



Università  
di Catania

## Super Resolution Networks applied on astronomic photographies

Università degli Studi di Catania  
Dipartimento di Matematica e Informatica  
Computer Vision (LM-18) 2022/2023  
Teachers: Sebastiano Battiato, Francesco Guarnera

**Michele Ferro** – 1000037665

July 2023

## **Abstract**

This project aims to demonstrate that Super Resolution deep networks can excel at difficult tasks, such as enhancing low resolution astronomical photos, even in the absence of specific training. These networks showcase their versatility by effectively generalizing across image domains and generating high-resolution outputs with enhanced details. With their potential to overcome limitations in image quality, SR deep networks have emerged as powerful tools applicable in numerous real-life tasks, fostering advancements in diverse industries and paving the way for future innovations.

So, three different kinds of Super Resolution networks will be tested on the dataset (without a preventive training) to see how the state-of-the-art solutions can handle a very specific task (and also a difficult one) like that of enhancing astronomic photographies.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	SRGAN . . . . .	2
1.2	ESRGAN . . . . .	2
1.3	Quality metrics . . . . .	3
1.3.1	Mean Squared Error (MSE) . . . . .	3
1.3.2	Peak Signal-to-Noise Ratio (PSNR) . . . . .	3
1.3.3	Structural Similarity (SSIM) . . . . .	3
<b>2</b>	<b>The dataset</b>	<b>5</b>
<b>3</b>	<b>Approaches</b>	<b>7</b>
3.1	BSRGAN . . . . .	7
3.2	Real-ESRGAN . . . . .	7
3.3	StableSR . . . . .	9
<b>4</b>	<b>Final results</b>	<b>10</b>
4.1	Used machine specifications . . . . .	10
4.2	Results comparison . . . . .	10
<b>5</b>	<b>Conclusions</b>	<b>15</b>

# 1 Introduction

**Single image super-resolution** (from now **SIRS** or simply **SR**) is a image enhancement process that aims to reconstruct the natural and sharp detailed **high-resolution (HR)** image from a **low-resolution (LR)** one.

Typically, the main supervised approach, known as **non-blind**, considers a **degradation model** that generates the LR image  $\mathbf{y}$  from an HR image  $\mathbf{x}$  in such a way:

$$\mathbf{y} = (\mathbf{x} \otimes \mathbf{k}) \downarrow_s + \mathbf{n}$$

This relation assumes that the LR image is obtained by first convolving the HR image with a Gaussian kernel  $\mathbf{k}$  to get a **blurry image**  $\mathbf{x} \otimes \mathbf{k}$ , followed by a **downsampling** operation  $\downarrow_s$  with scale factor  $s$  and an addition of **white Gaussian noise**  $\mathbf{n}$  with standard deviation  $\sigma$ .

Nevertheless, recent studies demonstrated a more realistic degradation model that also considers the artifact introduced by the JPEG algorithm compression.

In fact, here we have:

$$\mathbf{y} = \left( (\mathbf{x} \otimes \mathbf{k}) \downarrow_s + \mathbf{n} \right)_C$$

where  $C$  is the compression scheme.

There are a lot of Deep Learning networks that try to learn and understand the degradation model from a big set of data (some of these will be later seen), but this approach is often non-applicable because of the loss of HR images datasets from certain domains. Because of this, in real world applications are often used **blind unsupervised** approaches; we can divide these methods in different categories as we can see in the following table:

	Explicit Modelling	Implicit Modelling
External Dataset	Image-Specific Adaptation (without Kernel Estimation) SRMD, UDVD, DPSR, USRNet	Implicit Modelling with Data Distribution Learning CinCGAN, FS-SRGAN, DegradationGAN, FSSR, DASR, Pseudo-supervise
Single Image	Image-Specific Adaptation (with Kernel Estimation) IKC, DAN, VBSR, CBSR, KMSR, RealSR, DRL-DASR, KOALAnet, AMNet-RL	Single Image Modelling with Internal Statistics NPBSR, KernelGAN, ZSSR, FKP, DGML-SR

?

To gain a deeper comprehension about non-blind methods, we have to briefly introduce two different networks, namely SRGAN and ESRGAN.

## 1.1 SRGAN

The **SRGAN** [1] originates from the idea of applying the **ResNet (Residual Network)** architecture, fused with a GAN, to the SR domain.

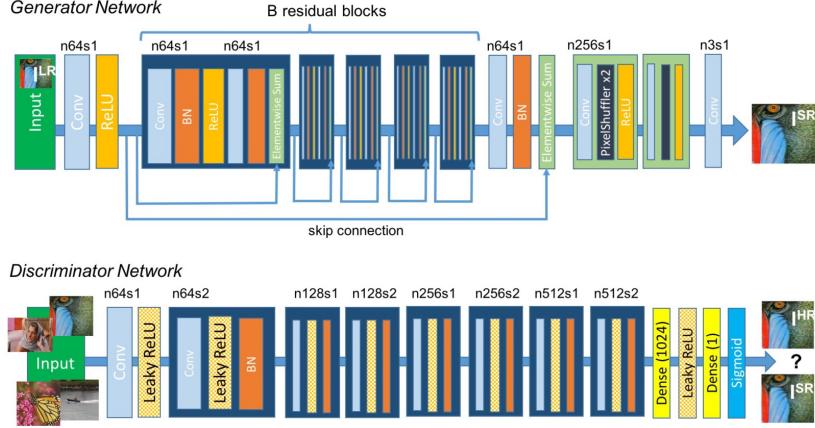


Figure 1: SRGAN architecture

## 1.2 ESRGAN

The **ESRGAN** [4] is an **Enhanced** version of the *SRGAN* (from this the E initial in the acronym). More specifically, this approach improves:

- the network architecture, as the batch normalization is replaced by a **Residual in Residual Dense Block (RRDB)**;
- the GAN's **adversarial loss**, which is modified to determine the relative "realness" of the image;
- the **perceptual loss** to enhance brightness and texture.

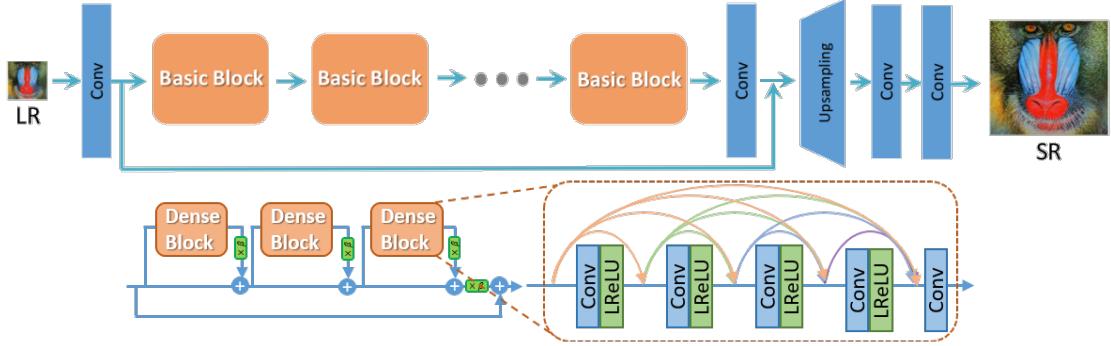


Figure 2: BSRGAN architecture

## 1.3 Quality metrics

In this section will be defined the **quality metrics** recorded to discuss the performances of the adopted approaches; all of these metrics are useful for predicting the perceived quality of the up-scaled images, after the SR process.

### 1.3.1 Mean Squared Error (MSE)

Also known as **Mean Squared Deviation (MSD)**, it measures the average of the squares of the errors between the estimated values and the actual value.

As it is derived from the square of Euclidean distance and the second moment of the error, not only it incorporates both the variance of the estimator and its bias, but it is also always a positive value that decreases as the error approaches to zero.

Let  $I$  and  $K$  respectfully the original and the SR monochrome images of  $M \times N$  pixels, we have:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [K(i, j) - I(i, j)]^2$$

### 1.3.2 Peak Signal-to-Noise Ratio (PSNR)

It defines the **ratio** between the maximum possible power of the signal and the power of the corrupting noise that affects the fidelity of its representation.

Given the MSE between two monochrome images  $I$  and  $K$ , the PSNR is defined (in dB) as:

$$PSNR = 10 * \log_{10} \left( \frac{MAX_I^2}{MSE} \right)$$

The higher is the obtained value, better is the similarity with the original image.

### 1.3.3 Structural Similarity (SSIM)

This metric is different from previous ones as it considers image degradation as perceived change in **structural information**; while MSE and PSNR estimate absolute errors, the SSIM (this is the acronym of the measure) predicts the loss of information about the structure of the objects in the visual scene.

The SSIM index is calculated on various windows of an image. The measure between two windows  $x$  and  $y$  of size  $N \times N$  is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

with:

- $\mu_x$  the pixel sample mean of  $x$ ;
- $\mu_y$  the pixel sample mean of  $y$ ;
- $\sigma_x^2$  the variance of  $x$ ;
- $\sigma_y^2$  the variance of  $y$ ;
- $\sigma_{xy}^2$  the cross-correlation of  $x$  and  $y$ ;

- $c_1 = (k_1 L)^2$ ,  $c_2 = (k_2 L)^2$  two variables to stabilize the division with weak dominator;
- $L$  the dynamic range of the pixel-values (typically  $2^{\#\text{bits per pixel}-1}$ );
- $k_1 = 0.01$  and  $k_2 = 0.03$  by default.

The higher is the obtained value, better is the structural similarity with the original image.

## 2 The dataset

**Galaxy10 DECals** [2] dataset is an improved version of *Galaxy10*, originally created with ***Galaxy Zoo Data Release 2***, where volunteers classified approximately 270k of SDSS galaxy images where up to 22k of those images were selected in **10 broad classes** using volunteer votes.

Galaxy Zoo later utilized images from ***DESI Legacy Imaging Surveys (DECals)*** with better resolution, and finally Galaxy10 DECals has combined all of the three datasets, resulting in a set of approximately 441k unique galaxies covered by DECals, about 18k of which were selected in 10 broad classes using volunteer votes with more rigorous filtering.

As this project involves the use of Super Resolution networks, the dataset we're going to use is **Galaxy10 SDSS**, a variation of the Galaxy10 Decals containing more than 21k  $69 \times 69$  pixels colored galaxy images, separated in 10 mutually exclusive classes, as we can see following:

Class ID	Label	Number of Examples
0	Disk, Face-on, No Spiral	3461
1	Smooth, Completely round	6997
2	Smooth, in-between round	6292
3	Smooth, Cigar shaped	394
4	Disk, Edge-on, Rounded Bulge	1534
5	Disk, Edge-on, Boxy Bulge	17
6	Disk, Edge-on, No Bulge	589
7	Disk, Face-on, Tight Spiral	1121
8	Disk, Face-on, Medium Spiral	906
9	Disk, Face-on, Loose Spiral	519

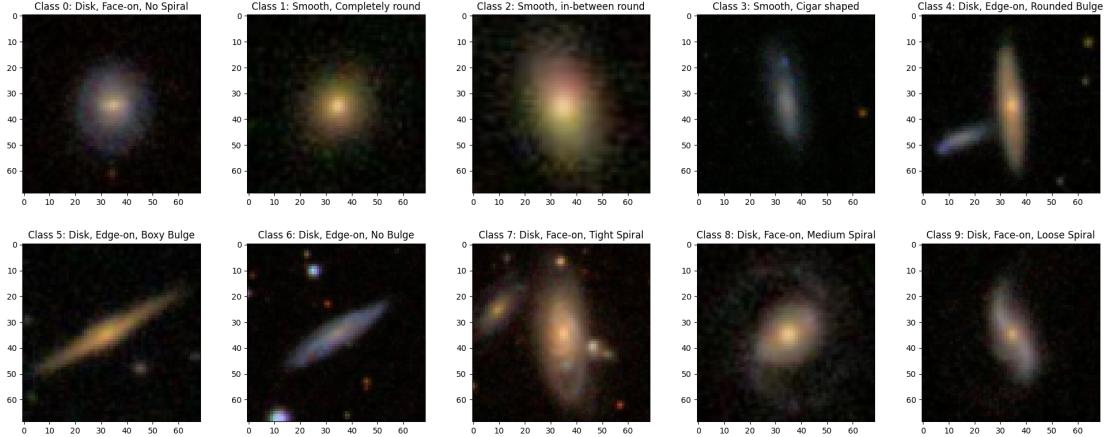


Figure 3: Galaxy10 SDSS examples for each class

Galaxy10 is born as **alternative** to MNIST or Cifar10 datasets as a deep learning toy for astronomers and astronomy passionates.

In order to obtain this particular low-resolution dataset, the authors took the original images  $424 \times 424$  pixels, cropped them to the center to obtain  $207 \times 207$  pixels images and finally downsampled via bilinear interpolation to  $69 \times 69$ .

However, the authors state that there is no guarantee on the accuracy of the labels and that Galaxy10 is not a balanced dataset: in its state, it should be used only for experimental and educational purposes, like in this case.

### 3 Approaches

#### 3.1 BSRGAN

BSRGAN [6] is acronym for ***Blind Super Resolution Generative Adversarial Network***, a deep model published in 2021 that uses the non-blind approach to learn a new degradation model to synthetize LR images for training.

More specifically, this degradation model is distinguished from the "traditional" one because of the following characteristics.

1. The blur, downsampling and noise are made more practical.

- **Blur:** reluts by two convolutions with isotropic and anistotropic Gaussian kernels from both the HR space and LR space;
- **Downsampling:** nearest, bilinear, bicubic, down-up-sampling;
- **Noise:** Gaussian noise, JPEG noise, processed camera noise are considered.

2. A **degradation shuffle** is introduced: instead of using the commonly-used blur/downsampling/noise-addition pipeline, the authors perform randomly shuffled degradations to synthesize LR images.

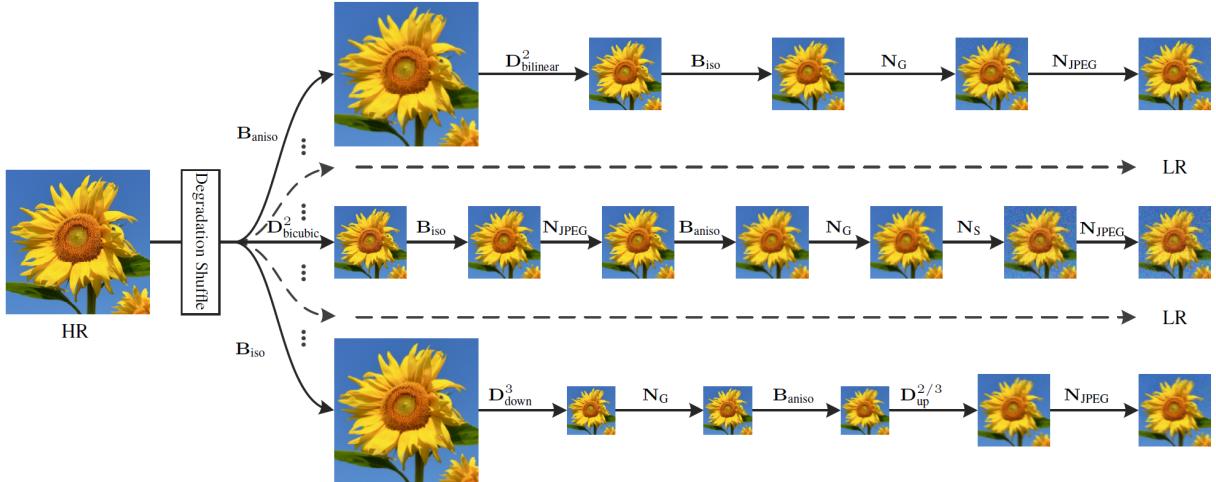


Figure 4: BSRGAN architecture

Tipically, the direct application consists into training a deep blind super-resolver with paired LR/HR images on a large dataset in order to produce unlimited perfectly aligned images.

This means that the degradation model tends to be unsuited to model a degraded LR image as it involves too many degradation parameters and also adopts a random shuffle strategy; because of these qualities, it can produce some degradation cases that rarely happen in real-world scenarios.

#### 3.2 Real-ESRGAN

**Real-ESRGAN** [5] aims at developing Practical Algorithms for General Image/Video Restoration. As the authors state, this network is a further enhancement of the ESRGAN network, in order to make it work on **real-world data** (here, the name of the model).

More specifically, by incorporating various degradation factors encountered in real-world scenarios, such as downsampling, noise, and compression artifacts, the model is trained to effectively handle real scenarios. In fact, compared to traditional single-image super-resolution methods that rely on handcrafted features and interpolation techniques, Real-ESRGAN’s deep learning approach allows it to learn complex image representations and capture fine-grained details.

While the specific performance of Real-ESRGAN may vary depending on the dataset and application, it has shown promising results in multiple benchmark datasets and outperforms several existing super-resolution methods in terms of both quantitative measures.

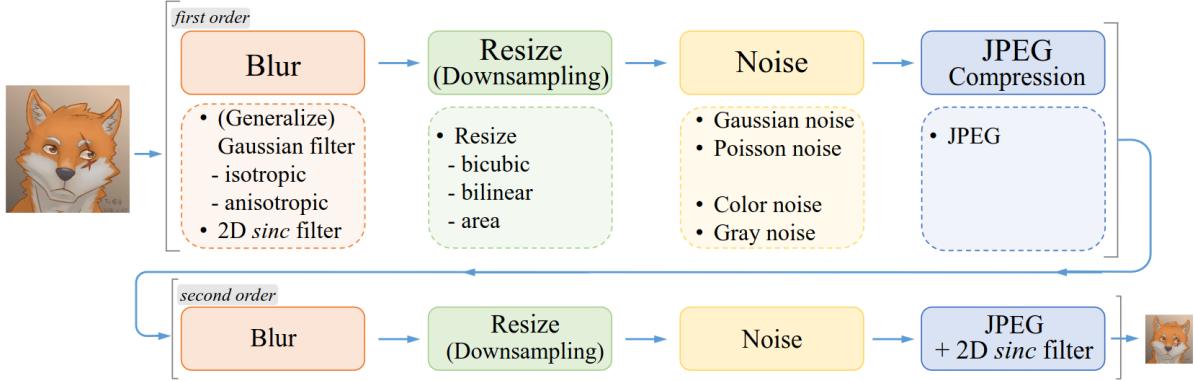


Figure 5: Real-ESRGAN degradation model

This network introduces a degradation model that simulates both first and second order degradation models, commonly encountered in real-world scenarios.

### 1. First-order degradation model

- **Downsampling:** The first-order degradation model includes downsampling, which reduces the resolution of the high-resolution image to obtain a low-resolution version. Downsampling can be performed using various techniques, such as bicubic or bilinear interpolation, which reduce the number of pixels in the image.
- **Additive Gaussian noise:** Another component of the first-order degradation model is the addition of Gaussian noise to the low-resolution image. This noise can simulate imperfections introduced during image acquisition, transmission, or compression.

### 2. Second-order degradation model

- **Compression artifacts:** The second-order degradation model in Real-ESRGAN accounts for compression artifacts. These artifacts are typically introduced when an image undergoes lossy compression, such as JPEG compression. Lossy compression techniques discard some image information to reduce file size, resulting in visual artifacts like blockiness and blurring.
- **Gaussian blur:** The second-order degradation model also includes Gaussian blurring. Gaussian blur simulates the blurring effect caused by various factors, such as camera lens imperfections or motion blur during image capture.

Thanks to the higher order components introduction during the training process, this kind of network learns to generate images that are visually more appealing and realistic than the approaches that just use the first-order degradation model.

### 3.3 StableSR

This kind of SR network originates from the authors realization about advancement in diffusion models for image synthesis tasks. Generative networks like Stable Diffusion can be applied to various content creation tasks, including image/video editing, and so image/video enhancement. In this approach, known as **StableSR** [3] (*Stable Super Resolution*), the authors preserve pre-trained diffusion priors without making explicit assumption about the degradation.

Instead of train a diffusion model from scratch, this network only needs to **fine-tune** a lightweight time-aware encoder and a few feature modulation layers for the SR task.

To suppress **randomness** inherited from the diffusion model as well as the information loss due to the encoding process of VQGAN, the authors introduced a controllable feature wrapping module with an adjustable coefficient to refine the outputs of the diffusion model during the decoding process of VQGAN.

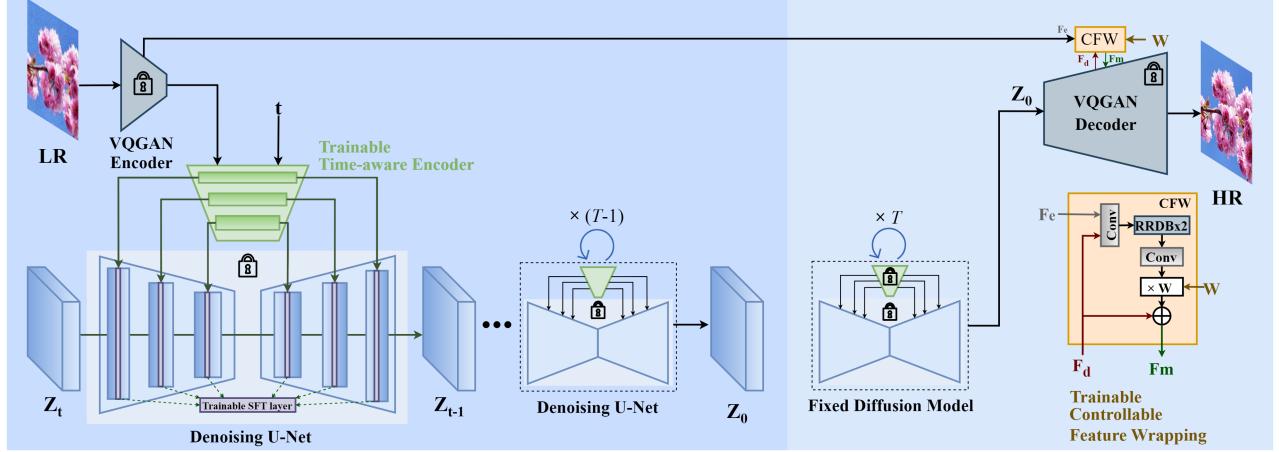


Figure 6: StableSR architecture

More in depth, first there is a fine-tuning of the **time-aware encoder**, attached to a fixed pre-trained **Stable Diffusion** model; features are combined with trainable **spatial feature transform (SFT)** layers. Then, the diffusion model is fixed. Introducing the **controllable feature wrapping (CFW)** module, the authors managed to obtain a tuned feature in a residual manner, given the additional information from LR features and features from the fixed VQGAN decoder. Finally, with an adjustable coefficient, CFW can trade between **quality and fidelity**; the **arbitrary-size super-resolution** is made possible applying an aggregation sampling strategy.

## 4 Final results

In the current chapter will be first showed and then discussed the results of every single network on each class, analyzing the values of the obtained quality measures.

### 4.1 Used machine specifications

All the networks have been tested on a machine provided by the **Pro Plan** of *Google Colab*.

The machine have the following specifications:

- **OS:** Ubuntu 20.04.6 LTS (Focal Fossa);
- **Kernel:** GNU/Linux 5.15.107+;
- **CPU:** Intel(R) Xeon(R) CPU @ 2.20GHz;
- **RAM:** 25 GB;
- **GPU:** NVIDIA Tesla T4;
- **VRAM:** 15 GB;
- **CUDA Driver:** Version 12.0.

### 4.2 Results comparison

First, let's see for each class the qualitative measure values recorded and then the actual predictions.

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	80.935	29.049	0.600
<i>Real-ESRGAN</i>	39.370	32.179	0.698
<i>StableSR</i>	70.704	29.636	0.734



Figure 7: Class 0: Disk, Face-on, No Spiral results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	96.721	28.275	0.768
<i>Real-ESRGAN</i>	52.028	30.968	0.801
<i>StableSR</i>	50.889	31.064	0.685

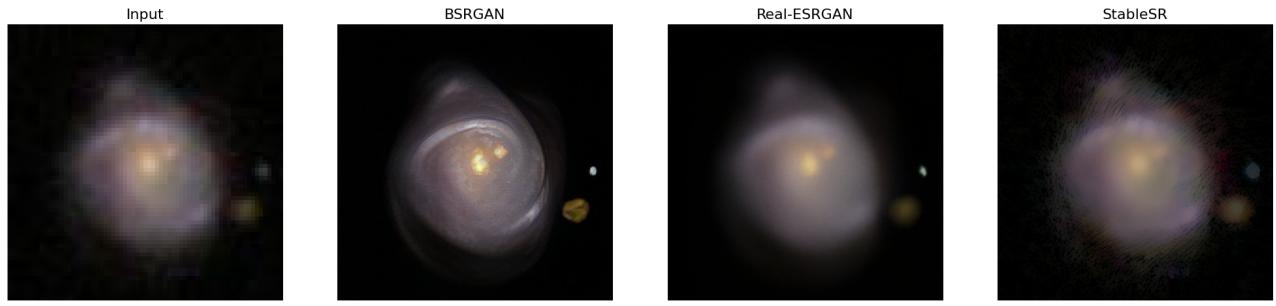


Figure 8: Class 1: Smooth, Completely round results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	77.746	29.223	0.608
<i>Real-ESRGAN</i>	32.691	32.986	0.606
<i>StableSR</i>	25.584	34.051	0.802

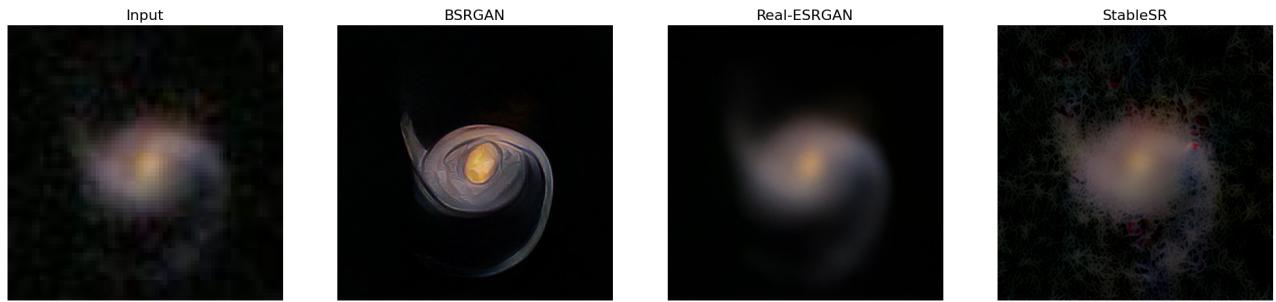


Figure 9: Class 2: Smooth, in-between round results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	354.315	22.636	0.625
<i>Real-ESRGAN</i>	19.961	35.128	0.702
<i>StableSR</i>	21.531	34.799	0.834

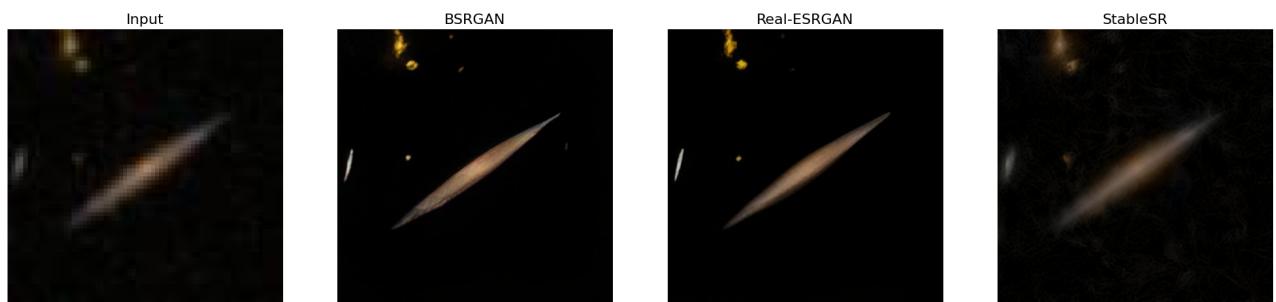


Figure 10: Class 3: Smooth, Cigar shaped results

<b>Network</b>	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
<i>BSRGAN</i>	61.981	30.208	0.682
<i>Real-ESRGAN</i>	37.954	32.338	0.698
<i>StableSR</i>	41.833	31.915	0.722



Figure 11: Class 4: Disk, Edge-on, Rounded Bulge results

<b>Network</b>	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
<i>BSRGAN</i>	133.640	26.871	0.538
<i>Real-ESRGAN</i>	51.112	31.045	0.527
<i>StableSR</i>	32.212	33.050	0.740



Figure 12: Class 5: Disk, Edge-on, Boxy Bulge results

<b>Network</b>	<b>MSE</b>	<b>PSNR</b>	<b>SSIM</b>
<i>BSRGAN</i>	99.449	28.154	0.465
<i>Real-ESRGAN</i>	94.677	28.368	0.368
<i>StableSR</i>	13.305	36.890	0.850



Figure 13: Class 6: Disk, Edge-on, No Bulge results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	33.973	32.819	0.765
<i>Real-ESRGAN</i>	19.092	35.322	0.757
<i>StableSR</i>	24.135	34.304	0.788

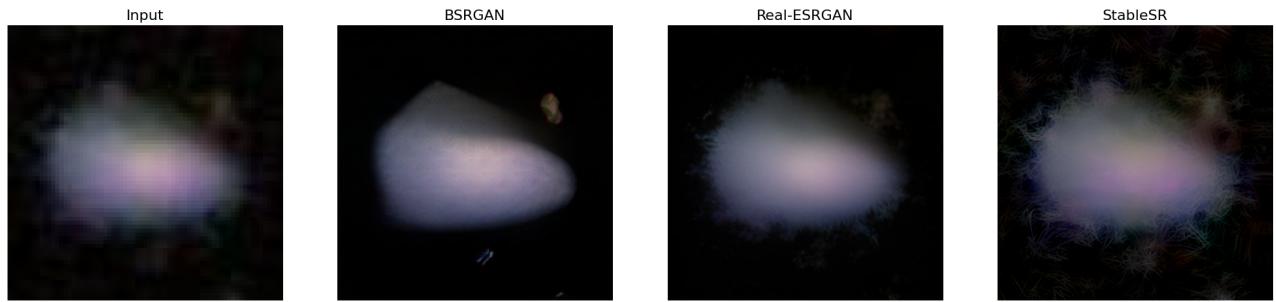


Figure 14: Class 7: Disk, Face-on, Tight Spiral results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	54.657	30.754	0.738
<i>Real-ESRGAN</i>	16.866	35.860	0.803
<i>StableSR</i>	27.384	33.755	0.810

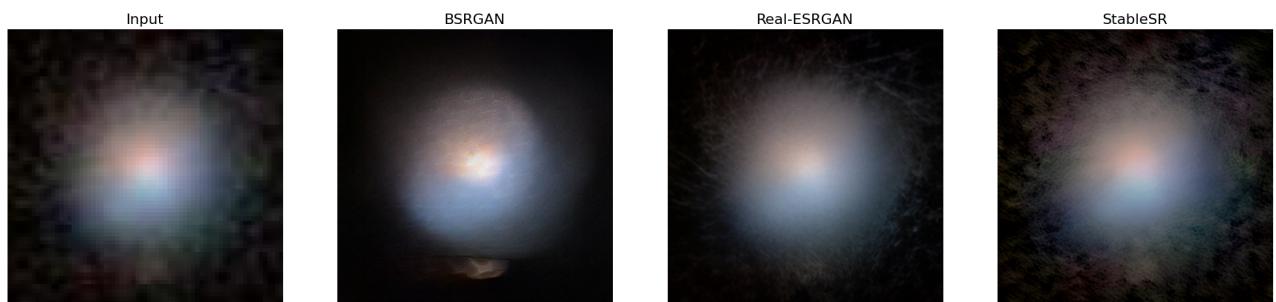


Figure 15: Class 8: Disk, Face-on, Medium Spiral results

Network	MSE	PSNR	SSIM
<i>BSRGAN</i>	161.91	26.038	0.542
<i>Real-ESRGAN</i>	27.244	33.778	0.712
<i>StableSR</i>	63.819	30.081	0.665

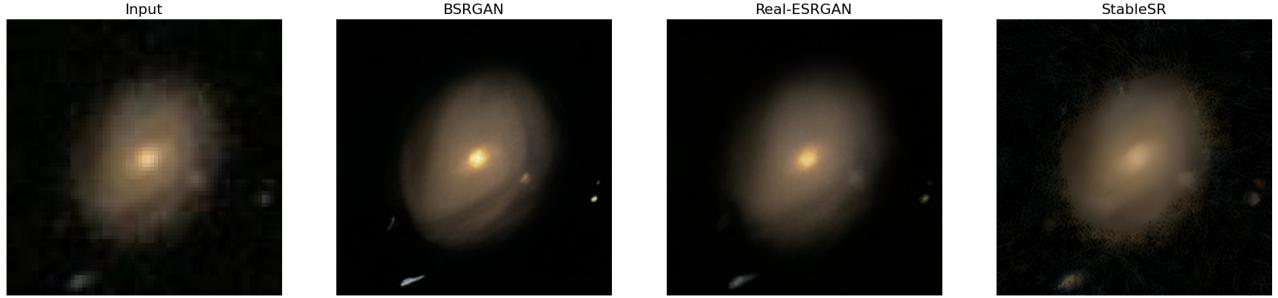


Figure 16: Class 9: Disk, Face-on, Loose Spiral results

As we can observe from the various tables, the *StableSR* network is the one on which have been recorded the higher SSIM value. However, this quality index describes how much the predicted image is structurally similar to the original one – even in its imperfections. From this point of view, we can clearly see that the *Stable Diffusion* backbone took part of the information to upscale from the blocketized parts of the original image: this means that, although the score is often higher on this network, the visual result actually isn't so pleasant.

More interesting are the results obtained with the first two kind of networks: *BSRGAN* and *Real-ESRGAN*. We can observe that the visual results are better in the images predicted by these two networks, as they are based on a non-blind approach that tries to apply an inverse transform with a degradation model (embedded in the network because of the pre-training); it's enthralling how, actually, the results depend from the example's provenience class – and so, from the image – more than from the used network.

In fact, in both of the cases we can see smoother and better appealing images. But anyway, there are some classes whose measures are colliding with the visual results; let's take the example of **Class 7 (Figure 14)**: here we can see how the PSNR value recorded with *Real-ESRGAN* is higher then the ones recorded with the other two networks, and so also the mean squared error is lower; nevertheless, the image predicted by the *BSRGAN* is visually more pleasant than the others, as it is smoother and free of artifacts introduced by the upscaling of the blocketized areas (which, in turn, are caused by the downscale that the authors applied on the original HR image).

On the other hand, let's observe the results of **Class 9 (Figure 16)**.

Here, the *Real-ESRGAN* network scored the higher SSIM value between the three (also, the MSE and PSNR values suggest a big similarity with the input image), and in fact we can see how the image is better looking than the ones predicted from its rival-networks: there are not artifacts and some fine details (like the stars in the right and on the bottom-left, which in BSRGAN prediction seems more like "color-stains") are not uselessly emphasized.

## 5 Conclusions

The results obtained from applying SR deep networks to enhance LR astronomical photos on the Galaxy10 dataset highlight the remarkable capabilities of these networks, even without specialized training for the task at hand. Despite the absence of domain-specific training, the SR deep networks successfully increased the resolution of the astronomical images. This success demonstrates the potential of SR networks to tackle challenging tasks and hints at their versatility in real-life applications.

Although the SR networks might not possess explicit knowledge about the intricacies of astronomical images, their capacity to generalize and extract meaningful features from any image domain allows them to adapt and produce satisfactory results.

The implications of this finding extend far beyond the realm of astronomical imaging. In various fields, such as medical imaging, surveillance, remote sensing, and digital forensics, there is often a dearth of high-resolution images due to hardware limitations or unfavorable environmental conditions. SR deep networks offer a solution by effectively enhancing low-quality images, providing crucial details and enabling more accurate analysis and decision-making processes.

It is worth noting that while SR deep networks exhibit promising potential, challenges and limitations persist. The performance of these networks heavily depends on the input image quality, and they may struggle with extremely low-resolution or heavily degraded images, as seen in the previous chapter. Additionally, training large-scale models and ensuring generalization across diverse image domains remain active areas of research.

## References

- [1] Christian Ledig et al. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. 2017. arXiv: 1609.04802 [cs.CV].
- [2] Henry Leung and Jo Bovy. *Galaxy10 SDSS dataset*. June 2023. URL: <https://astronn.readthedocs.io/en/latest/galaxy10sdss.html>.
- [3] Jianyi Wang et al. *Exploiting Diffusion Prior for Real-World Image Super-Resolution*. 2023. arXiv: 2305.07015 [cs.CV].
- [4] Xintao Wang et al. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. 2018. arXiv: 1809.00219 [cs.CV].
- [5] Xintao Wang et al. *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*. 2021. arXiv: 2107.10833 [eess.IV].
- [6] Kai Zhang et al. *Designing a Practical Degradation Model for Deep Blind Image Super-Resolution*. 2021. arXiv: 2103.14006 [eess.IV].