Ludwig-Maximilians-Universität München
MSc Data Science
Master Thesis

Summer Semester 23

Multi-Image Super Resolution and Domain Adaptation Techniques applied to Thermal Remote Sensing

June 30, 2023

Student:
Massaro, Patricio
<p.massaro@campus.lmu.de>

Acknowledgments

Contents

1 Introduction

This thesis delves into the intricate realm of MISR and Domain Adaptation techniques, tailored for the domain of thermal remote sensing, with an emphasis on the challenges and innovations within the unpaired dataset context. Thermal imagery, with its unique sensitivity to temperature variations, offers an invaluable perspective for phenomena such as wildfire tracking and climate change studies. However, the native resolution of such thermal images often falls short of the detail necessary for fine-grained analysis, propelling the need for advanced SR methods.

1.1 Storyline

- Forest fires are dynamic events that change quickly over time. They could be monitored using remote sensing LWIR -¿ LST
- Spatio-temporal trade off: Big missions have high resolution but bad revisit frequency. This is where forest plays a role. But can we improve the resolution using post-processing techniques?
- Super resolution is an ill-posed problem, supervised deep learning techniques are generally used in the current literature.
- MISR leverages on subpixel differences present in several images taken from the same scene, potentially giving more information to generate an SR image. The potential increase in performance comes with the cost a data-processing overhead.
- One of the biggest challenges in super resolution is to create proper datasets for model training. Usually the degradation model used to generate HR-LR pairs is very simplistic compared to real cases. This problem is called domain gap.
- To bridge the gap, techniques like domain adaptation using gans could be used to estimate the degradation process and generate more realistic datasets, that will translate in better production-ready models.

This thesis covers three main questions:

- How does the performance of Multi-Image Super Resolution (MISR) compare to Single-Image Super Resolution (SISR), considering the pre-processing burden associated with MISR?
- How do traditional baseline degradation models, such as gaussian blurring, compare against a probabilistic model that aligns the distribution of a source domain (e.g., ECOSTRESS) to our target domain (FOREST-2)?
- What impact do the different degradation models have during the inference stage, when the SR models are used in real data.

1.2 Wildfire monitoring using thermal remote sensing

1.2.1 Spatio-temporal trade-off

2 Super resolution

Single image Super-Resolution (SR) is the task of increasing the resolution of a given image as well as sharpening its content by predicting the high-frequency component and the missing information

Super resolution refers to an image processing process of recovering a corresponding high-resolution image from a low-resolution version of it, with applications that range from natural images [1], [2] to satellite [3] and medical imaging [4]. SR remains a challenging task in computer vision because it is considered an ill-posed problem: several HR images can generate exactly the same LR image.

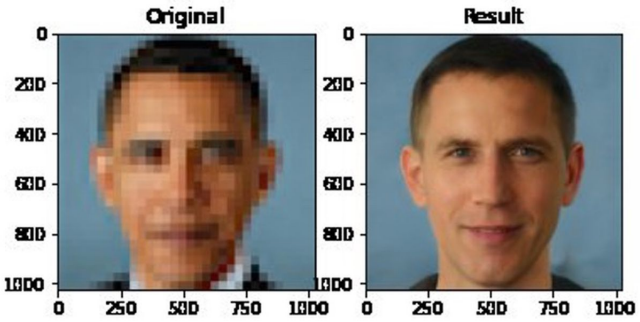

Figure 2.1: Example of super resolution as an ill posed problem. A blurry picture of Barack Obama can be generated from an HR image of another person.

Depending on the different number of low-resolution inputs, image super resolution reconstruction can be divided into single-image or multi-image. ASDFAGDSGADFS-GAFSFAD

Traditional interpolation-based methods for upsampling images were the first type of algorithms used for super resolution. The most common techniques are nearest-neighbor, bilinear and bicubic interpolation. Nearest-neighbor interpolation is the most straightforward algorithm, as the interpolated value is based on its nearest pixel values. While this method requires almost no calculations, the results are usually blocky because there are no interpolated smooth transitions. Bilinear and bicubic interpolation produces smoother transitions using linear or cubic interpolation in both axes. Bilinear interpolation needs a receptive fields of 2x2 and is usually faster, bicubic needs a receptive field of 4x4. THe latter is the most common baseline to understand the improvement of a super resolution network.

- what is super resolution?
- Interpolation based vs reconstruction-based vs learning-based

2.1 single-Image Super Resolution

In a typical SISR framework, the LR image I^{LR} is modeled as follows:

$$I^{LR} = (I^{HR} * k) \downarrow_s + n \tag{1}$$

Where $I^{HR} * k$ is the convolution between a blurring kernel k and the unknown HR image I^{HR}, $\downarrow_s$ is the downsampling operator with scaling factor s and n is a noise term. Super resolution objective is to solve this equation and obtain I^{HR} which as stated before, is an extremely ill-posed problem. Super resolution was first proposed in the 1960s, while the first use of multiple images dates of 1989. Machine learning was used for the first time in 2000. Deep learning appears as a branch of machine learning, emphasizing the use of multi-layer neural network cascade for feature exctraction and representation. The rise of the technology wave around 2010 changed the way of solving problems in different branches. Instead of piecing together individual functional modules to form a system, the focus is to optimize parameters by global training after the whole system is designed, what is called end-to-end training.

2.1.1 SR Resnet

2.2 Multi-Image Super Resolution

Multi-Image Super-Resolution (MISR) is the task of yielding HR images by fusing multiple LR observations of the same scene, which allows the achievement of higher reconstruction accuracy than relying on only one image. The development of this approach progressed at a slower pace due to the extesive pre-processing requirements imposed on the input, as this algorithms have a high sensibility to the input variability and their proper co-registration.

When the input images are of the same nature, but taken at different points in the temporal dimension, the problem is often called multi-image super resolution. On the other hand, when the images are taken at the same time but they come from different sensors and show different spectral bands, it is called multi-spectral super resolution, which will be further discussed.

In 2019, the European Space Agency (ESA) organized an SR challenge [5] based on real-world scenes acquired by the PROBA-V satellite, each of which contains an HR image (100m GSD) coupled with at least nine LR images that are not perfectly co-registered and they may be taken months apart. This challenge, with a not synthetically generated HR-LR image pairs, fostered a new generation of model architectures that are able to fuse the multiple LR images to create better reconstructions.

2.2.1 Multi-spectral super resolution

Also Referred to as "hyper-spectral super resolution" in the literature, The term "Multi-Spectral" emphasizes the use of multiple spectral bands, in contrast with the multi-image approach detailed previously. While the concept bears similarities to MISR, the key distinction lies in MSSR's use of a single scene captured with different spectral bands, as opposed to multiple images, to reconstruct a superior, super-resolved image.

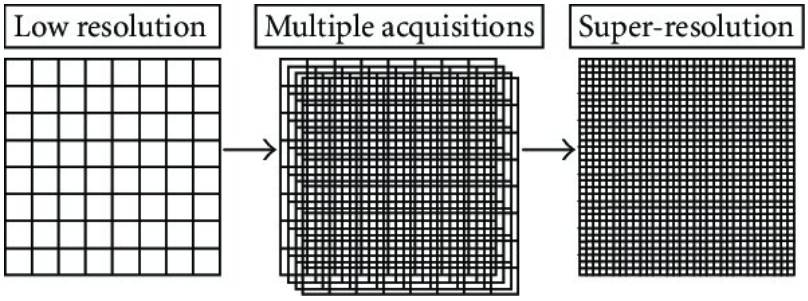

Figure 2.2: Multi-image super resolution algorithms combine multiple low-resolution image acquisitions into a high-resolution image. Source: [6]

In the context of MSSR, each spectral band, corresponding to a specific wavelenght range, provides unique information about the observed scene. Some of the spectral bands yield better resolution because of their physical properties and the costs related to their sensors. Using this higher resolution bands to increase the detail in the lower bands seems like a resaonable approach.

Traditional pan-sharpening algorithms could be considered as deterministic MSSR algorithms. It is usually used to increase the resolution of a multi-spectral RGB image using the panchromatic band. The overlap between the wavelengths of both wavelength ranges makes this algorithm straightforward. However, it is ill-suited for Thermal Infrared (TIR) data due to the disjointed spectral domains of the visible and TIR bands. In [7], A deep learning is trained assuming the presence of common information between low-resolution LWIR images and their higher resolution RGB counterparts, with the objective of creating a super-resolved product in the LWIR band by an effective fusion. This improved image retains the essential thermal information, while simultaneously incorporated enhanced spatial resolution details captured from the visible bands.

2.2.2 Importance of interframe correlation

2.2.3 RAMS

2.3 The domain gap problem

SR is a supervised problem, the super resolved image is compared to the HR ground truth and the pixel-by-pixel differences drive the gradients of the neural network to minimize the loss, in a fully supervised manner. Additionally, deep learning based SR methods are known to consume large quantities of training data. Most of the research in the field of SR is conducted by artificially producing HR-LR pairs by downscaling the HR images with known kernels, such as bicubic. However, this is rarely the case when using "non-ideal", real world images. In spite of their success on synthetic datasets, the poor generalization capacity of the trained SR networks limits their application in real scenarios, leading to blurry images and strange artifacts in the SR results [8].

The domain gap problem is the difference between the training data and the real-world data. The training data is usually generated synthetically, using bicubic downsampling, which is a very simple degradation model. However, real-world degradations are usually too complex to be modelled with an explicit combination of multiple degra-

dation types, as shown in Fig.3(c). Therefore, implicit modelling attempts to circumvent the explicit modelling function. Instead, it defines the degradation process f implicitly through data distribution, and all the existing approaches with implicit modelling require an external dataset for training. Typically, these methods utilize data distribution learning with Generative Adversarial Network (GAN) [16] to grasp the implicit degradation model possessed within training dataset, like CinCGAN [8]. Great attempts have been made in the last several years on the generation of real pairs of LR-HR images, but the process remains costly. Additionally, SR networks trained on the collected datasets are not able to generalize to images captured in other conditions.

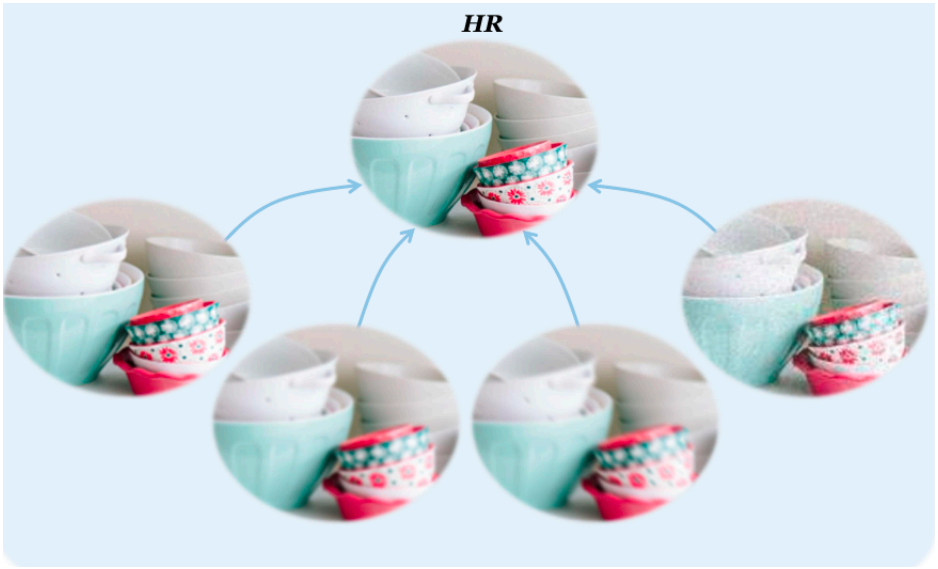

Figure 2.3: Effects of different degradation models on one HR image. Source: [9]

2.4 Blind image Super Resolution

The problem of SR with an unknown degradation process is known as blind SR. Growing attention has been paid to blind SR in recent years, towards filling the domain gap presented in 2.3. A schematic diagram of the problem is shown in Fig. 2.4. Non-blind SR assume that the degradation process is known, and maps the bicubic downsampled LR image to the natural HR image space. However, an arbitrary LR input image, as a scene captured by a satellite, is usually degraded by an unknown process, which is difficult to be modelled explicitly. The arbitrary LR input is not in the same domain as the bicubic downsamopled LR image, and thus the non-blind SR methods are not successful, moving to a different place in the HR space than the natural images. Blind SR methods, on the other hand, aim to learn the degradation process from the training data, and map the arbitrary LR input image to the natural HR image space.

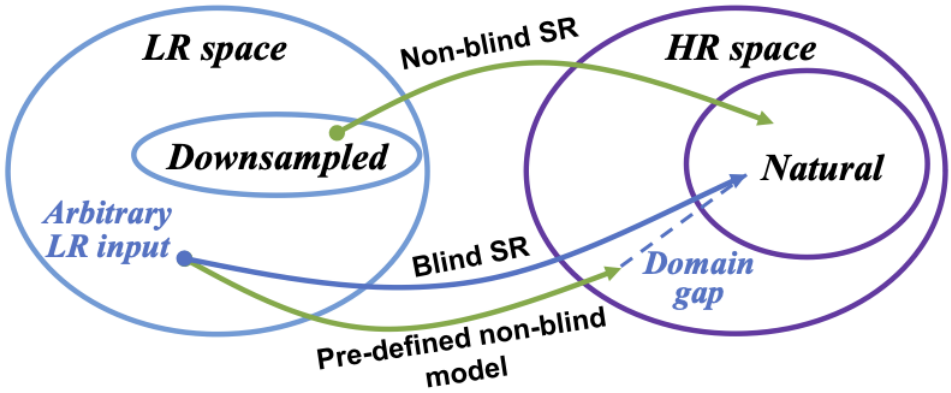

Figure 2.4: Domain interpretation of differences between non-blind and blind SR. Source: [9]

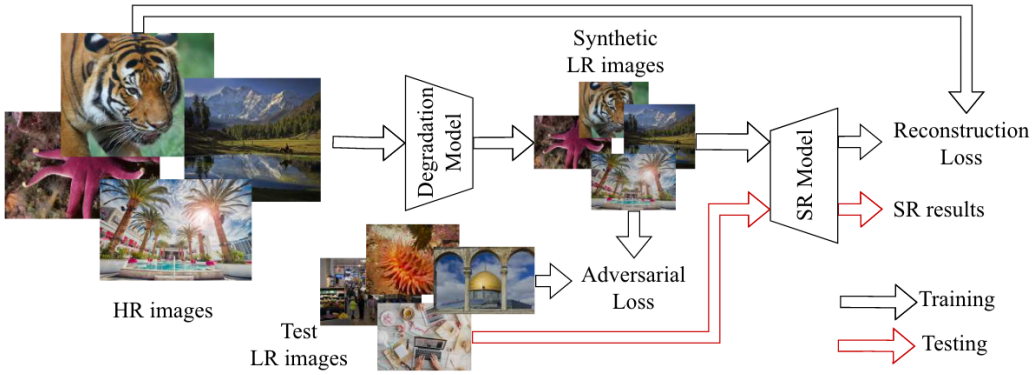

Figure 2.5: In Degradation-learning-based methods, adversarial training is used to encourage the degradation model to produce images in the same domain with the target domain (test images). Source [10].

3 Methodology

3.1 Baseline Degradation model

Early super-resolution methods commonly generated high-resolution (HR) to low-resolution (LR) samples using predefined degradation techniques, with bicubic downsampling being the most used setting [11]. This kind of synthetic data, while easy to obtain, often results in a domain gap problem, where the data used for training and assessing the model do not come from the same distribution as real data. This gap usually to performance drops when the models implemented in production environments. A possible solution is to synthesize samples with a stochastic degradation model, which includes a set of multiple blurring kernels and several random noises configurations. The larger degradation space grant these models better generalization capabilities and experts be part of the kernel definition process, based on prior knowledge of the degradation process. Unfortunately, the variety of predefined degradation's is still limited and still fail in most applications.

A degradation model like this one will be used as a baseline for this work.

3.1.1 Blurring Kernel

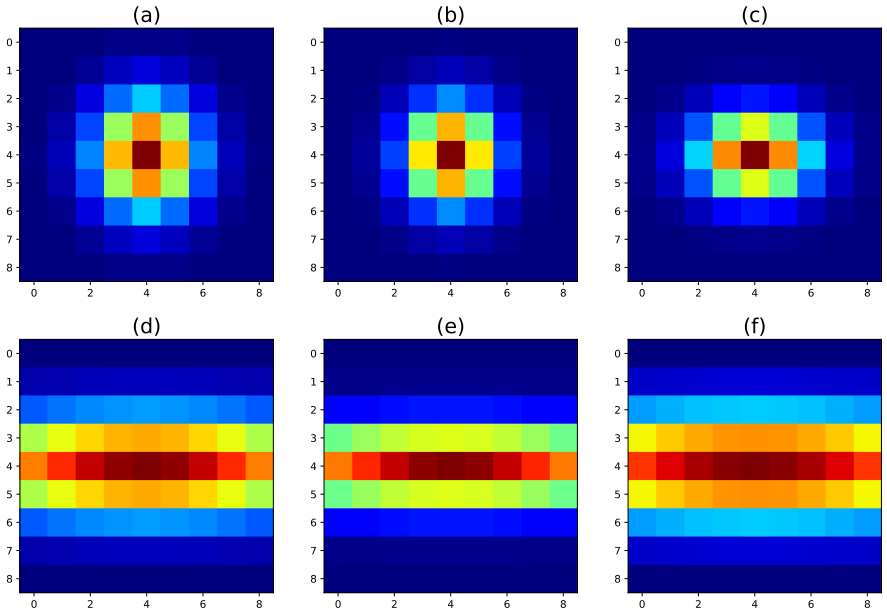

Figure 3.1: Example of kernels used in a stochastic degradation model. (a),(b) and (c) are generated using a symmetric variance on the x and y axis. (d) (e) and (f) are generated using an asymmetric variances, resulting in much more anisotropic kernels.

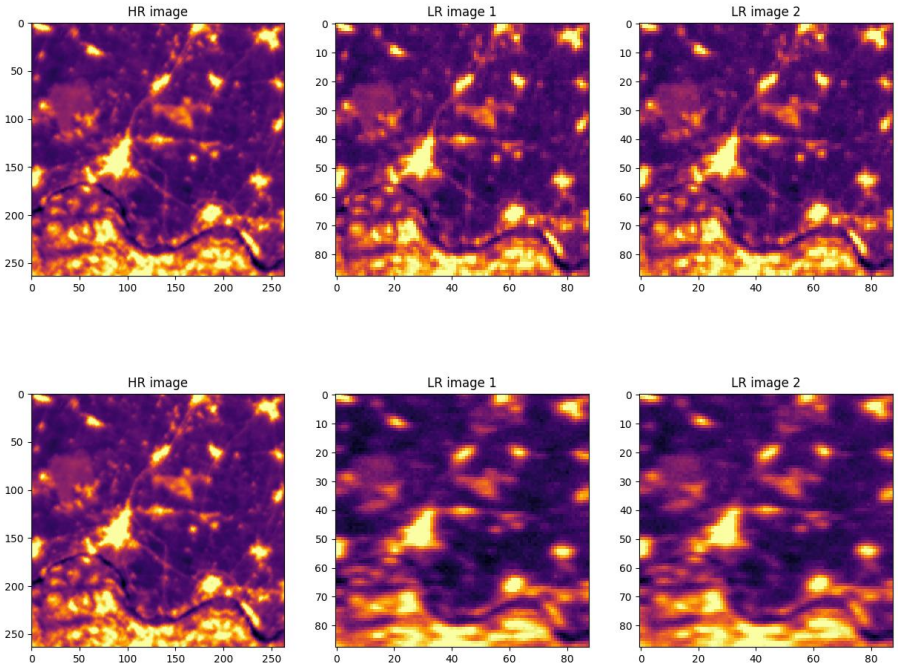

Figure 3.2: Effects of different blurring kernels on the HR-LR generation. The upper row contains images generated using blurring kernels with symmetric distributions. The lower rows contains images generated using asymmetric distributions for the variances, resulting in highly anisotropic kernels.

3.1.2 Radiometric error correction

As reported by the ECOSTRESS instrument sheet,[12] its nominal radiometric accuracy at 300K is 0.5K. FOREST-2 target radiometric accuracy is 1K. This difference in accuracy should be taken into account. To align these accuracies, we first calculate the additional error required using the following equation:

$$e_{\text{forest}} = \sqrt{e_{\text{eco}}^2 + e_{\text{extra}}^2} \tag{2}$$

where e_{eco} is the ECOSTRESS error, and e_{extra} is the additional error required for FOREST-2.

Using the above equation, we find that an additional radiometric error of approximately 0.8660K is needed. The next step involves converting this extra error into a radiance value. This requires calculating the derivative of the Planck equation at 300K, which is done numerically as follows:

$$\frac{\partial B}{\partial T} = \frac{B(\lambda, T + \delta T) - B(\lambda, T)}{\delta T} \tag{3}$$

By multiplying the results of equations 2 and 3, we can obtain the radiance error for both FOREST LWIR bands. The additional radiance errors for LWIR1 and LWIR2 bands are found to be 1.5472×10^{-1} W/sr/m^2/μm and 1.1444×10^{-1} W/sr/m^2/μm, respectively.

The difference in radiances will be split into two components. On one side, the cold Bias represents a systematic error in the measurement, this error acknowledges discrepancies that can be attributed to sensor calibration and atmospheric conditions. On the other side, the random noise accounts for unpredictable fluctuations in the measurement process. It could be due to a variety of sources like electronic noise in the sensor, random atmospheric disturbances, or other stochastic factors. As the extent of each component is not known and to give more variability to this basic degradation model, a random factor $\phi \in [0, 1]$ is introduced so that:

$$\begin{aligned} \varepsilon_{\text{final}} &= (1 - \phi) \times \varepsilon_{\text{radiance}} + \phi \times \eta \times \varepsilon_{\text{radiance}} \\ \eta &\sim \mathcal{N}(0, 1) \end{aligned} \tag{4}$$

The effects of the error correction is shown in Fig. 3.3. As the target radiometric error increases with respect to ECOSTRESS scenes, the loss of information is more noticeable.

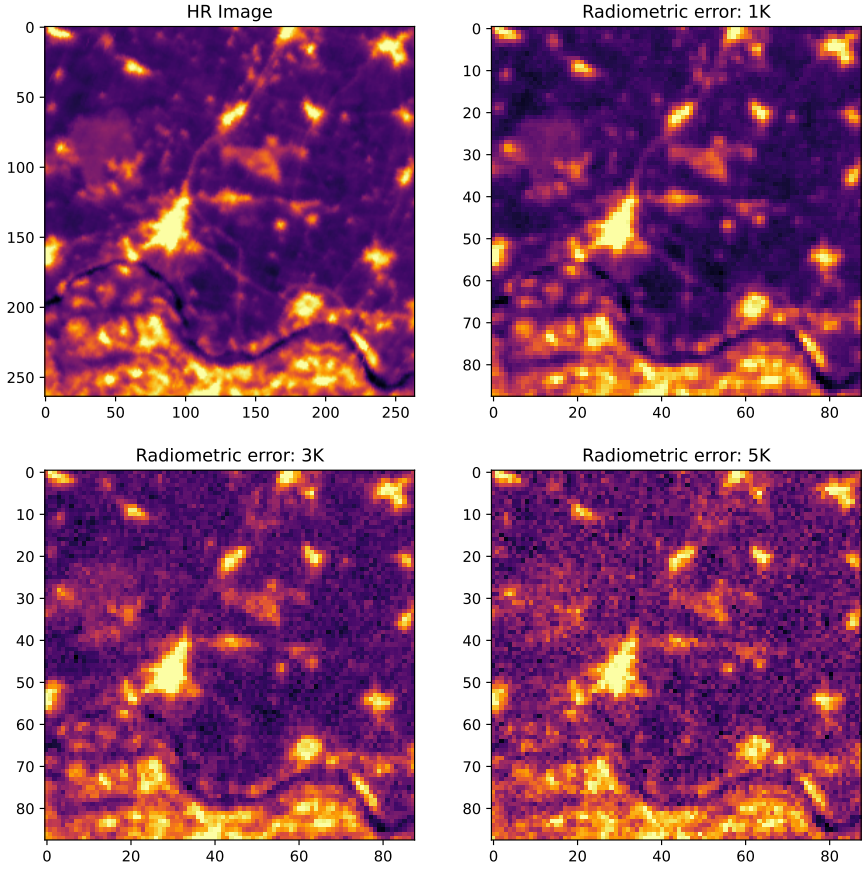

Figure 3.3: Effects of different radiometric error corrections on the HR-LR generation.

3.2 Models Architecture

3.2.1 SRResNet

Introduced in 2017 [13], SRResnet leverages on residual networks [14] that employ skip connections to solve the super resolution task. The architecture is detailed in Fig. **CITE**. Specifically, 16 residual blocks consisting two convolutional layers, followed by batch-normalization layers and ParametricReLU activation functions. The convolutional layers have 3x3 kernels and 64 feature maps. To increase the resolution of the input image, two trained sub-pixel convolution layers are used.

As this work focuses on having super resolved images with high physical consistency and not on the perceptual superiority of the images, improvements introduced in the publications like the Generative Adversarial Network (SRGAN) and the perceptual loss for gradient calculation are not used.

3.2.2 RAMS

3.2.3 Probabilistic Degradation Model

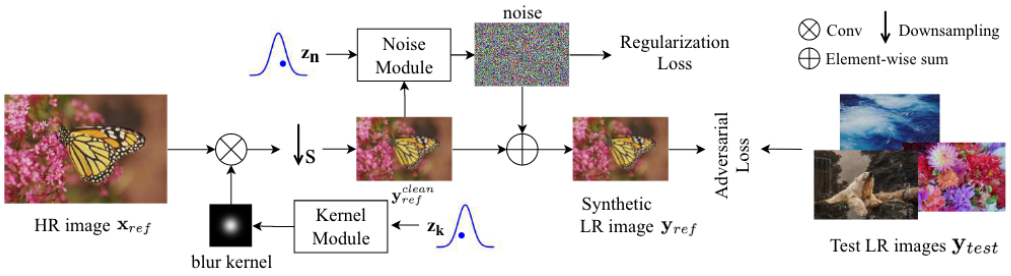

Figure 3.4: SADASDSA

Kernel Model

Noise Model

Discriminator

3.3 Referenced image quality metrics

When the ground truth high resolution image is available, the performance of a super-resolution algorithm can be evaluated using a variety of metrics. These metrics can be divided into two categories: pixel-based and perceptual-based. Pixel-based metrics are based on the pixel-wise comparison between the generated image and the ground truth. Perceptual-based metrics, on the other hand, are based on the perceptual similarity between the generated image and the ground truth. These metrics are build using a pre-trained deep neural network, which is usually trained on a large dataset of images. The following sections will describe the most commonly used metrics in the literature.

3.3.1 L_2 and L_1 Loss

The L_1 and L_2 losses are the most commonly used pixel-based metrics in the literature. Additionally, they are usually used as the loss function that drives the network gradients during training. In a general form, the L_1 and L_2 losses are defined as follows:

$$\mathcal{L}_{L_k} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|^k \tag{5}$$

Where y_i and $\hat{y}_i$ are the ground truth and the super resolved image, respectively, and k is the exponent of the loss function. The L_2 loss weights high-value differences higher than low-value differences due to the exponent of 2. This generates overly smooth for low values and a lot of variability in high values. For that reason, it is more common to see the L_1 loss being used in the literature.

3.3.2 Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) measures the quality of a reconstructed or super-resolved image in comparison to the original high-resolution image. It quantifies the amount of error or noise introduced during the image reconstruction process.

PSNR is calculated by first computing the Mean Squared Error (MSE) or $\mathcal{L}_2$ loss between the original and reconstructed images, and then taking the logarithmic ratio of the maximum possible pixel value squared. The PSNR value is usually expressed in decibels (dB).

The formula for PSNR is:

$$PSNR = 10 \cdot \log_{10}\left(\frac{I_{MAX}^2}{\mathcal{L}_2}\right) \tag{6}$$

where I_{MAX} is the maximum possible pixel value of the image, and $\mathcal{L}_2$ is the mean squared error between the original and the reconstructed image.

A higher PSNR value indicates better quality of the super-resolved image, as it signifies a lower level of noise or error. However, it's worth noting that it may not always align with human perceptual evaluations of image quality, as it focuses on physical consistency.

3.3.3 Structural Similarity Index (SSIM)

Structural Similarity Index Measure (SSIM) takes into consideration changes in structural information, luminance, and contrast. By doing that, it manages to reflect better the perceived changes in noise level and contrast. The SSIM index is calculated by dividing the image into windows of a certain size, and then comparing corresponding windows in the reference and target images. The SSIM index for a pair of windows, say x and y, is given by:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

where μ_x and μ_y are the average pixel values, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance of x and y, and c_1, c_2 are small constants to avoid division by zero. The final

SSIM score for the images is calculated by averaging the SSIM indices of all windows. An SSIM score of 1 indicates a perfect structural match between the two images, whereas a score of 0 indicates no structural similarities.

3.3.4 Learned perceptual image patch similarity (LPIPS)

LPIPS is a perceptual metric that leverages deep learning to compute perceptual differences between images. Specifically, it uses the activations of a pre-trained convolutional neural network (in this case, VGG [15]) to extract perceptual features from the images. Then, it calculates the Euclidean distance between these feature vectors to measure the perceptual difference. This measure has gained popularity in SR tasks due to its high correlation with human judgments of visual similarity.

The LPIPS score is given by:

$$LPIPS(I, I') = \sqrt{\sum_{i=1}^{N} w_i \| f_i(I) - f_i(I') \|^2} \tag{8}$$

where I and I' are the images being compared, $f_i(I)$ denotes the i-th layer activation when image I is input to the pre-trained network, N is the number of layers considered, and w_i is the learned weight for the i-th layer.

A lower LPIPS score indicates a lower distance between the feature vectors, and thus a greater perceptual similarity between the two images. Due to the fact that in this work we are interested in the physical consistency of the super-resolved images, this metric will be shown but will not drive any decision during the training process.

3.3.5 Adjusting measures to a multi-image framework

In order to calculate the losses and performance metrics, the generated test images (SR) are compared against the ground truth high resolution images (HR). Additional changes should be introduced in a MISR environment [5]. First, minor shifts on the contents of the pixels are expected and the metrics should have some tolerance to small pixel-translations in the high-resolution space by evaluating on a sliding cropped image. That means, looking for a displacement of SR by at most d pixels in each direction that minimizes the error. An example of how this is applied in a loss that needs to be minimized can be found in Eq. 9

$$\mathcal{L}^*(I^{HR}, I^{LR}, d) = \min_{u,v \in [0,2d]} \mathcal{L}(I^{HR}_{u,v}, I^{SR}_{u,v}) \tag{9}$$

Additionally, commonly used metrics punish biases as much as noise in the reconstruction. For example, if $I^{SR} = I^{HR} + \epsilon$, where ϵ is a constant bias, a perfect reconstruction of I^{SR} is possible if ϵ is known. A quality metric should award a high score in super-resolutions with this characteristics in comparison to the introduction of noise and information loss. Metrics like L2/L1 losses and PSNR do the exact opposite and should have a bias compensation like the following:

$$\mathcal{L}^*(I^{HR}, I^{LR}, d) = \min_{u,v\in[0,2d]} \mathcal{L}(I^{HR}_{u,v}, (I^{SR}_{u,v} + b))$$
$$b = \frac{1}{(W-d)(H-d)} \sum_{x,y} (I^{HR}_{u,v} - I^{SR}_{u,v}) \tag{10}$$

where W and H represent the width and height of the image, respectively.

3.4 Non-referenced Image quality metrics

No-Reference Image Quality Assessment (NR-IQA) aims to develop methods to measure image quality in alignment with human perception without the need for a high-quality reference image. Most of them are based on two steps: feature extraction and quality prediction using a regression module. They rely on the assumption that natural images share certain statistical information and that any distorsion may alter these statistics [16]. The results from any image an arbitrary image is compared to a default model trained on a large dataset of natural scenes. The difference between them is used to predict the quality of the image. In the last years, researchers relied on deep learning to perform the two steps in a single model. The workflow of these models is shown in Fig. 3.5.

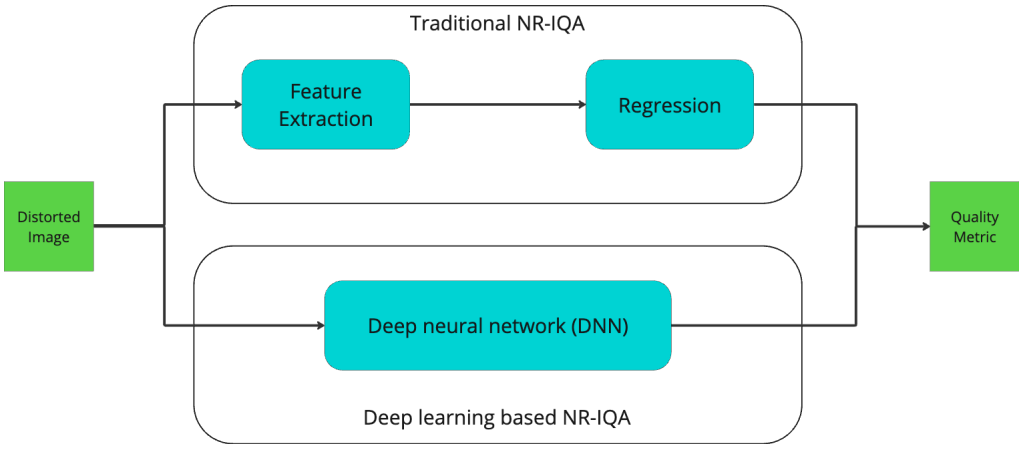

Figure 3.5: Workflow of a NR-IQA model.

3.4.1 Naturalness Image Quality Evaluator (NIQE)

The Naturalness Image Quality Evaluator (NIQE) [16] is a no-reference image quality assessment metric that quantifies the perceptual quality of images based on their naturalness. NIQE operates on the principle that pristine natural images exhibit specific statistical properties that can be quantified to establish a benchmark for quality assessment. NIQE employs a model based on a multivariate Gaussian distribution, characterized by a mean vector and covariance matrix, to represent the statistical attributes of a natural image's visual patterns. To assess the quality of an image, NIQE extracts a corresponding set of features and evaluates their deviation from this statistical model using the Mahalanobis distance. This distance measures the divergence of the image's features from those typical of high-quality natural images. A lower value suggests that

the image closely resembles the statistical properties of natural images, indicating higher perceived quality.

However, NIQE provides an objective measure of image quality that aligns with the naturalness of human visual perception, and is not able to quantify the physical consistency of a generated image.

3.4.2 Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)

3.4.3 Frequency Domain Analysis

The Fourier transform is widely used to analyze the frequency content in signals. It can be applied to multidimensional signals such as images, where the spatial variations of pixel-intensities have a unique representation in the frequency domain. Super-resolutions objective is to reconstruct missing high frequency components from a downscaled image. The expectation of a good SR algorithm is to amplify the high frequency components compared to a baseline like bicubic interpolation, while keeping noise at bay. The Fourier components provide global information about the image, as opposed to local information represented by pixel values in the spatial domain [17]. Using the Fast Fourier Transform (FFT), we convert the pixel intensity values of super-resolved images into a spectrum where each point represents a specific frequency contained in the spatial domain. The FFT is shifted so that the zero-frequency component is at the center of the spectrum. The resulting magnitude, after applying a logarithmic transformation, reveals the energy distribution across various frequencies. This is visualized in grayscale, where the intensity corresponds to the amplitude of the frequency components.

A radial profile of the FFT magnitude provides insights into how different spatial frequencies contribute to the image content in the vertical and horizontal direction. The radial profile is a function of the average intensity of frequencies at a given radius from the center of the Fourier transform. The average of the FFT magnitude is calculated for concentric circles of increasing radii, capturing a statistic of the frequency components in every direction. This metric serves as a benchmark for evaluating the performance of SR techniques against traditional interpolation methods such as bicubic interpolation.

Spatial frequency within an image context refers to the periodicity of the intensity variation over spatial dimensions, typically quantified in cycles per pixel. The central region of the frequency domain, after the shift operation, denotes the zero frequency. In contrast, the extremities of the domain delineate the highest frequencies, constrained by the image's discrete sampling rate. To quantitatively interpret these spatial frequencies, a radial-to-frequency mapping is necessary. This mapping accounts for the Nyquist frequency, which is delineated as half the sampling rate of the discrete imaging grid and acts as a threshold to prevent frequency aliasing. The conversion from a given radius in the FFT output to the corresponding spatial frequency is formalized as:

$$f(r) = \frac{r}{\frac{N}{2}} \cdot f_{\text{Nyquist}}, \tag{11}$$

where $f(r)$ signifies the spatial frequency associated with radius r, N represents the FFT image dimension, assuming a square configuration, and f_{Nyquist} the Nyquist frequency, which is 0.5 cycles per pixel in this case.

Through FFT we acquire a depiction of the frequency-based amplification or attenuation attributable to the SR techniques. Analyzing these profiles displays the ability of SR models for detail enhancement. However, it is important to note that this method does not account for any noise or artifacts generated by the SR, and should be used in combination to other supervised metrics.

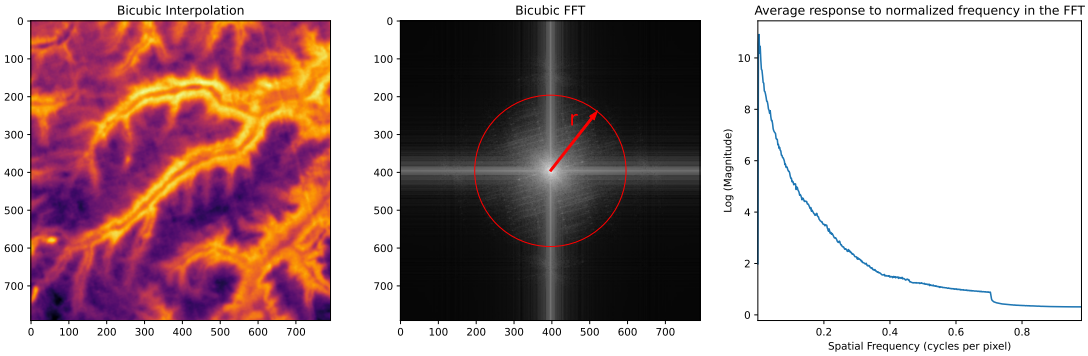

Figure 3.6: Steps of the frequency domain analysis. The Center image shows the log magnitude of the shifted FFT of a bicubic upsampled FOREST scene and an example of a radial profile, the average of all the points that have the same r is calculated. The right image displays the log magnitude obtained for every radial profile, translated into spatial frequency.

3.4.4 Gradient Distribution analysis

An alternative way of analyzing super-resolution results is by looking at the gradients of the images. HR images are sharper and thus each pixel, on average, has higher gradients magnitude with respect to both directions than their LR counterparts. A super-resolution algorithm should increase the sharpness of the edges, resulting in a gradient distribution that aligns more closely with that of the genuine HR image. An approximation of the gradients can be estimated by doing 2d convolutions between an image and the so called Sobel kernels displayed in Eq. 12 [18]. These kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid.

$$\hat{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \qquad \hat{G}_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{12}$$

The kernels can be applied separately to the input image to produce the component of the gradient in each orientation G_x and G_y. The magnitude of the gradient is given by:

$$|G| = \sqrt{G_x^2 + G_y^2} \tag{13}$$

The gradient magnitude histograms of the results of different super-resolution algorithms will be assessed, there by quantifying the enhancement in edge sharpness. This histogram provide insights into the frequency and intensity of the edges within an image. A better SR model should demonstrate a histogram with higher frequencies of larger

gradient magnitudes, indicating sharper edges. However, it is important to note that this analysis is unsupervised and disregards the effect of noise and artifacts introduced during the super-resolution process and should be considered in combination with other supervised metrics like PSNR.

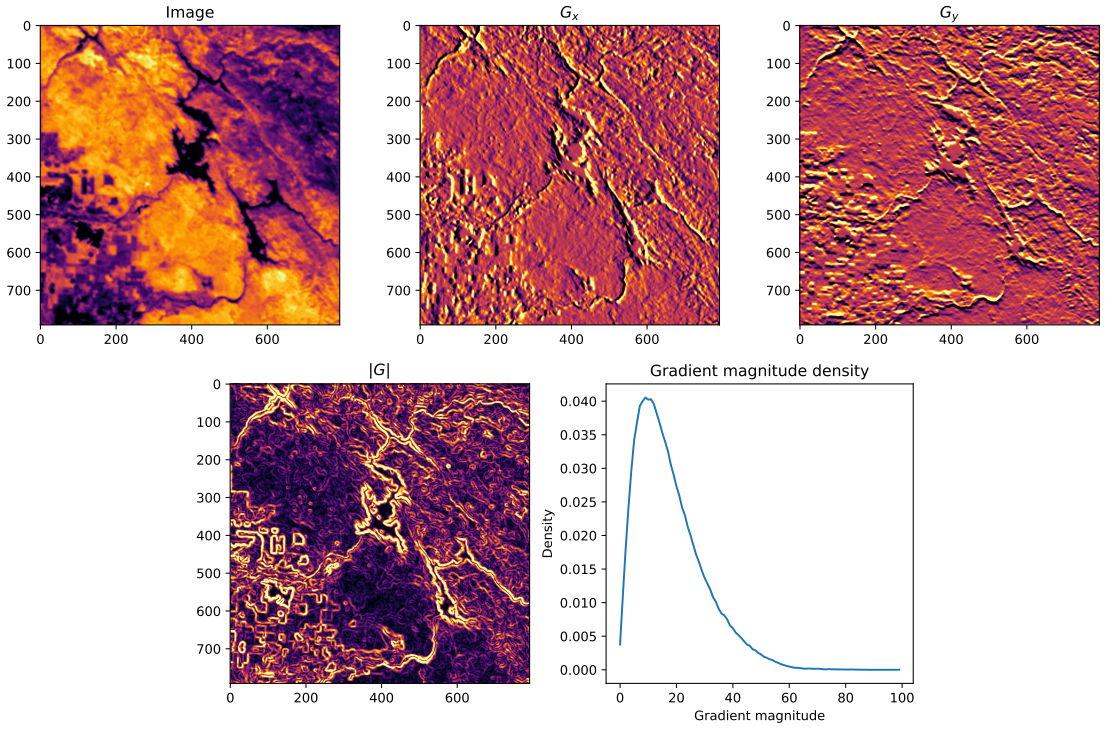

Figure 3.7: Steps to obtain a gradient magnitude density. Using the sobel operators, G_x and G_y are obtained from an image. The magnitude $|G|$ of each pixel is calculated using Eq. 13. The density can be estimated afterwards, using 100 bins in this case.

4 Experiment Setup

4.1 Obtaining a high resolution dataset

Super-resolution is inherently a supervised learning task that needs the availability of high-resolution (HR) data. In scenarios where HR data from sources like FOREST is unavailable, an alternative is to generate synthetic images that replicate the characteristics of a superior resolution FOREST dataset.

4.1.1 The ECOSTRESS mission

The NASA's ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission is designed to provide new insights the effects of the Earth's climate dynamics [19], with focus on the following scientific objectives:

1. Identify the critical thresholds of water use and water stress in key climate-sensitive biomes, typically by observing the transition zones between biomes.
2. Identify when plants stop taking up water over the course of a day.
3. Improve the accuracy of drought estimates based on agricultural water use in the continental United States.

ECOSTRESS employs thermal infrared radiometers, specifically Prototype HyspIRI Thermal Infrared Radiometer [20] to measure the radiation emitted from the Earth's surface. It provides a spatial resolution of 69 meters with a temperature sensitivity of a few tenths of a degree [19]. The swath size is 400x400 km. The detector separates the energy from five different wavelengths using filters attached to the detector, producing five separate image layers for each scene. The pixels represent the intensity of thermal infrared radiation emitted by the Earth's surface at each wavelength. The mission has a 4-day diurnal repeat cycle.

In the spatial domain, ECOSTRESS constitutes an excellent candidate for generating synthetic HR images, as it's resolution constitutes approximately a x3 increase compared to FOREST. In the spectral domain, it is important to confirm overlap between the missions bands. Given the narrower ECOSTRESS bands compared FOREST's, the strategy will be averaging the radiances to align the spectral properties. Fig. 4.1 shows this spectral band comparison. In the case of the LWIR1 FOREST band, the overlap is significant with the first three ECOSTRESS bands. Althouth the overlap is less pronounced in the LWIR2 band, the radiation spectrum of black-bodies at prevalent surface temperatures suggest the feasibility of constructing a synthetic LWIR2 from the last two ECOSTRESS bands. While FOREST's temporal resolution exceeds that of ECOSTRESS, allowing for the monitoring of new processes, this aspect is not the primary focus of the current study.

4.1.2 Donwloading ECOSTRESS Scenes

ECOSTRESS imagery is available via NASA's Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) [21]. This tool allows the request of area samples

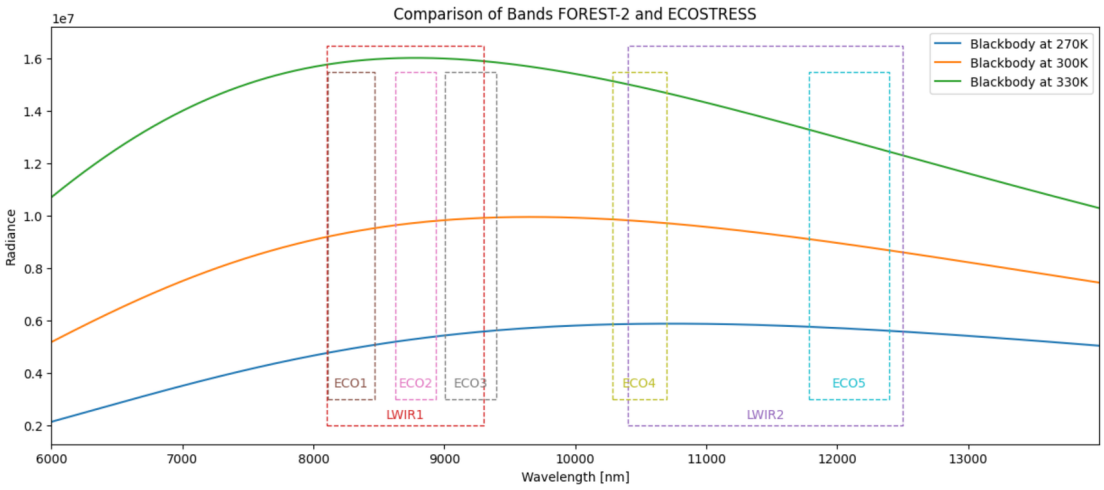

Figure 4.1: Wavelengths of the sensors in Ecostress and Forest satellites. The radiation spectrum of black-bodies at different temperatures are included for comparison.

via vector poligons. Using the product's API [22], Level 1 Mapped Radiance scenes of size 200x200 km with center on the locations provided in Fig. 4.1 were programmatically requested. Due to satellite hardware anomalies, certain spectral bands experienced acquisition gaps, needing a careful selection of date ranges to ensure the availability of all five bands [23].

Area	200 x 200 km
Products	Mapped Radiance (5 bands) Quality (5 Bands)
Dates	2018/08/20 - 2019/03/04 2023/05/01 - 2023/08/15

Table 4.1: Requests configuration

4.1.3 Selecting the best scenes

The AppEEARS platform returns multiple scenes that correspond to the specified area sample within the requested timeframe. This includes 5 mapped radiance measurements alongside their corresponding Quality Assurance (QA) bands. Additionally, a CSV file is provided, detailing quality statistics for each scene. The interface returns any scene that overlaps with the requested area. For that reason, some GeoTIFFs may be significantly smaller than others, with variances up to 90%. Moreover, an important number of these GeoTIFFs may contain a high percentage of bad quality pixels, rendering them unsuitable for model training. Furthermore, as highlighted in the ECOSTRESS frequently asked questions [24], the accuracy of radiance measurements is highly dependent on clear sky conditions; cloudy scenes typically yield negligible radiance emissions.

The dataset includes several GeoTIFFs for each scene. Downloading the entirety of this dataset is impractical due to its huge size. From the 50 scenes, each one is potentially replicated over 20 times over the 10 months request window. Such a dataset, given its

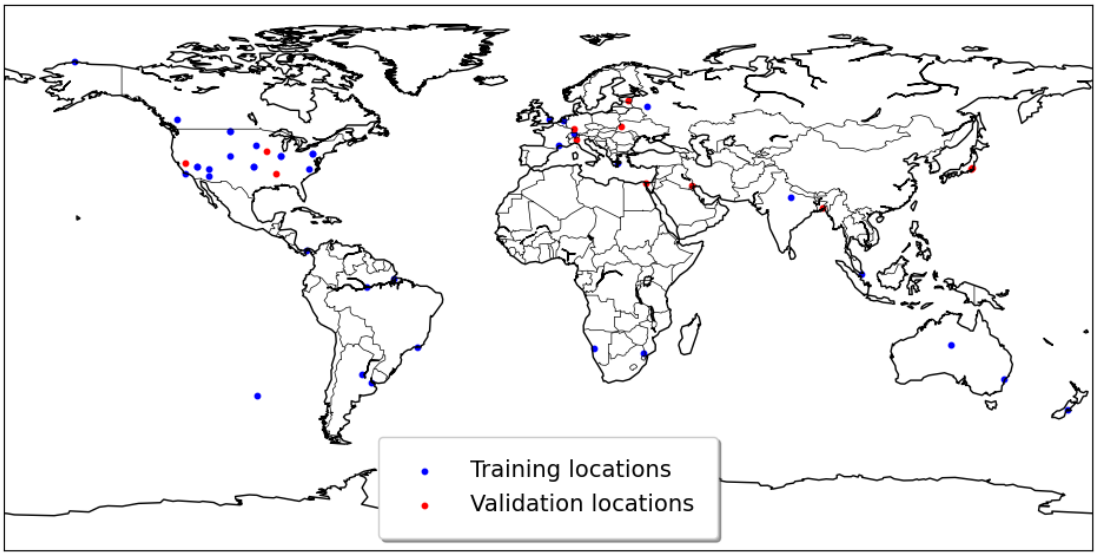

Figure 4.2: Location of the samples taken from ecostress.

magnitude, cannot be used for model training with the available hardware resources. Therefore, a procedure is developed to identify and select the most appropriate scene for each month, based on a predefined set of criteria:

1. Scenes should have a low proportion of bad quality pixels.
2. Scenes should have a considerable size so that many crops can be taken from it.
3. As clouds imply low radiance values, clear sky scenes will have high radiance values.

The procedure to get the best scene for each month is detailed below:

Algorithm 1 Process applied to the scenes returned from one area request.

QA statistics:
Get the average proportion of good pixels p_{gp} for the 5 radiances of the scene.
Discard scenes where $p_{gp} < 60$.
For each month, get the 3 scenes with the greatest p_{gp}.
Scene Statistics:
Get the biggest scene of each month.
Calculate the proportion between the size each scene and the biggest of the month.
Discard images which size proportion is smaller than 0.2.
Calculate the median of the radiance values of the scene.
Selecting the scene of the month:
Merge the QA statistics and the Scene statistics.
Select the scene that has the greatest median radiance value.

Applying this procedure, a dataset comprised of 5031 scenes taken from 50 area requests is reduced to 379 scenes.

4.1.4 Data Processing

In order to be able to use the data in a super-resolution algorithm, a set of processing steps must be performed on it.

The diagram in Fig. 4.3 displays the processing pipeline. The input are the 5 Mapped radiance and their respective quality bands.

Mapped radiances 1,2 and 3 are averaged to form the LWIR1 synthetic FOREST, mapped radiances 4 and 5 are averaged to form the LWIR2 synthetic FOREST. If any of the bands are missing, the corresponding LWIR synthetic forest is discarded.

The fill values in the mapped radiances and the data quality classes are used to create a binary mask for each spectral band. If a pixel is considered problematic, it is marked as a 1 in the binary mask. The QA band for a synthetic FOREST LWIR band is built using an OR operation on the corresponding ECOSTRESS spectral involved in it's construction. After being constructed, both the synthetic LWIR and the corresponding QA band are reprojected to the best utm epsg code, based on the latitude and longitude of the scene.

Value	Description
	Fill Value Classes
-9997	Pixel not seen
-9998	Missing data due to striping (not filled in)
-9999	Missing/bad data
	Data Quality Classes
0	Good
1	Missing stripe data, filled in
2	Missing stripe data, not filled in
3	Missing/bad data
4	Not seen

Table 4.2: Fill Value and Data Quality Classes

The synthetic LWIR are not suitable for the super-resolution task yet. They are too big to be kept in memory, and not all their values are of good quality. For that reason, for each scene, a number of random crops of size 264x264 pixels are taken. The random crop processor pipeline is displayed in Fig. 4.4. It is an iterative process where at each stage, crops that do not comply with the quality considerations (all pixels are of good quality and no stripe noise was detected) are discarded until the target number of crops per scene is achieved. Additionally, the Affine Transformation is translated so that the images can be georeferenced.

4.1.5 Obtaining FOREST-2 data

4.2 Datasets

For a better understanding of how the proposed architecture works, several datasets combinations are used. The implemented pytorch dataset class loads and yields samples

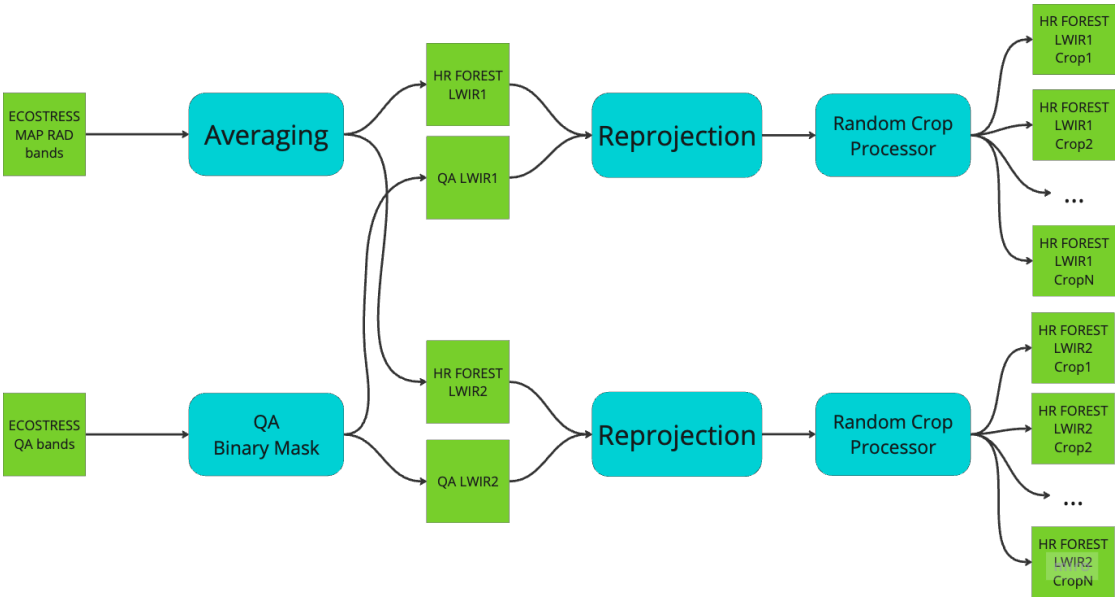

Figure 4.3: Data processing workflow

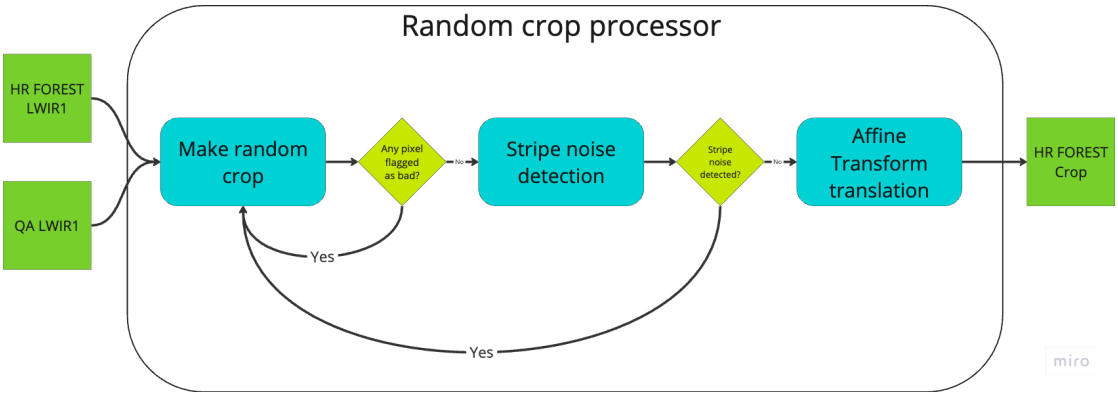

Figure 4.4: Random crop processor

from two different folders, one for the HR images (source domain) and one for the LR images (target domain). The HR images are the synthetic FOREST images produced from ECOSTRESS, while the LR images come from different souces that will be shown below. The samples are usually unpaired (that means, they are completely different scenes), but the dataset allows for paired datasets. In case any of the domains has less samples than the other, the class will bootstrap the smaller domain to match the size of the bigger one.

4.2.1 Training: Ecostress-DegradedEcostress

The dataset, $\mathcal{D}_{\text{eco}-\text{eco}}$ is built by taking the HR synthetic FOREST crops and applying the baseline degradation model proposed in 3.1. The 264x264 crops are reduced to 88x88. The training set is used to train the SR Resnet model, while the validation set is used to monitor the training process and avoid overfitting. Even though in this case the HR and LR version of the same scene is available, the training dataset is unpaired

by shuffling the samples. The validation set is not shuffled, and thus can be used to calculate supervised metrics like PSNR and SSIM.

4.2.2 Training: Ecostress-Forest (Unpaired)

The dataset $\mathcal{D}_{\text{eco}-\text{forest}}$ is composed of the 264x264 HR synthetic FOREST crops as the source domain and 88x88 LR FOREST crops as the target domain. Unfortunately, this dataset is not paired, as the HR and LR images are completely different scenes. Thus, no supervised metrics can be calculated on it. The metric used to determine the best model is the PSNR from the super resolution of the artificially degraded HR images.

4.2.3 Testing: Ecostress-Forest (Paired)

While the training dataset is the same as in the previous case, the validation dataset is composed of a limited amount paired scenes between ECOSTRESS and FOREST are available. This samples allow the calculation of supervised metrics like PSNR and SSIM.

5 Results and discussion

To serve as a baseline, a pipeline was trained using synthetic FOREST as the source domain and their artificially degraded version as the target domain. To obtain such data, the degradation model described in 3.1 was applied, blurring and adding noise to the synthetic FOREST images. This degradation model is stochastic, but with unknown parameters. The objective is to observe how the pipeline is able to mimic the degradation model used in the target domain.On the other side, the main pipeline was trained using synthetic FOREST as the source domain and real FOREST as the target domain. In this case, the degradation model is unknown and the objective is to observe how the pipeline is able to estimate it, generating LR versions of the synthetic FOREST images that come from the same distribution as the real FOREST images.

5.1 Source domain

This subsection will show the results from the experiments performed on the source domain. The process consists of degrading the synthetic HR FOREST images using the generator trained using adversarial learning and then super resolving it using the corresponding SR model used in the pipeline. This is the equivalent of the black arrows flow describes in fig. 2.5.

As in this case the ground truth is known, the performance of the SR model can be evaluated using metrics like PSNR and SSIM. Unfortunately, as the target and the source domain are not paired, the effects of the degradation model have to be evaluated using alternative methods.

Fig. 5.1 shows the results of both pipelines when applied to one sample from the source domain. For comparison, a simple gaussian blurring + downscaling is also shown. The main differences can be seen on the degraded LR images. While the baseline pipeline produces images very similar to gaussian blurring + downscaling, the adapted pipeline produces much more blurry images with more noise, suggesting that FOREST-2 produces less resolution than what was initially expected. However, both super resolved images yield better performance than just using bicubic interpolation, and they are very similar between them. This suggests that the super resolution model is able to recover the details lost during stronger degradation processes, but there seems to be a limit to the amount of detail that can be recovered. Even though the starting point is different (baseline LR is less blurry than adapted LR), the final result is very similar.

The estimated kernel and noise for both pipelines is also shown in fig. 5.1. While the baseline kernel is very simple and the noise is more or less uniform across the image, the adapted kernel is more complex and the noise seems to be strongly correlated with the image intensity. It is important not to overinterpret this result, as the kernel and noise are estimated using overparametrized models, and multiple combinations of kernel and noise may produce similar results. However, it is interesting to see that the adapted pipeline is able to estimate a more complex degradation model, which is closer to the real degradation model used in the target domain.

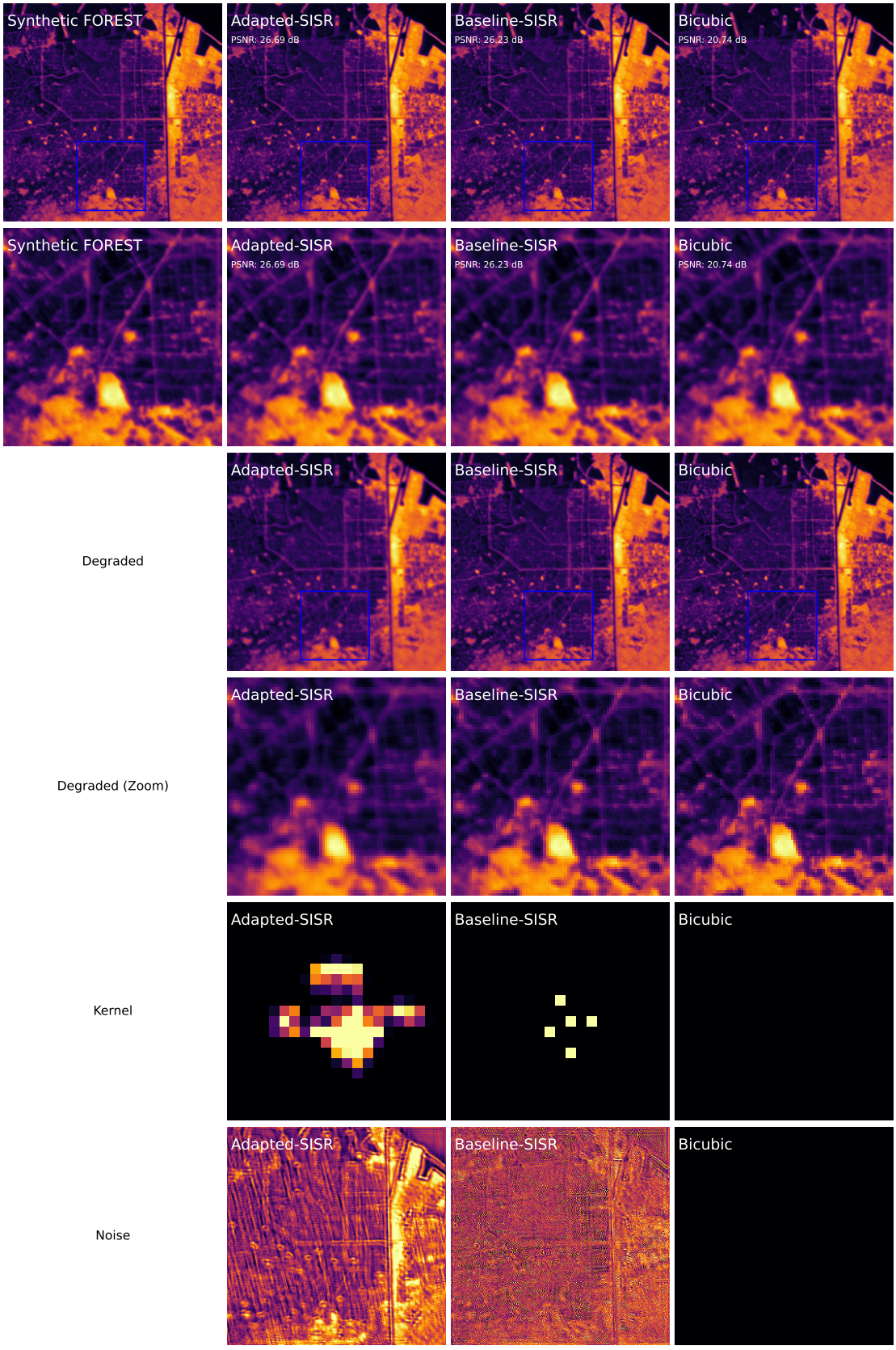

Figure 5.1: Applying different degradation models on an HR sample. In the Baseline SISR, the target domain used for training the generator is composed of sysnthetic HR FOREST-2 images degraded using the model detailed in 3.1. In the adapted SISR, the target domain used for training the generator is composed of real FOREST-2 images.

5.1.1 Effects of the degradation model

Fig. 5.2 shows the performance obtained by super resolving the degraded synthetic FOREST images for the whole validation dataset. In the upper row (a), the corresponding model is used to obtain the super resolved images. The performance, both in PSNR and SSIM, are very similar for both pipelines. The LPIPS shows a consistent behavior too. In the lower row (b), a simple bicubic interpolation is used to super resolve the degraded images. In this case, the baseline LR version yields better results than the adapted LR version, both in PSNR and SSIM. This suggests that the learned degradation model from FOREST-2 images loses more information than the baseline, resulting in a lower ground sampling distance than what is expected. Interestingly, the SR model is able to recover a big portion of the information, as the performance after using the SR models is very similar. This shows the relevance of the domain gap in super resolution, the SR model is able to apply the inverse of the degradation function in most cases, and the problem relies on that in the most common experiments, the wrong degradation model is shown to the model. Moreover, the LPIPS for the adapted LR version is better than the baseline LR version. FADSPOFKDASOFADOJSFJOADOJFAD

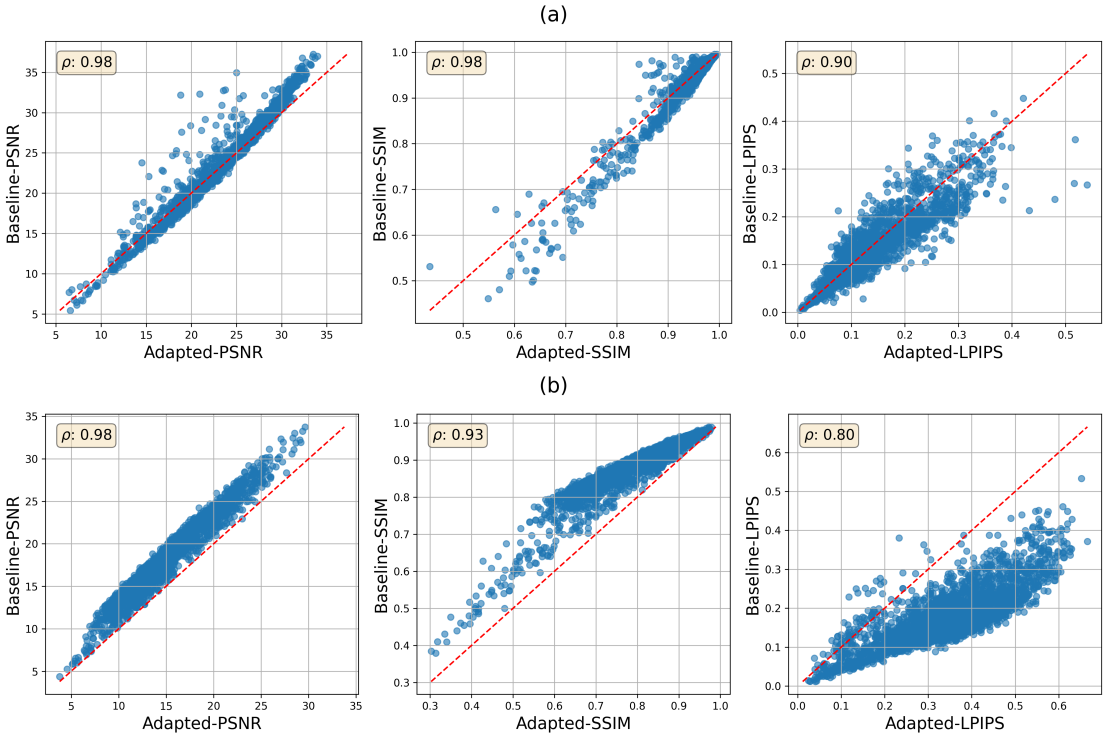

Figure 5.2: Performance obtained by super resolving the degraded synthetic FOREST images. In (a), the correspondingly trained SISR model is applied. In (b), a simple bicubic interpolation is used to super resolve the degraded images.

An alternative way to evaluate the lost frequency components due to the degradation model is by analyzing the frequency domain of the LR images. An example of the analysis of the whole validation dataset is shown in Fig. 5.3. In (a) the log magnitude of the FFT is across different spatial frequency values for the different degraded images is displayed. The spatial frequency is obtained from the radial distance to the center of the FFT, as

shown in 3.4.3. In (b), the amplification with respect to a simple gaussian blurring + downscaling is shown. While the baseline does not lose much information with respect to gaussian blurring + downscaling, the adapted pipeline consistently diminishes high frequency components starting at 0.1 cycles per pixel. It is important to note that 0.1 cycles per pixel at a 210m GSD corresponds to a cycle frequency of 21m. These results are consistent with the observations noted in Fig. 5.2, where the adapted pipeline yields worse LR images.

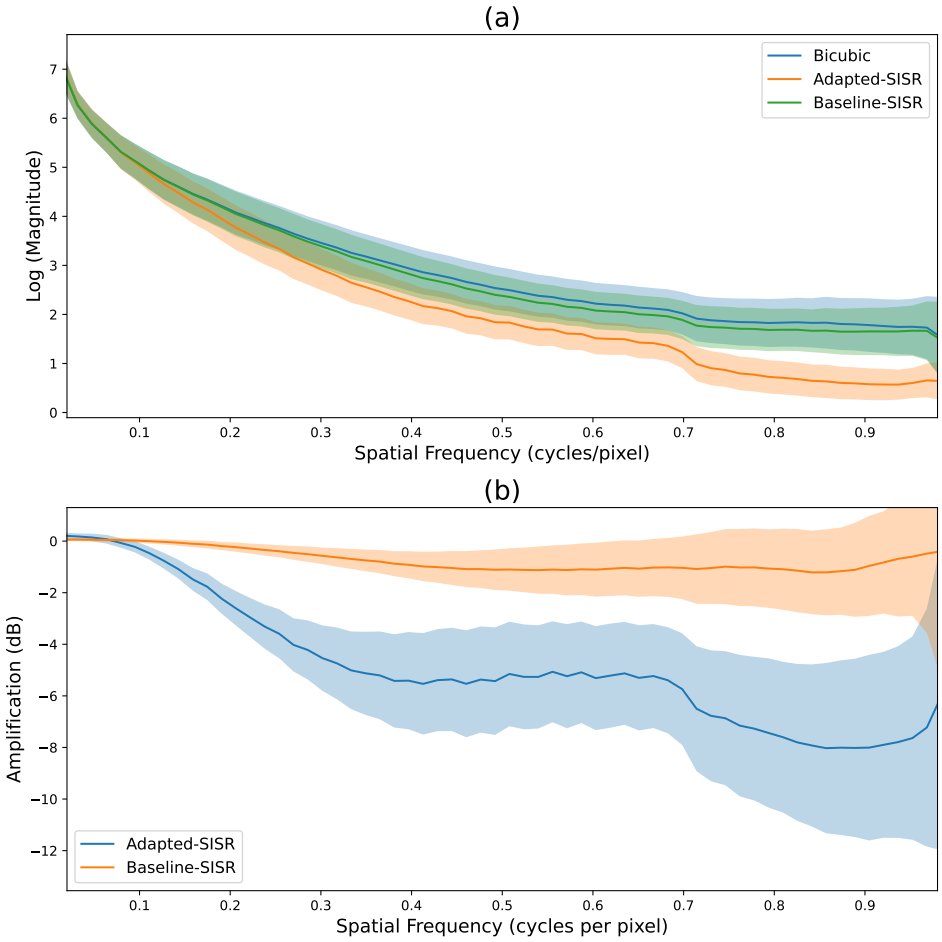

Figure 5.3: Frequency domain analysis of the LR images obtained by applying different degradation models on the HR sample displayed in Fig. 5.1. In (a), the log of the magnitude of the FFT for the LR images is shown, while in (b), the amplification with respect to a simple gaussian blurring + downscaling is shown.

5.2 Target domain

This subsection will show the results from the experiments performed on the target domain, this is the equivalent of the red arrows flow describes in fig. 2.5. In this case, the GAN trained for the degradation model is discarded and only the super resolution model is used. The input images are real FOREST-2 images, and the output images are super resolved versions of them. Due to the unpaired nature of the dataset, the performance of the SR model can not be evaluated using metrics like PSNR and SSIM. Other alternatives will be presented, and a qualitative analysis will be performed. Additionally, a quantitative analysis will be discussed using a very small sample of paired data obtained by synchronizing the overpass of FOREST-2 with the route of ECOSTRESS.

In Fig. 5.4, the super resolutions models were used with a 264x264 pixels crop of a real FOREST-2 image.

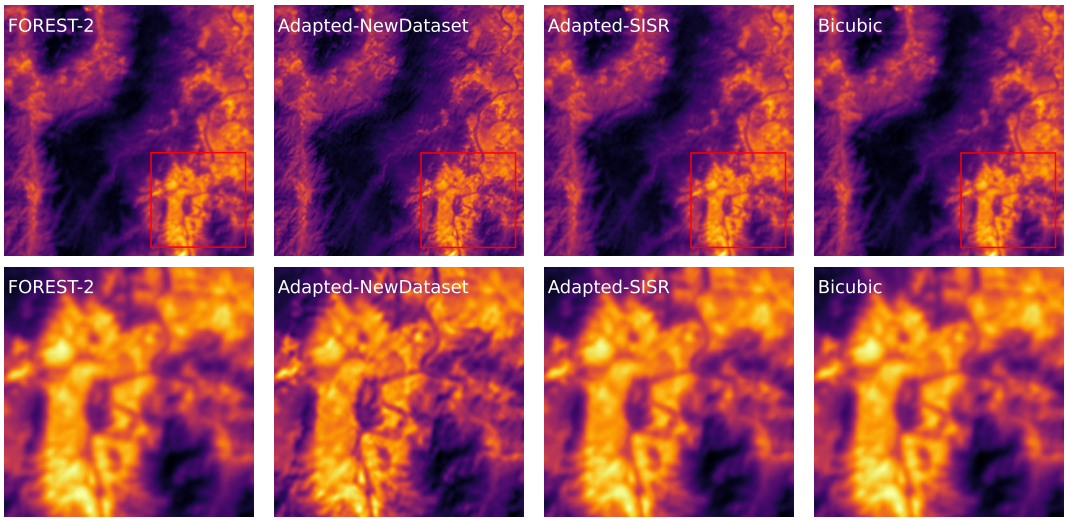

Figure 5.4: Super Resolved Forest-2 Scene using different SR models. In the upper row, the image is displayed. A detailed zoom is displayed below

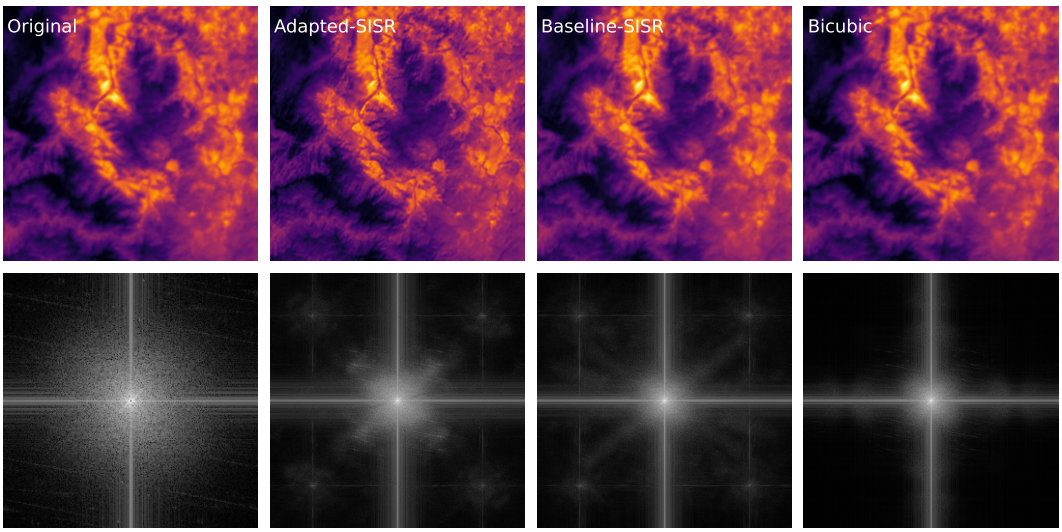

Figure 5.5: Super Resolved Forest-2 Scene using different SR models. In the upper row, the image is displayed. The log magnitude of the FFT is displayed below.

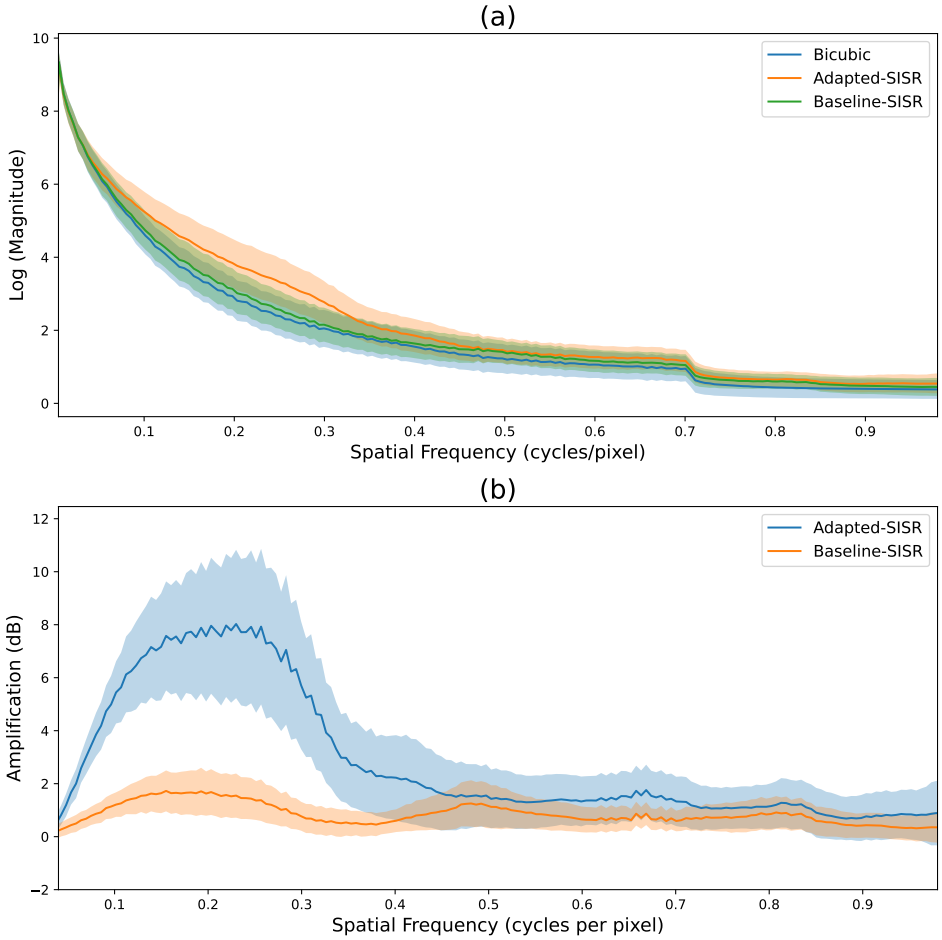

Figure 5.6: Frequency domain analysis of the SR images obtained by applying different SR models to the real FOREST-2 validation dataset. In (a), the log of the magnitude of the FFT for the SR images is shown, while in (b), the amplification with respect to a simple bicubic upsampling is displayed.

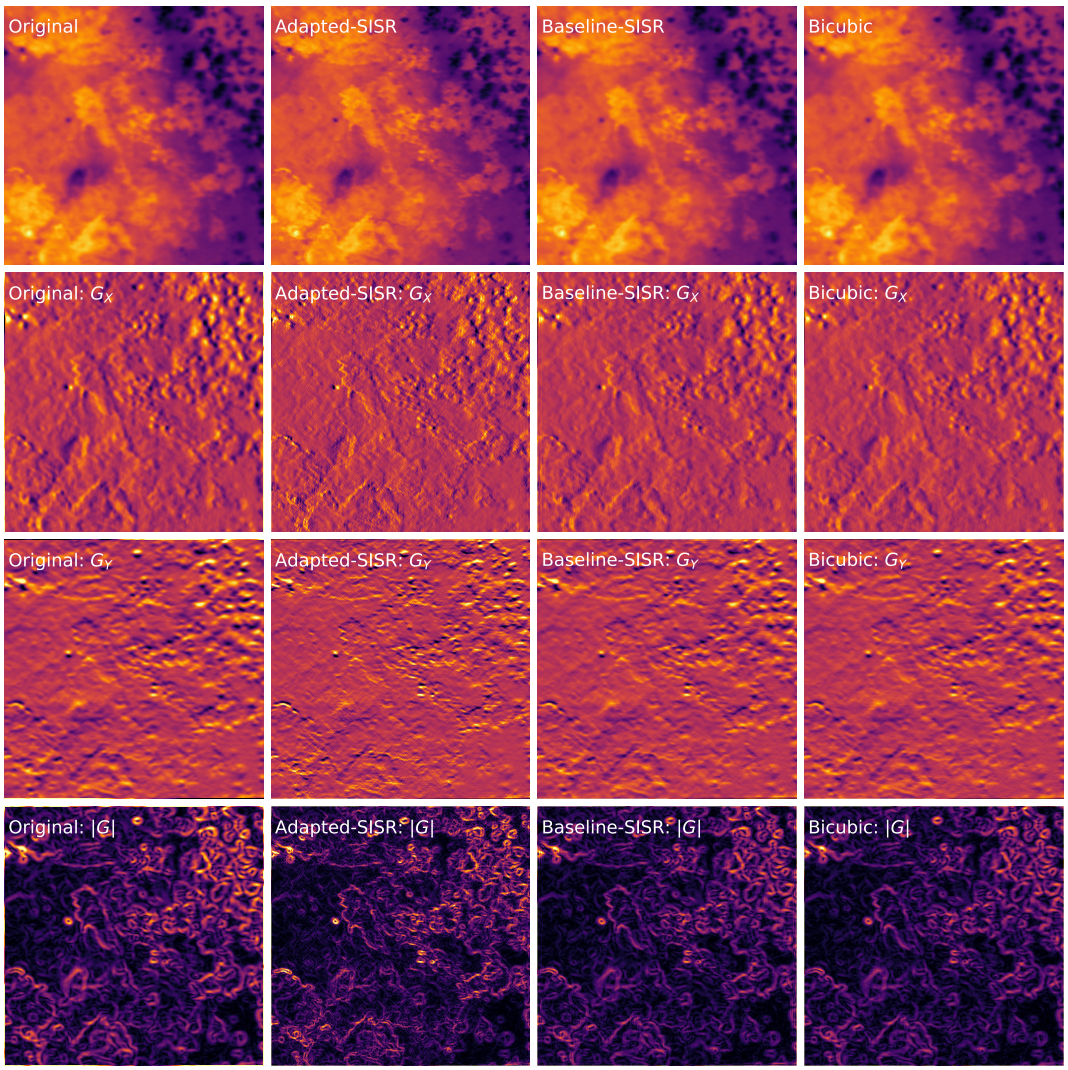

Figure 5.7: Gradient analysis of the super resolved images using different SR models for scenes coming from the real FOREST-2 validation dataset. In the upper row, the image is displayed. The gradients in the x and y direction (G_x and G_y respectiely) are displayed below. the gradient magnitude $|G|$ is displayed in the bottom row.

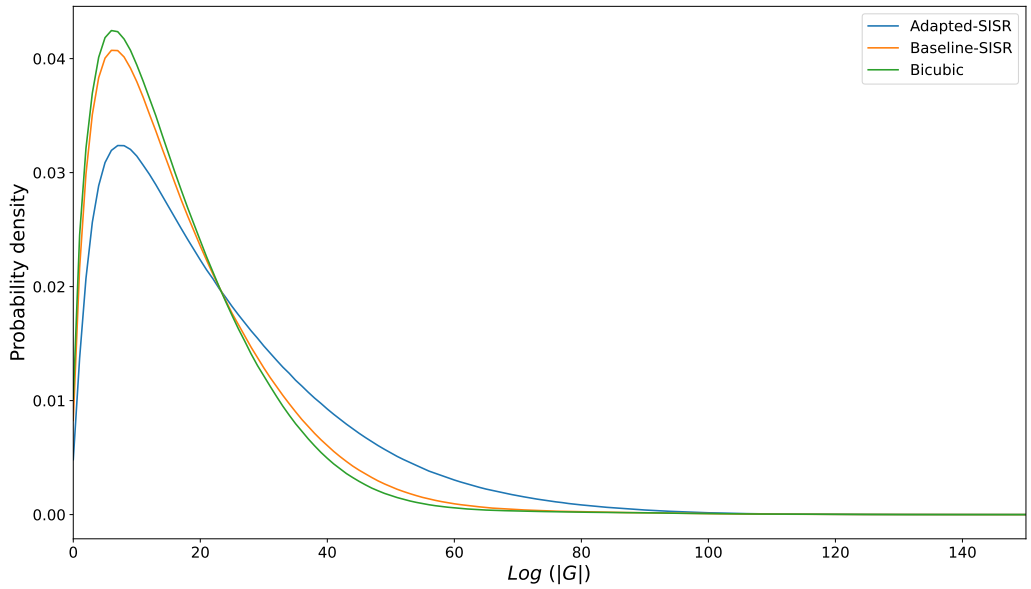

Figure 5.8: Histogram of the gradient magnitude $|G|$ for the whole validation real FOREST-2 dataset.

5.3 Effects of the domain gap

The combination of the probabilistic degradation model and the SR model were proven helpful to bridge the domain gap between the source and the target domain. However, it is important to understand what happens when the target domain used in training does not match the conditions that will occur in the real world. Two situations are identified, the first one is when the target domain used in training is more complex than what will be used in production, and the second one is when the target domain used in training is simpler than what will be used in production. The latter was largely adressed in this work, when the HR-LR pairs are generated using a baseline degradation model, which is overly optimistic. Another example of it would be that images taken from FOREST-2 in several years may not have the same distribution as current images, as the instruments may have degraded over time. However, the first one was not addressed, and is the focus of this subsection.

To simulate this situation, the model trained using real FOREST-2 images was used to super resolve synthetic FOREST images degraded using the baseline degradation model. The results are shown in Figs. 5.9 and 5.11. The performance of the adapted model is catastrophic, producing several artifacts and yielding a PSNR difference of approximately 10dB, which represent a 10x difference in terms of MSE.

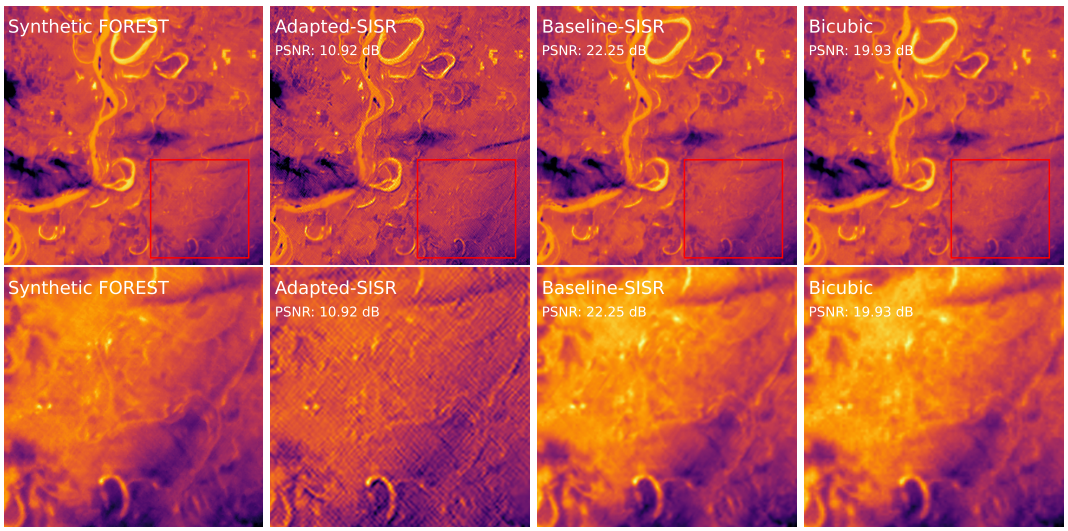

Figure 5.9: Effects of using a model trained with on different domain than at inference time. When using an Synthetic FOREST image degraded with the baseline degradation model as an input, the model trained using real FOREST-2 data as the target domain generates several artifacts and underperforms severly in terms of PSNR.

The performance results in terms of different metrics are shown in Fig. 5.10. In the conditions described above, the adapted super resolution model underperforms severly considered metric.

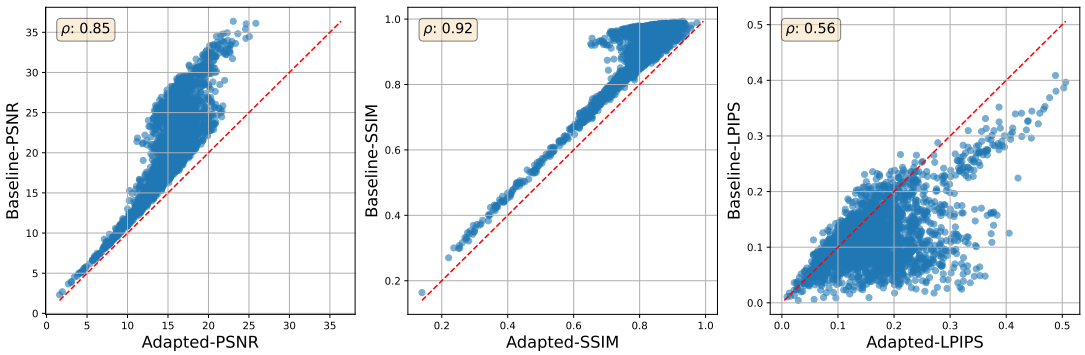

Figure 5.10: Performance obtained by super resolving the degraded synthetic FOREST images using different super resolution models.

In the frequency domain, the results are shown in Fig. 5.11. The adapted model adds amplification in the higher range of spatial frequency, related with noise and artifacts. The frequencies related to the proper signal are also amplified, suggesting an improvement in the resolution. This suggests that while the adapted model highlights edges and details, it also severly amplifies the noise and artifacts, resulting in a worse performance in terms of PSNR. It is important to note that the amplification curve does not show a considerable difference with respect to the one in Fig. 5.6, highlighting the fact that the frequency analysis is not suitable by itself to evaluate the performance of the SR model.

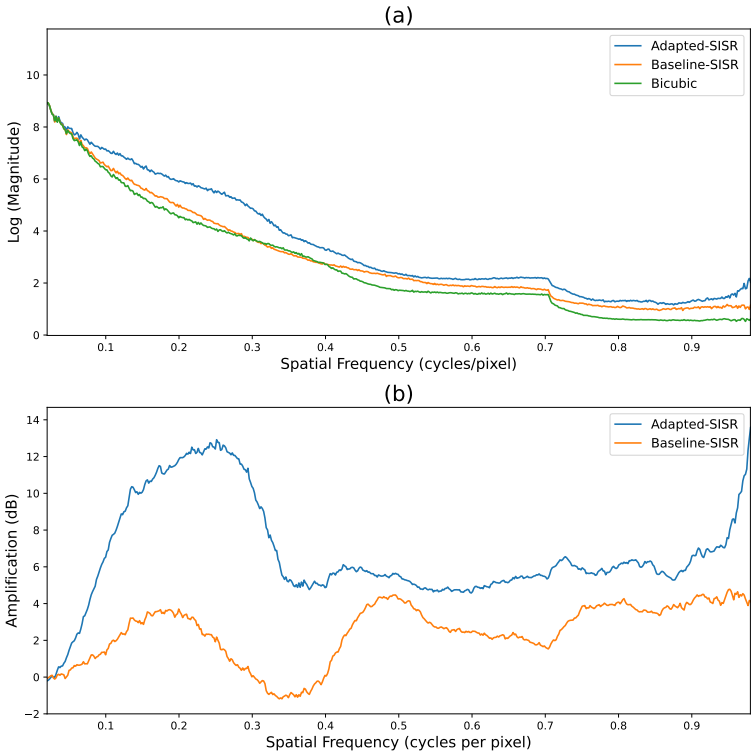

Figure 5.11: Effects of using a model trained with on different domain than at inference time. (a) shows the log magnitude of the radial average of the FFT for the SR images using different algorithms. (b) shows the amplification with respect to bicubic interpolation

This demonstrates that while this approach is very good to bridge a domain gap, it is not robust at all to domain shifts. This limitation is in sync with what is found in the literature, where GAN approaches are not able to generalize to arbitrary domains DASDASD.

As in the target domain the ground truth is not known due to the lack of a paired dataset, the performance of the SR model can not be evaluated using metrics like PSNR and SSIM. The result of the image quality assessment metrics were done by super resolving images coming from real FOREST-2 and also from the degraded synthetic FOREST-2 images and calculating their NIQE and BRISQUE scores. The results are shown in Fig. 5.12. For both metrics, a large gap is observed between the adapted model and the rest, when the input images come from real FOREST-2 images. Suggesting that the adapted model is able to produce more natural images than the rest. This behaviour does not exactly replicates when the input images come from synthetic FOREST-2 images, where the adapted model is not able to produce more natural images than the rest. Moreover, for the adapted model, both metrics tend to get worse when the input images come from synthetic FOREST-2 images. The contrary happens for the rest of the models. This suggests that the adapted model is able to produce more natural images than the rest, but only when the input images come from the same distribution as the target domain used in training. However, it is important to note that:

1. The images used for training the NIQE and BRISQUE models are not remote

sensing images, and therefore, the results may not be representative. This could be circumvented by training a NIQE/BRISQUE model using remote sensing images only.

2. NIQE and BRISQUE are objective measures of image quality, not of physical consistency or image fidelity.

While the comparison of results may help to understand the behaviour of the models, it is important to note that the results are not representative of the real world performance of the models.

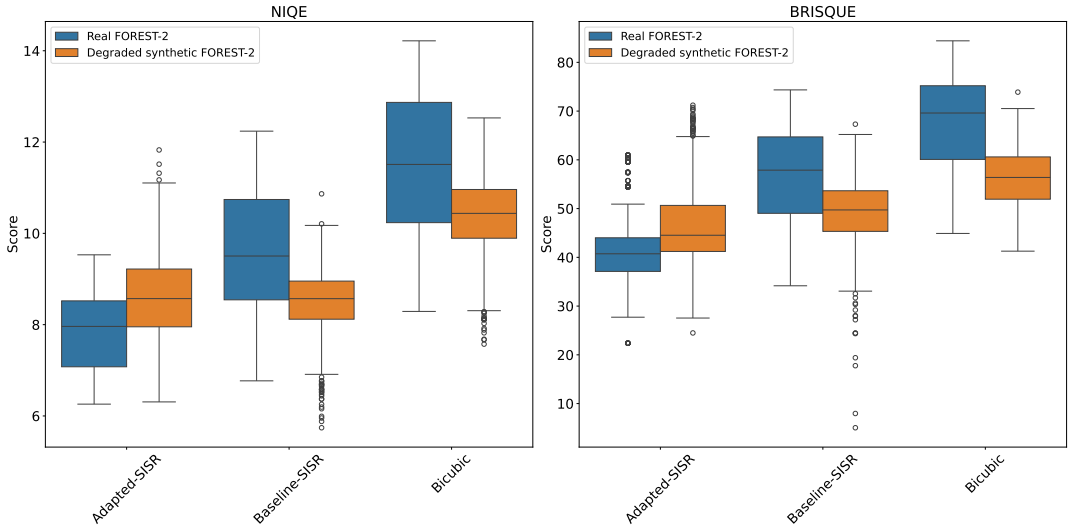

Figure 5.12: Image quality assessment metrics for the different SR models using different datasets as input. In both metrics, the lower the score, the better the image quality.

References

[1] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations, 2010.

[2] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, 2001.

[3] Diego Valsesia and Enrico Magli. Permutation invariance and uncertainty in multitemporal image super-resolution, 2021.

[4] Syed Muhammad Anwar Bashir, Yanning Wang, Murtaza Khan, and Yulei Niu. A comprehensive review of deep learning-based single image super-resolution, 2021.

[5] Marcus martens, Dario Izzo, Andrej Krzic, and Daniël Cox. Super-resolution of proba-v images using convolutional neural networks, 2019.

[6] John Kennedy, Ora Israel, Alex Frenkel, Rachel bar shalom, and Haim Azhari. Improved image fusion in pet/ct using hybrid image reconstruction and super-resolution, 01 2007.

[7] Christian Mollière, Julia Gottfriedsen, Martin Langer, Patricio Massaro, Christian Soraruf, and Matthias Schubert. Multi-spectral super-resolution of thermal infrared data products for urban heat applications, 2023.

[8] Andreas Lugmayr, Martin Danelljan, Radu Timofte, Namhyuk Ahn, Dongwoon Bai, Jie Cai, Yun Cao, Junyang Chen, Kaihua Cheng, SeYoung Chun, Wei Deng, Mostafa El-Khamy, Chiu Man Ho, Xiaozhong Ji, Amin Kheradmand, Gwantae Kim, Hanseok Ko, Kanghyu Lee, Jungwon Lee, Hao Li, Ziluan Liu, Zhi-Song Liu, Shuai Liu, Yunhua Lu, Zibo Meng, Pablo Navarrete Michelini, Christian Micheloni, Kalpesh Prajapati, Haoyu Ren, Yong Hyeok Seo, Wan-Chi Siu, Kyung-Ah Sohn, Ying Tai, Rao Muhammad Umer, Shuangquan Wang, Huibing Wang, Timothy Haoning Wu, Haoning Wu, Biao Yang, Fuzhi Yang, Jaejun Yoo, Tongtong Zhao, Yuanbo Zhou, Haijie Zhuo, Ziyao Zong, and Xueyi Zou. Ntire 2020 challenge on real-world image super-resolution: Methods and results, 2020.

[9] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond, 2021.

[10] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Learning the degradation distribution for blind image super-resolution, 2022.

[11] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution, 2018.

[12] Simon Hook and Gerardo Rivera. ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS). https://ecostress.jpl.nasa.gov/instrument, 2023. Accessed: 28-November-2023.

[13] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2017.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[16] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer, 2013.

[17] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution, 2021.

[18] Irwin Sobel and G. M. Feldman. An isotropic 3×3 image gradient operator, 1990.

[19] Jet Propulsion Laboratory. ECOSTRESS Fact Sheet, 2023. [Online accessed 28-November-2023].

[20] PhyTIR: Plant High Temperature InfraRed Viewer. https://phytir.jpl.nasa.gov/, 2023. [Online accessed 28-November-2023].

[21] Application for extracting and exploring analysis ready samples (AppEEARS). https://appeears.earthdatacloud.nasa.gov/, 2023. [Online; accessed 28-November-2023].

[22] Appeears api. https://appeears.earthdatacloud.nasa.gov/api/, 2023. [Online; accessed 28-November-2023].

[23] Land Processes Distributed Active Archive Center (LP DAAC). ECOSTRESS L1B Geolocated Radiance Data (ECO1BMAPRAD). https://lpdaac.usgs.gov/products/eco1bmapradv001/, 2023. [Online; accessed 28-November-2023].

[24] Ecostress faq. https://ecostress.jpl.nasa.gov/faq. Accessed: 2023-11-28.