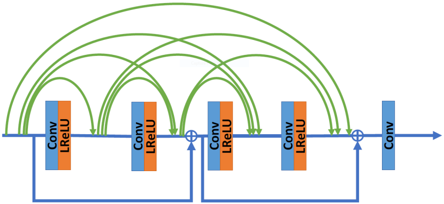


Figure: The basic block used in ESRGAN called Residual in Residual Dense Block (RRDB).



(a) Dense block



(b) Residual Dense block

Figure: **Top:** Dense block is the main path used in ESRGAN's RRDB. **Bottom:** Residuals are added every two layer in the Dense block.

Noise inputs

- Gaussian noise is added in order to benefit from stochastic variation.
- The resulting model is called nESRGAN+.
- Stochastic variation randomizes only certain local aspects of the generated images without changing our global perception of the images

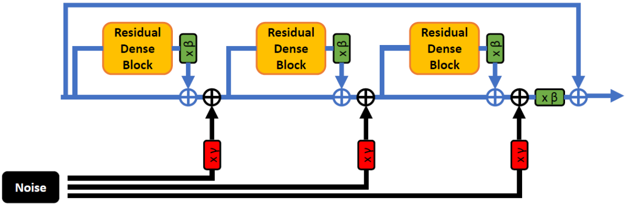


Figure: Gaussian noise is added after each residual along with a learned scaling-factor.

Alternative configurations for noise injection

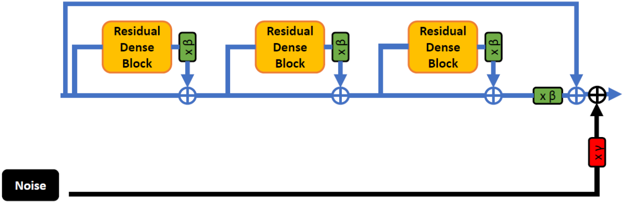


Figure: Gaussian noise is added at the end of the basic block.

Datasets

- The DIV2K dataset is used for training. Data augmentation are performed through random horizontal flips and 90 degree rotations.
- The models are evaluated on benchmark datasets:
  - PIRM datasets
  - Set5
  - Set14
  - Urban100
  - OST300

Image quality assessment

The super-resolution models are evaluated using:

- Peak signal-to-noise ratio (PSNR)
- Perceptual index which is a combination of the non-reference measures of Ma's score and NIQE.

$$perceptual\ index = \frac{1}{2}((10 - Ma) + NIQE)$$

Table: Quantitative evaluation (PSNR / Perceptual index) with other perception-based approaches.

	EnhanceNet	ESRGAN	ESRGAN+ (ours)	nESRGAN+ (ours)
Validation PIRM	25,06/2,68	25,17/2,55	24/<u>2,38</u>	24,32/**2,36**
Test PIRM	24,94/2,72	25,03/2,43	23,80/<u>2,31</u>	24,15/**2,29**
Urban100	23,54/**3,47**	24,36/3,77	23,28/<u>3,55</u>	23,22/3,55
OST300	24,37/2,82	24,64/2,49	23,84/**2,46**	23,80/<u>2,49</u>

Table: Quantitative evaluation (PSNR / Perceptual index) with other perception-based approaches including the different noise configurations.

	Validation PIRM	Test PIRM	Urban100	OST300
EnhanceNet	25,06/2,68	24,94/2,72	23,54/**3,47**	24,37/2,82
ESRGAN	25,17/2,55	25,03/2,43	24,36/3,77	24,64/2,49
nESRGAN+ (/3)	24,27/2,40	24,16/2,33	23,23/3,57	23,78/2,57
ESRGAN+	24/<u>2,38</u>	23,80/<u>2,31</u>	23,28/<u>3,55</u>	23,84/**2,46**
nESRGAN+ (/1)	24,32/**2,36**	24,15/**2,29**	23,22/3,55	23,80/<u>2,49</u>

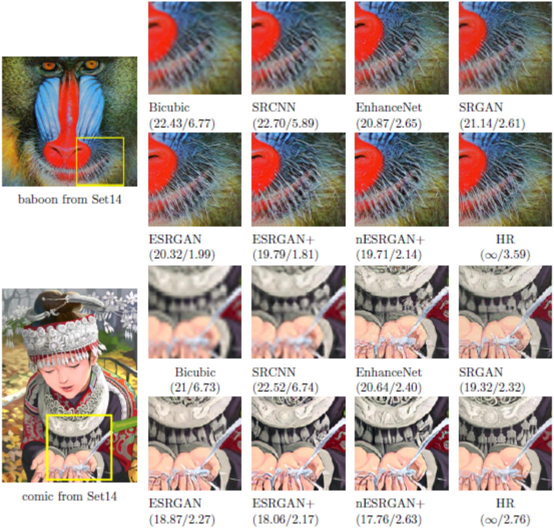


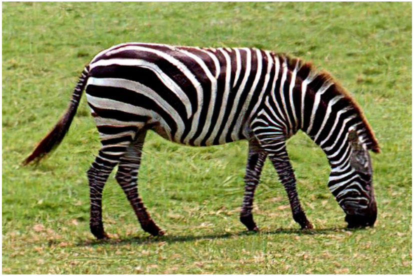
Figure: Qualitative evaluation (PSNR/ Perceptual index) with other approaches

Impact of the noise inputs

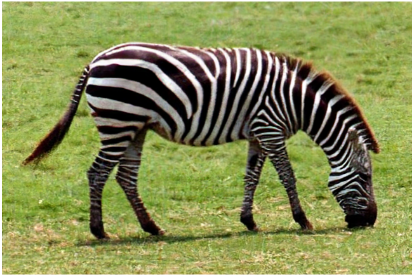
(a) ESRGAN+

(b) nESRGAN+

Figure: man from Set14.



(a) ESRGAN+



(b) nESRGAN+

Figure: zebra from Set14

Limitations of the noise inputs

(a) ESRGAN+
(22,18/**3,05**)

(b) nESRGAN+
(21,86/3,21)

Figure: coastguard from Set14

(a) ESRGAN+
(23,05/**2,95**)

(b) nESRGAN+
(22,54/3,08)

Figure: OST_012 from OST300

(a) ESRGAN+
(18,92/**2,36**)

(b) nESRGAN+
(19/2,67)

Figure: OST_118 from OST300

Conclusion

Our main contributions are the following:

- A novel basic block with an additional level of residual learning;
- Noise injection in the generator's architecture to benefit from stochastic variation.

Future works include:

- Exploit the noise inputs especially at test time;
- Apply multi-scale super-resolution;
- Find better image quality assessments.