Image Super-Resolution

Achyut Pillai
Abdullah Ashfaq

LR X4 Image 0878, Image Size: 510×435

ESRGAN X4 on Image 0878, Image Size: 2040×1740

1. **Problem Statement**
2. **Dataset**
3. **Model Architectures**
4. **Results**
5. **Conclusion**

1. Problem Statement
2. Dataset
3. Model Architectures
4. Results
5. Conclusion

How can high resolution images be accurately constructed given low resolution inputs, while maintaining overall structure and retrieving fine details?

We have addressed this using various deep learning methods to outperform traditional interpolation techniques.

1. Problem Statement
2. **Dataset**
3. Model Architectures
4. Results
5. Conclusion

DIV2K Dataset

The DIV2K dataset is a diverse collection of high resolution images specifically curated for image super-resolution that contains:

- 800 Training images
- 100 Validation images
- Format: 2K resolution PNG

Preprocessing:

- Used bicubic interpolation to downscale images by factors of 2, 4, and 8.
- Took a random patch of each image to use during training
- Applied random horizontal/vertical flips

Image: 0002.png

1. Problem Statement
2. Dataset
3. **Model Architectures**
4. Results
5. Conclusion

SRCNN, SRGAN, and ESRGAN

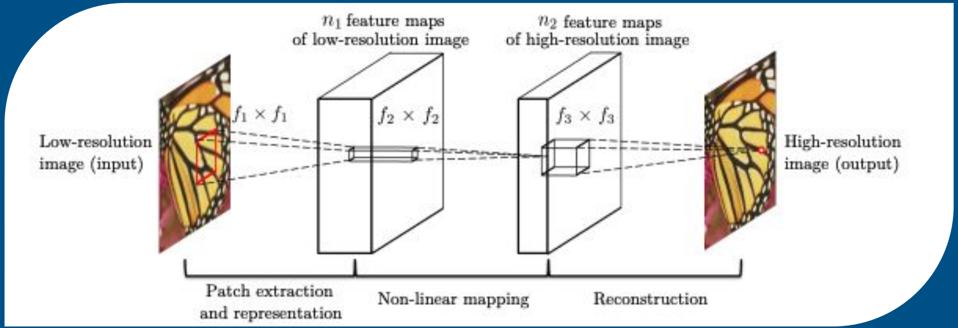
SRCNN Two Variants

SRCNN_v1:

- Three convolutional layers
- First layer: 9×9 kernel; next two: 5×5 kernels
- ReLU activation after each layer

SRCNN_v2:

- Increased capacity with 128 filters in the first layer
- Extra convolutional layer (3×3 kernel + ReLU) before the final layer



Architecture from from Dong et al. (2014)

SRGAN

Generator:

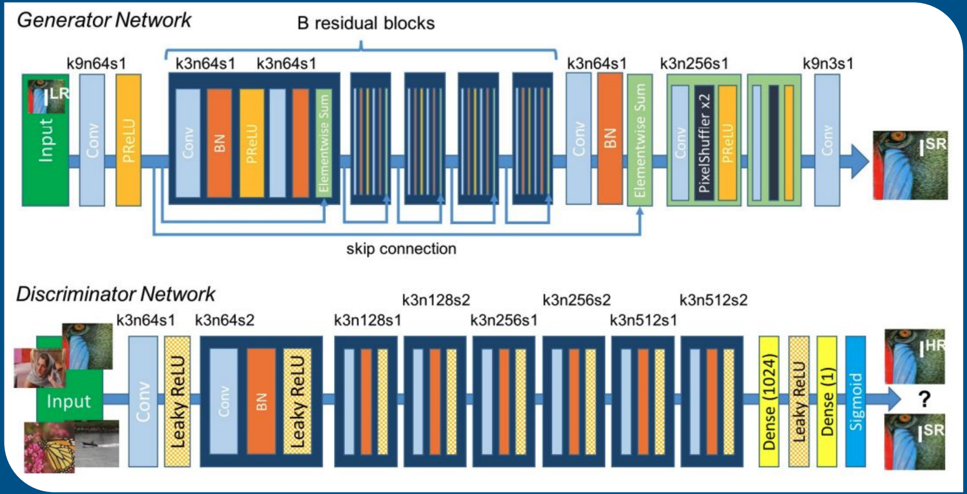
- Based on SRResNet architecture with an initial large-kernel convolutional layer
- Series of residual blocks (each with two conv layers, batch normalization, PReLU)
- Upsampling via sub-pixel convolution and final conv layer

Discriminator:

- Deep conv network to distinguish real vs. generated HR images
- Uses increasing feature dimensions, batch normalization, LeakyReLU, pooling, and dense layers

Perceptual Loss:

- VGG-based loss compares high-level feature representations



Architecture from from Ledig et al. (2017)

ESRGAN

Generator:

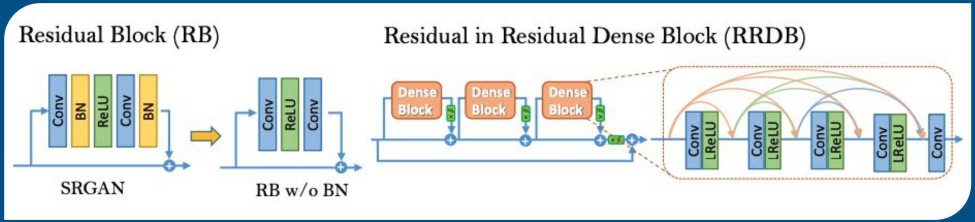
- Employs Residual in Residual Dense Blocks (RRDBs) with dense connections
- Initial convolutional layer followed by a stack of RRDBs
- Multi-stage upsampling with PixelShuffle and LeakyReLU
- Final convolutional layer reconstructs the high-resolution image

Discriminator:

- Same architecture as in SRGAN

Perceptual Loss:

- Utilizes the same VGG-based perceptual loss as SRGAN



Architecture from from Wang et al. (2018)

1. Problem Statement
2. Dataset
3. Model Architectures
4. **Results**
5. Conclusion

Metrics

- **Fréchet Inception Distance (FID)**: Compares the distributions of real and generated images, with lower scores reflecting a closer resemblance

- **Peak Signal-to-Noise Ratio (PSNR)**: Evaluates reconstruction quality, with higher scores indicating greater similarity to the ground truth.

- **Structural Similarity Index Measure (SSIM)**: Measures perceptual similarity between generated and ground-truth images, where a score of 1 indicates identical images.

FID Scores

Scale	Bicubic	$SRCNN_{v1}$	$SRCNN_{v2}$	SRGAN	ESRGAN
2	13.37	67.94	105.98	6.41	5.00
4	37.16	79.91	79.93	30.02	25.85
8	63.55	97.52	100.06	50.20	47.60

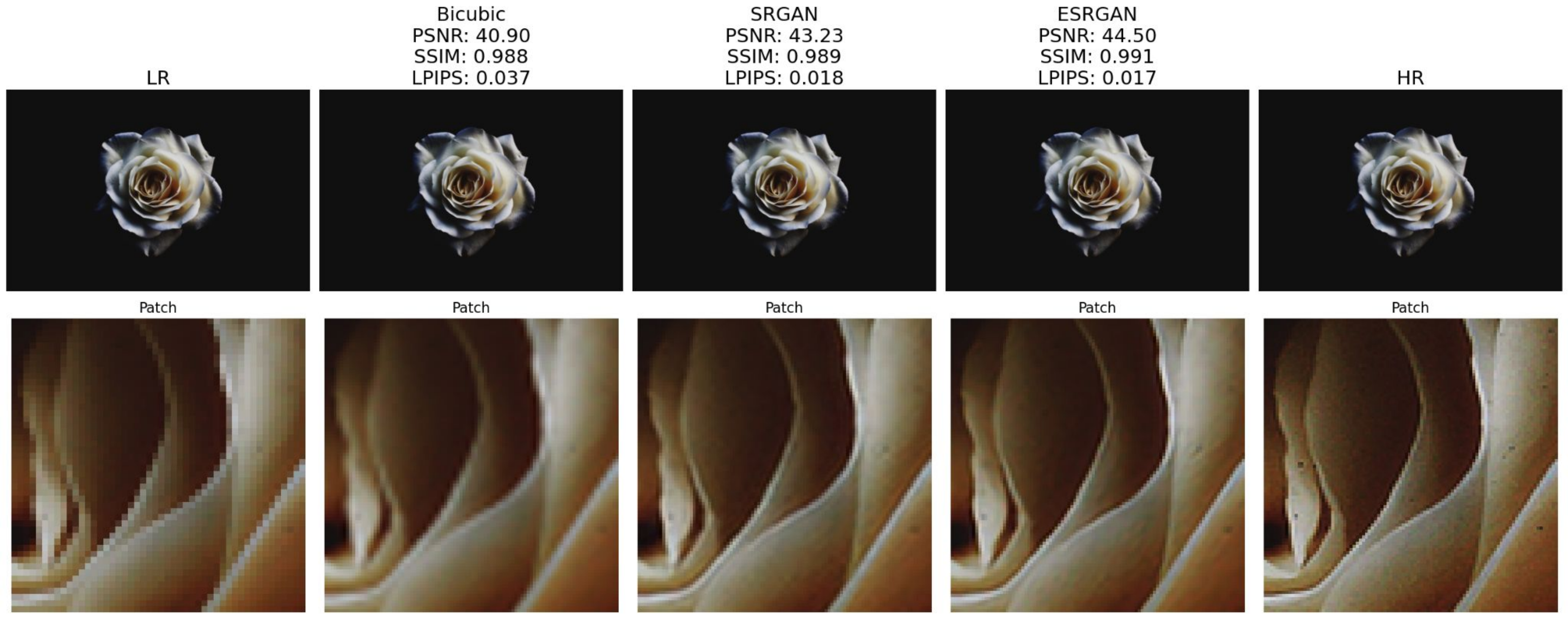
PSNR Scores

Scale	Bicubic	$SRCNN_{v1}$	$SRCNN_{v2}$	SRGAN	ESRGAN
2	30.99	25.67	21.26	32.83	33.53
4	26.64	24.38	24.39	27.19	27.24
8	23.09	22.07	22.0	23.19	23.43

SSIM Scores

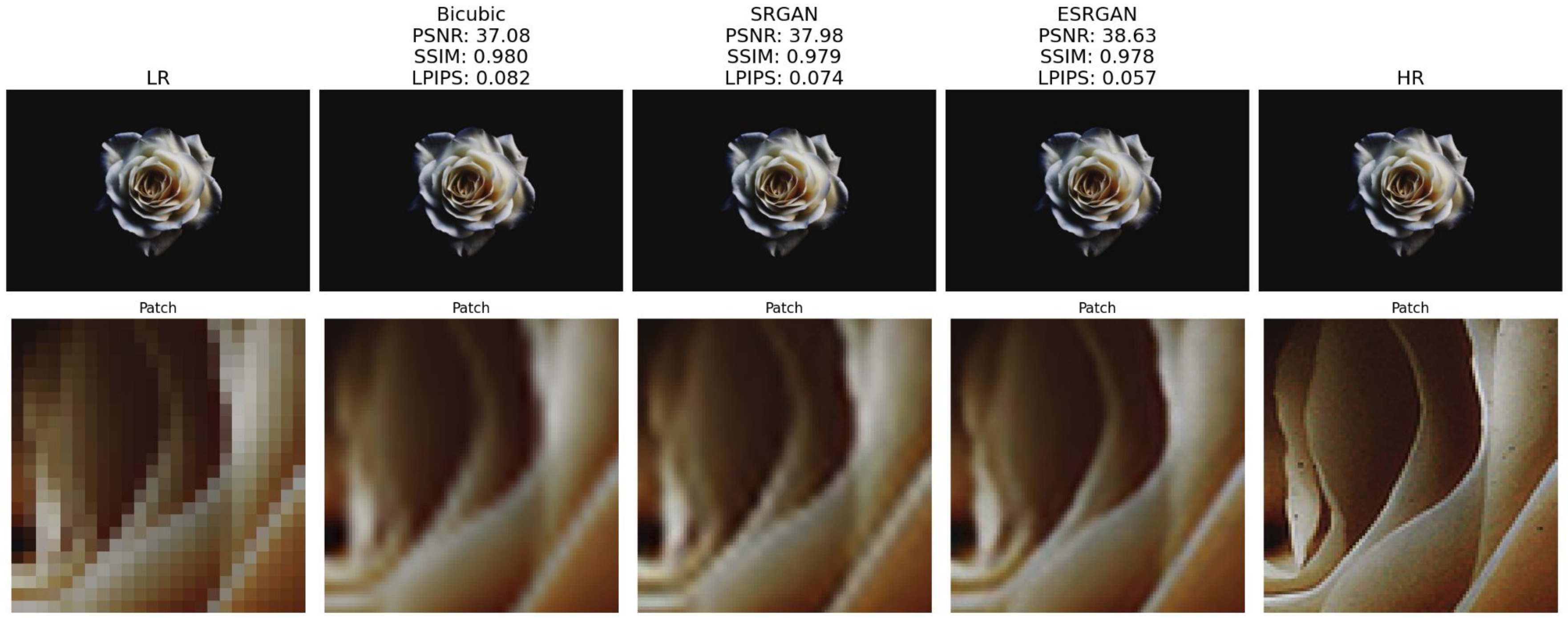
Scale	Bicubic	$SRCNN_{v1}$	$SRCNN_{v2}$	SRGAN	ESRGAN
2	0.9260	0.8005	0.4037	0.9432	0.9517
4	0.8551	0.7700	0.7678	0.8572	0.8622
8	0.8001	0.7237	0.7103	0.7908	0.8018

GAN Comparisons



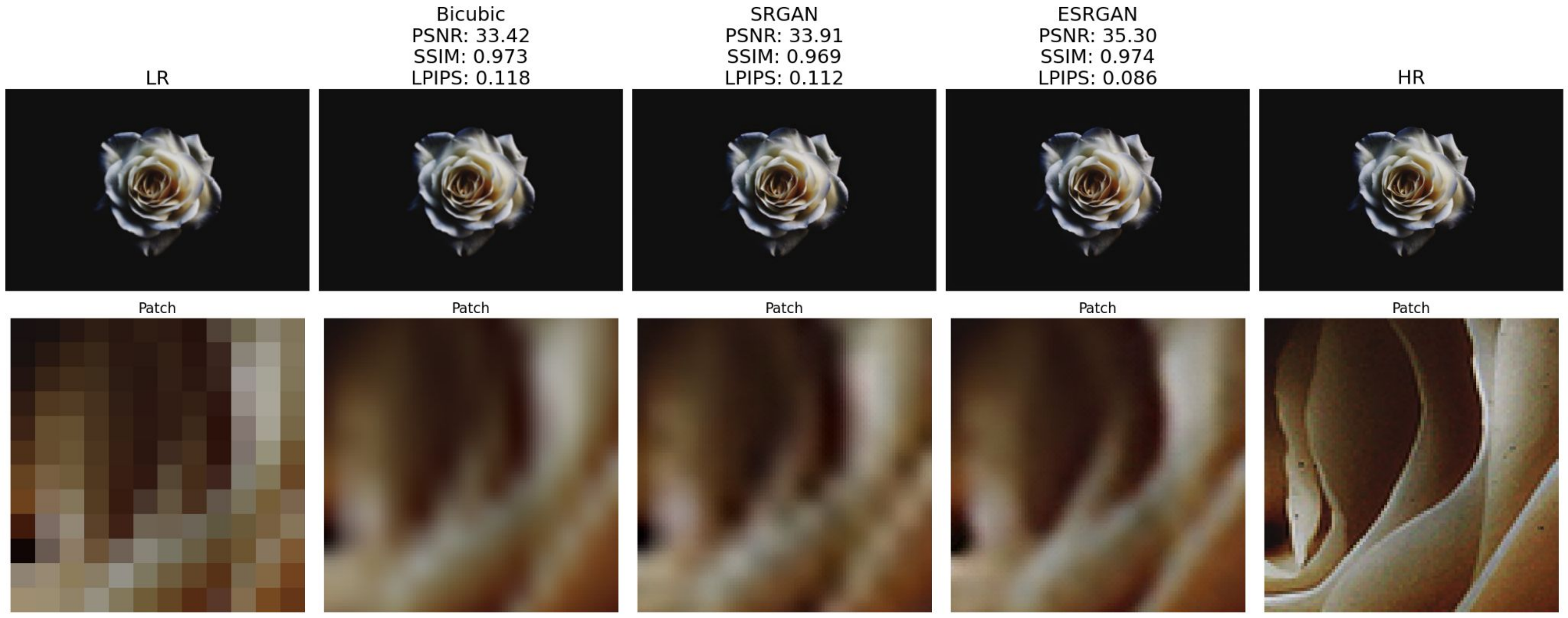
Comparison of Generated Images to Originals, Image 0843 for Scale Factor X2

GAN Comparisons



Comparison of Generated Images to Originals,
Image 0843 for Scale Factor X4

GAN Comparisons



Comparison of Generated Images to Originals,
Image 0843 for Scale Factor X8

Conclusion

The ESRGAN provides the best results with the SRGAN following closely behind. The SRCNN, while computationally efficient, provides only decent results, not any better than bicubic interpolation.

For all models, the lower the scale factor the better the results.

*LR X4 Image 0826,
Image Size: 510×384*

*ESRGAN X4 on Image 0826,
Image Size: 2040×1536*

References

[1] Dong, C., Loy, C., He, K. and Tang, X. (2014). Image Super-Resolution Using Deep Convolutional Networks. In European Conference on Computer Vision (ECCV).

[2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Aitken, A. P., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[3] Wang, X., Yu, K., Wu, S., Gu, C., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C. (2018). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In European Conference on Computer Vision (ECCV) Workshops.

[4] Agustsson, E. and Timofte, R. (2017, July). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

[5] Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., Zhang, L., Lim, B., & others. (2017, July). NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.