



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	1
1.1	Storyline	1
1.2	Wildfire monitoring using thermal remote sensing	2
1.2.1	Spatio-temporal trade-off	2
2	Super resolution	2
2.1	single-Image Super Resolution	2
2.1.1	SR Resnet	3
2.2	Multi-Image Super Resolution	3
2.2.1	Multi-spectral super resolution	4
2.2.2	Importance of interframe correlation	4
2.2.3	RAMS	4
2.3	Data requirements	4
2.3.1	Super Resolution as a supervised problem	4
2.3.2	Applying a basic degradation model	4
2.3.3	The domain gap problem	4
2.4	Blind image Super Resolution	4
3	Methodology	6
3.1	Baseline Degradation model	6
3.1.1	Blurring Kernel	6
3.1.2	Radiometric error correction	7
3.2	Models Architecture	9
3.2.1	SRResNet	9
3.2.2	RAMS	9
3.2.3	Probabilistic Degradation Model	9
3.3	Performance Metrics	9
3.3.1	Referenced image quality metrics	9
3.3.2	No-reference Image quality metrics	10
3.3.3	Frequency Domain Analysis	11
3.3.4	Gradient Distribution analysis	12
4	Experiment Setup	14
4.1	Obtaining a high resolution dataset	14
4.1.1	The ECOSTRESS mission	14
4.1.2	Donwloading ECOSTRESS Scenes	14
4.1.3	Selecting the best scenes	15
4.1.4	Data Processing	17
4.1.5	Obtaining FOREST-2 data	17
4.2	Datasets	17
4.2.1	Training: Ecostress-DegradedEcostress	18
4.2.2	Testing: Ecostress-DegradedEcostress (Different degradations)	19
4.2.3	Training: Ecostress-Forest (Unpaired)	19
4.2.4	Testing: Ecostress-Forest (Paired)	19

5 Results and discussion	20
5.1 Source domain	20
5.1.1 Effects of the degradation model	22
5.2 Target domain	24

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 -*i* 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

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.

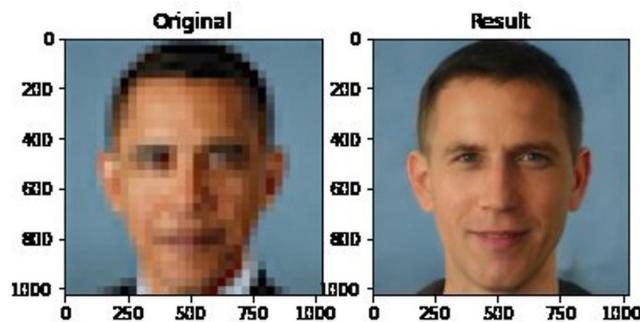


Figure 2.1: Example of super resolution as an ill posed problem. A blurry picture of Barack Obama can be generated from an HR image of another person.

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

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 2×2 and is usually faster, bicubic needs a receptive field of 4×4 . 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:

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

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 extraction 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 extensive 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.

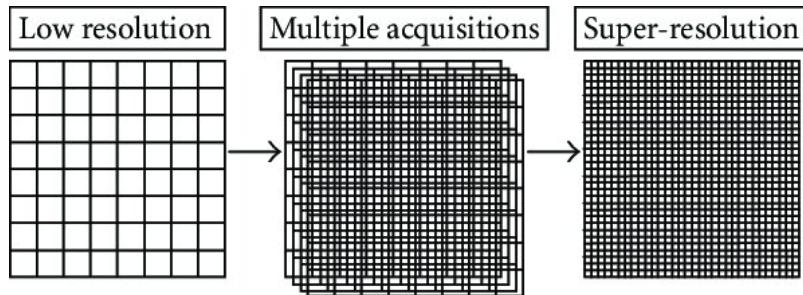


Figure 2.2: Multi-image super resolution algorithms combine multiple low-resolution image acquisitions into a high-resolution image. Source: [6]

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.

In the context of MSSR, each spectral band, corresponding to a specific wavelength 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 reasonable 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 Data requirements

2.3.1 Super Resolution as a supervised problem

SR is a supervised problem, the super resolved image is compared to the HR ground truth and the differences between them drives the gradients of the neural network to minimize the loss.

2.3.2 Applying a basic degradation model

2.3.3 The domain gap problem

2.4 Blind image Super Resolution

Blind SR for unknown degradation is proposed to bridge the domain gap present in the synthetic datasets generated using bicubic downsampling.

Nevertheless, real-world degradations are usually too complex to be modelled with an explicit combination of multiple degradation types, as shown in Fig.3(c). Therefore, implicit modelling attempts to circumvent the explicit modelling function. Instead, it defines the degradation process 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

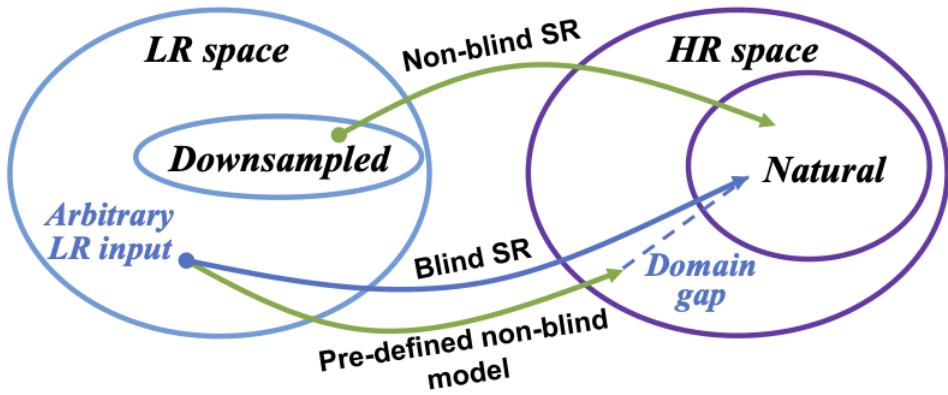


Figure 2.3: Domain interpretation of differences between non-blind and blind SR. Source: [8]

(GAN) [16] to grasp the implicit degradation model possessed within training dataset, like CinCGAN [8].

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 [9]. 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 leads 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 grants 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

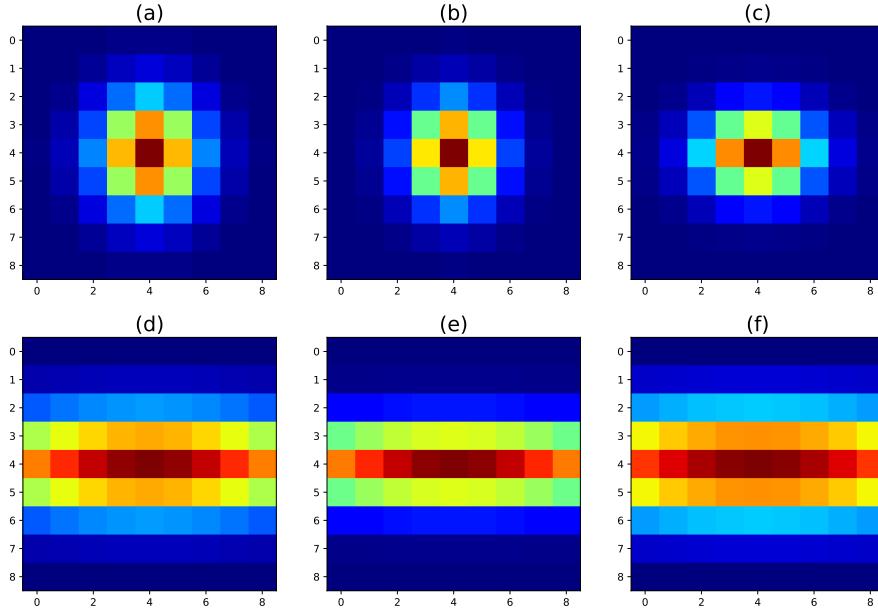


Figure 3.1: Examples of kernels used in a stochastic degradation model. (a),(b) and (c) are generated using a symmetric variance on the x and y axis. (d) (e) and (f) are generated using an asymmetric variance, resulting in much more anisotropic kernels.

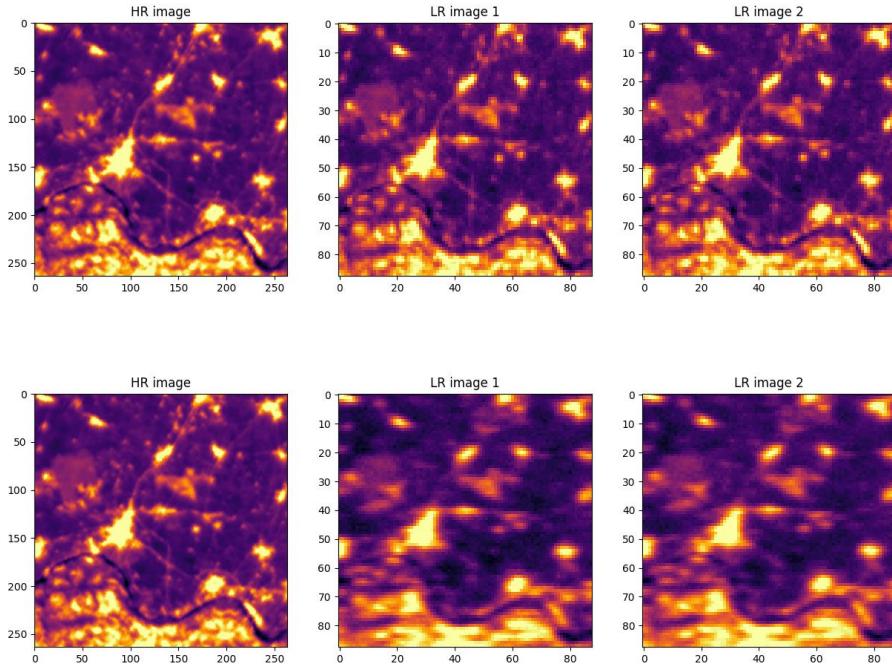


Figure 3.2: Effects of different blurring kernels on the HR-LR generation. The upper row contains images generated using blurring kernels with symmetric distributions. The lower rows contain images generated using asymmetric distributions for the variances, resulting in highly anisotropic kernels.

3.1.2 Radiometric error correction

As reported by the ECOSTRESS instrument sheet,[10] 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:

$$e_{\text{forest}} = \sqrt{e_{\text{eco}}^2 + e_{\text{extra}}^2} \quad (2)$$

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:

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

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/m²/μm and 1.1444×10^{-1} W/sr/m²/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:

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

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.

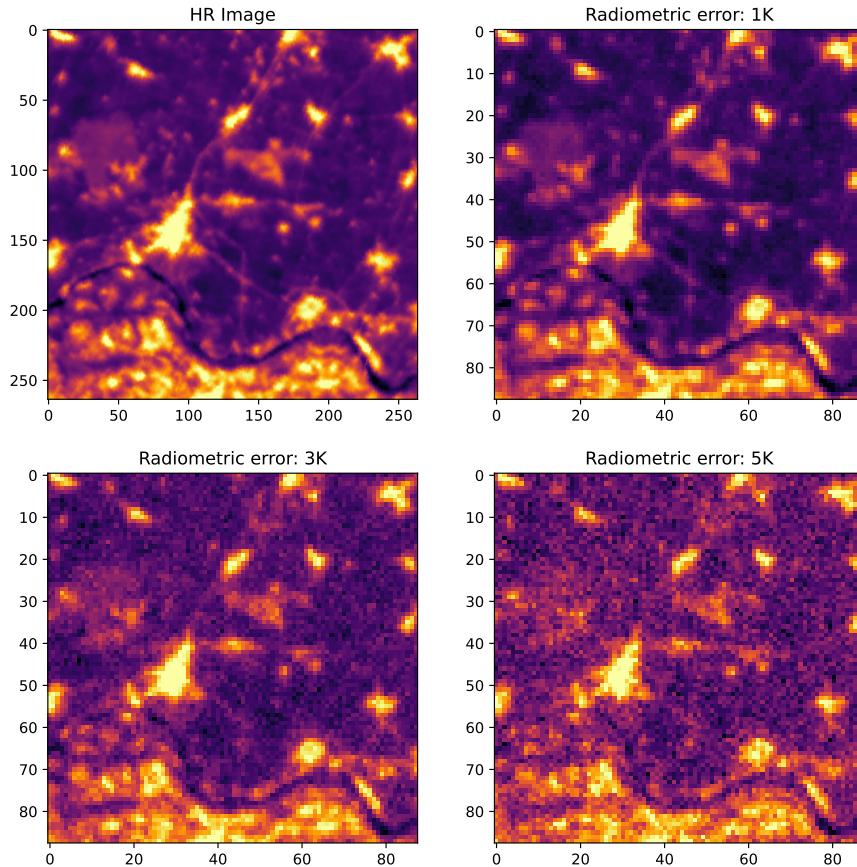


Figure 3.3: Effects of different radiometric error corrections on the HR-LR generation.

3.2 Models Architecture

3.2.1 SRResNet

Introduced in 2017 [11], SRResnet leverages on residual networks [12] 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

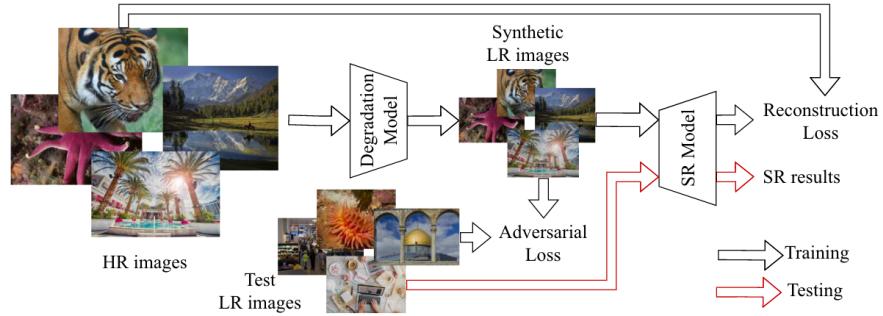


Figure 3.4: In Degradation-learning-based methods like this one, adversarial training is used to encourage the degradation model to produce images in the same domain with the target domain (test images). Source [13].

Kernel Model

Noise Model

Discriminator

3.3 Performance Metrics

3.3.1 Referenced image quality metrics

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 they 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. 5

$$\mathcal{L}^*(I^{HR}, I^{LR}, d) = \min_{u,v \in [0,2d]} \mathcal{L}(I_{u,v}^{HR}, I_{u,v}^{SR}) \quad (5)$$

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:

$$\begin{aligned} \mathcal{L}^*(I^{HR}, I^{LR}, d) &= \min_{u,v \in [0,2d]} \mathcal{L}(I_{u,v}^{HR}, (I_{u,v}^{SR} + b)) \\ b &= \frac{1}{(W-d)(H-d)} \sum_{x,y} (I_{u,v}^{HR} - I_{u,v}^{SR}) \end{aligned} \quad (6)$$

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

3.3.2 No-reference Image quality metrics

As a result, NIQE has found widespread use across various domains of image processing, including restoration, compression, and enhancement, where it serves as a standard for the performance of algorithms, particularly in the realm of image super-resolution.

Naturalness Image Quality Evaluator (NIQE)

The Naturalness Image Quality Evaluator (NIQE) [14] is a no-reference image quality assessment metric that quantifies the perceptual quality of images based on their naturalness. Unlike traditional metrics that require a reference image for comparison, 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.

BRISQUE

3.3.3 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-resolution objective is to reconstruct missing high frequency components from a downsampled 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 [15].

Utilizing 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 for 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:

$$f(r) = \frac{r}{\frac{N}{2}} \cdot f_{\text{Nyquist}}, \quad (7)$$

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.

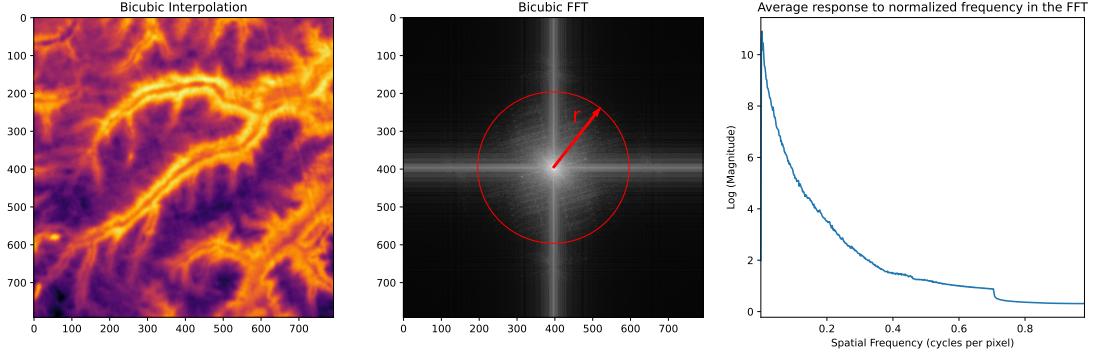


Figure 3.5: Steps of the frequency domain analysis. The Center image shows the log magnitude of the shifted FFT of a bicubic upsampled FOREST scene and an example of a radial profile, the average of all the points that have the same r is calculated. The right image displays the log magnitude obtained for every radial profile, translated into spatial frequency.

3.3.4 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 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. 8 [16]. These kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid.

$$\hat{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad \hat{G}_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (8)$$

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:

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (9)$$

The gradient magnitude histograms of the results of different super-resolution algorithms will be assessed, thereby 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. 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.

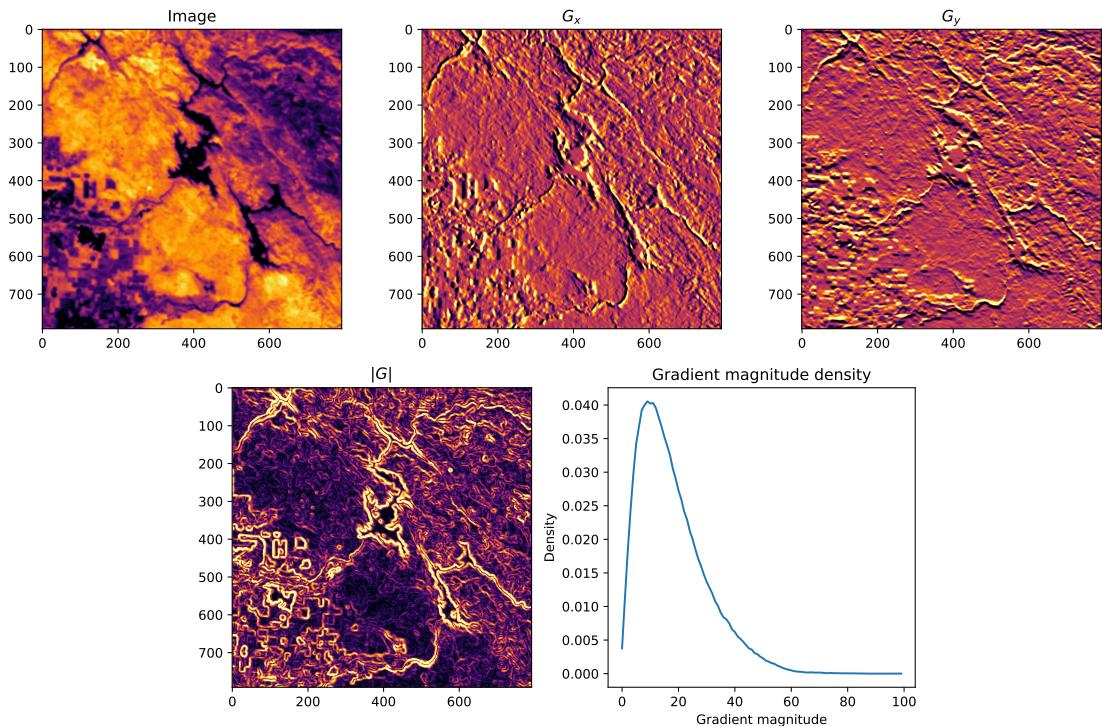


Figure 3.6: Steps to obtain a gradient magnitude density. Using the sobel operators, G_x and G_y are obtained from an image. The magnitude $|G|$ of each pixel is calculated using Eq. 9. The density can be estimated afterwards, using 100 bins in this case.

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 [17], 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 HypIRI Thermal Infrared Radiometer [18] 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 [17]. 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. Although 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 Downloading ECOSTRESS Scenes

ECOSTRESS imagery is available via NASA’s Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) [19]. This tool allows the request of area samples

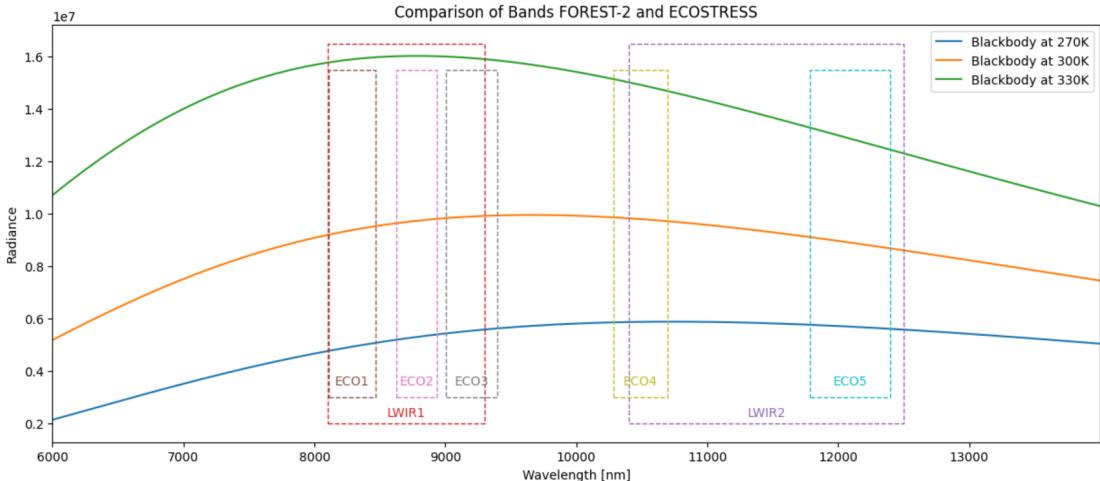


Figure 4.1: Wavelengths of the sensors in Ecostress and Forest satellites. The radiation spectrum of black-bodies at different temperatures are included for comparison.

via vector polygons. Using the product’s API [20], 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 [21].

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 [22], 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

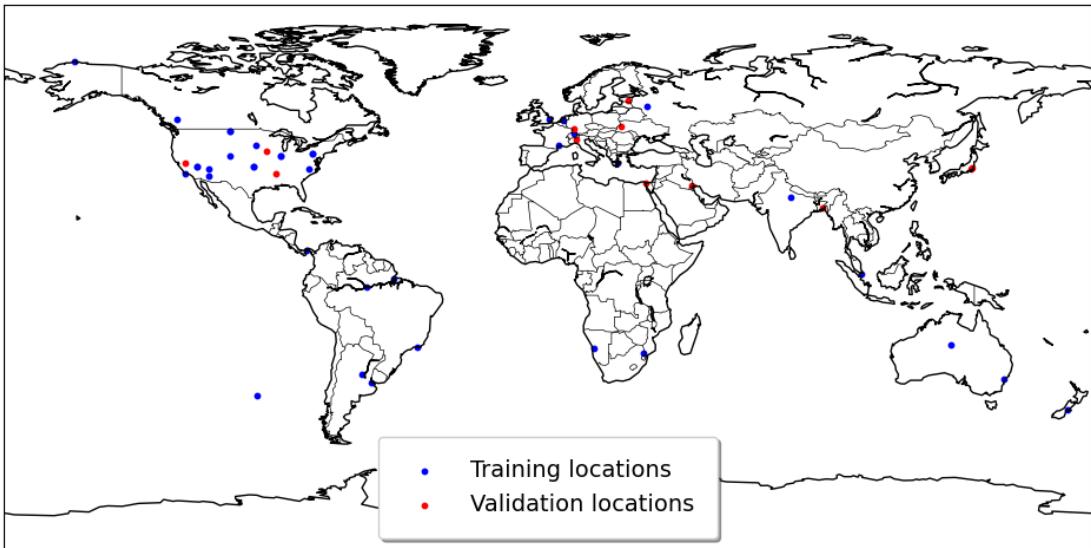


Figure 4.2: Location of the samples taken from ecostress.

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.

- 1: **QA statistics:**
 - 2: Get the average proportion of good pixels p_{gp} for the 5 radiances of the scene.
 - 3: Discard scenes where $p_{gp} < 60$.
 - 4: For each month, get the 3 scenes with the greatest p_{gp} .
 - 5: **Scene Statistics:**
 - 6: Get the biggest scene of each month.
 - 7: Calculate the proportion between the size each scene and the biggest of the month.
 - 8: Discard images which size proportion is smaller than 0.2.
 - 9: Calculate the median of the radiance values of the scene.
 - 10: **Selecting the scene of the month:**
 - 11: Merge the QA statistics and the Scene statistics.
 - 12: 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 its 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.

Table 4.2: Fill Value and Data Quality Classes

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

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

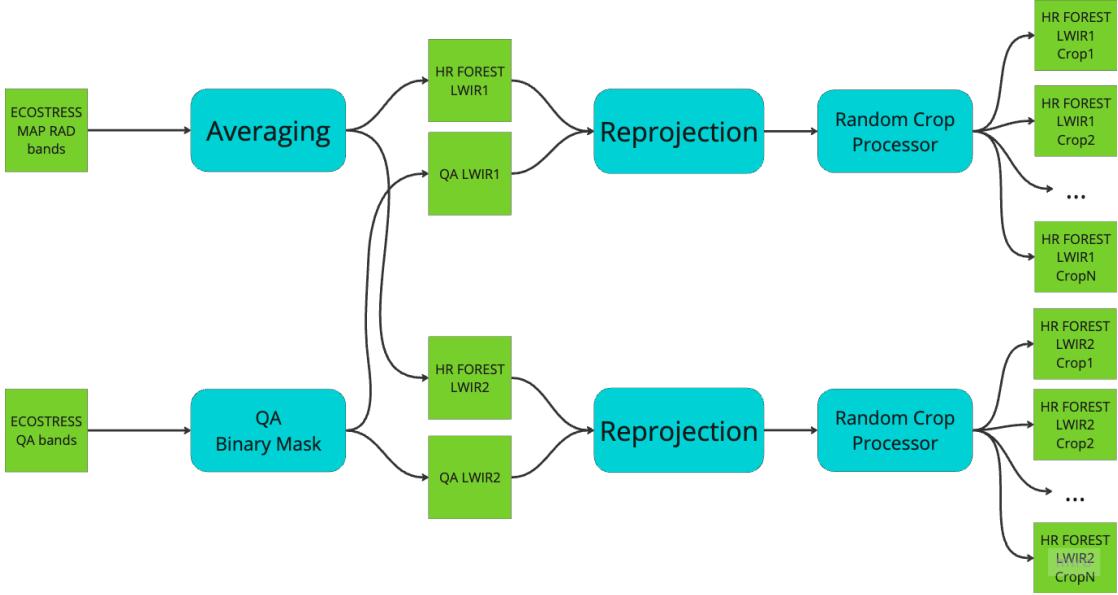


Figure 4.3: Data processing workflow

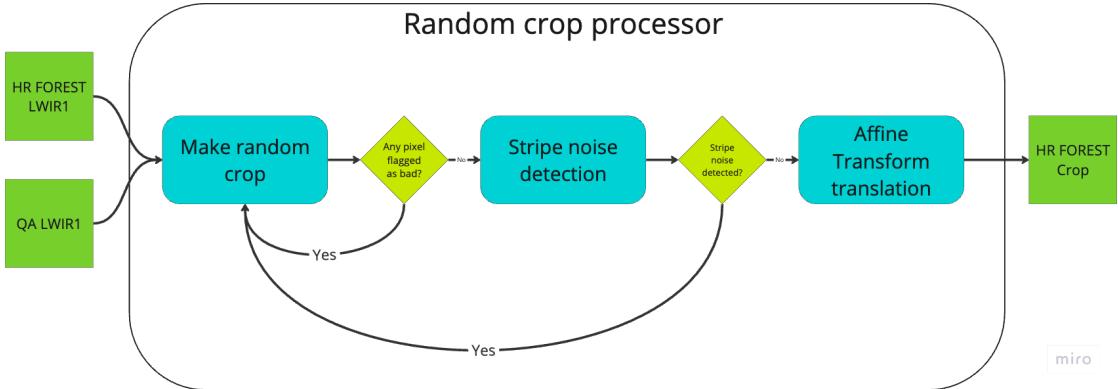


Figure 4.4: Random crop processor

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 sources 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, $\mathcal{D}_{\text{eco-eco}}$ 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 Testing: Ecostress-DegradedEcostress (Different degradations)

The dataset, $\mathcal{D}_{\text{eco-eco}}^{\text{anisotropic}}$ is built by taking the 264x264 HR synthetic FOREST crops and applying the baseline degradation model proposed in 3.1, resulting in 88x88 LR versions. However, the validation set is generated using a highly anisotropic degradation model, resulting in a forced domain gap between training and validation.

4.2.3 Training: Ecostress-Forest (Unpaired)

The dataset $\mathcal{D}_{\text{eco-forest}}$ 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.

4.2.4 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. 3.4.

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.

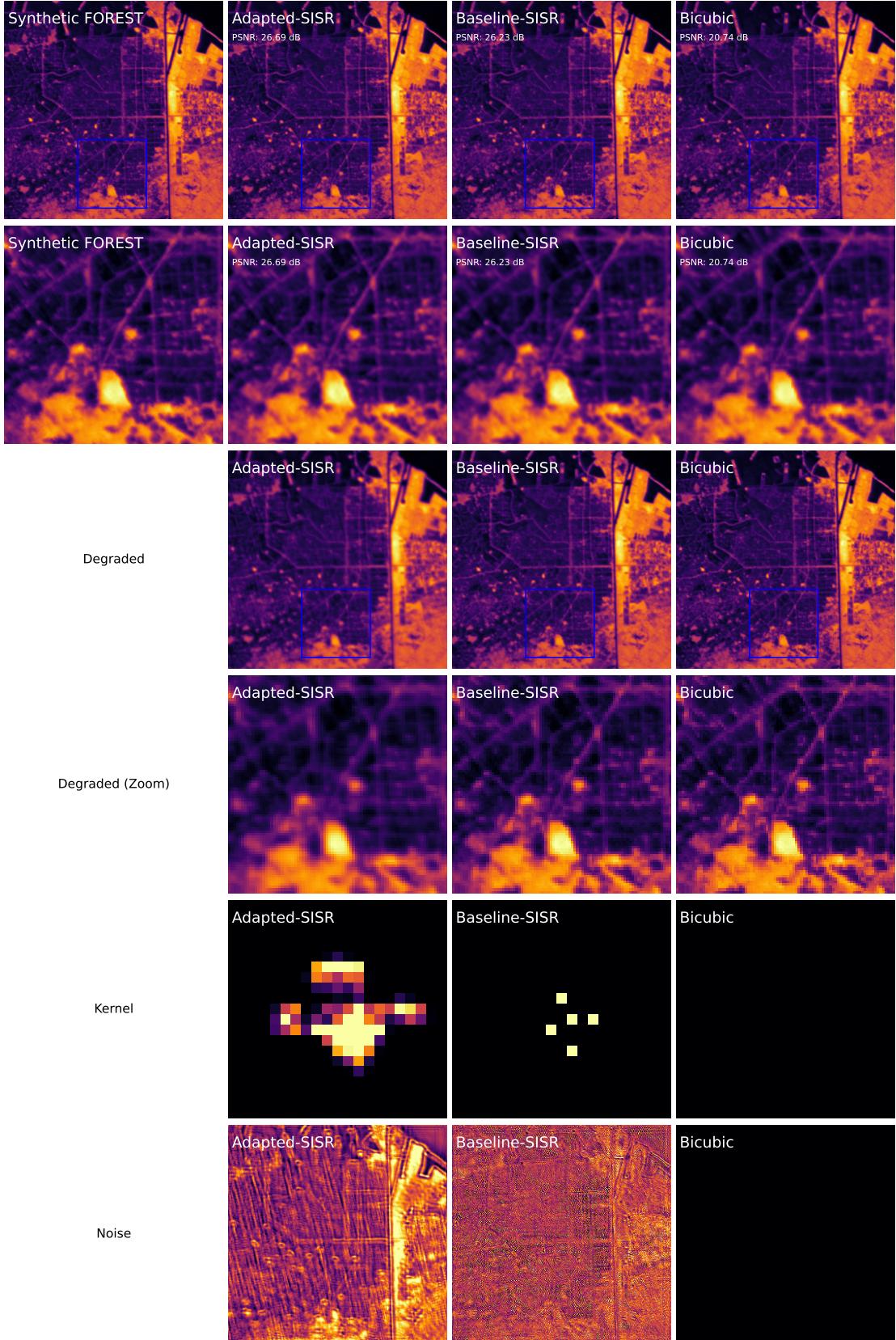


Figure 5.1: Applying different degradation models on an HR sample. In the Baseline SISR, the target domain used for training the generator is composed of sysnthetic HR FOREST-2 images degraded using the model detailed in 3.1. In the adapted SISR, the target domain used for training the generator is composed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

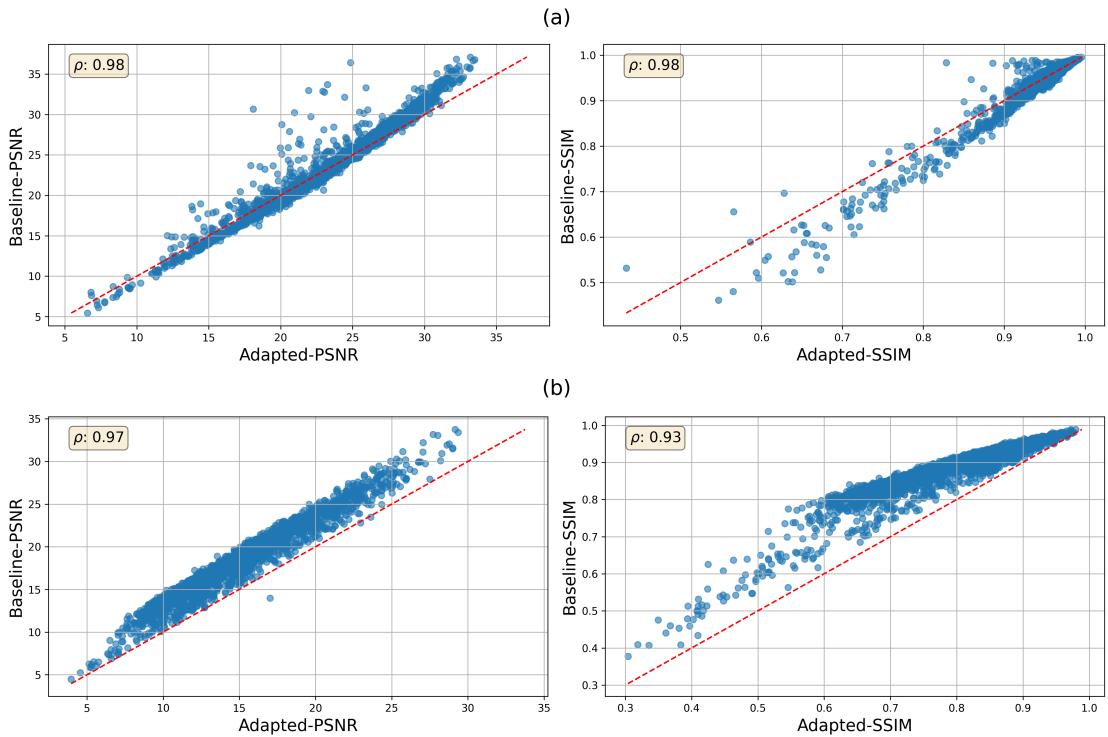


Figure 5.2: Performance obtained by super resolving the degraded synthetic FOREST images. In (a), the correspondingly trained SISR model is applied. In (b), a simple bicubic interpolation is used to super resolve the degraded images.

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 3.3.3. 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.

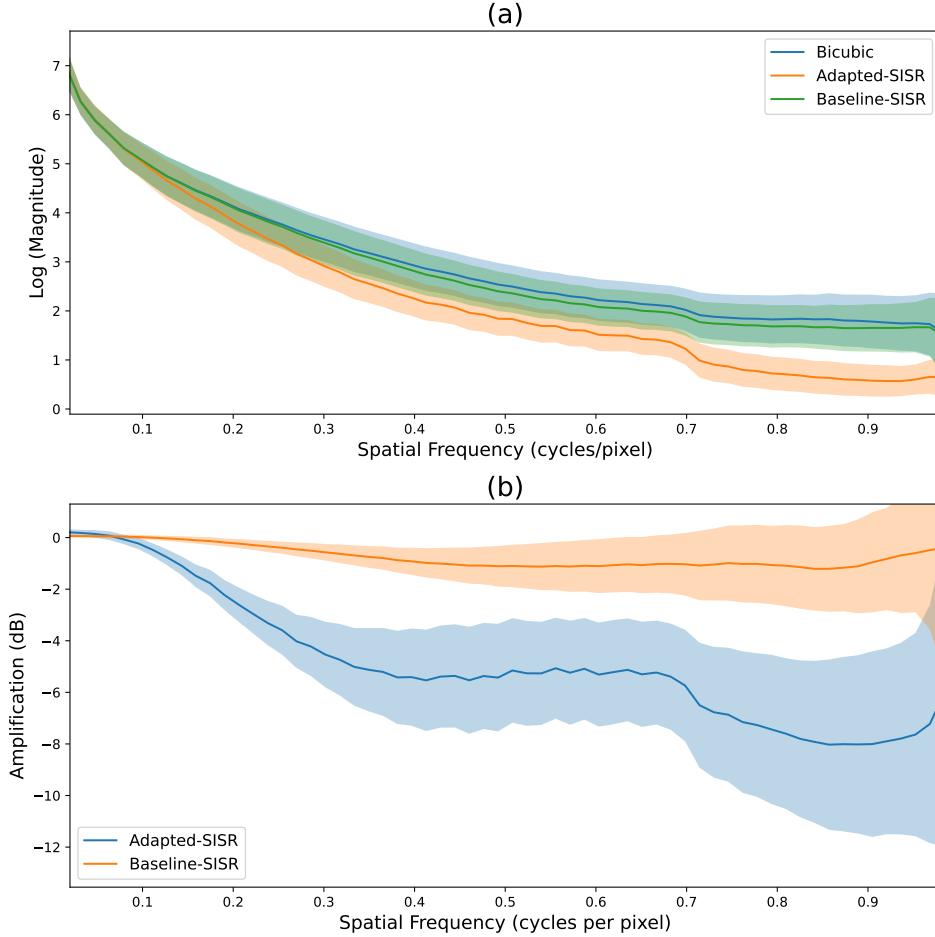


Figure 5.3: Frequency domain analysis of the LR images obtained by applying different degradation models on the HR sample displayed in Fig. 5.1. In (a), the log of the magnitude of the FFT for the LR images is shown, while in (b), the amplification with respect to a simple gaussian blurring + downscaling 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. 3.4. 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.

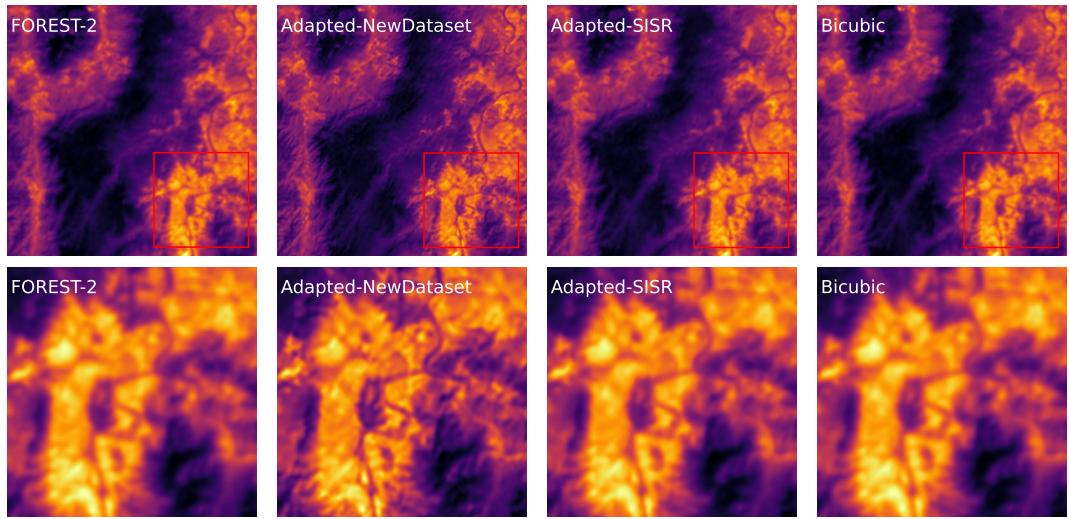


Figure 5.4: Sample of adapted SISR

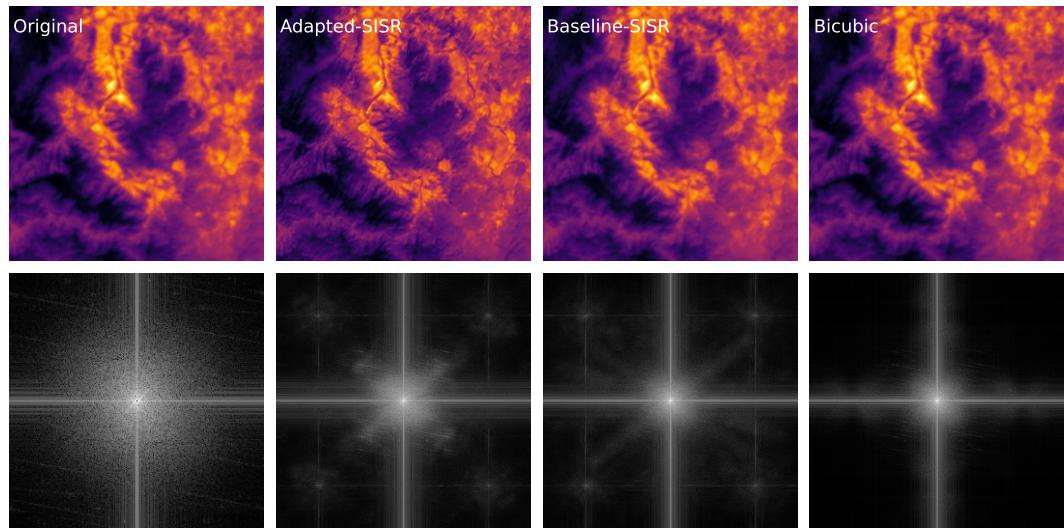


Figure 5.5: Super Resolved Forest-2 Scene using bicubic upsampling, the baseline blind SISR and the adapted approach.

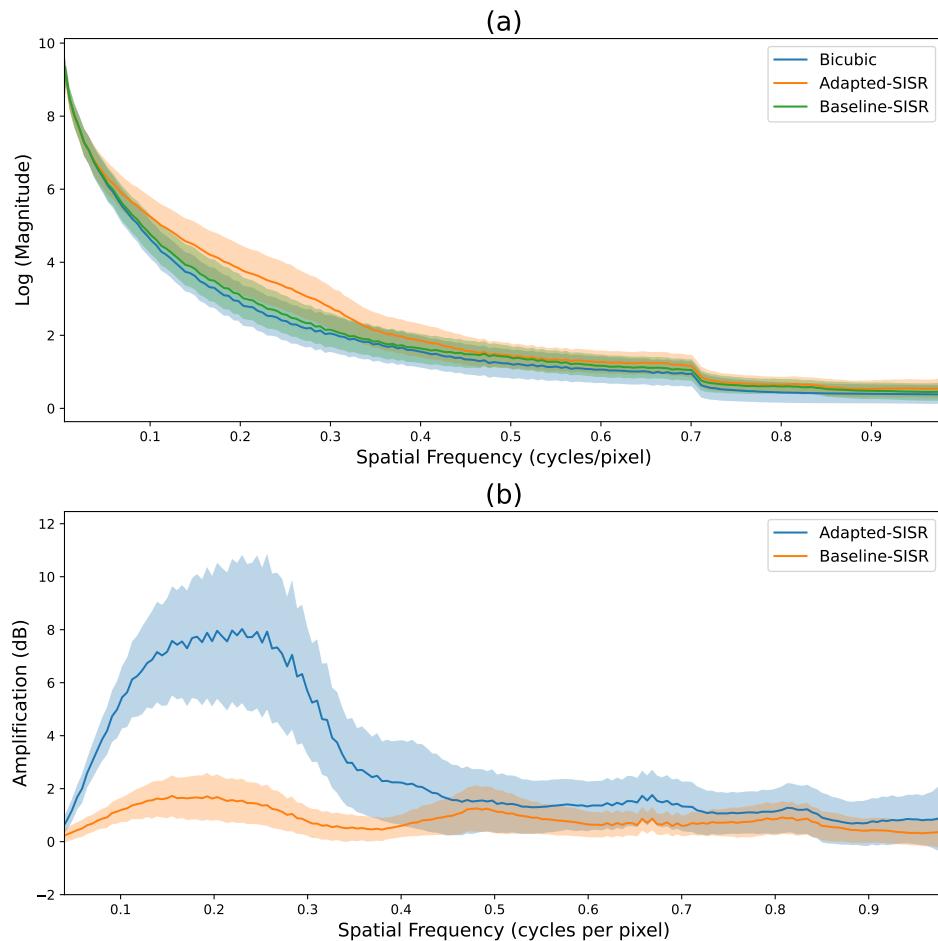


Figure 5.6

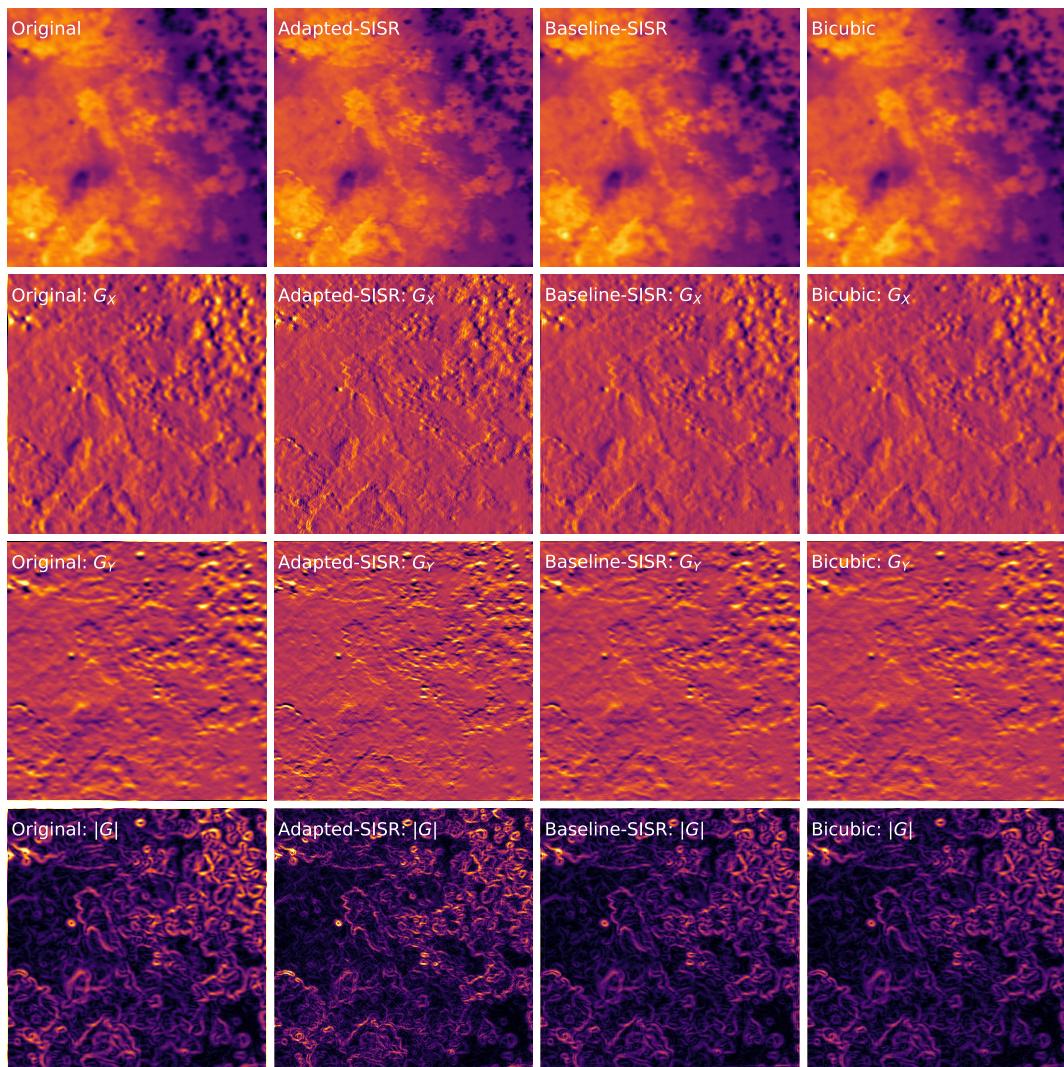


Figure 5.7: Sample of adapted SISR

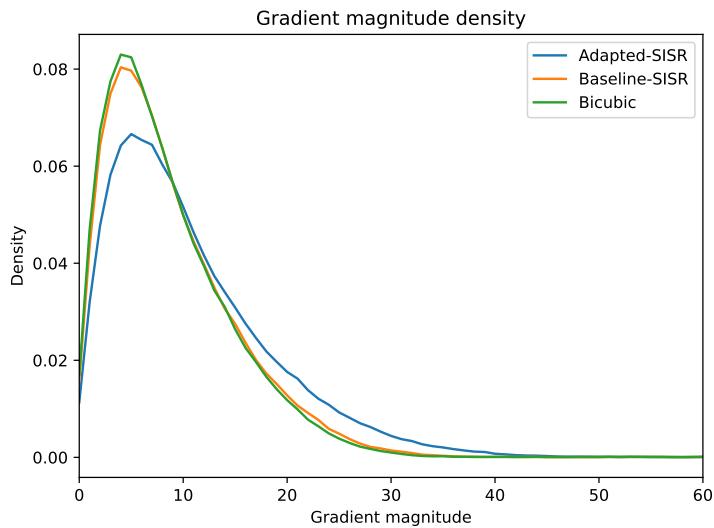


Figure 5.8: Histogram of the gradient magnitude $|G|$ for the sample displayed in Fig. 5.7

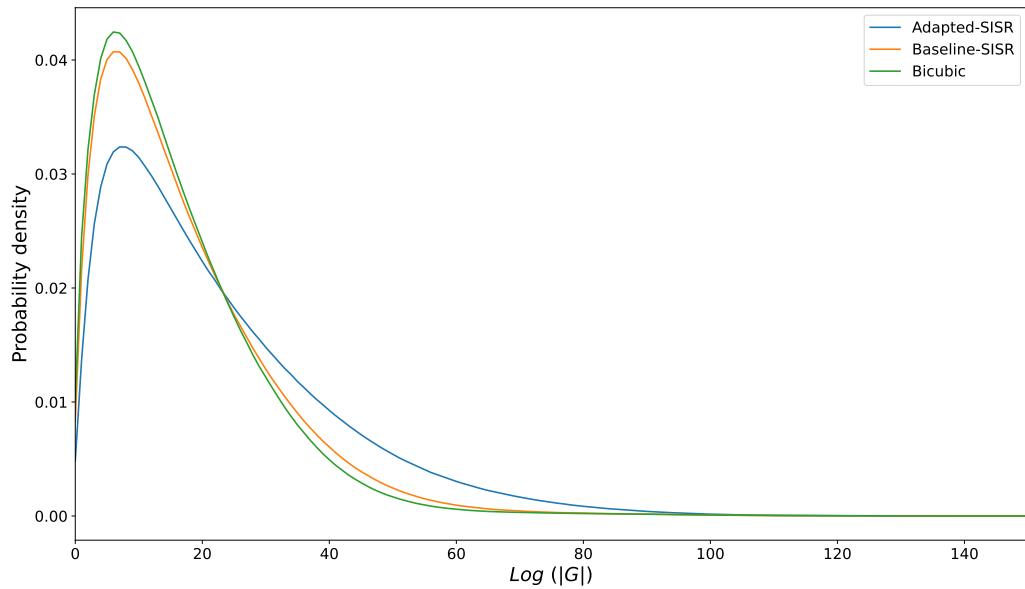


Figure 5.9

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 Gottfriesen, Martin Langer, Patricio Massaro, Christian Soraruf, and Matthias Schubert. Multi-spectral super-resolution of thermal infrared data products for urban heat applications, 2023.
- [8] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond, 2021.
- [9] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution, 2018.
- [10] 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.
- [11] 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.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [13] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Learning the degradation distribution for blind image super-resolution, 2022.
- [14] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer, 2013.
- [15] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution, 2021.

- [16] Irwin Sobel and G. M. Feldman. An isotropic 3×3 image gradient operator, 1990.
- [17] Jet Propulsion Laboratory. ECOSTRESS Fact Sheet, 2023. [Online accessed 28-November-2023].
- [18] PhyTIR: Plant High Temperature InfraRed Viewer. <https://phytir.jpl.nasa.gov/>, 2023. [Online accessed 28-November-2023].
- [19] Application for extracting and exploring analysis ready samples (AppEEARS). <https://appeears.earthdatacloud.nasa.gov/>, 2023. [Online; accessed 28-November-2023].
- [20] Appearers api. <https://appeears.earthdatacloud.nasa.gov/api/>, 2023. [Online; accessed 28-November-2023].
- [21] 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].
- [22] Ecostress faq. <https://ecostress.jpl.nasa.gov/faq>. Accessed: 2023-11-28.