University of Salerno

.DIEM

Department of Information and Electrical Engineering and Applied Mathematics

Ph.D. Program in Information Engineering
38th cycle

PH.D. THESIS

Single-image super-resolution of S5P radiance data through a PSF-aware and scalable spectral deep-learning-based approach

ALESSIA CARBONE

SUPERVISORS: **PROF. ROCCO RESTAINO, DR. GEMINE VIVONE**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE FOGGIA**

Year 2026

I dedicate this work to the one who knows me beyond my words, beyond my limits, and beyond my own understanding. I could never have dedicated this thesis to anyone if I were not truly, healthily, and enduringly in love. Yet, your presence has reshaped the way I see what is possible. Your love is a quiet force that steadies me, challenges me, and frees me all at once. Thank you for revealing that love has real, measurable consequences, yet it escapes any equation or law I know.

Abstract

Although the Sentinel-5 Precursor (S5P) satellite provides valuable global information on key air pollutants, its spatial resolution remains insufficient to capture their fine-scale distribution in confined areas such as cities, where concentrations are typically highest. Ground-based monitoring networks, while more detailed, are sparse or even absent in many regions, making satellite observations indispensable. Among current platforms, S5P offers the best compromise between spectral resolution — allowing the monitoring of multiple pollutants — spatial resolution, and global coverage. Nevertheless, physical constraints prevent improvements beyond its nominal spatial resolution, underscoring the relevance of image enhancement techniques.

This thesis presents the first application of a super-resolution approach — single-image super-resolution (SISR) — to S5P radiance data. We introduce S5Net, a lightweight deep-learning-based model that delivers excellent performance while remaining simpler than state-of-the-art architectures. In particular, we characterise S5P's payload by modelling its imaging acquisition process and generating realistic low-high resolution pairs, making the framework explicitly PSF-aware — an aspect often overlooked in remote sensing studies. Additionally, we propose a dynamic multi-directional cascade fine-tuning scheme that adaptively determines the number of iterations per channel from inter-band correlations. This approach preserves spec-

tral fidelity while optimising both spatial reconstruction and computational efficiency.

Our framework achieves excellent results. Across all orbits, PSF-aware SISR consistently outperforms existing methods, showing that explicitly modelling the sensor's degradation substantially improves spatial reconstruction. S5Net achieves the optimal balance between spatial detail and spectral fidelity, aligning quantitative metrics with visual quality. Furthermore, the proposed dynamic fine-tuning outperforms both PCA-based and static cascade approaches, achieving a superior balance between spatial enhancement, spectral consistency preservation, and computational efficiency. This results in particularly improved performance on spectrometer-partitioned subimages.

To demonstrate practical benefits, we additionally apply the framework to S5P imagery over a polluted urban area, extracting surface features from both original and super-resolved images to predict the oxidative potential (OP) of particulate matter. Predictions from super-resolved imagery consistently outperform those from nominal-resolution data. Since OP is a key indicator of particulate matter's toxicity, closely linked to adverse health effects, these results highlight the framework's potential for supporting urban exposure assessment.

By integrating sensor-specific characteristics and addressing the challenges of hyperspectral remote sensing, our PSF-aware and spectrally scalable SISR framework enhances S5P imagery in a problem-tailored way, enabling more accurate urban exposure assessments and supporting large-scale air quality monitoring.

Contents

List of Figures

List of Tables

List of acronyms

BD	band	26
CGA	conjugate gradient algorithm	98
CNN	convolutional neural network	43
DEM	detector electronics module	26
DL	deep-learning	42
ERGAS	erreur relative globale adimensionnelle de synthèse	94
ESA	European Space Agency	4
FOV	field of view	22
FR	full-resolution	50
GSD	ground sample distance	22
GSI	ground-projected sample interval	22
GT	ground-truth	43
HR	high-resolution	33
LR	low-resolution	33
MSE	mean square error	82
MTF	modulation transfer function	22
NIR	near-infrared	26
PCA	principal component analysis	52
PRF	pixel response function	28
PSF	point spread function	22
PSNR	peak signal-to-noise ratio	94
RMSE	root mean square error	94

Chapter 1

Introduction

This chapter introduces the overall context of the research, outlining the motivations behind the study, the research gaps it addresses, its objectives and original contributions, and the structure of the thesis. Section 1.1 provides a comprehensive overview of the research context, focusing on the relevance of air quality monitoring and the role of satellite observations, and highlights the motivation for enhancing their spatial resolution. Building on this, Section 1.2 examines in detail the main challenges and research gaps, with particular emphasis on the limitations of current super-resolution methods and the intrinsic difficulties of applying them to satellite data. Section 1.3 then presents the objectives and original contributions of this work, explaining the rationale for the chosen methodological approach and outlining the innovations introduced. Finally, Section 1.4 summarises the organization of the thesis, providing a clear roadmap for the subsequent chapters. In this way, the introduction connects the scientific context, the identified research gaps, and the research objectives with the background and literature review, as well as with the methodological and experimental work presented later. This highlights the coherent logic that underpins the

study and underscores the broader significance of its outcomes.

1.1 Context and motivation

According to the World Health Organization (WHO), air pollution is one of the leading environmental threats to human health, ecosystems, and climate. Exposure is increasing worldwide and is closely linked to ambient pollutant concentrations (World Health Organization, 2021).

The spatial and temporal distribution of air pollutants depends on multiple factors, including the geographical location of emission sources, their operational patterns (daily or seasonal cycles), the chemical and physical behaviour of each species (dispersion, deposition, and interactions), and meteorological conditions. While some pollutants are relatively homogeneous, concentrations are often highly heterogeneous and may vary considerably even across small spatial scales (World Health Organization, 2021).

Given the varied and critical impacts of air pollution, from health damages to economic losses and social inequalities, the availability of reliable and homogeneous monitoring systems is essential for developing effective policies. Without precise and continuous data on pollutant levels and their distribution, it is not possible to adequately assess population exposure or design emission reduction strategies (World Health Organization, 2021).

Population exposure data are generally obtained by local or national institutions using ground-level air-quality monitoring stations (AQMSs) (World Health Organization, 2021). These sites offer excellent temporal resolution, allowing for reliable hourly and daily averages of several contaminants. However, because they are costly, their distribution is uneven, especially

in rural areas and outside major cities (Siddique et al., 2024). As a result, even for the most frequently monitored pollutants, global coverage remains insufficient despite the increasing number of stations worldwide (World Health Organization, 2021).

Over the past two decades, satellite observations have become a cornerstone of air quality assessment. Satellite-derived estimates may diverge from surface measurements due to their column-integrated nature and the relatively coarse ground spatial resolution (or footprint) of current sensors (Siddique et al., 2024). Nevertheless, they are invaluable for assessing air quality in regions lacking dense monitoring networks and for investigating large-scale pollution patterns (Shaddick et al., 2018; Chang et al., 2019). The European Space Agency (ESA)'s Sentinel-5 Precursor (S5P) mission (European Space Agency, 2022) is specifically designed to monitor atmospheric composition and air quality on a global scale. With a footprint of 5.5×3.5 km^2 and a spectral range spanning from the ultraviolet to the short-wave infrared, S5P provides substantially higher spatial resolution and broader spectral coverage than earlier global-scale satellites, including NASA's Aura (Levelt et al., 2006), and ESA's ERS-2 (Burrows et al., 1999), MetOp (Munro et al., 2016), and ENVISAT (Bovensmann et al., 1999). Its resolution also approaches that of newer geostationary missions, such as GK-2B (Kim et al., 2020), NASA's TEMPO (Chance et al., 2019), and ESA's Sentinel-4 (Courrèges-Lacoste et al., 2008), which, however, do not provide global coverage. This combination of high spatial detail and capability for worldwide observations makes S5P the reference mission for high-resolution atmospheric monitoring.

Its data are routinely assimilated into major monitoring

systems and validated with ground-based stations (Inness et al., 2019; Borsdorff et al., 2018; Veefkind et al., 2012). Combining S5P data with statistical models improves surface-level pollutant estimates, outperforming previous satellites (Wu et al., 2021; Goldberg et al., 2021), and enables the study of seasonal and long-term air quality patterns (Riess et al., 2022; Mathew et al., 2024). Nevertheless, S5P cannot fully resolve intra-urban variability and local pollution hotspots, as each observation covers a footprint of about 5.5×3.5 km^2, reflecting an inherent trade-off between spatial and spectral resolution.

1.2 Gaps and challenges

Multiple studies have shown the potential advantages of higher-resolution satellite data. For instance, the use of fine-scale models can considerably enhance local air quality assessments (Neal et al., 2017). Likewise, capturing pollution patterns at finer scales allows for the identification of sources with the greatest health impacts, informing more effective mitigation measures (Goodkind et al., 2019). Furthermore, analyses indicate that the significance of spatial resolution is highest for local-scale studies. While coarse resolution may suffice for broad climate assessments, fine resolution is necessary to capture small-scale variations that are relevant for public health and urban planning (Hodnebrog et al., 2011; Siddique et al., 2024). These findings emphasise the need for methods that enhance the spatial resolution of current satellite observations, bridging the gap between their relatively coarse native resolution and the fine-scale detail required for precise and actionable air quality monitoring.

Advanced image processing techniques offer a potential

solution. Although they have proven effective in other domains, including remote sensing, their application to S5P data remains largely unexplored and challenging. Image enhancement can increase spatial detail, but it faces intrinsic limitations and technical difficulties. A prominent class of such techniques is super-resolution (SR), which encompasses algorithms that enhance an image's spatial resolution while preserving fine details. Essentially, SR simulates the result as if the image had been acquired by a sensor with higher nominal resolution, and is typically applied when no higher-resolution reference is available, as is the case for most satellite observations (Sdraka et al., 2022).

Most SR approaches lack a unique and stable solution though, and the absence of higher-resolution reference images further complicates performance assessment. Moreover, scaling factors in the satellite domain are often much larger than those in typical applications, and the combined spatial and spectral characteristics of satellite data can strongly influence model outcomes, rendering standard methods generally ineffective (Sdraka et al., 2022).

Dataset creation presents another critical challenge. Most training datasets are generated via simple downscaling of existing images, which does not capture the complex degradation processes inherent to satellite sensors. Models trained on such datasets may only reverse this artificial degradation, limiting their applicability to real-world scenarios. Addressing these issues requires models capable of handling diverse distortions, particularly for sensors with complex acquisition processes (Sdraka et al., 2022; Wang et al., 2021b). Although alternative training strategies have been investigated, they generally do not achieve the same level of accuracy and reliability as established approaches (Wang et al.,

2021b).

These challenges are further amplified in air quality monitoring, where imaging spans a very broad spectral range. Unlike natural images, these datasets comprise hundreds of spectral bands, each affected by band-specific noise, complicating feature extraction and reconstruction (Chen et al., 2023a). Enhancing such images requires maintaining spectral fidelity while improving spatial resolution, imposing substantial computational and methodological demands (Fernandez-Beltran et al., 2017).

1.3 Objectives and original contributions

Taken together, these considerations highlight the need for carefully designed approaches that enhance satellite data while preserving both spatial and spectral accuracy. Overcoming these limitations is essential for making higher-resolution satellite observations practically useful for environmental monitoring, policy-making, and public health studies, and directly motivates the objectives and the original contributions developed in this work.

This research aims to overcome the intrinsic limitations of S5P's observations for air quality monitoring by enhancing their spatial resolution while preserving spectral fidelity, taking advantage of its global coverage and comprehensive spectral range. This work represents the first application of these approaches to S5P data.

Also, it is the first to operate directly on Level-1b radiance measurements, rather than on Level-2 products representing near-surface pollutant densities. Focusing on Level-1b data

allows the study to generalise across various pollutants and to minimise the degradation introduced by the retrieval of Level-2 maps (European Space Agency, 2022).

As mentioned, real-world images are captured through complex, sensor-specific processes, and algorithms trained on artificially degraded datasets often fail when applied to actual satellite data (Wang et al., 2021b). To overcome this gap, our framework emphasises model-based techniques that incorporate the physics of the sensor's image acquisition, contrasting with much of the existing literature.

The high spectral dimensionality of S5P's data introduces additional challenges, including computational complexity and redundancy across hundreds of correlated bands. Traditional dimensionality reduction methods can either suppress low-probability features or compromise computational efficiency (Thenkabail and Lyon, 2016; Wang and Wang, 2014; Keshava and Mustard, 2002). Similarly, conventional band-wise similarity fine-tuning approaches often rely on a fixed number of iterations, ignoring the specific characteristics of the data (Guarino et al., 2024). To address this, we propose a fine-tuning strategy that dynamically adjusts the number of iterations according to the correlation among neighbouring channels, achieving a better trade-off between computational efficiency, spatial reconstruction quality, and spectral fidelity.

In addition, we develop a dedicated preprocessing pipeline for S5P radiance data, generating a dataset specifically tailored for learning-based SR. This component, motivated by the complexity and volume of hyperspectral measurements, addresses an aspect largely overlooked in previous studies and constitutes a valuable resource for subsequent research.

A preliminary generalization evaluation is included to verify the network's performance with synthetically generated

high-resolution references. A comprehensive evaluation is then performed both on reduced- and full-scale images, quantifying improvements in spatial and spectral accuracy and comparing the proposed framework with state-of-the-art methods and baselines. Special attention is given also to scenarios lacking reference data, highlighting the intrinsic challenges of applying these methods directly to operational satellite observations. The framework is further validated through a real-world case study, demonstrating its ability to enhance local air quality assessment and improve the prediction of health-relevant parameters, thereby supporting actionable insights for environmental monitoring and policy-making.

This research makes several original contributions to the field. First, it develops a deep-learning-based SR framework, specifically tailored for enhancing S5P's Level-1b data, addressing both spatial and spectral challenges. Second, it introduces an optimised tool capable of efficiently reconstructing high-resolution imagery while maintaining fidelity across multiple spectral channels, substantially reducing computational cost compared to processing each channel separately. Third, it implements a dedicated preprocessing pipeline for S5P radiance data, producing a novel dataset that supports learning-based SR and represents a resource for the community. Fourth, it employs a systematic evaluation protocol that accounts for the real-world characteristics of the sensor, including a complex degradation pattern, thus providing a more realistic assessment of performance than conventional benchmarks. Finally, it offers practical insights into the applicability and limitations of SR techniques for operational air quality monitoring, identifying scenarios where these methods yield significant benefits and outlining open challenges for future research.

1.4 Structure of the thesis

This thesis builds upon the findings of previous research and is structured to provide a coherent and progressive path from problem identification to methodological development, experimental evaluation, and practical application. The work advances both theoretical understanding and practical implementation of SR techniques for S5P data, tackling the unique challenges posed by satellite imagery, including high spectral dimensionality, sensor-specific degradation, and the need to preserve spectral fidelity while enhancing spatial resolution.

The thesis is organised as follows.

Chapter 2 introduces the scientific context, describing air pollution and monitoring limitations, satellite's characteristics and data and reviewing the literature in SR for remote sensing. This chapter highlights the main challenges in applying SR to satellite observations, such as the absence of reference images, the limitations of conventional dimensionality reduction approaches, and the computational demands of processing hundreds of spectral bands. It also identifies the gaps that this research aims to address, providing the foundation for the methodological choices made in the subsequent chapter.

Chapter 3 details the development of the proposed SR framework, focusing on model design, optimization strategies, and techniques for handling multi-channel data efficiently. Special attention is given to incorporating the physics of the S5P sensor into the framework, enabling more realistic reconstruction of images with higher spatial resolution and mitigating the effects of sensor-specific distortions. The chapter also discusses strategies for balancing computational

efficiency with reconstruction accuracy and the preservation of spectral fidelity, including a novel, parallelisable fine-tuning approach that dynamically adapts to the correlation structure across spectral channels.

Chapter 4 presents the experimental evaluation of the proposed framework. A comprehensive assessment protocol is employed, combining reduced-scale datasets with full-scale satellite observations to quantify improvements in both spatial and spectral accuracy. Comparisons with state-of-the-art methods and baseline models are provided to contextualise the performance of the proposed framework. The chapter also includes a real-world case study on urban air quality monitoring, demonstrating the practical applicability of the approach and its potential to support actionable insights for environmental monitoring and public health.

Chapter 2

Background and related work

In this chapter, we provide a comprehensive overview of the theoretical principles and prior work underpinning this thesis. First, in Section 2.1, the discussion begins with an exhaustive summary of air pollution, exploring both its impacts (Section 2.1.1) and the limitations of its monitoring (Section 2.1.2). In this context, the role of satellite data is highlighted as a critical tool, providing unique opportunities to quantify pollutant concentrations, although their effective use requires careful consideration. In Section 2.2, we discuss the fundamental aspects of satellite technology and sensor optics, introducing key concepts such as radiance measurement and instrument response (Section 2.2.1), as well as image acquisition and sensor resolution (Section 2.2.2). In Section 2.2.3, we focus on the S5P satellite platform, detailing its payload, and the associated image acquisition process. Building on this, the third part of the chapter (Section 2.3) reviews super-resolution techniques in remote sensing. We first formalise the problem (Section 2.3.1) and review the main state-of-the-art approaches, spanning interpolation-based

(Section 2.3.1.1), reconstruction-based (Section 2.3.1.2), and learning-based methodologies (Section 2.3.1.3), with particular emphasis on deep-learning-based techniques. We also discuss the specific challenges and limitations encountered when dealing with remotely sensed data, including sensor-specific degradation, image quality assessment, and dimensionality management for many spectral bands. This chapter thus provides the theoretical and literature background necessary to contextualise the methodologies and experimental analyses presented in the subsequent chapters.

2.1 Air pollution

The Clean Air Act (CAA) identifies six so-called criteria air pollutants for which the air quality standards are established: carbon monoxide (CO), sulfur dioxide (SO_2), particulate matter (PM)[1], classified in $PM_{2.5}$ and PM_{10}, according to the aerodynamic diameter of its particles in µm, lead (Pb), ozone (O_3), and nitrogen oxides (NO_x), like nitrogen dioxide (NO_2) (United States Congress, 1970).

Exposure to air pollution has increased particularly in low- and middle-income countries[2], and is closely linked to ambient pollutant concentrations (World Health Organization, 2021). People are exposed to air pollution in all the microenvironments in which they spend time, where a microenvironment refers to a three-dimensional space in which the pollutant concentration is assumed to be uniform at a

[1]PM, also called particle pollution, is a mixture of solid particles and tiny liquid drops floating in the air. Examples include dust, dirt, soot, and smoke (US Environmental Protection Agency, 2016).

[2]The grouping of countries in low, lower-middle, upper-middle, and high, based on their gross national income per capita, is decided by the World Bank (World Bank, 2021).

given time. Exposure can be conceptualised as the product of the pollutant concentration within a given microenvironment and the duration of contact with that pollutant. In some cases, concentrations may be relatively low, yet the overall contribution to exposure is substantial due to prolonged time spent in that environment. Conversely, in hotspots where pollutant concentrations are very high, even short periods of exposure can result in significant risks (World Health Organization, 2021).

Air pollution originates from both natural and anthropogenic sources, with the latter dominating in the era of industrialization. Combustion is the primary contributor: some pollutants are directly emitted, such as elemental carbon (the main constituent of PM), while others — nitrates, sulphates, organic carbon — form through secondary processes from gaseous precursors emitted by combustion, agriculture, or natural activity (World Health Organization, 2021). The major outdoor sources of combustion include transportation, industry, power generation, and biomass or waste burning. Other important contributors are the resuspension of dust, construction activity, and the long-range transport of pollutants from distant regions (World Health Organization, 2021).

At the global level, elevated pollutant concentrations are observed in regions with dense anthropogenic activity, such as South-East Asia, the Middle East, North America, and parts of Europe, as well as in arid areas, like North-Western Africa or the Gobi Desert, where dust plays a major role. In contrast, population-weighted exposure has decreased significantly in many countries, including North America and Western Europe (World Health Organization, 2021).

Different pollutants exhibit distinct spatial patterns. Nitrogen

dioxide (NO_2) levels are particularly high in Eastern Asia, the Middle East, North America and much of Europe (Larkin et al., 2017; Achakulwisut et al., 2019). Ozone (O_3) concentrations are generally lower than those of primary pollutants, with peaks reported in African megacities and the Middle East. Fine PM ($PM_{2.5}$) shows the widest global exceedance: in 2019, over 90% of the world's population was exposed to concentrations above the 2005 WHO guidelines (World Health Organization, 2021). The highest values occur in South-East Asia, the Eastern Mediterranean, and parts of Western Africa, especially due to the impact of Saharian dust, while the Americas and most of Europe report comparatively lower concentrations. Trends in $PM_{2.5}$ indicate a stable global mean population-weighted concentration, reflecting decreases in exposure in some countries and corresponding increases in others (World Health Organization, 2021).

Temporal variability is also marked. Seasonal and diurnal cycles as well as weekends versus weekdays patterns affect most pollutants, but meteorological drivers — such as wind, boundary layer height (i.e. the lowest altitude of the atmosphere directly affected by the Earth's surface), and precipitation — play a dominant role in shaping both local and regional distributions (World Health Organization, 2021).

2.1.1 Implications of air pollution

Air pollution contributes to millions of deaths annually and causes substantial loss of healthy life years (Burnett et al., 2018; Lelieveld et al., 2019; GBD 2019 Risk Factors Collaborators, 2020; Vohra et al., 2021). In 2019, mortality attributable to air pollution was estimated at 8.8 million per year globally, nearly double previous estimates (Lelieveld et al., 2019).

In particular, PM, nitrogen dioxide (NO_2), and ozone (O_3) are linked to mortality and non-communicable diseases[3], like lower respiratory infections, chronic obstructive pulmonary disease, ischaemic heart disease, lung cancer, and stroke (Cohen et al., 2017; Dominutti et al., 2023).

Emerging evidence links exposure to air pollution also to other types of diseases, such as type II diabetes, neonatal mortality, and neurological disorders, such as Alzheimer's disease (GBD 2019 Risk Factors Collaborators, 2020; Peters et al., 2019). Nitrogen dioxide (NO_2) alone accounts for a significant fraction of paediatric asthma, even in high-income countries (Achakulwisut et al., 2019).

Short-term exposure to PM triggers acute respiratory symptoms — such as cough, shortness of breath, and asthma — whereas long-term exposure increases the risk of chronic respiratory and cardiovascular diseases and reduces life expectancy (Lelieveld et al., 2019; Chen and Hoek, 2020). Additionally, PM can cause an imbalance in the antioxidant defence system, known as oxidative stress (Shankar and Mehendale, 2014), triggering inflammatory responses that may lead to tissue damage and cell death (Dominutti et al., 2023; Sies, 2018). To better capture these effects, novel health indicators have been proposed, such as the oxidative potential (OP) of PM, which reflects its capacity to induce oxidative stress in the lungs, the primary route of entry for its constituents. Several studies have shown that OP can, in some cases, be more predictive of adverse health outcomes than PM mass concentration alone (Bates et al., 2015; Yang et al., 2016; Abrams et al., 2017). Some studies have assessed the

[3]Non-communicable diseases, also known as chronic conditions, comprise all disorders related to environmental, behavioural, genetic, and physiological factors (World Health Organization, 2021).

correlation between PM and reduced foetal growth because of the oxidative stress it induces (Rashid et al., 2018). Similarly, some other researches have found that exposure to PM during pregnancy can damage the child's respiratory and immune systems and brain development (Johnson et al., 2021).

In general, the disease burden attributable to air pollution is comparable to that of other major global health risks, such as unhealthy diets and tobacco smoking, ranking among the top five out of 87 risk factors in the Global Burden of Disease assessment (GBD 2019 Risk Factors Collaborators, 2020).

The health-related economic consequences of air pollution are equally severe. Economic burdens arise from healthcare costs, damage to buildings, crops and infrastructures, climate change and environmental degradation, and losses in labour productivity (World Health Organization, 2021). In 2013, the World Bank estimated that global exposure to air pollution results in economic losses amounting to trillions of dollars annually. In the United States alone, $PM_{2.5}$ exposure was estimated to cause, in 2013, about 143 billion dollars in lost labour and 3.55 trillion dollars in welfare losses (World Bank, 2016).

The health burden of air pollution is highly variable across different people. Vulnerable groups, including children, the elderly, pregnant people, and individuals with pre-existing chronic conditions — such as asthma, chronic obstructive pulmonary disease, diabetes, heart failure, and ischaemic heart disease — experience disproportionately higher risks (World Health Organization, 2021).

Similarly, air pollution's impacts are also shaped by social inequalities. Populations with higher exposure often overlap with socio-economically disadvantaged groups, and, while high-income countries have generally reduced air pollution

since the 1990s, concentrations have increased in low- and middle-income countries, partly due to the production of goods for export to wealthier nations (Zhang et al., 2017). Consequently, air pollution not only directly harms human health but also hinders efforts to reduce social inequalities and to ensure healthy and safe environments, as advocated in Articles 1–3 of the Universal Declaration of Human Rights (United Nations, 1948).

2.1.2 The limitations of air pollution monitoring

Two major gaps remain in current air quality monitoring.

The first concerns the insufficient monitoring of air quality in many regions worldwide, particularly in low- and middle-income countries, where pollutant concentrations tend to be highest. Even in countries with extensive monitoring networks, rural areas often remain poorly covered. While monitoring of PM has been gradually expanding, coverage for other key pollutants — such as ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) — remains limited. This lack of comprehensive monitoring hinders the accurate assessment of population exposure and the identification of pollution hotspots (World Health Organization, 2021).

The second gap concerns the fine-scale spatial variability of certain gaseous pollutants within urban environments, which remains inadequately characterised. For example, concentrations of nitrogen dioxide (NO_2), black carbon — a major component of PM — and ultra-fine particles (UFP) (PM with aerodynamic diameter less than or equal to 0.1 µm) can vary by up to an order of magnitude over distances of only a few hundred metres (Karner et al., 2010). Such local variability

poses a significant challenge for exposure assessment (World Health Organization, 2021).

Since 2010, the integration of satellite retrievals with chemical transport models, which simulate the atmospheric transport and chemical transformation of pollutants, land-use information, and ground-based measurements has dramatically improved the ability to estimate pollutant concentrations on a global scale. These combined approaches have proven particularly effective in addressing the first gap — limited monitoring coverage — by providing consistent estimates in regions where ground-based measurements are sparse or unavailable (Shaddick et al., 2018; Chang et al., 2019).

To tackle the second gap, related to fine-scale spatial variability within urban areas, satellite data have increasingly been combined with land-use regression models, i.e., statistical models that relate ground-based pollutant measurements to spatial predictors. This integration allows for the capture of local pollution patterns by leveraging both high-resolution spatial predictors and remotely sensed observations, enabling a more detailed assessment of population exposure in complex urban environments. Yet, important challenges still remain in fully addressing this gap. While the integration of satellite observations with land-use regression models has markedly improved spatial resolution, major limitations persist due to data heterogeneity, the non-linear nature of the underlying relationships, and the limited availability of ground-based observations required to calibrate these models (Larkin et al., 2017).

2.2 Principles of satellite technology and optics

Remote sensing is the process of measuring features on the Earth's surface from a distance. Since direct contact with the object of interest is not possible, measurements rely on propagated signals, such as optical signals. While point-based measurements can be derived from remote sensing data, two-dimensional measurements — images — are of greater relevance. Indeed, the conventional, or more accurately, image-centred, approach to remote sensing places emphasis on the spatial relationships between ground features (Schowengerdt, 2007).

Satellites are remote sensors that orbit the Earth and are commonly classified as either passive or active. Active satellites, such as those using radar technology, transmit a signal and measure its return, whereas passive satellites detect electromagnetic radiation originating from the Sun and/or emitted or reflected by the Earth's surface and transmitted through the atmosphere (Schowengerdt, 2007).

2.2.1 Measurement chain and sensor responses

Although image sensors continue to improve, no device can capture a continuous physical signal with infinite precision. The combination of scanning in both the cross-track (orthogonal to the satellite motion) and in-track (along the satellite motion) directions yields a grid of discrete samples on the instrument's focal plane. The spectral radiance measured by passive satellites is converted by the sensor into an image of its spatial distribution. During this process, the signal

undergoes various modifications in its geometric, spatial, and radiometric properties, with each stage of the chain inevitably introducing distortions or losses. The overall workflow is illustrated in the block diagram in Figure 2.1. In particular, the continuous at-sensor radiance is first mapped onto the optical detectors (scanning; in blue), then converted into an analogue electronic signal (detection; in green), subsequently amplified and conditioned (electronic processing; in magenta), and finally digitised into discrete digital numbers representing the image pixels (A/D conversion; in yellow).

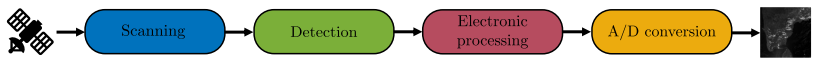

Figure 2.1: Block diagram representing the processing pipeline the radiance signal undergoes to become an image in a passive satellite.

Each pixel in the resulting image thus corresponds to a digital sample of the continuous radiance stream, whether in space (x, y), wavelength (λ), or time (t). The integrated signal is then stored as an integer value with a specified bit depth (q), providing the digital representation of the measured radiance (Schowengerdt, 2007).

The instrument must average over a non-zero integration time if the signal changes in time, a non-zero spectral bandwidth if the signal varies in wavelength, and a non-zero spatial distance if the signal varies in space. In other words, the output of most satellite instruments can be expressed as

$$o(z) = i(z) * r(z), \tag{2.1}$$

where $i(z)$ is the input signal, $r(z)$ is the instrument response, and $*$ denotes the convolution operation. The response refers to spatial, spectral, temporal, or radiometric integration, with z indicating the corresponding domain (Schowengerdt, 2007).

A key aspect in image processing is the instrument's spatial response. Fine details in the image are blurred compared to larger features, and this blurring is represented by the total sensor point spread function (PSF). The total PSF is the convolution of multiple components corresponding to different stages of the signal degradation (Figure 2.1): the optical, the detector's, the image motion's, and the electronic PSFs. Applying the Fourier transform to the total PSF and taking its amplitude component, the modulation transfer function (MTF) of the instrument can be obtained, which describes how the instrument transfers contrast at different spatial frequencies from the scene to the image, and therefore quantifies its ability to resolve fine spatial details (Schowengerdt, 2007).

2.2.2 Pixel properties of satellite images and the concept of resolution

The scanning of the instrument depends on the technology employed by the satellite: a line scanner uses a single detector to scan the entire scene; a whiskbroom scanner employs several detectors aligned in-track to achieve parallel scanning during each cycle; and a pushbroom scanner uses a linear array of detectors aligned cross-track that scan during platform motion. In any case, the angular cross-track coverage is called field of view (FOV), and the corresponding length on the ground is usually referred to as the swath. The spacing between pixels on the ground is called ground-projected sample interval (GSI), and its value, both in-track and cross-track, depends on the sampling rates and platform motion. The width of a pixel on the ground, known as ground sample distance (GSD), depends primarily on the sensor geometry and satellite

altitude. Due to the Earth's curvature, the effective GSD may vary across the swath, especially towards the edges. Satellites are usually designed so that GSI and GSD coincide to avoid overlapping pixels, but some degree of cross-track overlap is often introduced to improve data quality (Schowengerdt, 2007). Figure 2.2 schematically compares the three scanning technologies and illustrates how cross-track direction, in-track direction, swath, FOV, GSD, and GSI are defined in relation to the sensor geometry.

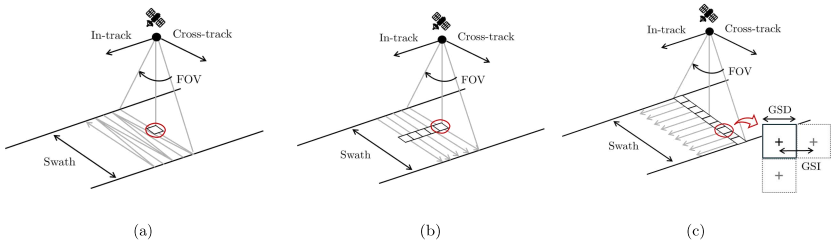

Figure 2.2: Illustration of the three main scanning technologies employed in passive satellite instruments — (a) line scanner, (b) whiskbroom scanner, and (c) pushbroom scanner — together with the representations of cross-track and in-track directions, swath, FOV, GSD, and GSI.

In summary, an image pixel in satellite data depends primarily on GSI, GSD, and the radiometric quantization level (q). Its spectral characteristics are determined by the specific wavelength ranges and bandwidths defined by the sensor, while its temporal characteristics are governed by the satellite's revisit time, that is, the interval required to pass over the same geographical location (Schowengerdt, 2007).

For these reasons, satellite imagery is characterised by its spatial, spectral, temporal, and radiometric resolutions. The smaller the area covered by a single pixel, the better the nominal spatial resolution of the sensor. Similarly, the narrower the frequency bands of the electromagnetic spectrum detected, the higher the nominal spectral resolution. The

shorter the revisit time of the satellite over the same location, the better its temporal resolution. The more bits used to represent a single pixel, the higher the radiometric resolution (Schowengerdt, 2007).

Even though image sensors are constantly upgraded, physical and engineering constraints remain (Fernandez-Beltran et al., 2017). In particular, enhancing one resolution often comes at the expense of another. For example, increasing spectral resolution by narrowing the bands or adding more channels requires longer integration times or more light to achieve a sufficient signal-to-noise ratio (SNR), since the incoming signal is divided among a larger number of bands. Such compensation is commonly achieved by enlarging the sampling area, which in turn degrades the spatial resolution. This trade-off is especially critical in hyperspectral imaging systems, namely, systems that can encompass hundreds or even thousands of spectral bands (Schowengerdt, 2007).

2.2.3 Overview of the Sentinel-5P satellite

Launched in 2017 within the Copernicus Programme[4], Sentinel-5 Precursor (S5P) is a single-satellite mission that measures most criteria air pollutants, namely, nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO), formaldehyde (HCHO), methane (CH_4), and aerosol, i.e., tiny solid or liquid particles suspended in the atmosphere (European Space Agency, 2022).

S5P is orbiting in a near-polar sun-synchronous orbit[5] at

[4]Copernicus is a collaborative programme of the European Union and the ESA, designed to ensure Europe's capacity for Earth observation (European Space Agency, 2022).

[5]A near-polar sun-synchronous orbit is an orbit that passes near the Earth's poles and crosses the same part of the Earth at approximately

an altitude of 824 km, completing an orbital cycle of approximately 16 days (European Space Agency, 2022).

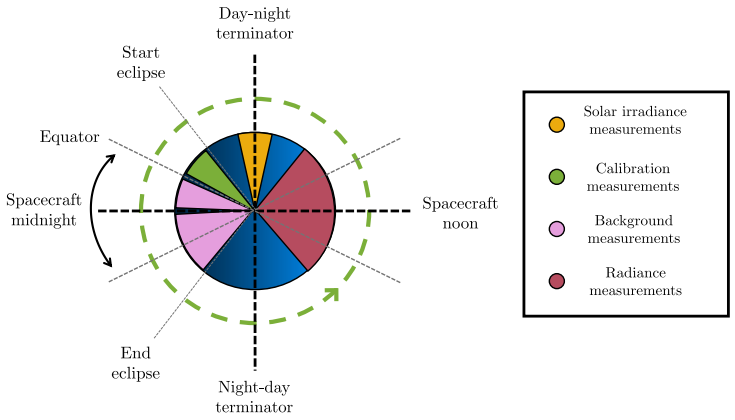

Figure 2.3: Graphical representation of the S5P orbit, defined from spacecraft midnight to spacecraft midnight. The diagram highlights the measurements' scheme along the orbit.

An S5P orbit, schematically illustrated in Figure 2.3, is defined from spacecraft midnight to spacecraft midnight. Since the orbits are defined without seasonal dependencies and certain measurements must be conducted in full eclipse conditions, the orbit measurements are performed as follows: Earth radiance measurements (in magenta) are acquired on the day side of the orbit, while near the day-night terminator, when the Sun is visible in the instrument's solar port (i.e., once per day, or after 15 orbits), a solar irradiance measurement (in yellow) is performed. Conversely, the night side of the orbit is used for calibration (in green) and background measurements (in pink) (European Space Agency, 2022).

2.2.3.1 TROPOspheric Monitoring Instrument

The TROPOspheric Monitoring Instrument (TROPOMI) is the sole payload of S5P and it operates in a pushbroom

the same local solar time (European Space Agency, 2022).

configuration, as represented in (c) in Figure 2.2. Its FOV spans 108° in the cross-track direction, corresponding to a swath of approximately 2600 km on the Earth's surface, and 48° in the along-track direction. The minimum GSD is 3.5×5.5 km^2 (European Space Agency, 2013).

The sources of light collected by the TROPOMI telescope include the Earth, the Sun, and three on-board calibration sources: a spectral line source, a white light source, and a common LED. These are fed through a calibration unit. Its telescope is a reflective two-mirror design that projects the target scene onto a rectangular slit serving as the entrance of the spectrometer system. The instrument comprises four spectrometers covering nearly 4000 overlapping spectral bands: (i) the ultraviolet (UV) spectrometer (0.270–0.320 µm), (ii) the ultraviolet and visible (UVIS) spectrometer (0.320–0.490 µm), (iii) the near-infrared (NIR) spectrometer (0.710–0.775 µm), and (iv) the short-wave infrared (SWIR) spectrometer (2.305–2.385 µm) (European Space Agency, 2013).

Each spectrometer is equipped with its own detector electronics module (DEM) (European Space Agency, 2013), a two-dimensional frame-transfer sensor with a split read-out register. The spectral and spatial dimensions of the spectrometer's output are mapped onto the columns and rows of the corresponding DEM, respectively. For a pre-defined number of steps, once a line enters the read-out register, its charge is added to the accumulated charge for a fixed number of repetitions. This irreversible process is referred to as row-binning. When the accumulated charge is finally shifted, it is read out pixel by pixel (column by column) through the DEM's output node. Each DEM has two output nodes, splitting the sensor into two halves known as bands (BDs), numbered from 1 to 8 (two per spectrometer). A fixed

number of consecutive DEM images are then co-added during a pre-defined co-addition period, further reducing the data volume to be transmitted to the ground. As with row-binning, this process is also irreversible (European Space Agency, 2013).

After read-out, the signal from each DEM's band passes through a video amplifier with a fixed gain and is then processed by a semiconductor integrated circuit (European Space Agency, 2013).

TROPOMI's Point Spread Function (PSF) and Pixel Response Funtion (PRF): To grasp the origin of TROPOMI's detected signal, it is essential to define the different functional areas of the instrument's observing system:

- The area that the optical system projects onto the DEM focal plane is known as the object space. It is described using spherical coordinates $(\phi, \epsilon, \lambda)$, where ϕ is the azimuth angle (observation angle in the cross-track direction), ϵ is the elevation angle (observation angle in the along-track direction), and λ is the wavelength.

- The DEM focal plane, referred to as the DEM space, is described by the row and column indices $[r, c]$, where rows correspond to the spatial dimension and columns to the spectral dimension.

- The optical system's focal plane is called image space. Geometric coordinates $[X, Y]$, with X in the direction of wavelength dispersion and Y perpendicular to X, are used to characterise this space (European Space Agency, 2013).

All three spaces are represented in Figure 2.4.

The projection of object space onto image space is referred to as spreading: a single point in object space is dispersed over

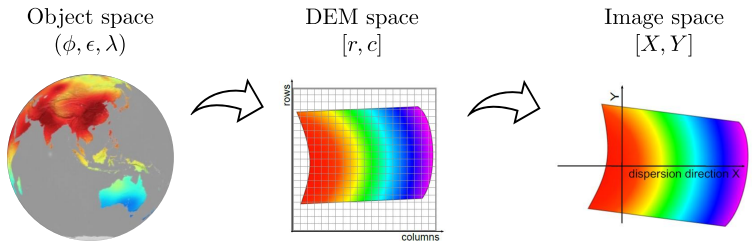

Figure 2.4: Schematic representation of the three reference spaces of TROPOMI: the object space, described in spherical coordinates $(\phi, \epsilon, \lambda)$; the DEM space, defined by row and column indices $[r, c]$; and the image space, characterised by geometric coordinates $[X, Y]$. The three illustrations are taken from European Space Agency (2013).

multiple locations in image space due to physical constraints. Conversely, the projection of image space onto object space is referred to as response: a single point in image space results from contributions from several points in object space (European Space Agency, 2013).

TROPOMI's pixel response function (PRF), $\Psi_{prf}[r, c](\phi, \epsilon, \lambda)$, is the function $\Psi(\cdot)$ describing the relative response of a single DEM pixel $[r, c]$ as a function of the object space dimensions $(\phi, \epsilon, \lambda)$ (European Space Agency, 2013).

The instrument transfer function provides a mathematical description of the mapping between unequal dimensions. As discussed in Section 2.2.1, in optical imaging systems, it is the Fourier transform of the PSF, which characterises the projection of object space $(\phi, \epsilon, \lambda)$ onto image space $[X, Y]$, which may differ from the DEM space $[r, c]$. The magnitude component of this complex function corresponds to the MTF (European Space Agency, 2013).

The image space does not necessarily coincide with the DEM space. Several processes thus influence the responses of image pixels for TROPOMI:

- Satellite motion: assuming a spherical Earth and constant satellite speed, this elongates the spatial response along-track (European Space Agency, 2013).

- Row-binning: the programmable summation of measurements in the cross-track direction. A single row in image space corresponds to the sum of multiple rows in DEM space. The ground sample interval in the cross-track direction is 1.75 km, but row-binning typically reduces the resolution, with a lower limit of 3.5 km for a binning factor of two (European Space Agency, 2013).

- Co-addition: the programmable summation of measurements along-track over a predefined co-addition period. Initially, light from the entire swath was co-added over approximately 1080 ms, during which the satellite travelled about 7.1 km along-track. The co-addition period was later set to 840 ms, resulting in a minimum along-track ground pixel size of 5.5 km (European Space Agency, 2013).

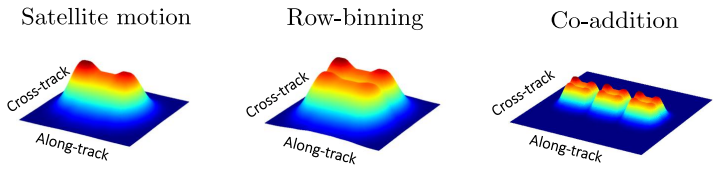

Figure 2.5: Illustration of how the spatial response in TROPOMI is affected by satellite motion (along-track smearing), row-binning (cross-track summation), and co-addition (along-track summation). The three illustrations are taken from European Space Agency (2013).

Therefore, to fully describe the PSF for TROPOMI, the PRFs of all individual DEM pixels $[r, c]$ contributing to each image

pixel $[X, Y]$ must be combined, incorporating all processes between DEM space and image space. This formalization defines the spatial response of a single image pixel during a co-addition period (European Space Agency, 2013). The combined effect of these three processes on the spatial response of TROPOMI is illustrated in Figure 2.5.

TROPOMI's Level-1b radiance images: The Level-1b radiance products of S5P's TROPOMI instrument have been freely available from 2019 onward through the Copernicus Data Space. These products are provided for each orbit and BD of each DEM in the netCDF-4 format[6] (European Space Agency, 2022).

A single digital number is generated from the radiance entering TROPOMI. The forward model, which combines physical principles with the instrument model, describes the generation of raw Level-0 data from the incoming signal. These Level-0 data are not publicly available.

The Level-0 to Level-1b (L01b) processor implements the reverse model, consisting of a series of sequential algorithms that convert the measured signal leaving the instrument back into the radiance entering it. This reverse model represents an attempt to invert the forward signal path entirely. However, some steps cannot be fully inverted due to information losses inherent in the forward model (European Space Agency, 2013).

Two examples of processes and instrument-specific effects that cannot be directly corrected by inverting the forward model are co-addition and row binning. In co-addition, several measurements are combined into a single value, while in row

[6]netCDF-4 (network common data form-version 4) is a widely used data format designed for the storage and sharing of large scientific datasets, particularly those organised in multiple dimensions (Rew and Davis, 1990).

binning, signals from adjacent DEM's rows are averaged. In both cases, the reconstruction assumes the average of the original image.

The Earth's spectral radiances, geolocated and radiometrically corrected, are provided as functions of four dimensions: measurement start time, cross-track index, along-track index, and spectral index. Each radiance value is accompanied by estimated latitudes and longitudes and metadata describing the instrument configuration during the orbit. Each unique set of programmable instrument settings — such as the number of measurements co-added and the row-binning factor — is identified by a configuration ID and version. This allows the instrument to operate in a wide variety of modes and configurations (European Space Agency, 2022).

The output of each ground pixel includes a quality indicator reflecting the status and reliability of the retrieval. The quality assurance value is a continuous variable ranging from 0 (error, resulting in no output) to 1 (no errors or warnings). It indicates whether the swath is cloud-covered or whether snow or ice is present on the surface. It is set to 0 if an error occurs at any point during processing (European Space Agency, 2022).

Each orbit spans the eight BDs (BD 1 to BD 8), each covering approximately 500 spectral channels and corresponding to half of a DEM — either UV, UVIS, NIR, or SWIR. In total, an orbit consists of 3,942 spectral channels, making TROPOMI's radiance images hyperspectral, with each pixel covering a minimum ground area of 3.5×5.5 km^2 (European Space Agency, 2022).

This limitation is directly linked to the second gap in air pollution monitoring discussed in Section 2.1.2, namely the difficulty satellites face in capturing fine-scale spatial variability in limited areas such as cities. As detailed in

Section 2.2.2, the nominal spatial resolution of hyperspectral sensors is constrained by physical limits (Schowengerdt, 2007). Although S5P represents one of the best compromises, this constraint highlights the importance of dedicated image processing techniques to enhance its effective spatial resolution.

2.3 Super-resolution techniques in remote sensing

Super-resolution (SR) refers to a class of algorithms designed to enhance the spatial resolution of images, generating higher-resolution outputs from lower-resolution inputs. In remote sensing, SR techniques are widely used to improve the spatial detail of satellite imagery (Fernandez-Beltran et al., 2017). Depending on the number of input images, SR methods can be classified into various categories. Three subfamilies exist:

- Single-image super-resolution (SISR) aims to extract a high-resolution (HR) version of a single low-resolution (LR) input image.
- Multi-image super-resolution (MISR) aims to synthesise a single HR image by taking different LR images of the same scene from various viewpoints and angles as input.
- Reference super-resolution (RefSR) combines auxiliary (reference) LR information with the model's input to offer additional details that aid in the process (Fernandez-Beltran et al., 2017).

Since the literature on each of the above categories is extensive, we will focus solely on SISR in this thesis, as it is the most

relevant one in our research. Indeed, SISR, in which only a single input image is used, is the only viable approach when no external information is available beyond an individual image captured at a specific location and time. This situation is common in remote sensing and is certainly the case for S5P (Fernandez-Beltran et al., 2017). From now on, we denote vectors by bold lowercase letters (e.g., $\mathbf{a}$), matrices by bold uppercase letters (e.g., $\mathbf{A}$), and scalars by italic letters (e.g., a).

2.3.1 Single-image super-resolution problem formalization

In mathematical terms, the SISR problem can be described as the attempt to recover an original high-resolution (HR) image $\mathbf{H} \in \mathbb{R}^{rH \times rW \times C}$, from the sole knowledge of one low-resolution (LR) image $\mathbf{L} \in \mathbb{R}^{H \times W \times C}$, where H, W, C are, respectively, the height, width and number of channels of the image, and r is the spatial scaling ratio between the LR and HR domains (Hu et al., 2022b).

The image given by the satellite $\mathbf{L}$ can be considered to be the result of the following process applied to an ideal HR image $\mathbf{H}$:

$$\mathbf{L} = \Phi(\mathbf{H}, \theta_\eta), \tag{2.2}$$

where $\Phi(\cdot)$ is the degradation process and θ_η are the parameters in the degradation process (Ye et al., 2023). The SISR task is to reconstruct a HR image $\widehat{\mathbf{H}} \in \mathbb{R}^{rH \times rW \times C}$ from the available LR image $\mathbf{L} \in \mathbb{R}^{H \times W \times C}$. This process can be expressed as follows:

$$\widehat{\mathbf{H}} = \Phi^{-1}(\mathbf{L}, \theta_\zeta), \tag{2.3}$$

where $\Phi^{-1}(\cdot)$ is the function solved backward from the degradation process, and θ_ζ are the parameters in the function

solved backward (Ye et al., 2023).

Certainly, the most challenging task in SISR is constructing the inverse problem. Most studies, however, represent the degradation function $\Phi(\cdot)$ as a sequence of consecutive operations that mimic the acquisition process of the sensing technology (Fernandez-Beltran et al., 2017). In this fashion, the c-th channel, with $c \in \{1, \ldots, C\}$, of the available LR image, $\mathbf{L}_c \in \mathbb{R}^{H \times W \times 1}$, is typically modelled as

$$\mathbf{L}_c = (\mathbf{H}_c * \mathbf{K}) \downarrow_r \ + n, \tag{2.4}$$

where $\mathbf{H}_c \in \mathbb{R}^{rH \times rW \times 1}$ is the corresponding channel of the HR image, $* \mathbf{K}$ is the blurring operation, i.e., the convolution $(*)$ with a blurring kernel $(\mathbf{K})$, $\downarrow_r$ is the down-scaling of a factor r, and n represents the noise (Fernandez-Beltran et al., 2017). In the specific context of remote sensing, as already explained in Section 2.2.1, the blurring kernel corresponds to the total point spread function (PSF) of the sensors (Schowengerdt, 2007). If the performance is not solely limited by diffraction and noise arises, the PSF can be modelled as a zero-centred Gaussian function, whose standard deviation depends on the specific sensors (Fernandez-Beltran et al., 2017). A graphical representation of the imaging model in Equation 2.4 is given in Figure 2.6.

The main aim of SISR is obtaining one estimated HR image, $\widehat{\mathbf{H}} \in \mathbb{R}^{rH \times rW \times C}$ from the available LR image $\mathbf{L} \in \mathbb{R}^{H \times W \times C}$. However, as the number of pixels in $\mathbf{L}$ is r times lower than the pixels in $\mathbf{H}$ the SISR problem is intrinsically ill-posed, namely, there exist countless HR images $\widehat{\mathbf{H}}$ that can obtain the same LR image $\mathbf{L}$ from a single imaging model (Fernandez-Beltran et al., 2017).

Over the years, a variety of techniques have been proposed to solve the SISR problem, and many relevant surveys

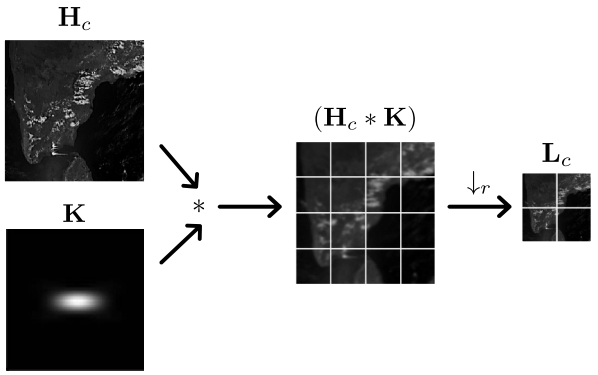

Figure 2.6: Graphical representation of the imaging model in Equation 2.4, showing how the c-th channel of the LR image results from the sensor's degradation of the corresponding c-th channel of an ideal HR image.

have been published, for all types of images (Wang et al., 2021b; Moser et al., 2023, 2024), for generally remotely sensed images (Fernandez-Beltran et al., 2017; Sdraka et al., 2022; Wang et al., 2022b) or specifically for hyperspectral images (Chen et al., 2023a). The research is thus quite active with about 100,105 papers published in all fields and 9,452 in remote sensing solely[7].

Proposed SISR methods are commonly classified according to the domain in which they operate (e.g., spatial or frequency), the theoretical framework used to formulate the problem, or specific process parameters. An example of the latter is the distinction between blind and non-blind SISR, depending on whether the blurring kernel $\mathbf{K}$ in Equation 2.4 is known. However, since many strategies incorporate traits from multiple categories, it is difficult to assign them unambiguously

[7]These numbers were obtained in 2025 from ScienceDirect database by querying it, firstly, through the keyword *single-image super-resolution* and, secondly, with the combination of keywords *single-image super-resolution* and *remote sensing*.

within existing taxonomies. We therefore adopt a flexible classification based on the functional characteristics of the methods, with the main categories presented in the following sections (Sections 2.3.1.1–2.3.1.3) (Fernandez-Beltran et al., 2017). Figure 2.7 provides a schematic representation of how these approaches operate.

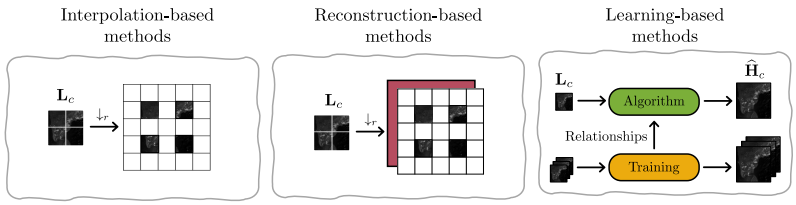

Figure 2.7: Schematic representation of the main categories of SISR approaches — interpolation-based, reconstruction-based and learning-based — illustrating their operational principles.

2.3.1.1 Interpolation-based methodologies

Initially developed techniques to handle SISR were interpolation-based approaches. Interpolation of images strives to expand an LR image on an HR grid and to estimate missing pixel values using an interpolation function, as represented in the first block of Figure 2.7. Interpolation is rarely regarded as a SR approach since it does not extract high-frequency information (Fernandez-Beltran et al., 2017). However, the HR image can be effectively determined by interpolating the expanded version of the LR image by a factor r (Aiazzi et al., 2013).

Indeed, from a theoretical standpoint (Shannon, 1948), if the original signal is band-limited and the sampling frequency exceeds the Nyquist rate, the missing pixels might be correctly recovered using a sinc function, i.e., an ideal low-pass filter. This requires infinitely large images and an endlessly long

interpolation filter, though. Since this is not feasible in real-world circumstances, various alternate techniques, such as local interpolators, are commonly used (Aiazzi et al., 2013).

The ideal sinc function can be approximated by piece-wise local polynomials passing through $p + 1$ samples, where p is the degree of the polynomial. Depending on this number, different interpolators are distinguished; some examples are the nearest-neighbour ($p = 0$) approximation, the linear ($p = 1$) interpolator, the quadratic ($p = 2$) kernel, the cubic ($p = 3$) approximation (Aiazzi et al., 2013; Keys, 1981; Li and Orchard, 2001), and a quite popular polynomial kernel with 23 taps (Vivone et al., 2021), which is applied $\log_2 r$ times through subsequent interpolations of factor 2. Since nearest-neighbour and quadratic interpolators have even symmetry, for interpolations with even factors, the interpolated images have fixed half-pixel shifts. Instead, regardless of the interpolation factor, all other interpolators are implemented as odd-length filters. As a consequence, they can produce images that are considerably closer to the originals (Aiazzi et al., 2013).

Ideal sinc interpolating functions may also be windowed, which would involve cutting off their infinite impulse response. Lanczos interpolator (Duchon, 1979) is an example of windowing in a certain interval $[-a, a]$ the sinc function. However, even if this kernel can be extremely useful for interpolation, the sinc function can easily result in an oscillating signal that causes Gibbs' ringing artefacts, i.e., sharp jumps, to arise in the resulting signal (Aiazzi et al., 2013).

Despite their speed and simplicity in providing higher-resolution results, interpolation algorithms cannot extract high-frequency information and they thus provide solutions with overly smooth textures (Fernandez-Beltran

et al., 2017).

2.3.1.2 Reconstruction-based methodologies

More efficient methods are reconstruction-based approaches. These techniques are intended to yield a final super-resolved image with greater detail while retaining all the visual characteristics of the LR image. In other words, reconstruction-based approaches seek to replicate the quality of the LR domain, followed by an initial interpolation, and to overcome the blurring and aliasing challenges generated by interpolation algorithms in the first place (Fernandez-Beltran et al., 2017), as represented in the second block of Figure 2.7.

This type of process usually consists of three phases:

1. The first step is the up-scaling of the LR image to the target resolution through a conventional interpolation method, such as bicubic or Lanczos interpolation (Section 2.3.1.1).

2. The second stage involves extracting chosen physical features from the LR image.

3. Finally, in the third phase, both the interpolated image and the extracted LR features are aggregated to provide the super-resolved output (Fernandez-Beltran et al., 2017).

The iterative back projection (Irani and Peleg, 1991) is the first example of a reconstruction-based method. It depends on iteratively reducing the reconstruction error between the LR input image and the LR version of the super-resolved predicted image (Fernandez-Beltran et al., 2017).

Another notable example is non-blind image deconvolution (Michaeli and Irani, 2013). The imaging

model described in Equation 2.4 is frequently reformulated in terms of a matrix representation, which can be obtained simply by expressing the involved images as column vectors (Eilers and Ruckebusch, 2022). Indeed, the acquisition process in Equation 2.4 can also be referred to as

$$\mathbf{l}_c = \mathbf{S}\mathbf{h}_c. \tag{2.5}$$

Here, $\mathbf{l}_c = \text{vec}(\mathbf{L}_c)$ and $\mathbf{h}_c = \text{vec}(\mathbf{H}_c)$ denote the vectorised representations of the c-th channel of the LR image ($\mathbf{L}_c$) and the HR image ($\mathbf{H}_c$), respectively. These vectors are obtained by stacking the columns of the corresponding matrices using the function vec$(\cdot)$. The matrix $\mathbf{S} = \mathbf{CP}$ implements both the blurring (via $\mathbf{P}$) and down-scaling (via $\mathbf{C}$) operations (Eilers and Ruckebusch, 2022).

The imaging model can be expressed as just showed if and only if the following hypotheses hold for these matrices:

- The blurring matrix $\mathbf{P}$ can be approximated by a block circulant with circulant block (BCCB) matrix, i.e., the matrix representation of the cyclic convolution operator. This hypothesis can be met if and only if the blurring kernel $\mathbf{K}$ in Equation 2.4 is shift-invariant and period boundary conditions are used for the convolution operator $*$.

- The matrix $\mathbf{C}$'s conjugate transpose interpolates the decimated image with zeros over the rows and columns, respectively (Campisi and Egiazarian, 2007; Zhao et al., 2016).

As the goal of SISR is inverting the acquisition process, if $\mathbf{S}$ is assumed to be known, and the problem is thus non-blind, given that Equation 2.5 is a system of linear equations, its solution existence depends on whether it is possible to find the inverse

matrix $\mathbf{S}^{-1}$ (Eilers and Ruckebusch, 2022). When $\mathbf{S}$ is square the system described has a unique solution corresponding to the solution of an optimization problem on the available LR image (Layton and Sussman, 2017). However, as the number of pixels available in $\mathbf{l}_c$ is lower than the number of pixels in $\mathbf{h}_c$, $\mathbf{S}$ is not square. So, to compensate for the instability of solutions, it is necessary to include a regularization factor in the optimization problem (Fernandez-Beltran et al., 2017).

As a result, the SISR goal can be easily achieved as the solution of:

$$\widehat{\mathbf{h}}_c = \arg\min_{\tilde{\mathbf{h}}_c} \left[\mathcal{L}\left(\mathbf{S}\tilde{\mathbf{h}}_c, \mathbf{l}_c\right) + \tau\Phi\left(\tilde{\mathbf{h}}_c\right) \right], \tag{2.6}$$

in which the regularization term $\Phi(\cdot)$ is weighted by the parameter τ and the loss function $\mathcal{L}(\cdot)$ is a consistency term between $\mathbf{S}\tilde{\mathbf{h}}_c$, namely, the c-th channel of the degraded version of the estimated HR image, $\tilde{\mathbf{h}}_c$, and the c-th channel of the available LR image, $\mathbf{l}_c$, both in column vector form (Fernandez-Beltran et al., 2017).

Finally, additional notable reconstruction-based approaches are gradient profile (Sun et al., 2008), projection onto convex sets (POCS) (Fan et al., 2017) and maximum a posteriori (MAP)-based (Schultz and Stevenson, 1996) methods.

Although reconstruction-based approaches produce significantly better results than interpolation-based techniques, the a priori they use is usually extremely complex, and these processes often need an extensive amount of knowledge and skill (Fernandez-Beltran et al., 2017).

2.3.1.3 Learning-based methodologies

Learning-based approaches are beyond any doubt the biggest area of study in the field of SISR nowadays. In order to generate the final super-resolved image, these methods first

learn from an external training set about potential connections between LR and HR domains (Freeman et al., 2000).

They generally consist of three steps, similarly to reconstruction-based approaches.

- The first step (or learning) relies on learning the LR-HR pair connections from a specific training set.
- The LR input image is utilised in the second step (estimation) to estimate the HR components that are linked to it.
- Eventually, the third stage (image generation) integrates the estimated HR components of the input image to produce the super-resolved product (Fernandez-Beltran et al., 2017).

Sparse coding is a popular traditional technique that takes use of the notion that real-world images often consist of a linear combination of small and sparse patches. Sparse coding approaches thus employ the substitution of LR patches with their corresponding HR counterparts to achieve sparse representation matching with the LR input image to produce HR outputs (Yang et al., 2008).

Another example is neighbourhood embedding, which proposes that small image patches from LR and HR images produce low-dimensional non-linear manifolds with similar local shape. This indicates that patches in the HR feature domain may be rebuilt with the same weights as those in the LR feature domain, as long as there is enough data (Chang et al., 2004).

Other famous techniques include the ANR method, which learns high- and low-resolution dictionaries from HR image blocks (Timofte et al., 2013), and the A+, which is an enhanced version of ANR (Cremers et al., 2015).

A subgroup of these methods includes mapping-based approaches, which see SISR as a regression problem between HR and LR domains. The primary idea is to learn a mapping function from input LR images to target HR images using a predefined training set. The learnt function is then used to create the final super-resolved output from the LR input image (Fernandez-Beltran et al., 2017), as represented in the third block of Figure 2.7. The literature has a range of strategies for conducting this regression. Bayesian networks, kernel-based algorithms, and neural networks are examples of commonly used methods for estimating the projection between LR and HR images. The basic principles of the process are the same, even though each of these techniques use different theoretical tools to achieve their objective. Early research, such as kernel Ridge regression (Kim and Kwon, 2010), was dependent on a single basic mapping function, but more modern approaches (Yang and Yang, 2013; He and Siu, 2011) use many of them.

The most recent techniques make use of deep neural networks as mapping functions (Fernandez-Beltran et al., 2017). Deep-learning (DL) (LeCun et al., 2015), based on multi-layer neural networks, has become a standard approach thanks to its ability to learn hierarchical data representations. Although it requires large and diverse datasets for training, the abundance and accessibility of Earth observation data nowadays create an ideal setting for the development of novel DL-based approaches (Sdraka et al., 2022).

Deep-learning-based approaches: A substantial body of work exists for the categories described earlier. Nonetheless, the prevailing trend is to adopt mapping-based strategies, particularly those founded on DL, which are the main focus of this literature review. Indeed, methods for SISR

that leverage DL have repeatedly demonstrated superior performance over time. This trend also informs the present study, which builds upon the recent advancements and core principles of DL-driven SISR in the context of remote sensing. The effectiveness of such techniques has been reinforced by international competitions such as NTIRE (new trends in image restoration and enhancement), PIRM (perceptual image restoration and manipulation), and AIM (advances in image manipulation), which have collectively pushed the development of more accurate, perceptually convincing, and computationally efficient SISR models (Ye et al., 2023).

In the case of remote sensing imagery, and hyperspectral datasets, like those provided by S5P, the challenges become more pronounced due to complex sensor-induced degradations, the high dimensionality of spectral channels, and the scarcity of ground-truth (GT) reference data for validating super-resolved outputs. These aspects are discussed further in Section 2.3.2. Nevertheless, DL-based SISR approaches offer a compelling alternative to traditional methods, providing significant improvements in reconstruction quality (Sdraka et al., 2022).

Despite differences across models, all DL-based SISR architectures share several essential elements, including upsampling layers, backbone network designs, learning strategies, loss functions, and other smaller yet influential components. Researchers typically combine these elements in various ways to construct models that satisfy the specific requirements of SISR, while attempting to optimise the trade-offs among competing performance factors (Wang et al., 2021b).

Initial lightweight deep neural networks designed for SISR include convolutional neural networks (CNNs) (LeCun et al., 2015) such as SRCNN (Dong et al., 2016a), FSRCNN (Dong

et al., 2016b), VDSR (Kim et al., 2016a), ESPCN (Shi et al., 2016), and DSRN (Han et al., 2018); recursive networks like DRRN (Tai et al., 2017), RFDN (Liu et al., 2020), and DRCN (Kim et al., 2016b); cascading-based networks such as CARN (Ahn et al., 2018) and CBPN (Zhu and Zhao, 2019); and multi-scale fusion networks including MSRN (Li et al., 2018) and MCSN (Li et al., 2020). While these early architectures achieved reasonable quantitative results, their perceptual quality often remained limited.

Studies have highlighted that distortion and perceptual attributes of images interact in a manner that is simultaneously complementary and competitive. The perception-distortion trade-off theorem formalises this, indicating that there exists a region in the perception-distortion space that cannot be reached. As a result, no SR method can achieve both minimal distortion and maximal perceptual quality simultaneously, even if some algorithms approach the theoretical boundary (Blau and Michaeli, 2018). Interestingly, generative adversarial network (GAN)-based models, such as SRGAN (Ledig et al., 2017), ESRGAN (Wang et al., 2019), and RankSRGAN (Zhang et al., 2022), tend to operate closer to this limit. However, the use of batch normalization in these early GAN architectures has been shown to degrade the perceptual quality of the reconstructed images.

Because remote sensing imagery exhibits properties quite distinct from natural images, research into SISR for this domain has evolved somewhat independently, while concurrently integrating increasingly sophisticated components. For example, some studies have proposed combining high-level semantic information with low-level detail extraction (Ahn et al., 2018; Tai et al., 2017; Tong et al., 2017), whereas others have incorporated local and

non-local attention mechanisms to capture long-range pixel dependencies.

Attention mechanisms are deployed in multiple ways in SISR: channel-attention is implemented in models such as RCAN (Zhang et al., 2018), SeNet (Hu et al., 2018), SESR (Cheng et al., 2018), and SAN (Dai et al., 2019); spatial-attention is used in DRLN (Anwar and Barnes, 2022) and RFANet (Liu et al., 2020); and hybrid attention appears in CSFM (Hu et al., 2020; Niu et al., 2020) and HAT (Chen et al., 2023b).

More recent innovations include non-local attention in RNAN (Zhang et al., 2019), pixel attention as in PAN (Zhao et al., 2020), and multi-scale attention mechanisms (Lu et al., 2019; Wang et al., 2020).

Additional approaches exploit diffusion processes (Saharia et al., 2023), transformer architectures (Chen et al., 2023b), residual learning (Huang et al., 2021; Zhang et al., 2018; Lim et al., 2017), or knowledge distillation strategies (Jiang et al., 2018; He et al., 2020; Hui et al., 2018; Zhang et al., 2021) to further enhance reconstruction fidelity and efficiency.

2.3.2 Challenges and limitations of super-resolution for remote sensing images

SISR in the context of remote sensing presents a number of inherent limitations and challenges that distinguish it from conventional computer vision applications, including the high dimensionality of hyperspectral data, the spectral redundancy across adjacent bands, and the lack of GT HR references. These aspects complicate both model development and training and the reliable assessment of performance.

SISR for remote sensing often builds on research in classical SISR, incorporating special enhancements to address the characteristics of remote sensing imagery (Wang et al., 2023). However, some published research, especially that concerning hyperspectral images, is particularly relevant to our work.

SISR is extremely useful for obtaining hyperspectral images with increased spatial resolution at a reduced cost. After only a few decades of research, a wide range of reconstruction approaches have arisen, with the classic methods falling into three main categories: (1) wavelet transform-based, (2) maximum a posteriori estimation-based, and (3) spectral-mixing-analysis-based. These older approaches, however, suffer from considerable shortcomings, including their computational complexity and the difficulty of obtaining reliable solutions (Chen et al., 2023a).

Thanks to the first-ever CNN-based model for SISR — SRCNN (Dong et al., 2016a) — proposed to solve this task, a series of works, from 3D-FCNN (Mei et al., 2017) to DPRPE (Wang et al., 2022a), mark the increasing presence of DL-based SISR methods for hyperspectral imaging in the current literature (Chen et al., 2023a).

A substantial amount of training data, i.e. HR labelled images, is required for DL-based hyperspectral SISR, particularly when employing supervised approaches. However, acquiring true HR hyperspectral data is constrained by the technical and economic limitations of current sensors, as previously discussed. This lack of HR references strongly affects both the design and the training of DL-based models. To mitigate these limitations, some works employ multispectral images as proxies for hyperspectral data, or adopt non-supervised training paradigms that do not rely on explicit HR references (Shocher et al., 2018). This data scarcity also motivates the

widespread use of synthetic degradation models to generate training pairs, whose realism and sensor dependence critically affect model performance (Sdraka et al., 2022). In addition, the unavailability of reliable HR references also hinders the objective assessment of reconstruction accuracy. Conventional full-reference metrics cannot be computed in the absence of references. As a result, the performance of hyperspectral SISR algorithms is often evaluated either on artificially degraded datasets, or by resorting to indirect quality indicators. Consequently, the lack of HR references represents a fundamental limitation that impacts the entire SISR pipeline — from model design and training to validation (Sdraka et al., 2022).

Additionally, hyperspectral data are made up of hundreds of spectral bands, which makes feature extraction even more complex. Moreover, hyperspectral SISR must ensure spectral fidelity in reconstructed images while also boosting spatial resolution for improved spectral decoding efforts (Chen et al., 2023a). Scholars have therefore been searching for better approaches for hyperspectral images since the emergence of DL-based SISR reconstruction techniques in 2016 (Fernandez-Beltran et al., 2017).

2.3.2.1 Lack of high-resolution ground-truth images

Among the various factors limiting the effectiveness of SISR for remote sensing images, the absence of HR GT data stands out as one of the most fundamental. In contrast to conventional computer vision datasets, where HR-LR pairs can be readily obtained through controlled downsampling, hyperspectral sensors do not provide multiple spatial resolutions of the same scene. As a result, it is typically impossible to acquire true HR counterparts for training or validating DL-based models. This

limitation permeates all stages of SISR.

Training paradigms: To address this challenge, several alternative paradigms have emerged, allowing models to be trained without requiring explicit HR supervision (Hosseiny et al., 2024).

One common strategy involves the use of proxy datasets, such as multispectral images that share similar spatial-spectral characteristics with hyperspectral data. Yet such datasets are often scarce, and, when available, merging data collected from heterogeneous sensors can lead to compatibility issues and reduced reliability (Sdraka et al., 2022).

Another widely adopted approach relies on synthetic degradation models, whereby HR-LR pairs are artificially generated by applying a predefined downscaling kernel to the available images. While such models enable the use of supervised learning frameworks, their performance often deteriorates when the real sensor conditions deviate from the simulated ones (Shocher et al., 2018; Sdraka et al., 2022). This dependence on the assumed degradation model severely limits the generalization capability of externally supervised networks (Shocher et al., 2018).

To overcome these shortcomings, recent research has shifted towards image-specific and non-supervised paradigms. A notable example is the zero-shot SR (ZSSR) framework (Shocher et al., 2018), which trains a compact CNN directly on the input image. The method exploits the internal recurrence of information within a single image, constructing HR-LR training pairs by downscaling the input and using these synthetic samples to learn image-specific mappings. At evaluation time, the available image is provided to the network to produce the super-resolved output.

This paradigm allows the network to adapt to the unique degradation and structural characteristics of each input image, enabling SR even when the acquisition process is unknown or non-ideal (Shocher et al., 2018). Such image-specific methods demonstrate that internal image statistics can often provide stronger predictive power than external ones, particularly under high uncertainty or complex degradations — as is typically the case in remote sensing (Shocher et al., 2018). When additional knowledge about the acquisition process, such as an estimated degradation kernel, is available, it can be explicitly incorporated into the training phase to further enhance reconstruction quality (Shocher et al., 2018).

Sensor-dependent degradations: The nature of the acquisition process itself plays a critical role in the applicability of SISR approaches to remote sensing data. Indeed, even image-specific or non-supervised approaches are influenced by the characteristics of the acquisition process. As highlighted by Shocher et al. (2018), the performance of supervised SR models tends to deteriorate when the real degradations differ from the synthetic ones used during training, due to the variability of sensor responses. As a result, even very complex models trained on simulated datasets perform poorly when applied in practice (Wang et al., 2021b; Shocher et al., 2018). A model assessed on such data effectively learns to "reverse" a different degradation process, and may therefore fail to generalise to other distortion schemes. As a result, in real-world settings, state-of-the-art SISR approaches often underperform. Interestingly, in most studies the degradation model employed to simulate image pairs is independent of the acquisition sensor, even in relatively complex models, and the issue remains largely unexplored (Sdraka et al., 2022).

Image quality assessment: Evaluating the performance of remote sensing SISR remains equally challenging. Although there are several methods to handle SISR, their success is primarily determined by the quality of the super-resolved images. The ideal scenario for image quality assessment is when a GT image is available for comparison with the generated image. However, this is rarely viable in remote sensing (Sdraka et al., 2022).

When GT images are not accessible, a reduced-resolution (RR) protocol can be employed (Wald et al., 1997). The available LR image can be scaled down using the imaging model, and the SISR algorithm may then be tasked to super-resolve it to its original resolution. In this way, a pseudo-GT image can be used to assess the model's accuracy. This protocol is based on the premise that model performance is scale-independent, and it must therefore satisfy two important properties:

- The consistency property, which requires that the LR version of the super-resolved image must be as comparable as possible to the original input LR image.
- The synthesis property, which requires that a coarser LR image, generated by reducing the original input according to the imaging model, can be super-resolved to yield a valid reference image (Fernandez-Beltran et al., 2017).

When a GT image is available, full-reference metrics are typically employed to quantify the distortion between the reference and the reconstructed image.

An alternative approach consists of employing the images at their original resolution, forgoing the use of a reference image. This methodology, known as the full-resolution (FR) protocol, has the benefit of testing algorithms at the resolution used in practice, assessing their ability to infer missing features

directly from the available data. When the GT image is not available, no-reference metrics are employed to assess the perceptual quality of the reconstructed image without relying on a reference. This approach is best used in conjunction with the RR protocol to mitigate the uncertainties caused by the scale-invariance assumption (Sdraka et al., 2022).

The two evaluation protocols are illustrated in Figure 2.8, which provides a schematic representation of their workflow.

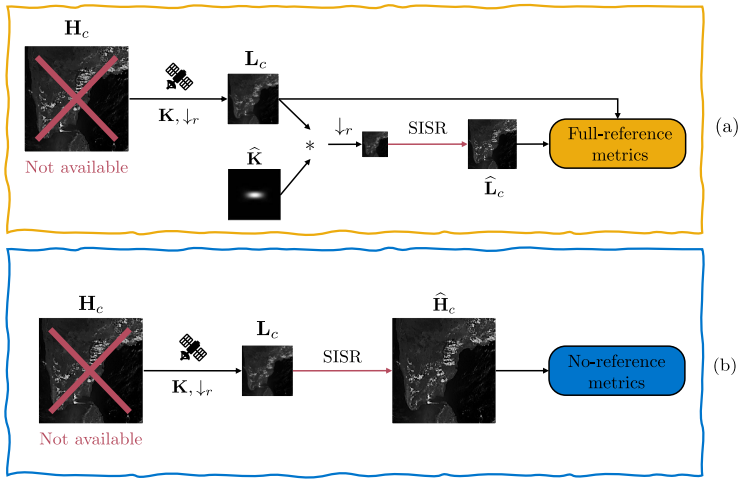

Figure 2.8: Schematic representation of the (a) RR and (b) FR protocols used to evaluate SISR in remote sensing. In (a), $\widehat{\mathbf{K}}$ denotes the assumed blurring kernel applied during the degradation process to simulate the creation of lower-resolution images. In (a) the availability of a GT image allows the use of full-reference metrics. In (b), where the GT image is not available, the performance is assessed through no-reference metrics.

2.3.2.2 Dimensionality management

A further challenge concerns the dimensionality of the input data when hyperspectral images are employed. Standard SISR algorithms are frequently directed at super-resolving grayscale images, i.e. single-channel inputs ($C = 1$). The standard approach for RGB images ($C = 3$) involves transforming them

into the YCbCr colour space, super-resolving the luminance channel (Y), and simply interpolating the chrominance channels (Cb and Cr). More recent methods in computer vision typically operate on all channels simultaneously, yet there is still little consensus on the most suitable strategy for multi-channel images (Fernandez-Beltran et al., 2017).

In the hyperspectral case, the issue is considerably more severe. Super-resolving each channel independently or all C channels jointly is computationally expensive and fails to exploit spectral redundancy (Akgun et al., 2005). Since spectral correlation is often high, most of the spectral information is usually contained within a sub-manifold of the overall spectral space. This makes dimensionality reduction a crucial step in hyperspectral data analysis, provided it is performed without discarding critical information (Shaw and Manolakis, 2002).

Because hyperspectral data comprise hundreds of spectral channels, hyperspectral SISR requires strategies that differ substantially from those used for panchromatic ($C = 1$) or multispectral ($C \approx 3$–10) data (Chen et al., 2023a; Aburaed et al., 2023). Hyperspectral sensors often capture reflection information across hundreds of successive narrow bands, providing richer spectral detail than natural images. First, feeding a full hyperspectral cube into a reconstruction network results in high computational costs and model training difficulties. Second, during the imaging process, noise unavoidably corrupts the acquired information, with variations across bands (Chen et al., 2023a). For these reasons, researchers have proposed SISR reconstruction techniques based on a subset of key bands, most often selected via principal component analysis (PCA) (Hu et al., 2017). PCA is an orthogonal linear transformation that identifies a reduced set of components capturing most of the variance in the data,

while also preserving detailed spatial features (Irani and Peleg, 1991; Michaeli and Irani, 2013). In hyperspectral SISR, the primary components are typically super-resolved, whereas the secondary ones — which contain less information and more noise — are interpolated before reconstructing the original image (Sun et al., 2008). From a theoretic perspective, PCA partitions the spectral space into correlated subgroups and selects the band with the highest entropy in each subset as representative (Chen et al., 2023a). This strategy is computationally efficient, although it may discard relevant spectral details.

Beyond PCA, several learning-based architectures have also been developed for hyperspectral SISR. For example, uSDN (Qu et al., 2018) implements hyperspectral SR using an encoder-decoder structure, while in Zheng et al. (2021) an autoencoder is used to tackle the pixel unmixing problem and improve band interaction through a learnable PSF. Other efforts include MHF-net (Xie et al., 2019), URSR (Wei et al., 2022), MIAE (Liu et al., 2022b), and GJTD-LR (Liu et al., 2022a). In Wang et al. (2021a), the authors developed a two-branch network to better exploit band similarity, combining a 2D branch for spatial information and a 3D branch for spectral correlation. Similarly, Zhu et al. (2021) presented a lightweight residual network to capture high-frequency band information, Arun et al. (2020) proposed a 3D conv-deconv framework with endmember constraints, and Chen et al. (2021) introduced SSRN, a self-supervised mapping between multispectral and hyperspectral domains. More recently, transformer-based architectures have also been employed, e.g. in Hu et al. (2022a), where global information is captured to achieve state-of-the-art results.

Finally, cascade fine-tuning approaches have been proposed

to reduce training complexity (Guarino et al., 2024). Yet, the majority of these techniques rely on statically predefined iteration counts, and their fine-tuning is neither parallelisable nor directly dependent on image characteristics. These limitations highlight the need for strategies that can account for spectral redundancy while improving computational efficiency.

Chapter 3

Methods

In this chapter, we will provide a detailed description of all the methods and algorithms employed in the experimental analysis. First, in Section 3.1, we will present the methodology employed to derive the sensor-specific degradation function from the instrument's calibration data, which forms the basis for characterising the sensor's spatial response. Subsequently, in Section 3.2, we introduce the proposed framework for applying SISR on S5P Level-1b radiance images. We describe in detail the architecture of the employed network (Section 3.2.1), including its design, the strategy for integrating the instrument's spatial response into training, and the optimization procedure with all training parameters. Building on this, we will then explain in detail the computationally efficient fine-tuning strategy specifically designed for hyperspectral data (Section 3.2.2), which allows the adaptation of the model across a large number of spectral channels while keeping the computational burden manageable. This chapter thus provides a foundation for understanding the rationale behind the methods applied, serving as a guide for interpreting the results presented in the following chapter.

3.1 Sentinel-5P PSF derivation from PRF

For each pixel in each DEM of Level-1b radiance data, ESA provides just the pixel response function (PRF), whose role is thoroughly described in Section 2.2.3.1. However, the imaging model to fully describe the acquisition of radiance images (Equation 2.4) requires the point spread function (PSF). Because DEM space and image space are not equivalent, we designed a method to derive PSFs from PRFs (Section 2.2.3.1).

The PRF products of TROPOMI instrument's four DEMs were retrieved during a calibration campaign from 10 January 2015 to 21 January 2015 in Liège, Belgium. They are fully available on TROPOMI's SentiWiki. Each product consists of a set of measurements of the PRF, described as the volume-normalised signal in two directions: azimuth and elevation angles. Each set of samples corresponds to one pixel, namely a single row (spatial dimension) and column (spectral dimension) of the DEM (European Space Agency, 2016).

During the calibration, both azimuth and elevation angles were varied with a step size of 0.05°, and their intervals were large enough to cover the entire PRF graph (European Space Agency, 2016). The number of measurements was fixed to the highest number of samples available across all pixels of each DEM to form a consistent grid. For pixels with fewer measurements, the remaining elements contain fill values. The number of samples is 406, 155, 158, and 1320 for the UV, UVIS, NIR, and SWIR DEM, respectively (European Space Agency, 2016).

The UV, UVIS, and NIR DEMs have 1025 rows and 1024 columns, while the SWIR DEM has 257 rows and 1000

columns. For each of the UV, UVIS, and NIR DEMs, columns 1–512 correspond to BD 1, BD 3, and BD 5, respectively, while columns 513–1024 correspond to BD 2, BD 4, and BD 6. For the SWIR DEM, columns 1–500 correspond to BD 7, and columns 501–1000 correspond to BD 8 (European Space Agency, 2016).

During the calibration, about 84% of the total pixels in each DEM were sufficiently illuminated. Within illuminated pixels, some PRFs were still not correctly determined and are provided as interpolated values computed using the PRFs of adjacent pixels (European Space Agency, 2016). The PSF for each DEM was derived from the preprocessed PRF at the centre of its illuminated pixels. The first illuminated rows' indices are 83, 82, 80, and 12, respectively for the UV, UVIS, NIR and SWIR DEM, while the last illuminated rows' indices are 939, 938, 937, and 227, for a total of 856, 857, 857, and 215 illuminated rows for each DEM. Similarly, the first illuminated columns' indices of the four DEMs are 187, 183, 183, and 12, respectively for the UV, UVIS, NIR, and SWIR DEM, while the last illuminated columns' indices are 837, 835, 837, 226, for a total of 650, 652, 654, and 214 illuminated columns for each DEM. The indices of the central illuminated pixels in the DEMs, expressed in the format (row; column), which we employ to formalise the PSFs, are $(511; 512)$ for the UV, $(510; 509)$ for the UVIS, $(508; 510)$ for the NIR, and $(120; 119)$ for the SWIR.

Each set of employed PRF measurements is subsequently preprocessed. Specifically, the data are centred by subtracting their barycentres in the azimuth and elevation directions (Equation 3.1).

$$b_\alpha = \frac{\sum_{i=1}^{N} (\alpha_i - PRF_i)}{\sum_{i=1}^{N} PRF_i}, \tag{3.1}$$

where b_α is the barycentre along the azimuth (ϕ) or elevation

(ϵ) direction, α_i is the i-th measured angular position in that direction, and PRF_i is the i-th PRF measurement among the N available values in the set. Then, the PRF measurements are resampled through bilinear interpolation (Section 2.3.1.1) with a step size of 0.002 onto predefined grids spanning intervals of $[-0.40, 0.40]$, $[-0.35, 0.35]$, $[-0.30, 0.30]$, and $[-0.90, 0.90]$ for the UV, UVIS, NIR, and SWIR DEM, respectively.

Each set of preprocessed PRF measurements is then fit to a two-dimensional Gaussian function (Equation 3.2) convolved with a rectangular window of a certain width, depending on the DEM.

$$f(x, y) = A \cdot \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right), \tag{3.2}$$

where A is the amplitude of the function and σ_x and σ_y are the standard deviations along the x and y axes, respectively. In our specific case, the axes x and y are ϵ and ϕ. The fitting procedure is performed using non-linear least squares optimization (Scales, 1985).

The fitting of the considered PRF is optimised by maximising the R-squared (Equation 3.3) between the fitted and empirical PRF samples:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{\mathbf{y}})^2}, \tag{3.3}$$

where y_i are the empirical values, $\hat{y}_i$ are the fitted values, and $\bar{\mathbf{y}}$ is the mean of all empirical values, i.e, $\mathbf{y}$. An R-squared score close to 1 indicates a high degree of agreement between observed and modelled data, suggesting that the modelling step being evaluated explains a large proportion of the variance in the observed values.

The convolution window width is varied in steps of 0.002 over the intervals $[0.475, 0.485]$, $[0.220, 0.230]$, and $[0.220, 0.230]$ for the UV, UVIS, and NIR DEM, respectively, and over

the interval $\{0.01, 0.1, 0.2, 0.3, 0.4, 0.5\}$ for the SWIR DEM. A preliminary interval was determined by visually inspecting the PRF graph and centring the range around the observed R-squared peak. Subsequently, the intervals were refined to improve resolution around their respective maxima. The resulting R-squared as a function of the convolution width between fitted and empirical PRFs samples, along with surface and contour plots of the preprocessed and fitted PRF samples, are shown in Figures 3.1–3.4 for the UV, UVIS, NIR, and SWIR DEM, respectively.

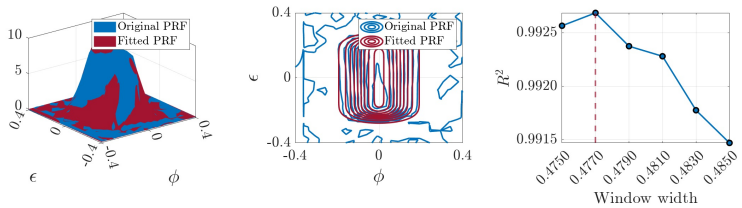

Figure 3.1: The figure represents the surface and contour plots of the original and fitted PRF for UV DEM. The function fitted to a two-dimensional centred asymmetric Gaussian function convolved with a rectangular window is obtained with the best convolution window width (0.477). The third plot represents the R-squared values when the convolution window width varies in the interval $[0.475, 0.485]$ with step size 0.002. The best value is indicated with a red dotted line.

The plots demonstrate that the adopted fitting model is optimised to provide the best approximation of the preprocessed PRFs for all DEMs, identifying the optimal convolution width that maximises the fit quality in terms of the R-squared metric.

The curves show a distinct peak for each DEM. In particular, UVIS and NIR DEM have a narrow convolution window width (0.226), as confirmed by their smaller PRFs (contour plots of Figures 3.2). Also, they share the best fitting function, which further confirms the similarity of their PRFs. Conversely,

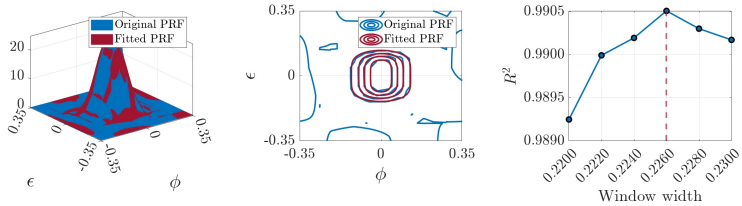

Figure 3.2: The figure represents the surface and contour plots of the original and fitted PRF for UVIS DEM. The function fitted to a two-dimensional centred asymmetric Gaussian function convolved with a rectangular window is obtained with the best convolution window width (0.226). The third plot represents the R-squared values when the convolution window width varies in the interval $[0.220, 0.230]$ with step size 0.002. The best value is indicated with a red dotted line.

UV DEM has a bigger convolution window width (0.477) as confirmed by the bigger PRF (contour plot of Figure 3.1). Differently from all others, SWIR DEM is characterised by a very small convolution window width (0.100) despite having a very big PRF (contour plot of Figure 3.4). This is due to the fact that the shape of its PRF is different from the others. This is further confirmed by the lower R-squared obtained for the best convolution window width (0.9589) compared with the other DEMs. The plots generally indicate that the fitting is sensitive to the choice of the convolution window width, and that it can be tuned precisely. The resulting fits reproduce well the original PRFs (surface plots in Figures 3.1–3.4), validating their use in the subsequent derivation of the PSFs and the MTFs. The best values obtained are 0.477, 0.226, 0.226, and 0.100 for the UV, UVIS, NIR, and SWIR DEM, respectively.

The set of fitted PRF measurements is then mapped onto the image space by simulating all processes occurring between DEM and image space (Section 2.2.3.1).

- For row binning, we simulated the programmable sum of measurements along the azimuth angle's direction.

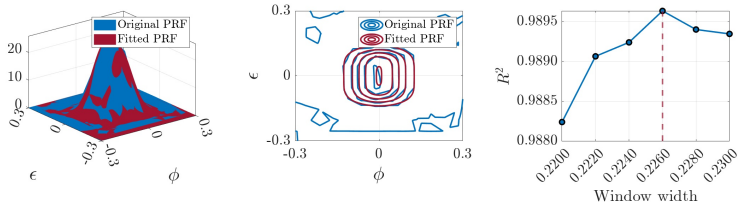

Figure 3.3: The figure represents the surface and contour plots of the original and fitted PRF for NIR DEM. The function fitted to a two-dimensional centred asymmetric Gaussian function convolved with a rectangular window is obtained with the best convolution window width (0.226). The third plot represents the R-squared values when the convolution window width varies in the interval $[0.220, 0.230]$ with step size 0.002. The best value is indicated with a red dotted line.

The number of PRF measurement sets involved in row binning is fixed for each pixel within a single DEM, being two for all DEMs except the SWIR, where no row binning is applied. This procedure was formalised by (i) computing the shift between the barycentres of the sets along the azimuth angle's direction, (ii) resampling the sets accordingly, and (iii) finally summing them after fitting each to a two-dimensional Gaussian function (Equation 3.2) convolved with a rectangular window of optimal width defined for the corresponding DEM.

- As regards satellite motion, under the assumptions of constant satellite velocity and a perfectly spherical Earth, it was modelled as the convolution between the row-binned signal and a rectangular window in the along-track direction. The window length corresponds to the angular displacement occurring during the exposure time, which in turn depends on the number of co-added measurements, normalised by the sampling step of the oversampled grid.

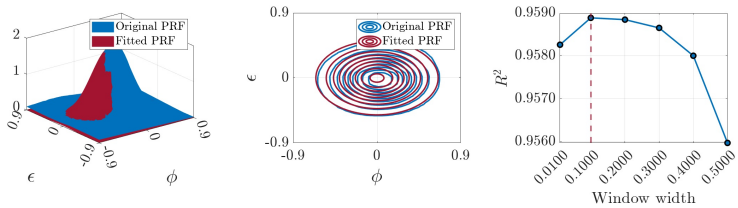

Figure 3.4: The figure represents the surface and contour plots of the original and fitted PRF for SWIR DEM. The function fitted to a two-dimensional centred asymmetric Gaussian function convolved with a rectangular window is obtained with the best convolution window width (0.100). The third plot represents the R-squared values when the convolution window width varies in the interval $\{0.01, 0.1, 0.2, 0.3, 0.4, 0.5\}$. The best value is indicated with a red dotted line.

- Furthermore, TROPOMI co-adds several consecutive measurements before transmission. As for row binning, the number of the PRF sets of measurements that are involved in co-addition is fixed for each pixel in a single DEM. For the considered pixel — namely, the central illuminated pixel in each DEM — the co-addition factor is 3, 14, 14, and 5 for the UV, UVIS, NIR, and SWIR DEMs, respectively. The effect of co-addition was simulated by performing a convolution between the involved PRFs, once row binning and satellite motion have been applied, and a rectangular window whose width corresponds to the total angular displacement accumulated during the exposure time in the elevation angle's direction, normalised by the sampling step of the oversampled grid.

The result of these steps is a response of a single pixel in image space that incorporates all effects that happen at DEM-level. Figures 3.5–3.8 show the contour plots of the intermediate functions obtained at each of these steps for the UV, UVIS,

NIR, and SWIR DEM, respectively.

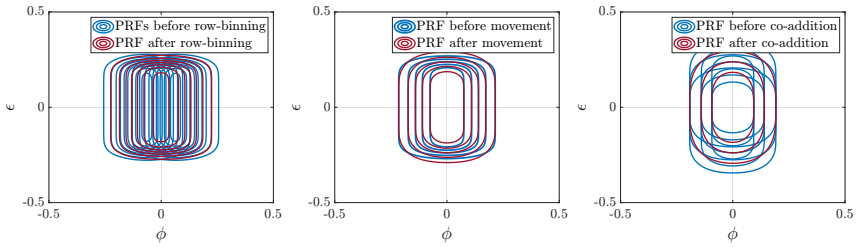

Figure 3.5: The figure shows the contour plots of the intermediate functions before and after each main step of the PSF derivation process, namely, row binning, platform motion compensation, and co-addition with factor 3 for the UV DEM.

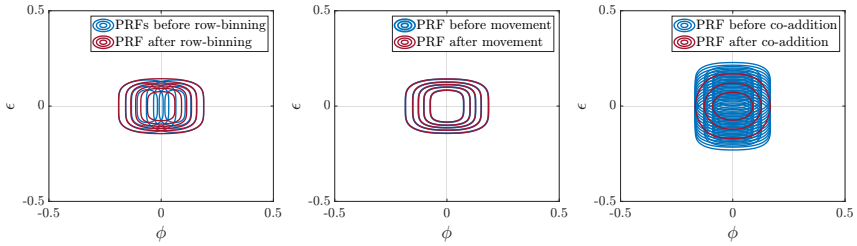

Figure 3.6: The figure shows the contour plots of the intermediate functions before and after each main step of the PSF derivation process, namely, row binning, platform motion compensation, and co-addition with factor 14 for the UVIS DEM.

As expected, the row binning operation, which is not applied to the SWIR DEM, introduces a first spatial smoothing in the ϕ direction. Similarly, the platform motion correction slightly distorts the PRF along the ϵ direction. However, the most substantial effect is observed after the co-addition step, whose impact is especially visible in UVIS and NIR DEM due to their high co-addition factor, i.e., 14 for both. The representation of the intermediate responses confirms the cumulative effect of the processing pipeline and justifies the need to model it explicitly for PSF derivation.

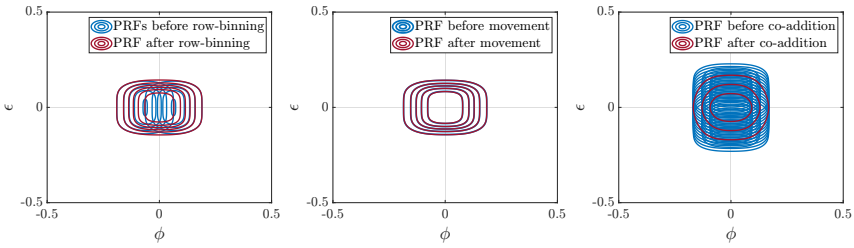

Figure 3.7: The figure shows the contour plots of the intermediate functions before and after each main step of the PSF derivation process, namely, row binning, platform motion compensation, and co-addition with factor 14 for the NIR DEM.

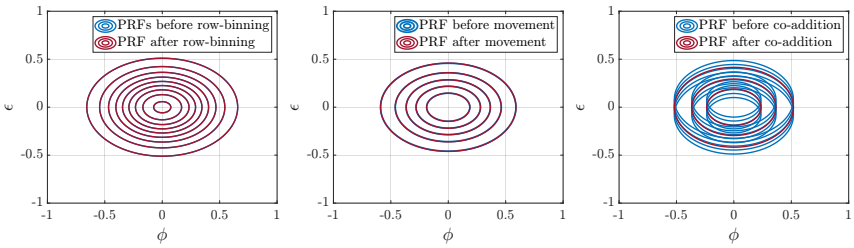

Figure 3.8: The figure shows the contour plots of the intermediate functions before and after each main step of the PSF derivation process, namely, row binning, platform motion compensation, and co-addition with factor 5 for the SWIR DEM. As evident from the first plot, the row binning operation has no effect, as it is not applied to this DEM.

Once the result of these steps is obtained, it can be sampled — depending on the scaling ratio considered — to estimate the PSF. The final PSFs can be analytically approximated by a centred two-dimensional Gaussian function with distinct standard deviations along the azimuth (ϕ) and elevation (ϵ) directions (formalised in Equation 3.2), as discussed in Section 2.3.1.

All fitted PSFs derived for all DEMs and sampled according to a scaling ratio of 4, are shown in Figure 3.9. The surface plots in Figure 3.9 reveal distinct spatial characteristics for each DEM. While the PSFs of the UVIS, NIR, and SWIR DEMs are elongated along ϕ, the PSF of the UV DEM appears

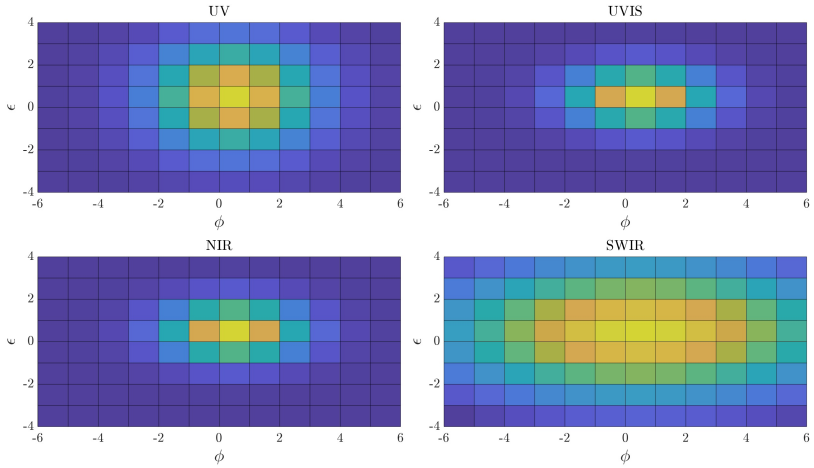

Figure 3.9: The figure shows the surface plots of the fitted PSFs for UV, UVIS, NIR, and SWIR DEM, viewed from above. Each PSF is normalised so that the sum of its values equals one and it is fitted to a Gaussian function. The oversampling ratio is set to 4. In the plots, ϕ refers to the azimuth angle's direction and ϵ to the elevation angle's direction.

more symmetric between ϕ and ϵ directions. Even if the along-track resolution is the same (5.5 km) for all DEMs, the UV DEM shows a broader PSF in the ϵ direction. This is likely due to a wider empirical PRF shape, which also results in a larger interpolation window for the preprocessing. In contrast, SWIR PSF is considerably broader in ϕ direction, indicating a reduction in spatial detail. This effect is attributable to the absence of row binning, resulting in a lower resolution in the across-track direction (7.1 km) if compared to the other DEMs (3.5 km). Apart from the PSFs for UVIS and NIR DEM, that are very similar, the pronounced differences with the others highlight the importance of modelling the PSF individually for each DEM.

Figure 3.10 presents a comparison between the original (non-fitted) PSF values and the corresponding analytical approximation obtained through fitting for the UV, UVIS,

NIR, and SWIR DEM. The R-squared values (Equation 3.3) between the two profiles are also reported. The high R-squared

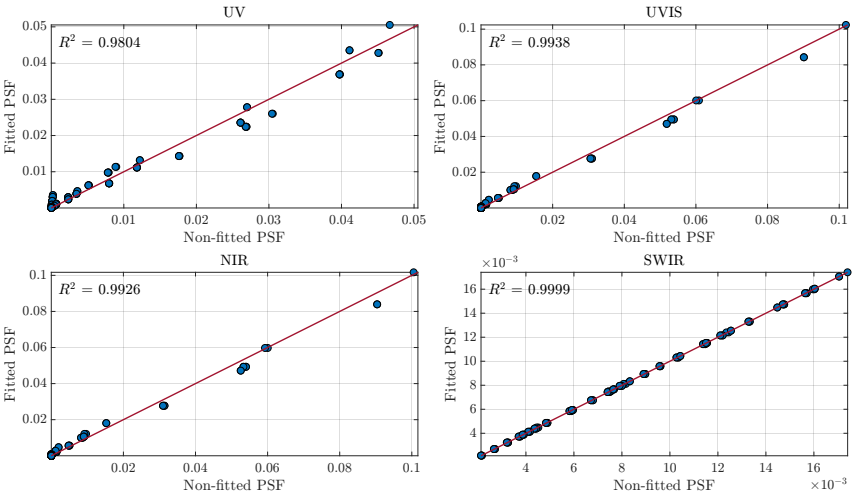

Figure 3.10: Comparison between the fitted and non-fitted PSF values for the UV, UVIS, NIR, and SWIR DEM. The R-squared values are also reported.

values confirm that the two-dimensional asymmetric Gaussian function closely reproduces the original PSF. This allows the PSF to be easily derived from the MTF gains at the Nyquist frequency (Section 2.2.1), i.e., the frequency equal to half of the sampling rate, in both ϕ and ϵ directions, simplifying the analysis.

Once the PSF has been obtained, it is first normalised to unit energy. The MTF is then derived as the modulus of its two-dimensional Fourier transform. From the resulting two-dimensional MTF, azimuthal and elevation profiles are extracted. From these profiles, the MTF at the Nyquist frequency can be obtained. The gains at the Nyquist frequency are extracted separately for the azimuth and elevation angles' directions and are used as compact descriptors of the spatial response of the instrument at image level. The one-dimensional MTF slices along ϕ and ϵ, obtained from the respective fitted

PSFs, are shown in Figure 3.11 for all DEMs. The gains at the Nyquist frequency of the MTFs in both ϕ and ϵ directions are summarised in Table 3.1.

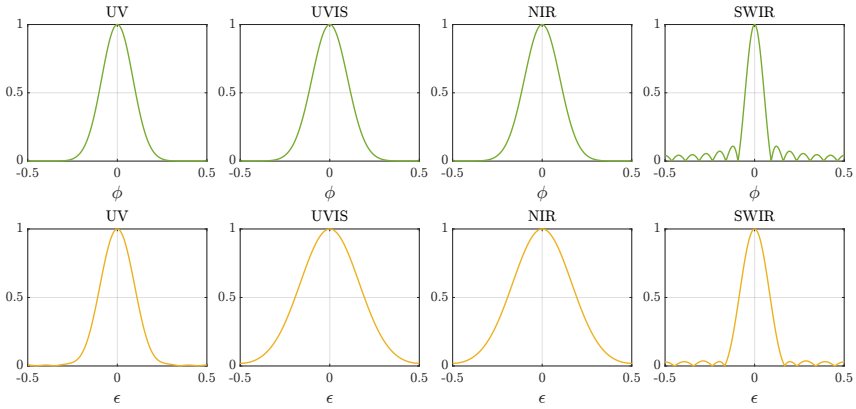

Figure 3.11: The figure illustrates the first slices along both the azimuth angle's direction (ϕ) and the elevation angle's direction (ϵ), respectively in the first and second rows of the figure, of the two-dimensional MTFs for UV, UVIS, NIR, and SWIR DEM for ratio 4.

	UV	**UVIS**	**NIR**	**SWIR**
ϕ	0.37	0.44	0.45	0.15
ϵ	0.36	0.74	0.74	0.20

Table 3.1: In this table, the gains at the Nyquist frequency for both the azimuth angle's direction (ϕ) and the elevation angle's direction (ϵ) obtained from the MTFs of UV, UVIS, NIR, and SWIR DEM are listed for ratio 4.

From both Figure 3.11 and Table 3.1 it is clear that the ϵ direction exhibits higher gains than the ϕ direction for UVIS and NIR DEMs, as expected from the strong asymmetry of their PSFs in Figure 3.9 and difference in shape of their MTFs' slices in Figure 3.11. These DEMs also achieve the highest gains, indicating a better image quality. The very similar plots for these DEMs also confirm the fact that their PSFs are very

similar to each other, as clear from Figure 3.9. Conversely, the UV and SWIR DEM show the lowest gains in both directions, consistent with their wider PSFs. The symmetry between the ϕ and ϵ directions in the PSF of UV DEM is further confirmed in the plots of the derived MTF's slices in Figure 3.11. As regard SWIR DEM, it exhibits the narrowest MTF's slices, especially along ϕ, where the gain at the Nyquist frequency is lower. This behaviour is consistent with its very large and elongated PSF along ϕ and lower spatial resolution across-track due to the lack of row binning. Also, if compared to the other DEMs, SWIR DEM's MTF's slices have many oscillations, especially along ϕ, suggesting strong smoothing effects.

The whole methodology herein described is summarised in the block diagram reported in Figure 3.12.

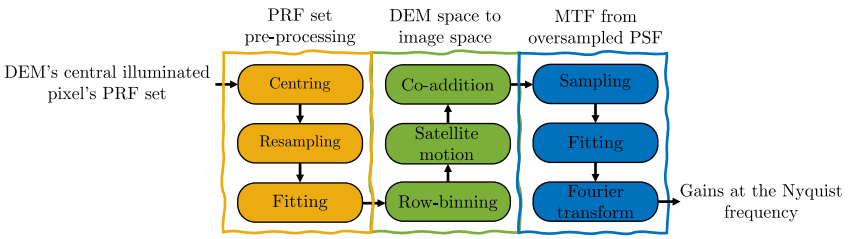

Figure 3.12: Block diagram of the proposed methodology for obtaining the PSF formalization from TROPOMI PRF data. The diagram illustrates how the MTF gains at the Nyquist frequency are derived from the set of PRF measurements for one DEM, from which a formal representation of the PSF can be obtained.

3.1.1 Sensitivity analysis of the derivation method

In this section, additional sensitivity analyses are presented and discussed. In particular, the following aspects are investigated: the influence of the specific set of PRF measurements used in

the derivation process; the impact of fitting the PRF; and the influence of selected model parameters.

No experiments were performed on varying binning factors, because they are always equal to two for all DEMs except SWIR, where binning is not applied. Instead, we analysed PRFs sensitivity to changes in the co-addition factor, which varies across the considered region.

These additional experiments will be presented only for the UVIS DEM, for simplicity, as similar conclusions can be drawn for the other DEMs.

3.1.1.1 Comparison between central and row-specific PRFs

The PSFs and MTFs are derived from the PRF data of the central illuminated pixel of each DEM. However, each pixel in a DEM is associated with a different set of PRF measurements. In this paragraph, this aspect is analysed and discussed for the DEM's rows (spatial dimension).

Figure 3.13 shows R-squared values (Equation 3.3) between the empirical PRF of each row in the UVIS DEM and: (i) the fitted PRF of the same row (green) or (ii) the fitted PRF of the central row (cyan). The column index (spectral dimension) is fixed at the centre.

As shown in the figure, the central PRF can be considered representative of all pixels within the central region of the spatial dimension (rows) of the DEM, where the R-squared of both curves closely match. This behaviour arises because, at the swath's centre, distortions are minimal and the sensor's PRFs are nearly uniform across adjacent pixels. In contrast, at the swath edges, the signal is affected by geometric distortions due to the Earth's curvature, leading to greater discrepancies

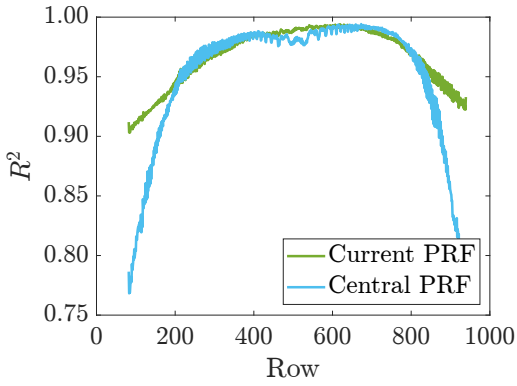

Figure 3.13: The figure shows the R-squared values between the empirical PRF associated with the current row of UVIS DEM and either the fitted PRF corresponding to the same row (in green) or the fitted PRF corresponding to the central row (in cyan). The column is chosen to be at the centre.

between the local and central PRFs. This is evident from the increasing divergence between the two curves toward the edges of the swath.

In all subsequent analyses, the Level-1b data from S5P are cropped to regions near the nadir, where the spatial resolution across-track is at most 8 km, and are subsequently resampled onto uniform spatial grids with a resolution of 3.5 km across-track. As such, the central PRF is considered a valid representative for the entire DEM.

3.1.1.2 Effect of averaging multiple central PRFs

Similarly to the previous analysis, it is now discussed the possibility of averaging multiple PRFs along the considered central rows of the UVIS DEM. A range of averages in the interval $\{1, ..., 11\}$ with a step of 2 is considered, and the central PRF is always included in the averaging process. The column index is fixed at the centre in this analysis too.

Figure 3.14 shows the R-squared between the preprocessed

PRF of the current row (indicated on the x-axis; only the central interval $\{200, ..., 700\}$ with a step size of 20 is considered) and the fitted averaged PRF values for UVIS DEM. In the plot, the rows corresponding to the averaged PRFs are indicated with markers. The plot of the R-squared

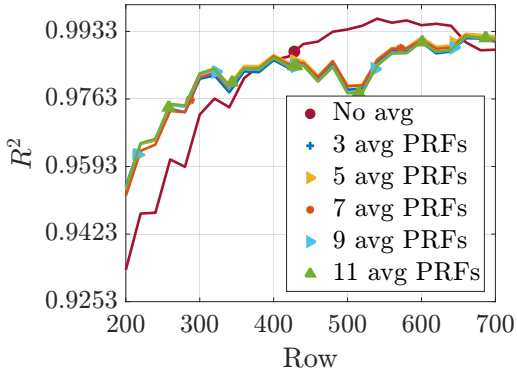

Figure 3.14: Analysis for the UVIS DEM showing the R-squared values between the preprocessed PRFs for each row and the fitted averaged PRFs for UVIS DEM. The markers indicate the rows considered in the average.

values between the current empirical PRF and the fitted averaged PRF over an increasing number of PRFs shows that the R-squared remains consistently high, with only minor variations as the number of averages increases. This indicates that averaging multiple PRFs does not significantly improve the fit quality compared to using the central PRF alone.

In practical terms, this suggests that the central PRF is already sufficiently representative for the PSF derivation method, as anticipated in the preceding paragraph, and averaging additional PRFs does not provide substantial benefits.

3.1.1.3 Influence of column-dependent PRFs on gain variations

A further analysis was conducted on five columns of the DEM — specifically, columns 183 (first illuminated column),

346, 509 (central column), 672, and 835 (last illuminated column) — to evaluate potential variations along the columns (spectral dimension) of the DEM. In this analysis, the row index was fixed at the centre.

Figure 3.15 shows line plots of the gains at the Nyquist frequency in both the ϕ and ϵ directions, obtained from PRFs corresponding to different columns. The analysis of

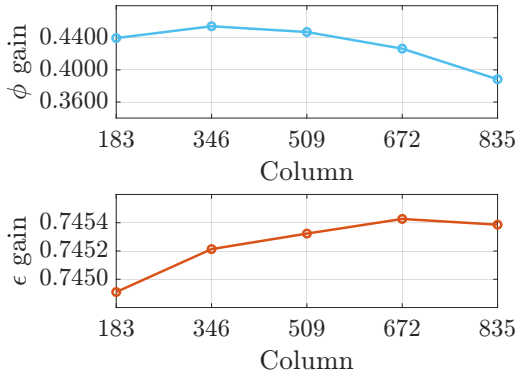

Figure 3.15: Line plots of the corresponding gains at the Nyquist frequency along the azimuth angle's direction (ϕ) and elevation angle's direction (ϵ) for the selected columns of the UVIS DEM.

the PSFs corresponding to different columns of the UVIS DEM in Figure 3.15 reveals that variations along the column dimension are relatively minor. Specifically, the line plots of the gains at the Nyquist frequency show minimal differences among the columns considered, despite corresponding to very different spectral frequencies. This is especially evident in the ϵ direction, where the values vary at the fourth decimal place. This finding suggests that the PRFs of the sensor are largely homogeneous across the spectral dimension of a given DEM. Consequently, for subsequent analyses and final modelling, it was deemed reasonable to fix the column at the centre.

All conducted analyses regarding the influence of the PRF choice on PSF derivation indicate that selecting the central

preprocessed PRF — both along the rows (spatial dimension) and columns (spectral dimension) of each DEM — is a justified approach for deriving the PSFs and MTFs.

3.1.1.4 Impact of PRF fitting

This paragraph presents the results of the comparison between the derivation of the PSF directly from the preprocessed PRF and the PSF obtained from the PRF approximated to a known analytical function.

Figure 3.16 shows: (i) a plot comparing the fitted PSF samples obtained directly from the preprocessed PRF (x-axis) with those obtained from the analytically approximated PRF (y-axis), including the corresponding R-squared value (Equation 3.3); (ii) the MTF slice along the ϕ direction of the two-dimensional MTFs derived from both fitted PSFs; and (iii) the MTF slice along the ϵ direction, also derived from both fitted PSFs. The figure refers to the UVIS DEM. As it can be

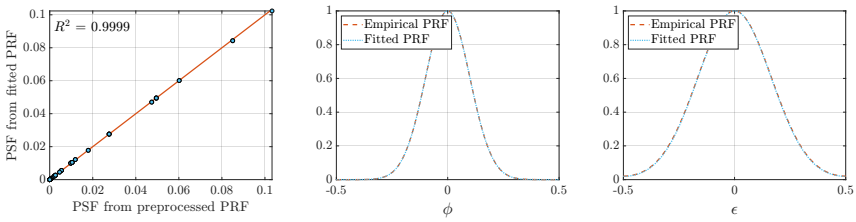

Figure 3.16: The plots represent, from left to right, (i) the fitted PSF values obtained directly from the preprocessed PRF (x-axis) and the fitted PSF values obtained through the PSF derivation method described in Section 3.1 (y-axis); (ii) the MTF slice in azimuth angle's direction (ϕ) derived from the two PSFs; and (iii) the MTF slice in elevation angle's direction (ϵ) derived from the two PSFs. The plots refer to UVIS DEM.

observed in the first plot, there is a strong correlation between the fitted PSF samples derived directly from the preprocessed PRF and the fitted PSF samples derived from the modelled

PRF, as confirmed by the high R-squared (0.9999). Similarly, the MTFs' slices along ϕ and ϵ are nearly completely overlying.

Using a fitted PRF does not significantly change the resulting PSF compared to the preprocessed PRF. Nevertheless, we adopt the approximation to achieve a smoother representation, reduce measurement noise, and provide a more reproducible PRF for future analyses.

3.1.1.5 Influence of different PRF fitting functions

In this paragraph, the results of the comparison between different analytical functions used to fit the preprocessed PRF are discussed. In particular, the final fitting function — namely, a centred asymmetric two-dimensional Gaussian (Equation 3.2) convolved with a rectangular window of optimal width — is compared with both a single centred asymmetric two-dimensional Gaussian and a weighted mixture of two such functions.

Figure 3.17 shows the surface plots of the difference between the preprocessed PRF and the fitting function for all three analytical functions and the plot of the empirical PRF samples (x-axis) and the fitted ones (y-axis) with the respective R-squared (Equation 3.3) values. The figure refers to the UVIS DEM. It is clear by observing the surface plots of the differences in Figure 3.17 that the Gaussian function convolved with a rectangular window in its best configuration yields the best approximation for the preprocessed PRF. It is indeed more strongly correlated to the empirical PRF samples than the single Gaussian (+4% on the R-squared) and the mixture of two Gaussian functions (+1% on the R-squared) as shown in the fourth plot in Figure 3.17.

For this reason, in the PSF derivation method the preprocessed

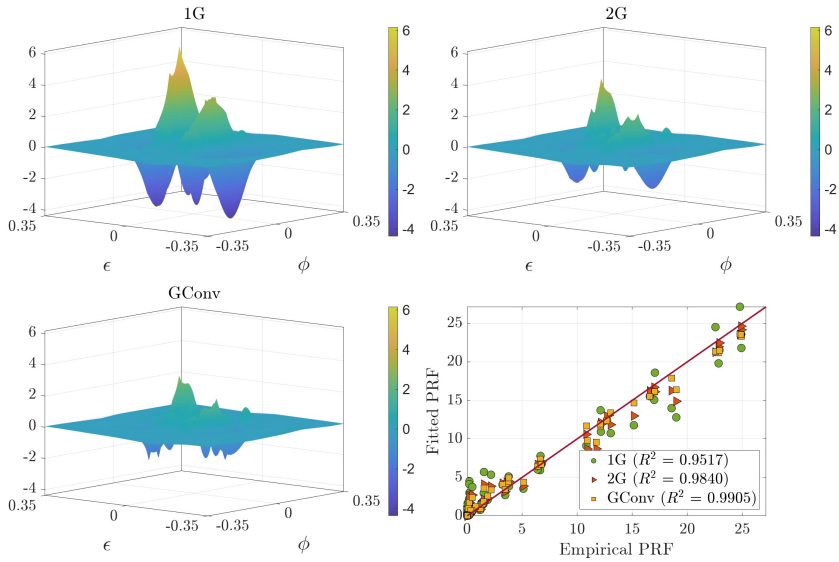

Figure 3.17: Surface plots of the difference between the preprocessed PRF and the fitted PRF for all three analytical functions. In particular, 1G refers to the single Gaussian function, 2G to the mixture of two Gaussian functions, and GConv to the Gaussian function convolved with a rectangular window of optimal width, i.e., 0.226. The last plot shows the empirical PRF samples (x-axis) and the fitted PRF samples (y-axis) for all three fittings with the respective R-squared values. The figure refers to the UVIS DEM.

PRF is fitted to a centred asymmetric Gaussian function convolved with a rectangular window.

3.1.1.6 Effect of the co-addition factor

The impact of the co-addition factor on the derived PSF is evaluated in this section. Although the PRF selected for PSF derivation is assumed to correspond to the response of the central pixel of the DEM, where the co-addition factor is fixed — and the implications of this assumption have already been discussed — we further investigate how different co-addition factors across the UVIS DEM affect the resulting

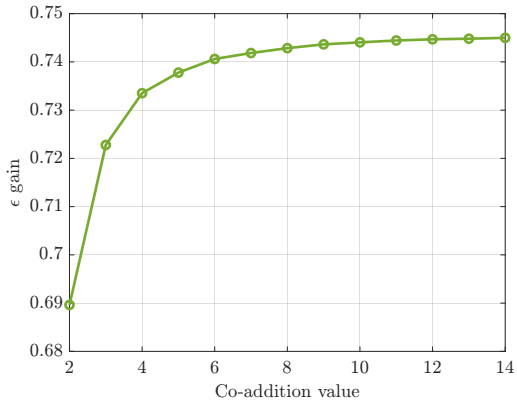

Figure 3.18: Variation of the gains at the Nyquist frequency along the elevation angle's direction (ϵ) as a function of the co-addition factor, ranging from 2 to 14, for the UVIS DEM.

PSF.

Figure 3.18 shows the variation of the gains at the Nyquist frequency along the elevation direction (ϵ) as the co-addition factor varies in the interval $\{2, ..., 14\}$ with steps of 1. The considered interval was selected by taking into account all co-addition factors of the UVIS DEM pixels. Only the elevation direction is presented because the co-addition affects exclusively this axis.

As evident from the plot, the gain increases with the co-addition factor. This occurs because summing multiple responses along the ϵ direction amplifies the cumulative broadening effect, which enhances the response's sensitivity and thus the relative gain. The gain corresponding to a co-addition factor of 2 is especially lower. However, this factor only occurs at pixels located along the edges of the DEM, which — as discussed before — are excluded from further analyses. In contrast, the gains for all other co-addition values remain approximately constant within the region of interest. It is therefore reasonable to consider only the co-addition factor of the central pixel of the DEM.

3.2 Proposed super-resolution framework

We have proposed and employed a DL-based algorithm, called S5Net, which is specifically designed to conceptually mirror the sequence of phases that characterise the transformation between the LR and the HR domains in learning-based SISR approaches (Section 2.3.1.3).

To address the gaps identified in the literature on this topic (Section 2.3.2), particularly the lack of GT references, the standard nature of image degradations, and the challenges related to hyperspectral image processing, we proposed two contributions. First, we introduced a dedicated image-specific training process that explicitly incorporates the spatial response of the imaging instrument, formalised through the methodology described in Section 3.1, ensuring that the learned model remains physically consistent with the acquisition system. Second, we developed a computationally efficient fine-tuning strategy tailored to hyperspectral imagery, enabling the adaptation of the network to a large number of channels while keeping processing times manageable and preserving spectral consistency.

3.2.1 S5Net: a deep-learning architecture for PSF-aware super-resolution

In this subsection, we introduce PSF-aware S5Net, the DL-based architecture that we developed in Carbone et al. (2024b) to perform SISR of S5P Level-1b images while explicitly accounting for the instrument's spatial response. We indeed describe also how prior knowledge of the PSF is formally incorporated into the training procedure, ensuring that the

model learns reconstructions that are physically consistent with the imaging system.

3.2.1.1 Core network design

As explained in Section 2.3.1.3, mapping-based learning approaches handle the SISR problem by observing a training set of pairs of LR and HR images, and they attempt to find a mapping function between low- and high-resolution domains.

We have thus defined a neural network, which we will refer to as S5Net, shown in Figure 3.19, that conceptually matches the four phases that make up this transformation — image upsampling, image representation, non-linear mapping, and image reconstruction. The second, third and fourth parts can be represented as convolution layers, while, for the first one, we choose a transposed convolution layer (Zhang et al., 2023). For this reason, S5Net can be considered a CNN. For clarity, we denote a convolution layer as $Conv_i(f_i, n_i, c_i)$, and the transposed convolution layer as $TransConv_i(f_i, n_i, c_i)$, where the variables f_i, n_i, c_i represent the filter size, the number of filters and the number of channels in the i-th phase, with $i = \{1, ..., 4\}$, respectively.

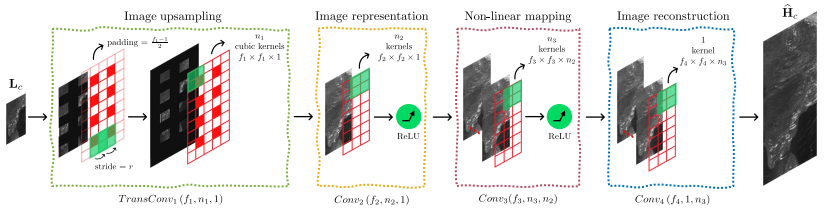

Figure 3.19: A graphical representation of S5Net consisting of four blocks: image upsampling, image representation, non-linear mapping, and image reconstruction.

- Image upsampling: as image upsampling particularly affects the task, a transposed convolution layer

$TransConv_1(f_1, n_1, 1)$ is herein employed (Zhang et al., 2023). Given a scaling ratio of r between the low- and high-resolution domains, transposed convolution increases the size of the input by inserting $r - 1$ zeros between two consecutive pixels and performing zero padding. It then convolves the resulting image n_1 times with kernels of size $f_1 \times f_1 \times 1$.

- Image representation: this process takes each patch in the upsampled image and represents it as a high-dimensional vector in a given basis (Aharon et al., 2006). A linear representation is obtained by applying the convolutional layer $Conv_2(f_2, n_2, 1)$, which convolves the upsampled image n_2 times with filters of size $f_2 \times f_2 \times 1$, applying appropriate padding (here, border pixels are replicated to handle edge effects), and optionally adding biases.

- Non-linear mapping: each n_2-dimensional vector is non-linearly mapped onto an n_3-dimensional vector. This is implemented as a convolution with n_3 filters of size $f_3 \times f_3 \times n_2$, namely $Conv_3(f_3, n_3, n_2)$. Additional layers can be stacked at this stage to increase non-linearity if needed.

- Image reconstruction: the patch-wise HR representation is combined to produce the final super-resolved image. In the neural network, this corresponds to a convolution with one filter of size $f_4 \times f_4$, applied to the n_3-dimensional representation, i.e., $Conv_4(f_4, 1, n_3)$. The output has one channel, matching the original input image.

We combined the four steps to create the neural network depicted in Figure 3.19. The rectified linear unit

(ReLU) activation function is applied after each intermediate convolution to introduce non-linearity (Nair and Hinton, 2010). Overall, the network can be formally expressed as:

$$\widehat{\mathbf{H}}_c = Conv_4\big(\sigma(Conv_3(\sigma(Conv_2(TransConv_1(\mathbf{L}_c))))\big) \quad (3.4)$$

where $\sigma(\cdot)$ denotes the ReLU activation function, and $\mathbf{L}_c$ and $\widehat{\mathbf{H}}_c$ denote the c-th channel of the LR input image and the HR reconstructed image, respectively, with $c \in \{1, \ldots, C\}$ and C the total number of channels in both images.

It is important to note that S5Net works on a per-channel basis: each spectral channel $c \in \{1, ..., C\}$ is processed independently by the network, which learns to reconstruct the corresponding HR counterpart. This choice implies that both training and testing must be repeated for all C channels.

3.2.1.2 Optimization and training scheduling

We train S5Net using stochastic gradient descent (SGD) (Robbins and Monro, 1951). The optimization configuration depends on whether the transposed convolutional layer is active.

- Without transposed convolution: only the three convolutional layers are trained. The learning rate is set to 10^{-2} for the first two convolutional layers, while the last convolutional layer is trained with a reduced learning rate (10^{-3}).
- With transposed convolution: training is performed in two phases. In the first phase, only the convolutional layers are updated as for the previous case, while in the second phase the convolutional layers are frozen and the transposed convolutional layer is optimised with a learning rate 10^{-2}.

The model is always trained by minimising the mean square error (MSE) between the network's output and the c-th channel of the HR GT image for a fixed number of iterations i_s. The MSE loss function is defined as:

$$MSE(\mathbf{H}_c, \widehat{\mathbf{H}}_c) = \frac{1}{N} \sum_i \left(H_{c,i} - \widehat{H}_{c,i} \right)^2, \tag{3.5}$$

where $H_{c,i}$ and $\widehat{H}_{c,i}$ denote the i-th pixel values of the c-th channel of the reference image $\mathbf{H}$ and the reconstructed image $\widehat{\mathbf{H}}$, respectively, and N is the number of pixels. Lower values of the MSE indicate better reconstruction accuracy.

To avoid border effects that could bias the learning process, a customised MSE function is herein employed. In particular, when the fixed parameter d is greater than zero, a border of $d - 1$ pixels is cropped from both the reconstructed and the reference images before computing the MSE. This ensures that only the central, reliable region of the images contributes to the optimization process. When $d = 0$, the full images are compared. This strategy is particularly beneficial when dealing with upsampling operations, where edge artefacts are more likely to occur.

It is worth noting that, due to the lack of real HR GT data for the TROPOMI instrument, as discussed in Section 2.3.2, training is performed in a non-supervised, image-specific fashion. Each image is used to generate its own degraded counterpart, which serves as input during optimization, while the original image is subsequently employed during assessment, as shown in Figure 3.20. This strategy is consistent with state-of-the-art approaches in remote sensing SR (Sdraka et al., 2022; Hosseiny et al., 2024), such as zero-shot SR (Shocher et al., 2018), and enables the network to exploit the internal spatial redundancies within each image rather than relying on cross-image generalization.

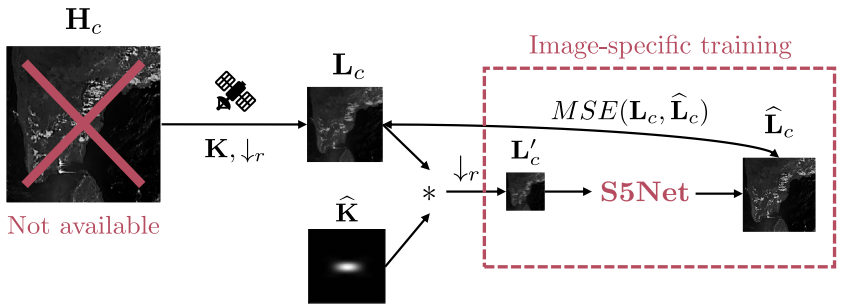

Figure 3.20: Overview of the non-supervised, image-specific training strategy adopted for S5Net due to the lack of real GT data for the TROPOMI instrument. Each available image $\mathbf{L}_c$ is used to generate its own degraded counterpart $\mathbf{L}'_c$ through the assumed blurring kernel $\widehat{\mathbf{K}}$. This image serves as the input during optimization, while $\mathbf{L}_c$ is subsequently employed during assessment. During training, the network minimises the MSE loss (as defined in Equation 3.5) between the original image $\mathbf{L}_c$ and its reconstruction $\widehat{\mathbf{L}}_c$.

The network is thus initialised with the two-dimensional cubic polynomial kernel (Keys, 1981), discussed in Section 2.3.1.1, for the first transposed convolution channel, and with the state-of-the-art SRCNN (Dong et al., 2016a) weights for the remaining convolutional channels, and then fine-tuned in a non-supervised fashion, following the strategy described above.

3.2.1.3 Integration of PSF knowledge into the training process

A critical aspect of the proposed architecture and training strategy is the explicit integration of TROPOMI's PSF into the training process of S5Net. Instead of relying on generic downsampling or idealised degradation models, the spatial response of the instrument is directly embedded in the generation of the training data.

Specifically, LR available images are artificially degraded by convolving them with the formalised PSF of the sensor under

study and then decimated according to the target resolution, as formalised in Equation 2.4. The original LR images are then used as GTs. In this way, the training set consists of realistic pairs that mimic the actual acquisition process.

This strategy allows the network to learn a mapping that compensates for the specific distortions introduced by the instrument's PSF. As a result, the reconstructed super-resolved images are physically consistent with the sensor's imaging model and better adapted to the characteristics of the original dataset. The approach ensures that the network is optimised for the same type of degradation encountered during image acquisition, thereby reducing the gap that would otherwise arise if synthetic LR images were generated using generic kernels, as discussed in Section 2.3.2.

3.2.2 Dynamic multidirectional cascade fine-tuning algorithm for hyperspectral data

In the standard training procedure adopted in Carbone et al. (2024b), S5Net is fine-tuned independently for each spectral channel. That is, each channel of the available LR image in the training set is retrained from the same initial weights $\mathbf{W}_s$ and biases $\mathbf{b}_s$. During this procedure, the network iteratively estimates one possible HR version of the corresponding LR image in the training set by minimising the loss function between the estimation and the related GT image (according to Equation 3.5) for a fixed number i_s of iterations. The learned weights $\mathbf{W}_c$ and biases $\mathbf{b}_c$ $\forall\ c \in \{1, ..., C\}$, where C is the number of total channels in the imagery employed, are then used by the network to estimate in the test phase a super-resolved image.

The main issue of this process is that the fine-tuning completely restarts from the same set of weights and biases for each channel, costing an extensive amount of fine-tuning iterations, i.e., $C \cdot i_s$.

3.2.2.1 Description of the proposed efficient fine-tuning strategy

The dynamic multi-directional cascade fine-tuning algorithm we propose in Carbone et al. (2024a), depicted in Figure 3.21, starts with the initial set of weights $\mathbf{W}_s$ and biases $\mathbf{b}_s$ from the monochromatic image $\mathbf{L}'_{\frac{C}{2}}$, i.e., the central channel in the training input image $\mathbf{L}'$. Then, the tuning iteratively proceeds in two different directions: from the set of weights $\mathbf{W}_{c-1}$ and biases $\mathbf{b}_{c-1}$ $\forall\, c \in \left\{\frac{C}{2}+1, \ldots, C\right\}$ and from the set of weights $\mathbf{W}_{c+1}$ and biases $\mathbf{b}_{c+1}$ $\forall\, c \in \left\{1, \ldots, \frac{C}{2}-1\right\}$, in both cases for i_c iterations.

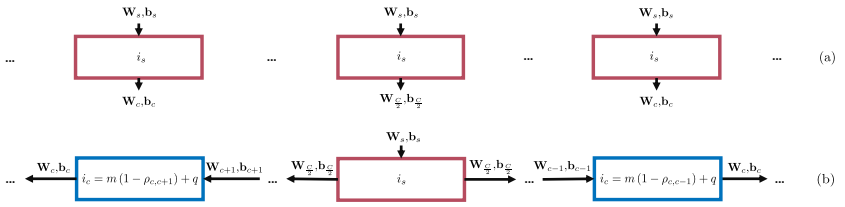

Figure 3.21: An overall graphical description of the proposed dynamic multi-directional cascade fine-tuning algorithm (b) compared to the traditional fine-tuning (a).

With a few exceptions, we observed that the correlation between consecutive channels is often quite high and that far-away ones have relatively low correlations, particularly when their spectral frequencies are significantly different. As a result, we argue that the more uncorrelated the involved channels at each step, the more iterations would be required for the tuning of the current model from the adjacent one, and vice

versa, in contrast to most approaches in the literature, which typically rely on static training paradigms (Section 2.3.2).
So, during the procedure, when $c = \frac{C}{2}$ the number of iterations is fixed to i_s, and, for each $c \neq \frac{C}{2}$, it is dynamically computed as the result of the linear relation formalised in Equation 3.6:

$$\begin{cases} m \cdot (1 - \rho_{c,c+1}) + q & \text{if } c \in \left\{1, \ldots, \frac{C}{2} - 1\right\} \\ m \cdot (1 - \rho_{c,c-1}) + q & \text{if } c \in \left\{\frac{C}{2} + 1, \ldots, C\right\} \end{cases}, \tag{3.6}$$

where $\rho_{c,c\pm1}$ is the Pearson correlation coefficient between $\mathbf{L}'_c$ and $\mathbf{L}'_{c\pm1}$. To estimate the values of the other two parameters, i.e., m and q, we employed a static multi-directional cascade fine-tuning on the network, executing a fixed number of iterations $i_c \in \{1, \ldots, i_s\}$ per channel. We subsequently determined the iteration i_c^τ at which, during this process, each model is τ away from achieving its optimal performance in terms of a chosen metric. Then, we found the best line fitting between these values and all corresponding $(1 - \rho_{c,c\pm1})$ $\forall\, c \in \{1, \ldots, C\}$. The two parameters, namely, m and q, are respectively the slope and intercept of the straight line obtained as a result.

Chapter 4

Experimental results

In this chapter, we present, analyse, and critically discuss all experimental results obtained throughout the research using the proposed SISR methodology. We begin by describing the experimental setup of the whole assessment (Section 4.1), including details on the datasets used, and the preprocessing steps applied (Section 4.1.1). We then introduce the evaluation metrics employed to assess performance (Section 4.1.2) and the specific setting of the image quality assessment protocols (Section 4.1.3). The whole experimental setup for SISR evaluation is detailed (Section 4.1.4), followed by an overview of key implementation and training details (Section 4.1.5). The experimental evaluation is structured into two parts. Section 4.2 is focused on the full assessment of the proposed approach, which constitutes the core of the evaluation of the proposed framework. It is organised into three parts: (i) the full optimization of the proposed SISR approach; (ii) the evaluation of the impact of the PSF on the PSF-aware SISR methodology applied to S5P radiance data; (iii) the results of the proposed dynamic multi-directional cascade fine-tuning algorithm, integrated into the SISR assessment, along with comprehensive final results on S5P data. On the other hand,

Section 4.3 illustrates an applicative case study focusing on air pollution monitoring. This chapter serves to contextualise all experimental findings, providing the necessary basis for interpreting the effectiveness and implications of the proposed framework.

4.1 Experimental setup

In this section, we describe the experimental setup. Specifically, we present the datasets used and detail the preprocessing steps applied across the experimental assessment. We also outline the evaluation metrics, the setting of the image quality assessment protocols and the experimental setup for SISR evaluation. Finally, we provide relevant implementation and training details to ensure reproducibility.

4.1.1 Sentinel-5P Level-1b dataset and preprocessing for super-resolution assessment

In all SISR analyses we used Level-1b radiance data of S5P's TROPOMI instrument as described in Section 2.2.3.1. In particular, ten representative orbits are employed. Table 4.1 summarises their primary characteristics.

By a deep analysis, it resulted that:

- BD 1 radiance values are characterised by a very low SNR, usually lower than 50 decibels (dB), and a very low spatial resolution, equal to minimum 10 km, which may cause issues in the analyses.
- BD 7 and BD 8 radiance values have a minimum spatial resolution of 7.1 $\times$ 5.5 km^2 due to the SWIR DEM's

Tag	Date	Orbit	Central location
AG	23-06-2020	13962	(32° N, 5° E) - Algeria
CG	23-06-2020	13962	(1° S, 15° E) - Congo Republic
CS	23-06-2020	13960	(40° N, 56° E) - Caspian Sea
EE	23-06-2020	13961	(48° N, 22° E) - Romania
EG	04-08-2024	35285	(26° N, 29° E) - Egypt
FR	23-06-2020	13962	(44° N, 3° E) - France
IN	01-04-2023	28317	(17° N, 80° E) - India
IT	04-08-2019	09365	(42° N, 12° E) - Italy
RS	04-08-2019	09364	(17° N, 42° E) - Red Sea
US	09-07-2023	29729	(36° N, 114° W) - Arizona

Table 4.1: Overview of the ten selected orbits used in the SISR analyses, including their tag, acquisition date, orbit number, central coordinates, and the corresponding central location. The orbits are ordered alphabetically by their tag.

inability to perform row binning.

- All radiance values in a single file are not equally distributed on the geographical grids that characterise them, particularly towards the swath's edges where the Earth's curvature is most noticeable. In other words, the resolution is not uniform across the pixels, but it is quite comparable in the middle of the swath.

- All radiance values in a single file are not defined in a straight rectangular grid. As a result, latitudes are not constant across-track and longitudes are not constant along-track.

- Different files (BDs) corresponding to the same orbit are not defined on the same geographical grid, despite being captured simultaneously by the sensors.

- The correlation matrices between adjacent channels

revealed consistent upward and downward peaks, which could not be fully explained by the official documentation (European Space Agency, 2022). These peaks occurred at the same spectral channels across different orbits, suggesting a possible link to instrument calibration. Notably, only 2–4% of channels exhibited negative peaks.

To overcome these limitations, we designed a preprocessing pipeline aimed at producing consistent, spatially aligned hyperspectral cubes for each orbit. The pipeline proceeds as follows:

- All netCDF-4 files corresponding to a single orbit, defined by acquisition date and overpass time, are read simultaneously. The BD 1 file is excluded due to its poor SNR and coarser spatial resolution.

- The radiance matrices are flipped along-track to account for the satellite's South-to-North flight direction. A cross-track cropping is then performed to retain only the region with spatial resolution better than or equal to 8 km. Along-track cropping is also applied to restrict the analysis to the area of interest centred on the location specified in Table 4.1. Within this region, a row binning factor of two is applied uniformly, except for the SWIR DEM, for which no row binning is performed. In contrast, the co-addition factor varies across channels.

- Values equal to the official fill value or lower than zero are considered invalid. These values are estimated through interpolation (Section 2.3.1.1) of the neighbouring pixels. This step ensures the continuity of the dataset and avoids artefacts.

- Because the native grids of different BDs are not coincident, straight and evenly spaced geographical grids of resolution 3.5×5.5 km^2 are generated using the latitude and longitude values provided in the metadata as coordinates. All radiance values are resampled onto the newly defined grids using MATLAB's natural neighbour interpolation.

- The resampled radiance matrices of all valid channels — amounting to 3445 channels per orbit — are then stacked along the spectral dimension, producing a coherent three-dimensional cube with two spatial and one spectral dimension for each orbit. This arrangement guarantees that all subsequent analyses and SISR experiments are performed on datasets that are both spatially aligned and spectrally complete.

The whole preprocessing pipeline herein described is summarised in the block diagram reported in Figure 4.1.

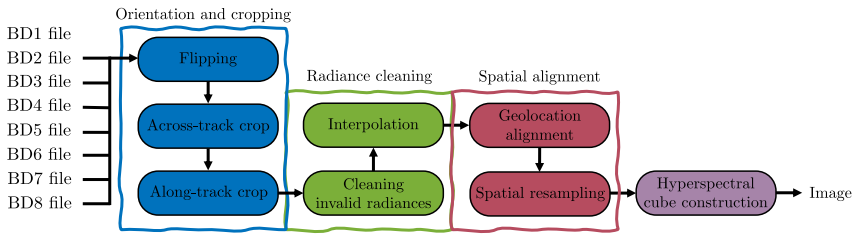

Figure 4.1: The diagram illustrates the complete preprocessing workflow applied to S5P Level-1b images, including all steps required to prepare the data for subsequent analysis.

A false-colour representation of the employed orbits (Table 4.1) is given in Figure 4.2. Each image is displayed using channels 3050, 2250, and 1700 for red, green, and blue, respectively.

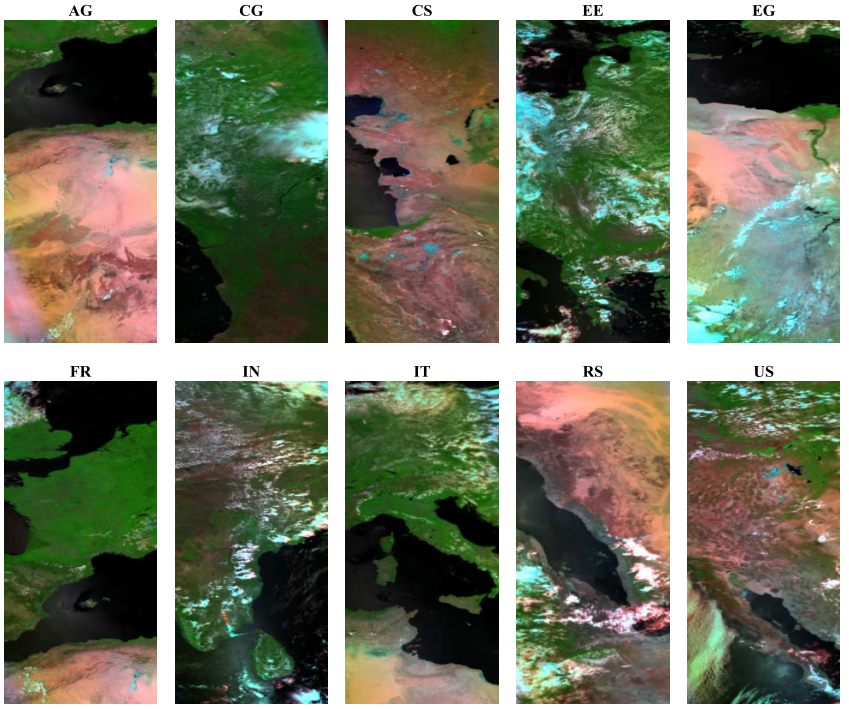

Figure 4.2: False-colour representation of all orbits indicated by their tag (Table 4.1). The images are displayed using channels 3050 (red), 2250 (green), and 1700 (blue) and are ordered alphabetically.

4.1.2 Evaluation metrics for super-resolution assessment

To assess the performance of the proposed PSF-aware SISR method, a variety of quantitative quality metrics are employed. These metrics are traditionally categorised, depending on the availability of a GT image — a limitation discussed in Section 2.3.2 — into full-reference and no-reference quality metrics.

In the whole experimental evaluation of images, in order to compute the metrics, the images were cropped by 15 pixels around the edges so as to avoid border artefacts that could bias the results.

In this study, we evaluate super-resolved images $\widehat{\mathbf{H}}$ using a GT image $\mathbf{H}$ with four full-reference metrics: peak signal-to-noise ratio (PSNR), erreur relative globale adimensionnelle de synthèse (ERGAS)[1], universal image quality index (UIQI), and spectral angle mapper (SAM).

The peak signal-to-noise ratio (PSNR) is a pixel-based error metric that expresses the ratio between the maximum possible pixel value in $\mathbf{H}$ and its distortion from $\widehat{\mathbf{H}}$. In this thesis, the PSNR is consistently computed as the mean of the values obtained for each channel.

$$PSNR(\mathbf{H}, \widehat{\mathbf{H}}) = \frac{1}{C} \sum_{c} \left[20 \log_{10} \left(\frac{\max(\mathbf{H}_c)}{RMSE(\mathbf{H}_c, \widehat{\mathbf{H}}_c)} \right) \right], \tag{4.1}$$

where $\mathbf{H}_c$ and $\widehat{\mathbf{H}}_c$ are the c-th of C channels of $\mathbf{H}$ and $\widehat{\mathbf{H}}$, $\max(\mathbf{H}_c)$ is the maximum pixel value of $\mathbf{H}_c$, and $RMSE(\mathbf{H}_c, \widehat{\mathbf{H}}_c)$ is the distortion between $\mathbf{H}_c$ and $\widehat{\mathbf{H}}_c$ in terms of root mean square error (RMSE). The RMSE is defined as the square root of the MSE (Equation 3.5). As lower values of the MSE, and consequently of the RMSE, indicate better performance, higher values of the PSNR, expressed in dB, correspond to better image quality.

The erreur relative globale adimensionnelle de synthèse (ERGAS) (Wald, 2002) is a pixel-based error metric that quantifies the normalised average error per spectral channel c between $\mathbf{H}$ and $\widehat{\mathbf{H}}$.

$$ERGAS(\mathbf{H}, \widehat{\mathbf{H}}) = \frac{100}{r} \sqrt{\frac{1}{C} \sum_{c} \left(\frac{RMSE(\mathbf{H}_c, \widehat{\mathbf{H}}_c)}{\mu_{\mathbf{H}_c}} \right)^2}, \tag{4.2}$$

where $\mu_{\mathbf{H}_c}$ is the mean value of $\mathbf{H}_c$. A lower value of the ERGAS indicates higher image quality.

[1]From the French is commonly translated as *relative dimensionless global error in synthesis*.

The universal image quality index (UIQI) (Wang and Bovik, 2002) is a structural similarity metric that combines loss of correlation, luminance distortion, and contrast distortion between $\mathbf{H}$ and $\widehat{\mathbf{H}}$, represented, respectively, by the three terms in Equation 4.3.

$$Q(\mathbf{H}, \widehat{\mathbf{H}}) = \frac{1}{C} \sum_{c} \frac{\sigma_{\mathbf{H}_c \widehat{\mathbf{H}}_c}}{\sigma_{\mathbf{H}_c} \sigma_{\widehat{\mathbf{H}}_c}} \cdot \frac{2\mu_{\mathbf{H}_c} \mu_{\widehat{\mathbf{H}}_c}}{\mu_{\mathbf{H}_c}^2 + \mu_{\widehat{\mathbf{H}}_c}^2} \cdot \frac{2\sigma_{\mathbf{H}_c} \sigma_{\widehat{\mathbf{H}}_c}}{\sigma_{\mathbf{H}_c}^2 + \sigma_{\widehat{\mathbf{H}}_c}^2}, \quad (4.3)$$

where $\sigma_{\mathbf{X}}$ denotes the standard deviation of $\mathbf{X}$, either $\mathbf{H}_c$ or $\widehat{\mathbf{H}}_c$, and $\sigma_{\mathbf{H}_c \widehat{\mathbf{H}}_c}$ is the covariance between $\mathbf{H}_c$ and $\widehat{\mathbf{H}}_c$. The UIQI index ranges from -1 to 1, where 1 indicates perfect similarity. Following the original paper on UIQI (Wang and Bovik, 2002), since image quality is often spatially variant, we apply Equation 4.3 to local regions of size $w \times w$, with $w = 32$, sliding pixel by pixel both horizontally and vertically over the image. The overall UIQI is then computed as the mean of all local values. To ensure numerical stability, following previous works such as the research on structural similarity index measure (SSIM) (Wang et al., 2004) — the successor of UIQI — we introduce two small constants in Equation 4.3, $K_1 = K_2 = 10^{-9}$, to avoid divisions by zero.

The spectral angle mapper (SAM) (Yuhas et al., 1992) measures spectral similarity by computing the angle between the spectral vectors of each pixel in the LR image $\mathbf{L}$ and the corresponding pixel in the downsampled version of the super-resolved image $\widehat{\mathbf{H}}$ with scaling ratio r, denoted as $\widehat{\mathbf{H}}_{\downarrow r}$. The overall SAM is obtained as the average of these angles over all pixels, providing a global measure of spectral agreement:

$$SAM(\mathbf{L}, \widehat{\mathbf{H}}_{\downarrow r}) = \frac{1}{N} \sum_{i} \arccos\left(\frac{L_i \cdot \widehat{H}_{\downarrow r, i}}{\|L_i\| \|\widehat{H}_{\downarrow r, i}\|} \right), \quad (4.4)$$

where L_i is the value of the i-th pixel in the LR image $\mathbf{L}$, $\widehat{H}_{\downarrow r, i}$ denotes the i-th pixel value of the downsampled version of the

reconstructed image $\widehat{\mathbf{H}}$ with scaling ratio r, N is the number of pixels in both images, and $||\cdot||$ denotes the Euclidean norm. Values of SAM range from 0 (perfect spectral agreement) to π.

In this thesis, the evaluation metrics are referred to as PSNR, ERGAS, UIQI, and SAM in the text, and denoted by $PSNR$, $ERGAS$, Q, and SAM when reporting their numerical values (e.g. in tables and plots).

4.1.3 Image quality assessment protocols

As regards image quality assessment, although the ideal scenario involves access to a HR GT image, this is often unfeasible in remote sensing, where the primary motivation for SR approaches lies in the lack of HR imagery. To address this, we adopted two complementary evaluation protocols (Section 2.3.2): the RR protocol and the FR protocol.

As GT data are not available in our case, the reduced resolution (RR) protocol is adopted (Section 2.3.2.1). Following the protocol, the available LR image is down-scaled using the acquisition model in Equation 2.4. The SISR algorithm is then applied to super-resolve the down-scaled image back to its original resolution, allowing the original LR image to serve as a reference for quantitative evaluation.

The full resolution (FR) protocol is also considered, in which the super-resolved output is assessed directly at the original resolution without relying on a reference image (Section 2.3.2.1). This protocol complements the RR evaluation by providing an assessment of the model's ability to reconstruct high-frequency content in the absence of scale-invariance assumptions.

4.1.4 Experimental setting for super-resolution evaluation

This section describes all configurations of the proposed S5Net and its training strategies, as well as the benchmark approaches and algorithms adopted to assess SISR performance on S5P data in Section 4.2.

4.1.4.1 Ablation study of S5Net

Regarding S5Net, the architecture that gave the best results and was therefore used in the experiments corresponds to the parameter selection $n_1 = c, n_2 = 64, n_3 = 32, f_1 = 15, f_2 = 9, f_3 = 5, f_4 = 5$, which reflects a common design choice.

Given the importance of the upsampling phase in SISR techniques, an ablation study on S5Net was conducted focusing on its transposed convolution layer in the first subsection of Section 4.2.1.1.

In particular, a variant of S5Net without the initial transposed convolution layer (S5Net_cubic) was trained and evaluated under the same conditions, replacing the learned upsampling with a fixed bicubic interpolation technique (see Section 2.3.1.1).

4.1.4.2 Setup of benchmark approaches for super-resolution

For comparative purposes, a non-blind deconvolution reconstruction-based method (Section 2.3.1.2) was included in the evaluation and thoroughly optimised. In particular, we built on the approach proposed in (Eilers and Ruckebusch, 2022) in which regularization is employed to handle SISR as a reverse problem, as already formalised in Equation 2.6.

The Euclidean norm is used as the loss function $\mathcal{L}(\cdot)$ and two regularizations are also considered:

- A shrinking parameter μ is used to weigh the norm of the estimate $\tilde{\mathbf{h}}_c$. However, this regularization alone can easily result in Gibbs artefacts.
- To overcome the issue introduced by the first term, a smoothing parameter λ is used to weigh the prime differences of the estimate (Eilers and Ruckebusch, 2022).

In short, the objective function is formalised as follows:

$$Q = |\mathbf{l}_c - \mathbf{S}\tilde{\mathbf{h}}_c|^2 + \mu|\tilde{\mathbf{h}}_c|^2 + \lambda|\mathbf{D}\tilde{\mathbf{h}}_c|^2, \tag{4.5}$$

where $|x|$ denotes the Euclidean norm of a generic vector x and $\mathbf{D}$ is the matrix implementing the prime differences. The solution $\tilde{\mathbf{h}}_c$ is computed via an iterative algorithm, which runs until either the residual norm falls below a threshold δ or the maximum number of iterations is reached (Eilers and Ruckebusch, 2022), specifically the conjugate gradient algorithm (CGA) (Hestenes and Stiefel, 1952).

Two separate analyses were conducted to optimise this benchmark approach: (i) tuning the parameters μ and λ in the objective function (Equation 4.5); and (ii) testing different initialization strategies for the starting image $\widehat{\mathbf{H}}$. The threshold parameter δ and the maximum number of iterations were set accordingly.

A baseline bicubic interpolation method (Section 2.3.1.1) (Aiazzi et al., 2013), the optimised reconstruction-based approach, and four DL-based methods — namely, two CNNs, SRCNN (Dong et al., 2016a) and VDSR (Kim et al., 2016a), a residual network, EDSR (Lim et al., 2017), and a pixel-attention network,

PAN (Zhao et al., 2020) (Section 2.3.1.3) — were all considered under the same experimental conditions and on the same dataset as the proposed S5Net.

4.1.4.3 Training strategy and training data preparation for S5Net

All experiments involving S5Net included a fine-tuning phase to adapt the network to the S5P images in an image-specific fashion. As discussed in Wang et al. (2021b) and Hosseiny et al. (2024), many recent approaches, like the zero-shot SR in Shocher et al. (2018), thus adopt online learning strategies, where a pre-trained network is further refined on a downsampled version of the image to be super-resolved, assuming scale invariance. In line with this paradigm, fine-tuning was performed independently for each test image as specified in Section 3.2.1.2.

During this training phase, the same acquisition model used in the RR assessment protocol was employed to generate synthetic LR-HR training pairs. For FR protocol assessment, the available image was used as the training target, and the input was obtained via a single degradation using the chosen scaling factor. For RR protocol evaluation, the available LR image was used as the network output during testing, while its degraded version was used as the input. This intermediate image was also adopted as the training target, and the network training input was generated via an additional degradation step.

In both protocols, each spectral channel of the training image is fine-tuned independently of the others. The convolutional layers are trained for 2000 epochs, and when applicable an additional phase of 1000 epochs is used for the transposed convolutional layer. The total number of fine-tuning epochs

per spectral channel is thus set to $i_s = 3000$. The parameter d in the loss function, whose role was explained in Section 3.2.1.2, was set to $d = 8$. When applicable, this approach is referred to as S5Net_all.

Preliminary evaluation with conventional supervised training: The objective of the just described training strategy is not to achieve cross-image generalization, but to recover spatial details consistent with the instrument's degradation model. This image-specific fine-tuning ensures that the reconstruction remains physically grounded and free from biases introduced by external datasets. In order to assess the generalization capability of the proposed network, we perform a preliminary evaluation using simulated HR references in Section 4.2.2.3.

In this setting, conventional supervised training is applied on synthetically degraded images, providing a complementary check on the network's ability to recover fine-scale spatial structures when a GT is available. The dataset is split into training, validation, and test sets to ensure proper model assessment. This test is intended solely as a verification and does not replace the primary image-specific training protocol.

Effect of PSF knowledge in S5Net training: First, a sensitivity analysis (see Section 4.2.2.2) evaluates the robustness of S5Net to moderate mismatches between the assumed and true PSF. Then, to investigate the role of the knowledge of sensor-specific PSFs into the proposed SISR model, two variants were considered for S5Net in Section 4.2.2.4:

- S5Net_match, where the degradation filter to generate the synthetic training image was derived from the gains at the Nyquist frequency (Table 3.1) of the sensor-specific

PSF model (as formalised in Section 3.1);

- S5Net_nomatch, where the filter was defined using fixed symmetrical gains of 0.3, a common assumption in the literature.

Managing dimensionality in the training process of S5Net: For the dynamic multi-directional cascade fine-tuning algorithm (S5Net_dynamic), the central channel ($c = C/2$) was fine-tuned for $i_s = 3000$ iterations. All remaining channels ($c \neq C/2$) were fine-tuned for i_c iterations, initialised from the model of the adjacent channel as formalised in Equation 3.6.

Specifically, the linear fitting parameters m and q were estimated once from a static variant of the algorithm, in which a fixed number of iterations per channel was adopted. For each channel, the first epoch at which PSNR decreased by more than $\tau = 0.01$ from its maximum was identified, yielding the threshold $i_c^{0.01}$. These thresholds were then regressed against $1 - \rho_{c,c\pm1}$ to obtain m and q. The same configuration was applied both to the full image (global SR, GSR) and to the four DEM-based sub-images (DEM-wise SR, DSR). All results reported in the third subsection of Section 4.2.3 refer to this setup.

To assess the importance of the dynamic setting in reducing the computational complexity (Section 2.3.2.2), the proposed fine-tuning algorithm for S5Net is benchmarked against the baseline S5Net_all fine-tuning approach, as well as two state-of-the-art approaches, (i) S5Net_PCA, where the image is decomposed into principal components via PCA (Section 2.3.2.2), the first $k = 9$ of which are fine-tuned and super-resolved by the network, while the remaining components are upscaled via bicubic interpolation

(Section 2.3.1.1); and (ii) S5Net_static, which applies a static fine-tuning strategy starting with $i_s = 3000$ iterations on the central channel, followed by $i_c = 75$ iterations on all other channels starting from the model of the adjacent channels. In particular, we built upon the computationally efficient fine-tuning methodology presented in Guarino et al. (2024), where the number of iterations i_c is statically set based on the spectral distance between wavelengths. However, since the channels of the S5P radiance images are uniformly distributed, the number of iterations was kept fixed across all channels. When possible, these benchmark approaches were applied both in the GSR and DSR settings.

The number of principal components k in S5Net_PCA, the number of static iterations i_c in S5Net_static and the parameter τ in S5Net_dynamic were all optimised to fairly compare all the approaches.

4.1.5 Software and hardware environment

To ensure reproducibility, the software and hardware environments used in the experiments are detailed. All experiments involving neural networks were implemented in PyTorch version 2.0.0 and executed on an NVIDIA GTX 1080 Ti GPU.

4.2 Assessment of PSF-aware and spectrally scalable super-resolution

In this section, we present a comprehensive evaluation of the proposed PSF-aware SISR model and all algorithms employed

for super-resolving S5P radiance imagery. We first optimise the proposed approach, then we assess the impact of the derived PSF (Section 3.1) on the performance of the PSF-aware SISR model, and finally conduct a full evaluation of the SISR task on S5P radiance imagery, including analyses of the proposed fine-tuning algorithm.

4.2.1 Optimization of the proposed super-resolution framework

In this section, we discuss the optimization, sensitivity analyses, and ablation study of the proposed architecture.

4.2.1.1 Ablation study of the model

In this section, an ablation study is conducted on the proposed PSF-aware SISR approach, namely S5Net, to assess the need for its architectural components as described in Section 4.1.4.1. For this analysis, we selected four representative channels corresponding to the four DEMs in the S5P Level-1b dataset — specifically, the first channel of the second BD for each DEM — which we refer to as the respective DEM: UV, UVIS, NIR, and SWIR.

In Table 4.2, the results of the ablation study are presented for four monochromatic bands of the ten considered orbits (Table 4.1) once averaged at a scaling ratio of 4. The RR protocol was applied for this analysis, and performance was assessed using the UIQI (Equation 4.3), ERGAS (Equation 4.2), and PSNR (Equation 4.1) metrics (Section 4.1.2).

The network trained with the transposed convolution layer (S5Net) consistently outperforms the version trained with the bicubic interpolator (S5Net_cubic). More precisely, we

	Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$
UV	S5Net_cubic	0.9145	0.9295	40.3332
	S5Net	**0.9148**	**0.9279**	**40.3477**
UVIS	S5Net_cubic	0.9386	3.5691	35.5718
	S5Net	**0.9390**	**3.5657**	**35.5783**
NIR	S5Net_cubic	0.9255	4.9705	34.5615
	S5Net	**0.9257**	**4.9648**	**34.5669**
SWIR	S5Net_cubic	0.9061	3.7882	33.6397
	S5Net	**0.9091**	**3.7317**	**33.7771**
Mean	S5Net_cubic	0.9212	3.3143	36.0266
	S5Net	**0.9222**	**3.2975**	**36.0675**

Table 4.2: Performance of S5Net with the transposed convolution layer (S5Net) and with the standard bicubic interpolator (S5Net_cubic) across four monochromatic bands, averaged over the ten orbits listed in Table 4.1, for a scaling ratio of 4. The best result in each sub-group is highlighted in bold.

observe a notable improvement in every quality index, with an average increase of +0.04 dB in terms of PSNR. This confirms the importance of the transposed convolution layer, which — unlike the static bicubic interpolator — learns its weights during the training phase, thereby providing a decisive advantage in the SISR task.

The most pronounced gains are observed for the SWIR band (+0.14 dB in terms of PSNR), where the ability of the learnable upsampling layer proves particularly beneficial. This result is likely influenced by the lower native spatial resolution across-track of the SWIR channel, which makes it more reliant on the upsampling step to recover fine details.

4.2.1.2 Optimization of the fine-tuning algorithm

To optimise the proposed dynamic multi-directional cascade fine-tuning algorithm (Section 3.2.2), we selected only one entire preprocessed orbit (Section 4.1.1) in order to limit computational complexity, namely IN (Table 4.1).

For the proposed dynamic multi-directional cascade fine-tuning algorithm for S5Net (S5Net_dynamic), Table 4.3 reports the resulting PSNR on the full reconstructed image using S5Net_dynamic, where i_c is dynamically computed through the parameters m and q obtained for different threshold values $\tau \in \{0.01, 0.03, 0.05\}$. Results are shown both for the image considered as a whole (GSR) and for the image partitioned according to the DEM (DSR). The S5Net_all case is used here as a baseline. To reduce the computational load, a subset of 102 spectral channels was uniformly sampled from the entire preprocessed IN orbit to preserve the original channel distribution. The analysis was conducted following the RR protocol with a scaling ratio of 4.

In both the GSR and DSR scenarios, increasing the threshold τ leads to a systematic decrease in the final PSNR, in line with the corresponding reduction in the average number of training epochs per channel. This confirms that an overly aggressive criterion — induced by larger τ values — can negatively affect reconstruction quality. Since the resulting number of epochs remains relatively low compared to S5Net_all in all cases, the configuration corresponding to the highest PSNR was selected, striking a favourable balance between performance and computational efficiency. For this reason, we selected $\tau = 0.01$.

In both cases, an attempt was made to exclude from the analysis all points where the number of epochs was exactly zero — that is, those for which the PSNR curve consistently

	Method	τ	$PSNR\uparrow$	Iterations
GSR	S5Net_dynamic	0.01	*30.2585*	30013
		0.03	30.2350	25817
		0.05	30.2232	23808
DSR	S5Net_dynamic	0.01	*30.5416*	55539
		0.03	30.5328	33946
		0.05	30.5283	26067
	S5Net_all	-	**30.5933**	306000

Table 4.3: PSNR scores and number of fine-tuning iterations achieved by S5Net_dynamic for different threshold values $\tau \in \{0.01, 0.03, 0.05\}$. Results are reported for both the GSR and DSR scenarios. S5Net_all is used as a baseline. The analysis employs 102 uniformly sampled spectral channels from the IN image, with a scaling ratio of 4 under the RR protocol. In the table, the best-performing model overall is highlighted in bold, while the best-performing models for S5Net_dynamic in the GSR and DSR cases are highlighted in italics.

worsened during training. This resulted in a gain of 0.062 dB for the GSR scenario, and 0.004 dB for the DSR scenario in terms of PSNR. Consequently, points with zero epochs were excluded from the fitting procedure.

The resulting linear fittings for $\tau = 0.01$, excluding zero points, between $i_c^{0.01}$ and the Pearson correlation coefficient between the current channel c and the adjacent channel $c \pm 1$ ($\rho_{c,c\pm1}$) for all considered channels $c \in \{1, \ldots, C\}$ in the image ($C = 102$) are shown in Figure 4.3 for both the GSR and DSR scenarios. These fittings were used to derive the parameters m and q, which are employed in S5Net_dynamic to dynamically determine the number of training epochs i_c. The parameters representing the slope and intercept of the linear fittings are $m = 98.32$ and $q = 7.71$ for the GSR case, and $m = 112.29$ and $q = 4.90$ for the DSR case, respectively.

Most points are clustered on the left side of the plot because

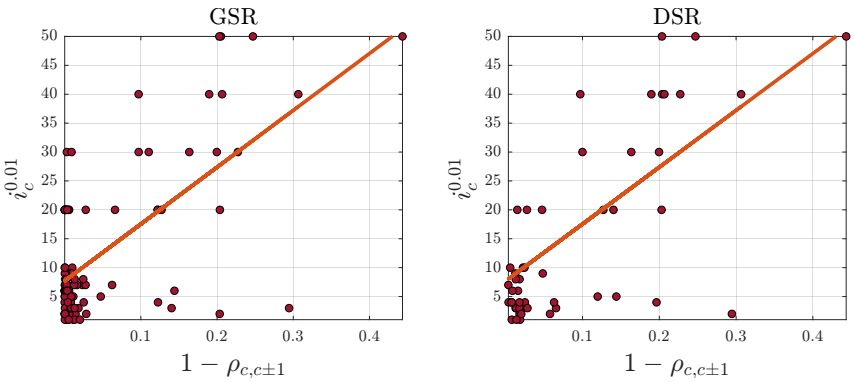

Figure 4.3: The figure shows scatter plots of the number of epochs $i_c^{0.01}$ at which S5Net_static exceeds a threshold corresponding to 1% below the maximum PSNR value for each channel c, plotted against the residual correlation between the current channel c and its adjacent channel $c \pm 1$ (i.e., $1 - \rho_{c,c\pm1}$). Linear fits for both the GSR and DSR scenarios are also shown.

the correlations between adjacent channels are generally high. A small percentage (2–4%) of channels show low correlation values (points on the right), which, as discussed in Section 4.1.1, is probably due to calibration issues. Despite these outliers, a correlation exists between the residual correlation $1 - \rho_{c,c\pm1}$ and the number of epochs $i_c^{0.01}$, as quantified by the Pearson correlation coefficients of 0.66 for GSR and 0.73 for DSR. This validates and justifies the derived dynamic rule for dynamically computing the number of training epochs i_c.

Although the optimization of the dynamic rule was performed on a single representative orbit (IN) and a uniform channel sampling, this choice was guided by computational considerations and supported by the observation that the inter-channel correlation structure is consistent across orbits. As shown in subsequent experiments (last subsection of Section 4.2.3), the dynamic rule generalises effectively to the whole set of channels and all orbits, confirming the robustness

of the estimated parameters m and q.

4.2.2 Evaluation of the impact of PSF knowledge on super-resolution

In this section, we evaluate the impact of incorporating the PSF knowledge — derived and analysed in Section 3.1 — in the proposed approach for SISR of S5P radiance images.

First, the benchmark deconvolution-based approach, solved through the CGA, is optimised to provide a consistent reference for comparison. A sensitivity analysis is performed to evaluate the robustness of the proposed S5Net framework to mismatches between the assumed and true PSFs. Then, in order to validate the effectiveness of the proposed S5Net framework, we conducted a generalization capability analysis simulating the presence of GTs. In the end, a comprehensive quantitative and qualitative assessment is conducted using both RR and FR protocols via the original image-specific training strategy.

All the results are compared with state-of-the-art methodologies to contextualize performance and computational complexity.

For these preliminary validations of the model, we selected a single representative channel for each of the four DEMs in the whole dataset (Section 4.1.1). Specifically, we used the first channel of the second BD from each DEM, which corresponds to the central channel of the whole DEM, in accordance with the results for the derivation of the PSFs in Section 3.1. This resulted in four monochromatic images for each orbit: first channel of either BD 2 (UV), BD 4 (UVIS), BD 6 (NIR), or BD 8 (SWIR), for a total of 40 monochromatic images.

4.2.2.1 Optimization of benchmark deconvolution-based approach solved through CGA

In this section, the deconvolution-based benchmark SISR approach solved with the CGA is optimised as described in Section 4.1.4.2.

As regards the first analysis, both μ and λ (Equation 4.5) were varied over the discrete set $\{0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 100\}$. The averaged PSNR values (Equation 4.1) — computed on the four monochromatic images and subsequently averaged over all ten orbits (Table 4.1) are reported in Figure 4.4 for scaling ratio 4. The RR protocol is employed. In the figure, all reported PSNR values were raised to the power of $\gamma = 5$ in order to stretch the dynamic range of the visualised data. The values of λ and μ are expressed on a logarithmic scale.

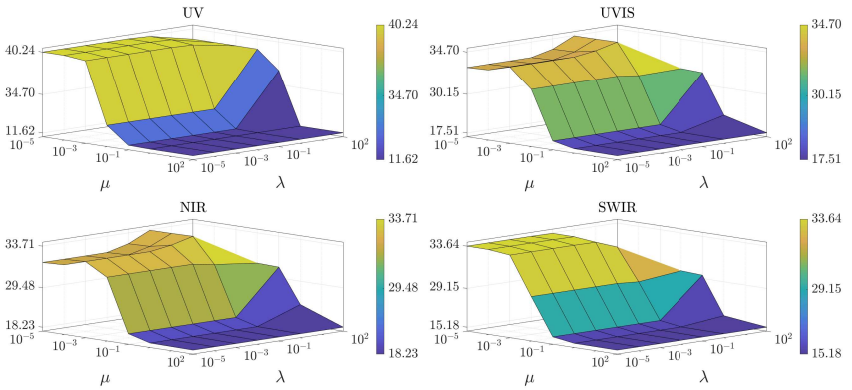

Figure 4.4: Averaged PSNR values (Equation 4.1) obtained for the benchmark deconvolution-based approach solved with the CGA by varying the parameters μ and λ over the discrete set $\{0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 100\}$. Results were computed on the four monochromatic images and subsequently averaged over all ten orbits (Table 4.1) for scaling ratio 4 and the RR protocol.

This analysis demonstrated that, when μ varies in the interval

$[0.00001, 0.001]$, the values do not change significantly with any value of λ for UV, for values of λ in the interval $[0.00001, 1]$ for SWIR, and for values $\{0.1, 1\}$ of λ for UVIS and NIR. This behaviour indicates that the method is generally insensitive to variations of λ within the considered range, provided that μ remains within $[0.00001, 0.001]$. Such robustness is desirable, as it reduces the need for tuning the parameters for each DEM. Consequently, the values of $\mu = 0.00005$ and $\lambda = 0.1$ were selected to guarantee a stable and consistent behaviour across all monochromatic images.

As for the second analysis, different initialization strategies — namely nearest neighbour, bilinear, as in Eilers and Ruckebusch (2022), bicubic, Lanczos, and 23-tap interpolation (Section 2.3.1.1) — are compared under the optimal setting of $\mu = 0.00005$ and $\lambda = 0.1$. The PSNRs obtained for each strategy are first averaged across the four monochromatic images and then further averaged over the ten considered orbits (Table 4.1). The final results, reported in Table 4.4, refer to a scaling ratio of 4 and the RR protocol.

Initialization	$PSNR \uparrow$
Nearest neighbour	34.6880
Bilinear	35.0829
Bicubic	35.1164
Lanczos	<u>35.1183</u>
23-tap	**35.1325**

Table 4.4: Comparison of different image interpolation methods used for initialization in the deconvolution-based SISR framework, evaluated at a scaling ratio of 4 under optimal parameters ($\mu = 0.00005$, $\lambda = 0.1$) and the RR protocol. For each method, the PSNR is averaged first across the four monochromatic images and then across the ten orbits (see Table 4.1). The bolded and underlined values indicate, respectively, the best and second-best performing strategies.

Among all tested interpolation strategies, 23-tap interpolation achieves the best performance, surpassing even bilinear interpolation — the approach originally adopted in the reference implementation of the deconvolution-based method (Eilers and Ruckebusch, 2022). Based on these results, 23-tap interpolation was selected as the preferred initialization strategy for all subsequent analyses.

The threshold parameter was set to $\delta = 0.0001$ times the norm of the LR image and the maximum number of iterations was chosen to be equal to 200 to avoid a high temporal complexity.

4.2.2.2 Sensitivity analysis to PSF mismodeling

In practical scenarios, the exact sensor's PSF may not be perfectly known, and a certain degree of mismatch between the assumed and the actual PSF is therefore expected. For this reason, we evaluate the sensitivity of the proposed model to PSF mismodeling by analysing how the reconstruction performance varies when the PSF parameters used during fine-tuning differ from those used to degrade the test images.

Specifically, we consider Gaussian PSFs with equal gains along the two spatial directions and analyse the reconstruction PSNR (Equation 4.1) obtained by our proposed S5Net when the gain at the Nyquist frequency used for fine-tuning varies in the range $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, while the gain used to degrade the test images is kept fixed. Three representative cases are considered, corresponding to degradation gains at the Nyquist frequency equal to 0.2, 0.4, and 0.6, whose results are shown in Figure 4.5, Figure 4.6, and Figure 4.7, respectively. The analysis is performed under the RR protocol with scaling ratio 4. In each plot, every curve corresponds to a different orbit for all ten orbits (Table 4.1), while the dashed red vertical line marks the matching condition, i.e., when the gains used

for degradation and fine-tuning coincide.

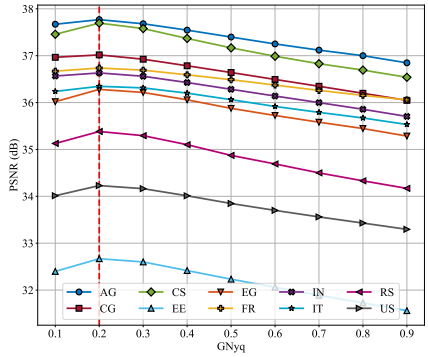

Figure 4.5: Sensitivity of S5Net PSNR to PSF mismodeling for a Nyquist-frequency gain of 0.2. PSNR is reported under the RR protocol and scaling ratio 4 as a function of the fine-tuning gain (0.1–0.9). The dashed red line marks the matching condition between degradation and fine-tuning gains.

As expected, the highest PSNR values are achieved when the assumed and actual PSF gains match. More importantly, however, the curves exhibit a relatively flat region around this optimal point in all three figures, indicating that the reconstruction performance degrades only marginally for moderate PSF mismatches. This behaviour suggests that the proposed S5Net model is robust to small inaccuracies in the PSF parameters used during fine-tuning. The presence of this area of minimal sensitivity around the optimal gain is particularly relevant in practical applications, where the true sensor PSF may only be approximately known, yet the model is still able to maintain high reconstruction fidelity.

4.2.2.3 Generalization capability evaluation

As discussed in Section 2.3.2.1, real HR references are unavailable for S5P data, which precludes the use of conventional computer vision training strategies. Nonetheless,

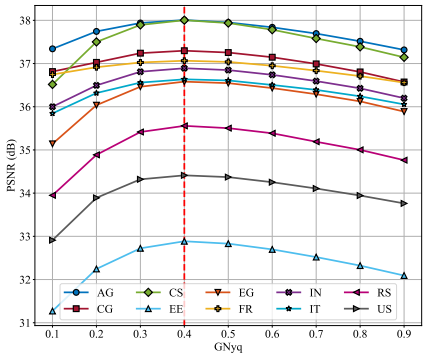

Figure 4.6: Sensitivity of S5Net PSNR to PSF mismodeling for a Nyquist-frequency gain of 0.4. PSNR is reported under the RR protocol and scaling ratio 4 as a function of the fine-tuning gain (0.1–0.9). The dashed red line marks the matching condition between degradation and fine-tuning gains.

to demonstrate the importance of adapting the network parameters to the specific acquisition sensor, without taking into account the peculiarities of the individual image being considered, we tested the generalization capability of the proposed approach through synthetic HR-LR pairs obtained by artificially degrading the available images.

Given the dataset of 10 S5P images (described in Table 4.1), we performed a conventional training/validation/testing split by generating five different combinations inspired by a cross-validation logic, as follows:

1. Training on AG, CG, CS, EE, EG, and FR; validation on IN and IT; and test on RS and US.

2. Training on AG, CG, RS, US, IN, and IT; validation on CS and EE; and test on EG and FR.

3. Training on AG, CG, CS, EE, EG, and FR; validation on RS and US; and test on IN and IT.

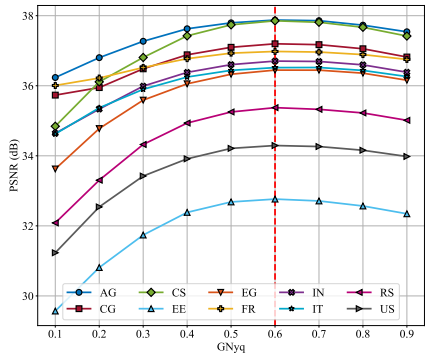

Figure 4.7: Sensitivity of S5Net PSNR to PSF mismodeling for a Nyquist-frequency gain of 0.6. PSNR is reported under the RR protocol and scaling ratio 4 as a function of the fine-tuning gain (0.1–0.9). The dashed red line marks the matching condition between degradation and fine-tuning gains.

4. Training on IN, IT, RS, US, EG, and FR; validation on AG and CG; and test on CS and EE.

5. Training on IN, IT, RS, US, CS, and EE; validation on EG and FR; and test on AG and CG.

This procedure was repeated for three different scaling ratios (i.e., 2, 3, and 4) and for the four selected spectral channels, i.e., US, UVIS, NIR, and SWIR. Results were then averaged across the five splits.

Tables 4.5–4.7 report the quantitative performance of S5Net compared with the chosen baseline (bicubic interpolation) and five state-of-the-art methods — deconvolution solved with the CGA, once optimised, SRCNN (Dong et al., 2016a), VDSR (Kim et al., 2016a), EDSR (Lim et al., 2017), and PAN (Zhao et al., 2020) — in terms of the three selected image quality metrics (Section 4.1.2), namely, UIQI (Equation 4.3), ERGAS (Equation 4.2) and PSNR (Equation 4.1), across all channels, respectively for the three scaling ratios. The prediction time in seconds (s) is reported as well.

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9503	4.8341	38.4310	0.0098
CGA	0.9556	4.9916	38.3793	1.1530
SRCNN	<u>0.9572</u>	<u>4.4810</u>	<u>39.0575</u>	0.0860
VDSR	0.9537	4.9223	38.3349	0.0165
EDSR	0.9402	5.7815	37.1389	0.0320
PAN	0.9403	5.7561	37.1667	0.0283
S5Net	**0.9623**	**4.1775**	**39.5788**	0.0050

Table 4.5: Quantitative evaluation of the generalization capability of the proposed SR methodology with a scaling factor of 2. S5Net is trained using five distinct train/validation/test combinations (as described in Section 4.2.2.3), and the reported values correspond to the average of the four monochromatic channels over all resulting test sets. The best overall result for each metric is highlighted in bold, while the second-best is underlined.

As shown in Tables 4.5–4.7, S5Net consistently achieves the best performance across all image quality metrics and for all three scaling ratios. The improvement is evident when compared not only with the baseline (bicubic interpolation) but also with all five state-of-the-art methods considered in this study.

For a scaling factor of 2, S5Net attains the highest values of UIQI and PSNR (0.96 and 39.58 dB, respectively), and the lowest ERGAS (4.18), outperforming the second-best method (SRCNN) by a clear margin. Similar behaviour is observed for a scaling factor of 3: S5Net again ranks first in every metric, with a PSNR gain of approximately 0.37 dB over VDSR. The trend persists for the most challenging setting, scale 4 (Table 4.7), where S5Net achieves a PSNR of 36.19 dB, exceeding the second-best method by more than 0.27 dB, and obtains both the highest UIQI (0.92) and the lowest ERGAS (3.25).

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9248	4.1364	36.5157	0.0111
CGA	0.9298	4.3872	36.1950	13.8307
SRCNN	0.9337	3.9139	36.9758	0.1719
VDSR	<u>0.9349</u>	<u>3.8732</u>	<u>37.0512</u>	0.0179
EDSR	0.9334	3.9455	36.9273	0.0296
PAN	0.9342	3.8820	37.0288	0.0290
S5Net	**0.9414**	**3.6885**	**37.4180**	0.0048

Table 4.6: Quantitative evaluation of the generalization capability of the proposed SR methodology with a scaling factor of 3. S5Net is trained using five distinct train/validation/test combinations (as described in Section 4.2.2.3), and the reported values correspond to the average of the four monochromatic channels over all resulting test sets. The best overall result for each metric is highlighted in bold, while the second-best is underlined.

These results highlight that the advantage of S5Net is not restricted to a specific scale or metric: the network consistently exhibits stable and superior performance across all experimental configurations. This confirms that, when embedded in a conventional supervised framework, S5Net has a good generalization ability and is able to capture and reconstruct fine-scale spatial structures that other externally supervised deep networks fail to recover.

Some super-resolved images are shown in Figures 4.8 and 4.9. The figures present false-colour representations using the SWIR, NIR, and UVIS monochromatic images mapped to red, green, and blue channels, respectively. They both include the GT image, the outputs of three representative benchmark methods — namely, bicubic interpolation, SRCNN (Dong et al., 2016a), and PAN (Zhao et al., 2020) — and the results of our S5Net for a scaling ratio of 4. All ten orbits (Figure 4.2) are shown, in particular IN, US, EG, RS, and CS are shown in

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9038	3.5645	35.4515	0.0235
CGA	0.9109	3.7940	35.1325	2.1539
SRCNN	0.9138	3.4052	35.8312	0.0398
VDSR	<u>0.9147</u>	<u>3.3681</u>	<u>35.9058</u>	0.0274
EDSR	0.9077	3.6263	35.3821	0.0269
PAN	0.9074	3.6139	35.4060	0.0225
S5Net	**0.9234**	**3.2481**	**36.1855**	0.0051

Table 4.7: Quantitative evaluation of the generalization capability of the proposed SR methodology with a scaling factor of 4. S5Net is trained using five distinct train/validation/test combinations (as described in Section 4.2.2.3), and the reported values correspond to the average of the four monochromatic channels over all resulting test sets. The best overall result for each metric is highlighted in bold, while the second-best is underlined.

Figure 4.8 and AG, EE, FR, IT, and CG are shown Figure 4.9, with a 128 × 128 pixel region, selected to highlight differences in pixel-level reconstruction quality.

By inspecting Figures 4.8 and 4.9, it becomes evident that, across all orbits, S5Net delivers visibly sharper and more faithful reconstructions, particularly noticeable along complex edge structures such as coastlines. Examples include the Caspian Sea and the Red Sea in the CS and RS orbits, respectively, the Italian coastline in IT, and the Gulf of Mexico in US. The precise delineation of these contours emphasises the method's superior ability to recover spatial details that remain challenging for other approaches. Bicubic interpolation produces overly smoothed outputs, with fine structures largely suppressed, whereas state-of-the-art DL methods such as PAN (Zhao et al., 2020) yield a softer, blended appearance and still exhibit residual blur, especially around high-frequency transitions.

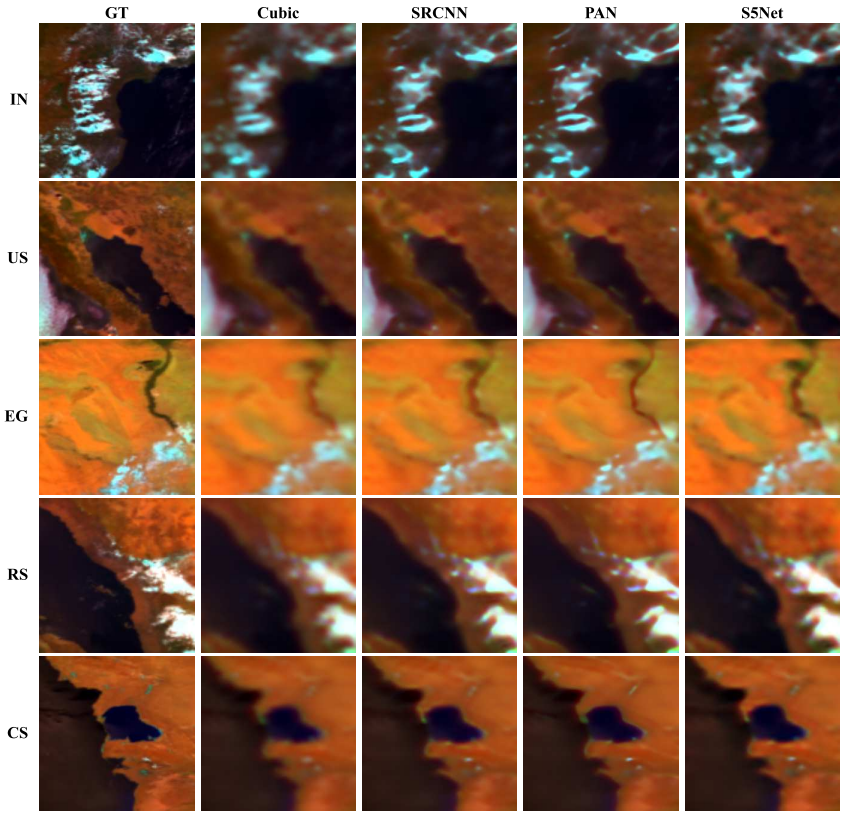

Figure 4.8: False-colour visualisation of super-resolved images for IN, US, EG, RS, and CS orbits for scaling ratio 4 used to assess generalization capability. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 128×128 pixel crop is shown to highlight local reconstruction differences. From left to right: the GT, the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PAN (Zhao et al., 2020), and our S5Net.

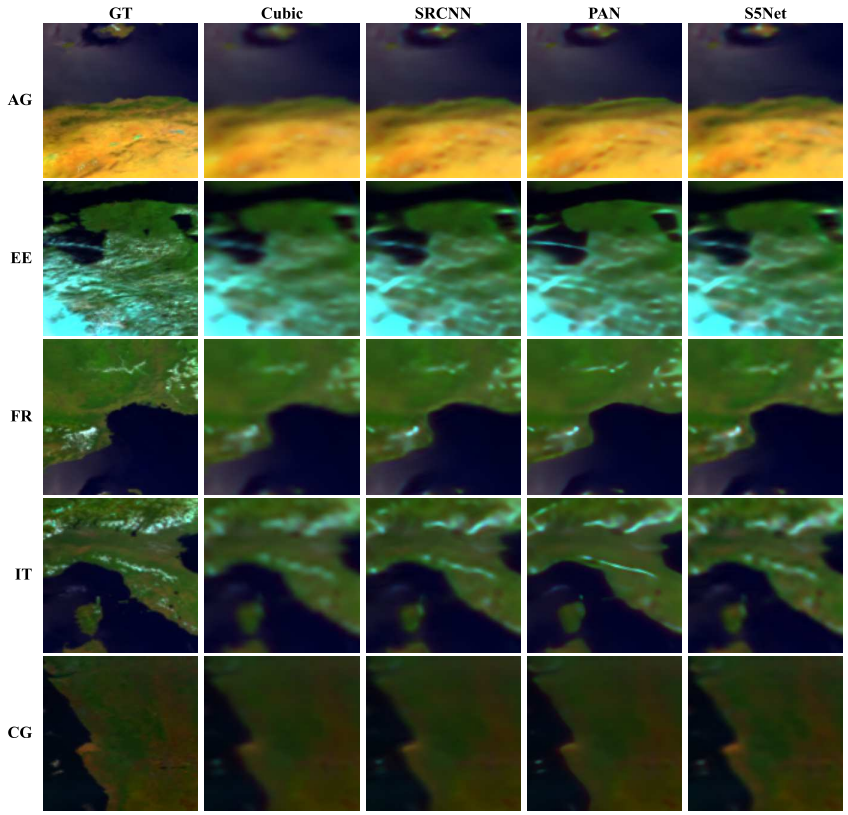

Figure 4.9: False-colour visualisation of super-resolved images for AG, EE, FR, IT, and CG orbits for scaling ratio 4 used to assess generalization capability. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 128×128 pixel crop is shown to highlight local reconstruction differences. From left to right: the GT, the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PAN (Zhao et al., 2020), and our S5Net.

It is important to emphasise that this evaluation relies on artificially generated HR-LR pairs and, therefore, cannot be replicated in operational scenarios where the GT is unavailable. Nevertheless, these results provide a quantitative and qualitatively demonstration that, if HR references were accessible, S5Net would outperform state-of-the-art methods and maintain its effectiveness across different scaling ratios and channels. This analysis complements the image-specific experiments described in the following sections, providing further evidence of the network's robustness and adaptability to the unique characteristics of S5P imagery.

4.2.2.4 Non-supervised super-resolution performance evaluation

This section evaluates the performance of the proposed S5Net in the image-specific, non-supervised training configuration. The analysis is structured into two components. First, the performance of the proposed SISR model is assessed at reduced resolution, enabling a controlled evaluation against known reference data. Then, the method is tested at full resolution on real S5P imagery to characterise its behaviour under real-world conditions.

Reduced resolution assessment: The quantitative evaluation of the impact of PSF knowledge on the proposed SISR method (S5Net) at RR is presented in Tables 4.8–4.10 for scaling factors of 2, 3, and 4, respectively. For each factor, the evaluation includes three selected image quality metrics (Section 4.1.2), namely, UIQI (Equation 4.3), ERGAS (Equation 4.2) and PSNR (Equation 4.1). The prediction time in seconds (s) is reported as well. Results are reported as averages over the four monochromatic images, aggregated

across all orbits listed in Table 4.1.

The performance of our PSF-aware S5Net (S5Net_match) was compared to its PSF-unaware counterpart, as described in Section 4.1.4.3, to the chosen baseline (bicubic interpolation) and five state-of-the-art methods — deconvolution solved with the CGA, once optimised, SRCNN (Dong et al., 2016a), VDSR (Kim et al., 2016a), EDSR (Lim et al., 2017), and PAN (Zhao et al., 2020).

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9503	4.8341	38.4310	0.0098
CGA	0.9556	4.9916	38.3793	1.1530
SRCNN	<u>0.9572</u>	<u>4.4810</u>	<u>39.0575</u>	0.0860
VDSR	0.9537	4.9223	38.3349	0.0165
EDSR	0.9402	5.7815	37.1389	0.0320
PAN	0.9403	5.7561	37.1667	0.0283
S5Net_nomatch	0.9565	4.7283	38.7048	0.0062
S5Net_match	**0.9599**	**4.3149**	**39.3077**	0.0058

Table 4.8: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at RR with a scaling factor of 2. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

As can be observed in Tables 4.8–4.10, there is a general tendency for the proposed method to outperform all the presented state-of-the-art approaches, including both well-established and more recent ones in the field. Notably, while the performance naturally degrades as the SR ratio increases — due to the increasing ill-posedness of the inverse problem (Section 2.3.1) — our method consistently maintains a clear advantage across all tested scaling ratios. This confirms its robustness and generalization capability, even

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9248	4.1364	36.5157	0.0111
CGA	0.9298	4.3872	36.1950	13.8307
SRCNN	0.9337	3.9139	36.9758	0.1719
VDSR	<u>0.9349</u>	<u>3.8732</u>	<u>37.0512</u>	0.0179
EDSR	0.9334	3.9455	36.9273	0.0296
PAN	0.9342	3.8820	37.0288	0.0290
S5Net_nomatch	0.9338	4.1831	36.5750	0.0055
S5Net_match	**0.9395**	**3.7459**	**37.2971**	0.0053

Table 4.9: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at RR with a scaling factor of 3. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

in more demanding scenarios where fine details need to be reconstructed from increasingly degraded observations. The progressive drop in absolute performance with higher ratios is expected and reflects the intrinsic difficulty of reconstructing details from LR images; however, the relative performance gain with respect to the baseline and state-of-the-art methods remains evident, highlighting the effectiveness of our PSF-aware modelling approach.

Furthermore, as the reported results suggest, when developing a SR technique, the proper design of the PSF used to generate the simulated dataset is crucial. This necessity arises from the fact that the PSF formalises the physical processes involved in image acquisition, as discussed in Section 2.2. Consequently, if image acquisition is not accurately simulated, the resulting inverse problem becomes significantly more challenging, potentially leading to misleading solutions (Section 2.3.1). This issue becomes particularly critical when satellite images

Method	$Q \uparrow$	$ERGAS \downarrow$	$PSNR \uparrow$	**Time**
Cubic	0.9038	3.5645	35.4515	0.0235
CGA	0.9109	3.7940	35.1325	2.1539
SRCNN	0.9138	3.4052	35.8312	0.0398
VDSR	0.9147	<u>3.3681</u>	<u>35.9058</u>	0.0274
EDSR	0.9077	3.6263	35.3821	0.0269
PAN	0.9074	3.6139	35.4060	0.0225
S5Net_nomatch	<u>0.9162</u>	3.6529	35.3972	0.0052
S5Net_match	**0.9222**	**3.2975**	**36.0675**	0.0043

Table 4.10: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at RR with a scaling factor of 4. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

are employed for quantitative analyses, where each pixel value carries meaningful physical information. This is especially true for S5P radiance data, which are used to generate highly accurate air pollution maps.

We also want to underline that the distinction between PSF-aware and PSF-unaware scenarios — typically overlooked in the literature (Sdraka et al., 2022) — becomes especially evident after $i_s = 3000$ epochs of fine-tuning for each channel. It is crucial that the simulated data remain consistent with the real data used for inference, in order to ensure reliable evaluation. This aspect becomes even more relevant in the case of S5P, where the gain at the Nyquist frequency (Table 3.1) tends to deviate significantly from the commonly assumed value (i.e., 0.3), particularly for the UVIS and NIR DEMs. This highlights the need for precise sensor-specific modelling of the PSFs to avoid introducing errors in the subsequent SR and downstream tasks, such as atmospheric pollutants retrieval

and monitoring.

An average gain of approximately 0.67 dB in terms of PSNR is achieved by properly integrating PSF knowledge into the S5Net model. This improvement is consistently observed across all monochromatic images and orbits (see Table 4.1 for details on the considered orbits), as illustrated in Figure 4.10 for the IN, US, EG, RS, and CS orbits, and in Figure 4.11 for the AG, EE, FR, IT, and CG orbits. These figures display the arctangent of the difference between the MSE (Equation 3.5) maps obtained with the PSF-unaware (S5Net_nomatch) and PSF-aware (S5Net_match) configurations, for a scaling factor of 4 and an area of 128 $\times$ 128 pixels of the original image (Figure 4.2). The same colour scale is used across all maps. Green regions indicate where the PSF-aware approach achieves lower reconstruction error. This behaviour is consistent across different scaling factors; maps for scaling ratios 2 and 3 are therefore omitted here for brevity. These visual findings are corroborated by the PSNR (Equation 4.1) and ERGAS (Equation 4.2) values reported in Table 4.10, both of which are directly related to the MSE.

As can be observed in Figures 4.10 and 4.11, the performance difference between the two configurations is less pronounced when the filters involved exhibit similar PSFs. This is particularly evident for the UV detector, where the filters used in the PSF-aware and PSF-unaware scenarios display nearly identical gains at the Nyquist frequency, resulting in minimal differences across the corresponding maps. A similar behaviour is observed for the SWIR DEM, although less markedly, due to the minimal mismatch in gains and the correspondingly reduced improvement in the maps. In contrast, for the UVIS and NIR DEMs, the gains at the Nyquist frequency in the PSF-aware configuration differ greatly from those typically

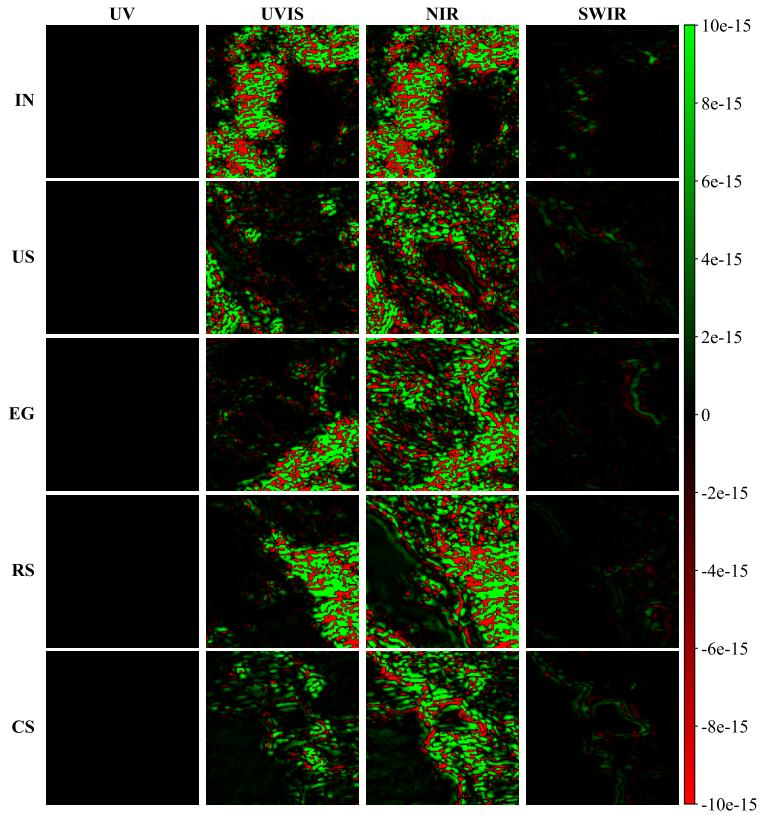

Figure 4.10: Arctangent of the difference between the MSE (Equation 3.5) maps obtained with the PSF-unaware (S5Net_nomatch) and PSF-aware (S5Net_match) configurations, for a scaling factor of 4. Each row corresponds to one of five orbits, either IN, US, EG, RS, or CS, while the four columns represent the spectrally aligned monochromatic images (UV, UVIS, NIR, SWIR) over the same 128×128 pixels. Green tones indicate regions where the PSF-aware approach achieves lower reconstruction error.

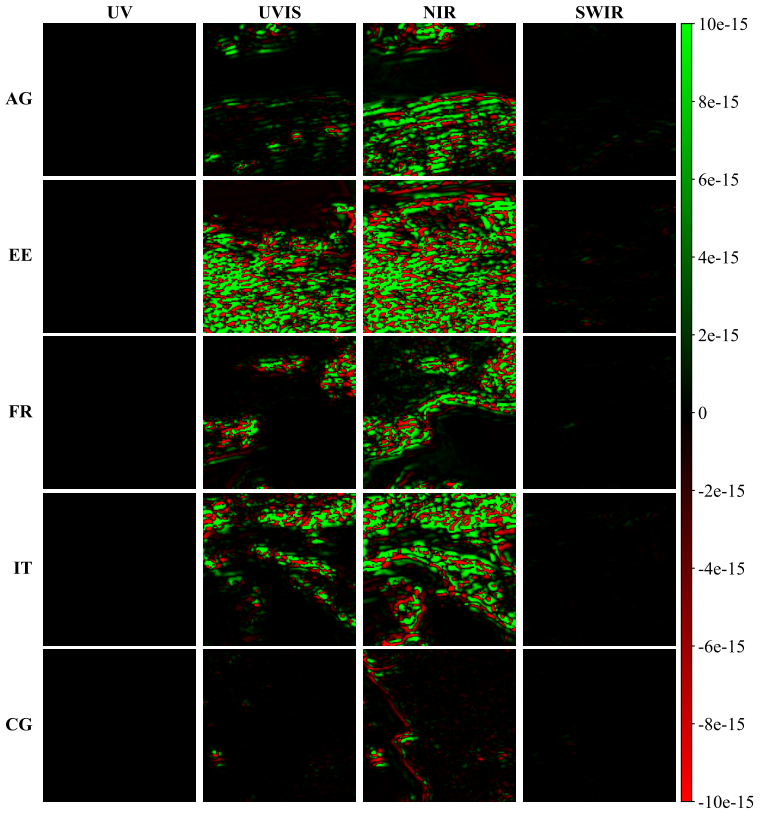

Figure 4.11: Arctangent of the difference between the MSE (Equation 3.5) maps obtained with the PSF-unaware (S5Net_nomatch) and PSF-aware (S5Net_match) configurations, for a scaling factor of 4. Each row corresponds to one of five orbits, either AG, EE, FR, IT, or CG, while the four columns represent the spectrally aligned monochromatic images (UV, UVIS, NIR, SWIR) over the same 128×128 pixels. Green tones indicate regions where the PSF-aware approach achieves lower reconstruction error.

assumed in state-of-the-art SR models, i.e., 0.3 in both directions. This discrepancy results in a notable reduction of reconstruction error in several regions, as highlighted by the more extensive green areas in the maps in Figures 4.10 and 4.11. Overall, the results consistently confirm the superior performance of the PSF-aware configuration, reinforcing the conclusion that explicitly modelling and integrating the actual sensor-dependent PSF leads to improved image quality across all monochromatic images and orbits.

Some super-resolved images are shown in Figure 4.12 and 4.13. The figures present false-colour representations using the SWIR, NIR, and UVIS monochromatic images mapped to red, green, and blue channels, respectively. They both include the GT image, the outputs of two representative benchmark methods — namely, bicubic interpolation and SRCNN (Dong et al., 2016a) — and the results of the PSF-unaware S5Net (S5Net_nomatch) and the PSF-aware S5Net (S5Net_match) for a scaling ratio of 4. All ten orbits (Table 4.1) are shown, in particular IN, US, EG, RS, and CS are shown in Figure 4.12 and AG, EE, FR, IT, and CG are shown Figure 4.13, with the same 128 $\times$ 128 pixel region selected for Figure 4.10 and Figure 4.11 to highlight differences in pixel-level reconstruction quality.

A qualitative comparison clearly shows that S5Net generally yields visually superior results, not only in comparison with bicubic interpolation, which tends to produce overly smooth and blurred images, but also with respect to the SRCNN (Dong et al., 2016a) state-of-the-art SR method. S5Net is able to better preserve details, edges, and patterns, leading to sharper reconstructions across all representative orbits. This improvement is particularly evident in textured areas, where both the interpolated and SRCNN outputs show a visible

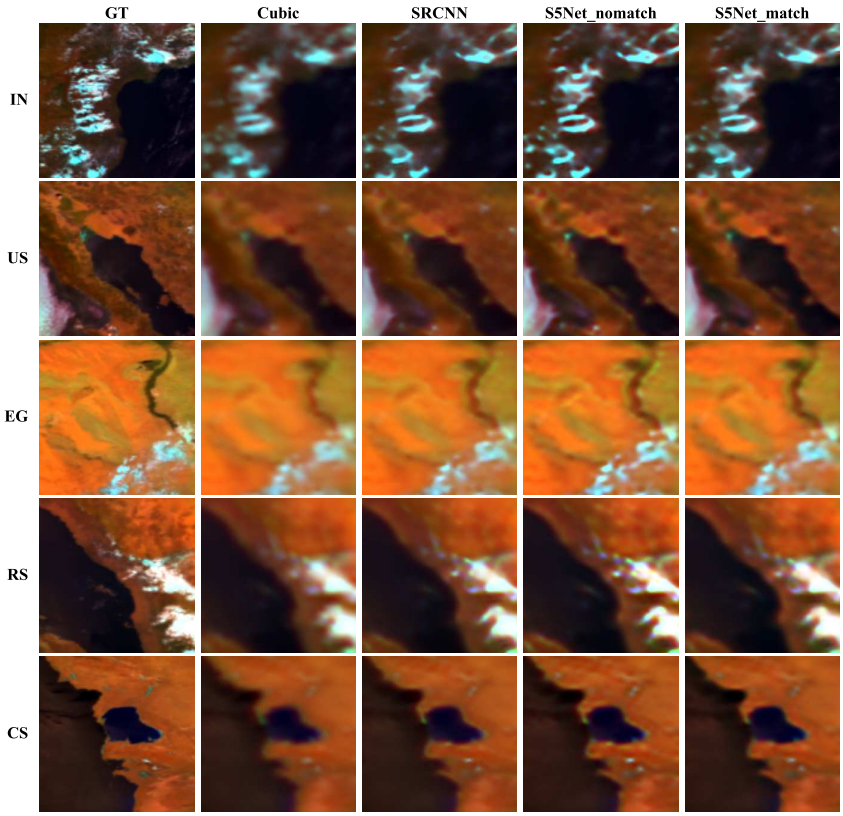

Figure 4.12: False-colour visualisation of super-resolved images for IN, US, EG, RS, and CS orbits for scaling ratio 4 at RR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 128 × 128 pixel crop is shown to highlight local reconstruction differences. From left to right: the GT, the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

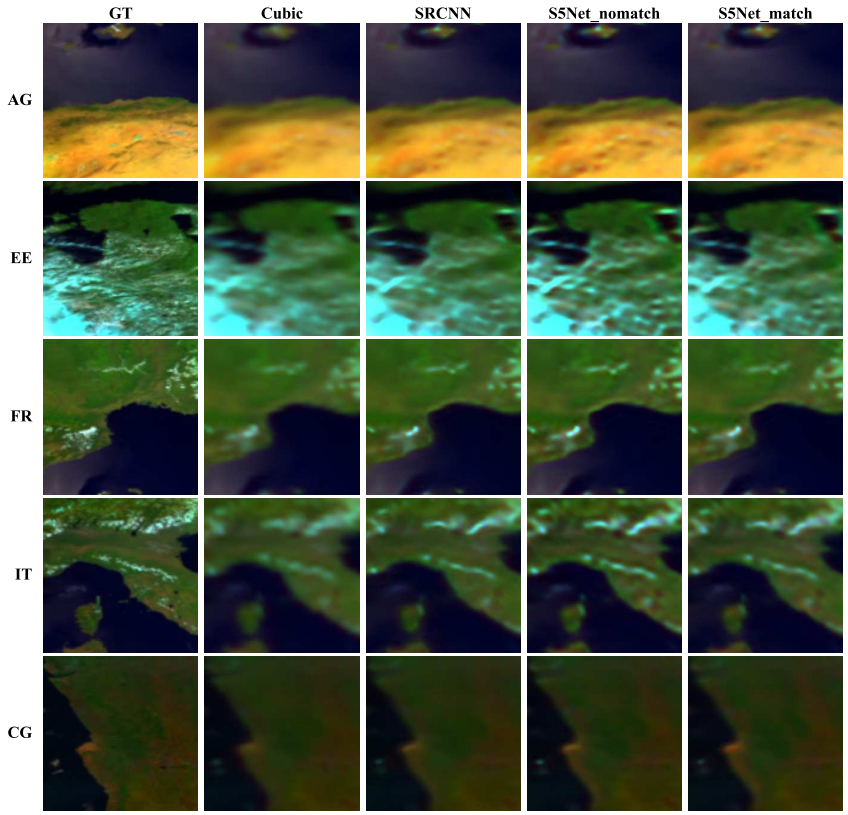

Figure 4.13: False-colour visualisation of super-resolved images for AG, EE, FR, IT, and CG orbits for scaling ratio 4 at RR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 128 × 128 pixel crop is shown to highlight local reconstruction differences. From left to right: the GT, the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

loss of definition — for example, along the Nile River in the EG section, at the coastlines of the Caspian Sea and the Gulf of Mexico in the CS and US patches, respectively, and at the coastlines of the Mediterranean sea in the AG patches. These visual observations confirm the benefit of incorporating sensor-specific PSF information into the reconstruction process. While the difference between the PSF-aware and PSF-unaware variants of S5Net (S5Net_match vs. S5Net_nomatch) is not always immediately perceptible, it becomes evident along certain edges — especially in regions where land meets water — where S5Net_match more accurately preserves edge sharpness. However, per-channel pixel-wise differences are better highlighted for all orbits in Figure 4.10 and Figure 4.11, which provide a more direct proof of the discrepancies introduced by neglecting the sensor's PSF.

To further highlight differences in pixel-level reconstruction quality and to analyse the spatial distribution of the reconstruction error produced by the proposed framework, Figures 4.14 and 4.15 are reported. Each figure includes the GT images, the outputs of a representative benchmark method — namely, SRCNN (Dong et al., 2016a) — the SSIM (see Section 4.1.2) maps computed between the SRCNN results and the corresponding GT images (averaged over the three monochromatic channels used in the false-colour visualization, i.e., SWIR, NIR, and UVIS), the reconstruction obtained with the proposed PSF-aware S5Net (S5Net_match), and the corresponding SSIM maps computed between S5Net_match and the GT images, again averaged over the same three monochromatic channels. All ten orbits listed in Table 4.1 are shown: in particular, the IN, US, EG, RS, and CS orbits are reported in Figure 4.14, whereas the AG, EE, FR, IT, and CG orbits are shown in Figure 4.15. For each orbit, the

same 128 × 128 pixel region selected in Figures 4.12 and 4.13 is displayed in order to highlight differences in pixel-level reconstruction quality.

The SSIM maps are visualised using a colour scale ranging from red (low SSIM, indicating poorer reconstruction quality) to blue (high SSIM, indicating higher structural similarity with the GT image). Consequently, darker colours correspond to lower reconstruction accuracy. When analysed individually, the maps exhibit typical and spatially coherent patterns. In particular, higher SSIM values are generally observed over homogeneous regions, whereas lower values are mainly concentrated along sharp spatial transitions and highly textured areas. These regions include coastlines (e.g., in IN, US, RS, CS, AG, FR, and IT), river boundaries such as along the Nile in EG, and desert areas with strong spatial variability (e.g., in EG and AG). This behaviour is expected, as such structures contain high-frequency spatial information that makes the reconstruction task inherently more challenging.

Although the overall spatial patterns are similar, a closer inspection reveals consistent qualitative differences between SRCNN and the proposed PSF-aware method (S5Net_match), particularly along sharp boundaries. In the first set of orbits (IN, US, EG, RS, and CS) in Figure 4.14, clearer differences can be observed along the coastlines, especially in the IN, US, and CS scenes. In these areas, the SSIM maps associated with S5Net_match appear noticeably lighter, with reduced dark blue regions along the coastal edges, indicating improved reconstruction of sharp transitions. A similar behaviour can be observed in the second set of orbits (AG, EE, FR, IT, and CG) in Figure 4.15. In particular, in AG, FR, and IT the sea areas close to the coastline appear significantly lighter in the SSIM maps of S5Net_match compared to SRCNN, suggesting

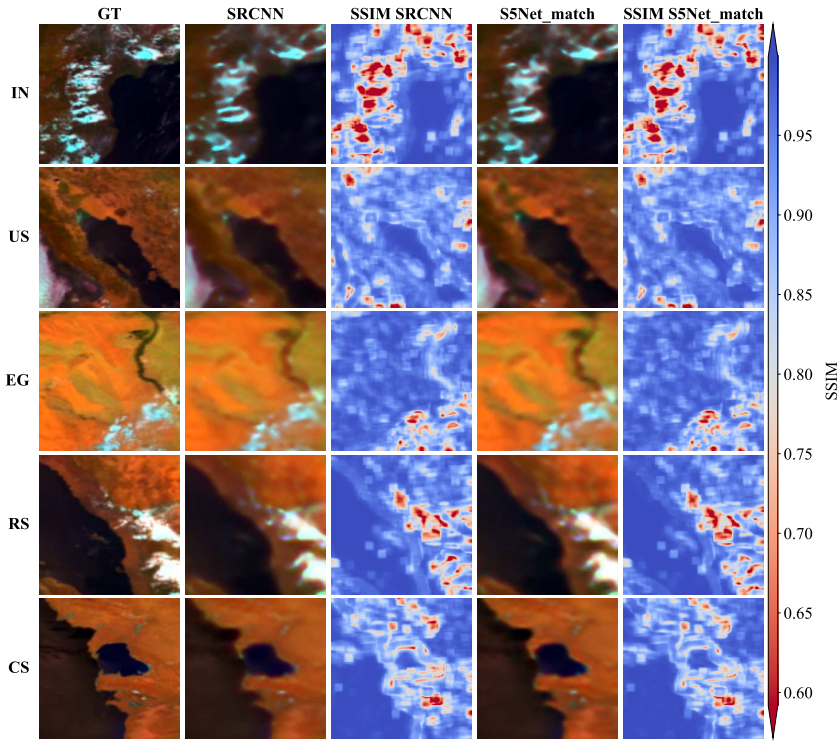

Figure 4.14: GT images, reconstructions obtained with SRCNN (Dong et al., 2016a) and with the proposed PSF-aware method (S5Net_match), and the corresponding SSIM maps for the IN, US, EG, RS, and CS orbits at RR with scaling ratio 4. The SSIM maps are computed between each reconstruction and the corresponding GT image, averaging over the SWIR, NIR, and UVIS monochromatic bands. For each orbit, a 128 $\times$ 128 pixel crop is shown to highlight the spatial distribution of the reconstruction quality.

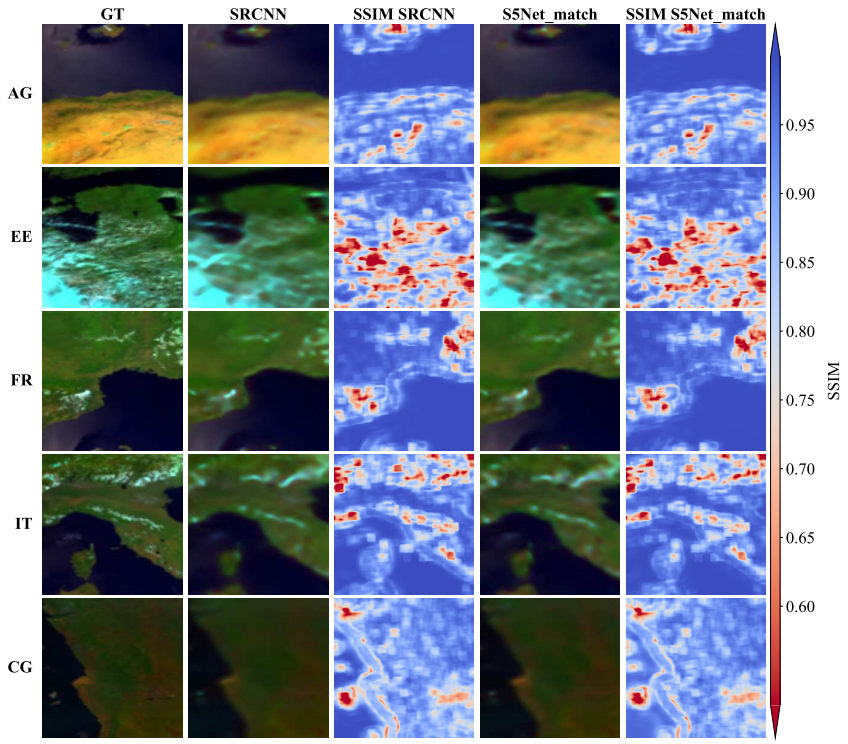

Figure 4.15: GT images, reconstructions obtained with SRCNN (Dong et al., 2016a) and with the proposed PSF-aware method (S5Net_match), and the corresponding SSIM maps for the AG, EE, FR, IT, and CG orbits at RR with scaling ratio 4. The SSIM maps are computed between each reconstruction and the corresponding GT image, averaging over the SWIR, NIR, and UVIS monochromatic bands. For each orbit, a 128×128 pixel crop is shown to highlight the spatial distribution of the reconstruction quality.

that the proposed method is better able to preserve fine spatial structures and accurately reconstruct the land-sea boundaries. This effect is particularly evident in some scenes, such as RS and CS in Figure 4.14, and IT in Figure 4.15, where the dark red areas, i.e., those characterised by lower similarity, appear less extended for S5Net_match, resulting in visibly fewer regions with low SSIM values in the corresponding maps.

Overall, all the quantitative and qualitative observations indicate that the proposed PSF-aware model (S5Net_match) achieves a more accurate reconstruction of fine spatial structures compared to the benchmark methods.

Full resolution assessment: It was seen in the previous paragraph that a good modelling of the PSF can significantly improve all SR techniques that depend on it. When there is no GT image available, this is even more vital.

To ensure that the numerical results were not biased by the matching acquisition procedure required in the RR protocol for the S5Net approach, we also validated the integration of PSF knowledge into the SISR method against its PSF-unaware counterpart, as well as the same baseline (bicubic interpolation) and state-of-the-art methods — deconvolution solved with the CGA, once optimised, SRCNN (Dong et al., 2016a), VDSR (Kim et al., 2016a), EDSR (Lim et al., 2017), and PAN (Zhao et al., 2020) — under the FR protocol.

We assess the consistency between the super-resolved images — once downsampled using the corresponding scaling factor — and the available LR images in terms of PSNR (Equation 4.1) and UIQI (Equation 4.3) metrics, which both provide reference-based measures of either pixel-based fidelity or structural similarity, respectively.

Tables 4.11–4.13 provide the metrics of all methods averaged

over the four monochromatic images and overall the ten orbits (Table 4.1) for scaling ratios of 2, 3, and 4, respectively. The time in seconds taken by each algorithm is shown too.

Method	$Q \uparrow$	$PSNR \uparrow$	**Time**
Cubic	0.9881	46.0345	0.0235
CGA	0.9932	46.6522	2.1539
SRCNN	0.9943	<u>49.4262</u>	0.0398
VDSR	0.9912	45.7261	0.0274
EDSR	0.9837	42.7888	0.0269
PAN	0.9832	42.7397	0.0225
S5Net_nomatch	<u>0.9944</u>	48.2065	0.0052
S5Net_match	**0.9956**	**50.3144**	0.0043

Table 4.11: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at FR with a scaling factor of 2. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

As expected, and similarly to what was observed for the RR protocol, increasing the scaling ratio leads to a degradation in the metrics, due to the increased difficulty in reconstructing the image.

Our model consistently outperforms the benchmark methods in terms of pixel-wise consistency with the LR image, as demonstrated by systematically higher PSNR and UIQI values. Moreover, the PSF-aware S5Net provides a clear advantage over its PSF-unaware counterpart across all scaling ratios, with an average gain of +3.32 dB in terms of PSNR. This confirms that incorporating the PSF into the training process enhances the alignment between the reconstructed and original observations, even when the reference image is not available, i.e., in real-world scenarios. The pixel-wise consistency with

Method	$Q \uparrow$	$PSNR \uparrow$	**Time**
Cubic	0.9880	45.9823	0.0235
CGA	0.9911	45.6840	2.1539
SRCNN	0.9945	49.7398	0.0398
VDSR	<u>0.9948</u>	<u>50.0822</u>	0.0274
EDSR	<u>0.9948</u>	49.3014	0.0269
PAN	0.9942	49.4399	0.0225
S5Net_nomatch	<u>0.9948</u>	48.0161	0.0052
S5Net_match	**0.9967**	**51.6872**	0.0043

Table 4.12: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at FR with a scaling factor of 3. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

the LR inputs is particularly crucial for applications where accurate pixel-level reconstruction is fundamental.

Some super-resolved images are displayed in Figures 4.16–4.20. As seen for the RR protocol, all figures use false-colour representations, with the SWIR, NIR, and UVIS monochromatic images mapped to the red, green, and blue channels, respectively. Each figure includes the outputs of two representative benchmark methods — namely, bicubic interpolation and SRCNN (Dong et al., 2016a) — as well as the results produced by the PSF-unaware (S5Net_nomatch) and PSF-aware (S5Net_match) versions of the proposed S5Net for a scaling ratio of 4. Two of the ten orbits are displayed in each figure. For every orbit, a 512 × 512 pixel region of the original image (Figure 4.2) is shown in the first row, while a 128 × 128 pixel region from the same orbit is shown in the second row. The reduction in area corresponds to the chosen scaling ratio and is introduced to enable a more detailed

Method	$Q \uparrow$	$PSNR \uparrow$	**Time**
Cubic	0.9880	45.9750	0.0235
CGA	0.9908	45.5820	2.1539
SRCNN	0.9934	49.1493	0.0398
VDSR	<u>0.9940</u>	<u>49.8600</u>	0.0274
EDSR	0.9913	45.1644	0.0269
PAN	0.9910	45.8883	0.0225
S5Net_nomatch	0.9933	46.7525	0.0052
S5Net_match	**0.9960**	**50.9220**	0.0043

Table 4.13: Quantitative evaluation of the effect of PSF knowledge for our proposed SR methodology at FR with a scaling factor of 4. Quality metrics are reported for the average of the four monochromatic images averaged over all test orbits. The overall best result is highlighted in bold, while the second-best result is underlined for all image quality metrics.

assessment of the reconstruction quality. The 512 × 512 regions match those presented in Figures 4.12 and 4.13.

The qualitative comparison for the FR protocol in Figures 4.16–4.20 highlights the importance of visual analysis in this setting, since the absence of a GT prevents the metrics alone from fully capturing the differences between methods. In particular, our proposed SISR approach, S5Net, produces reconstructions that are visibly sharper and more faithful than those obtained with bicubic interpolation or SRCNN (Dong et al., 2016a). This advantage is particularly evident in the 128 × 128 pixel crops, which are specifically selected to highlight such differences. Clear improvements can be observed along fine geographical structures, such as the Italian, French, and Algerian coasts (IT, FR, AG), the Nile river in Egypt (EG), the coastline of India (IN), the shores of the Gulf of Mexico (US), and the contours of the Caspian Sea (CS) and the Red Sea (RS). These findings

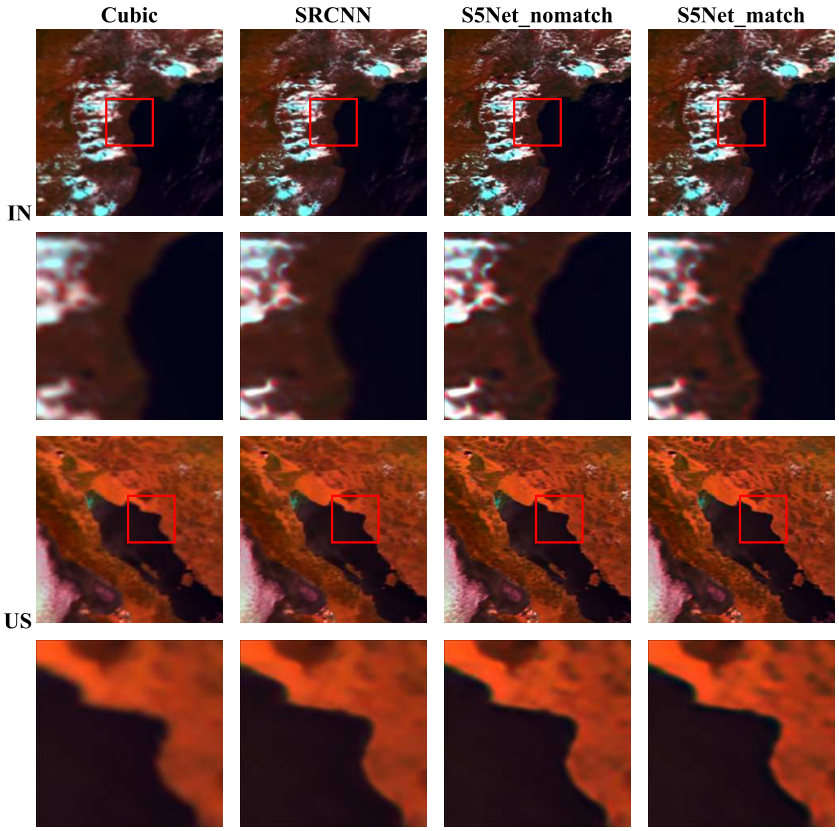

Figure 4.16: False-colour visualisation of super-resolved images for IN and US orbits for scaling ratio 4 at FR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 512 × 512 pixel crop and a 128 × 128 pixel crop are shown in the first and second rows, respectively. From left to right: the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

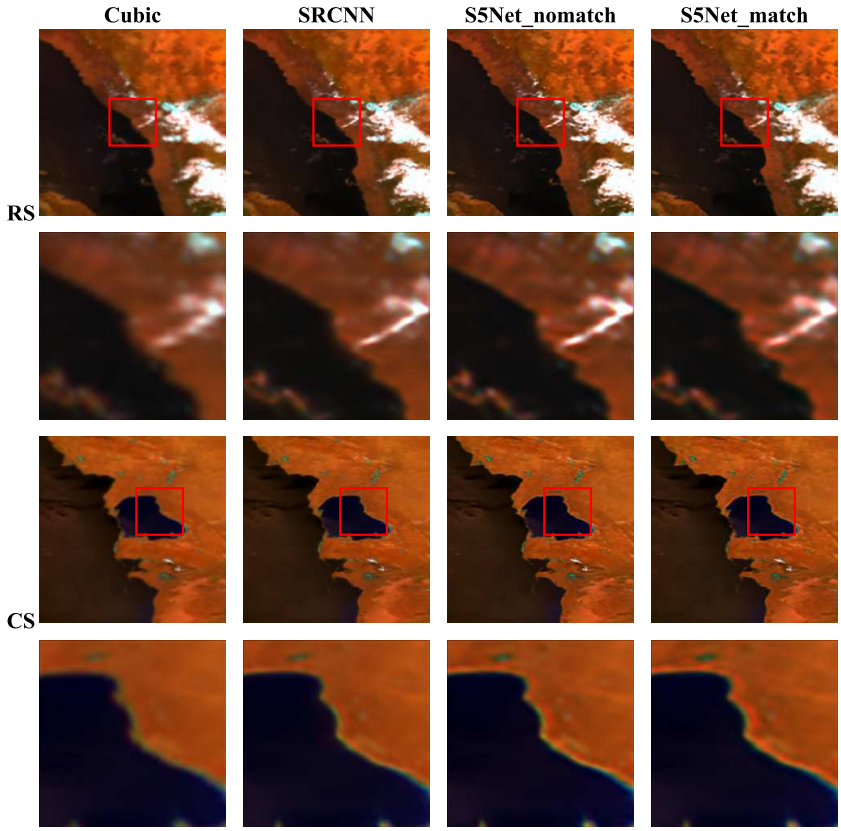

Figure 4.17: False-colour visualisation of super-resolved images for RS and CS orbits for scaling ratio 4 at FR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 512 × 512 pixel crop and a 128 × 128 pixel crop are shown in the first and second rows, respectively. From left to right: the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

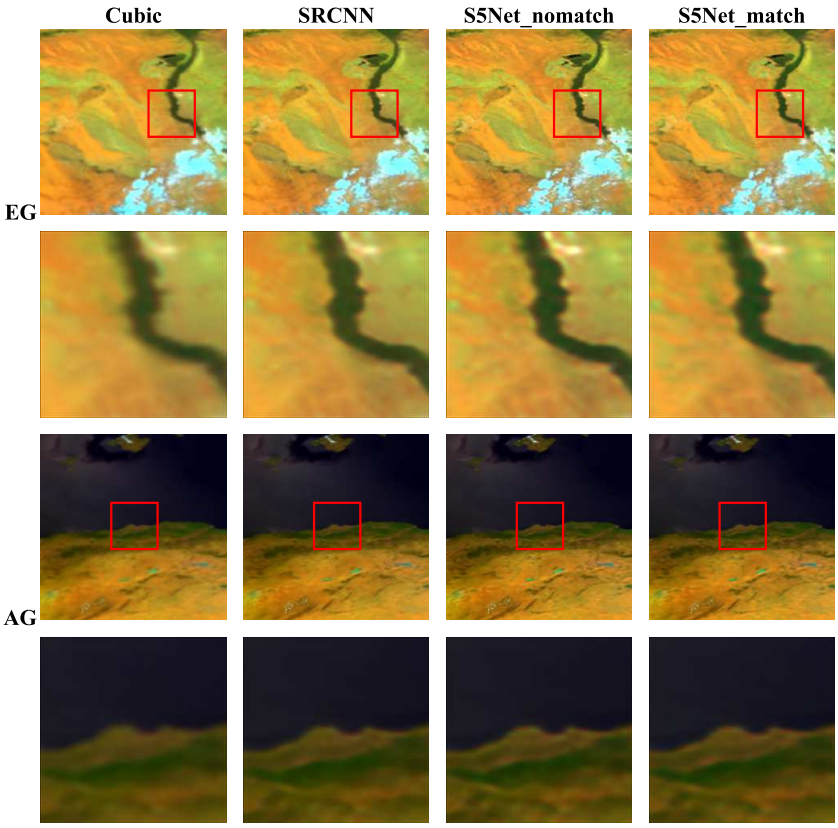

Figure 4.18: False-colour visualisation of super-resolved images for EG and AG orbits for scaling ratio 4 at FR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 512 × 512 pixel crop and a 128 × 128 pixel crop are shown in the first and second rows, respectively. From left to right: the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

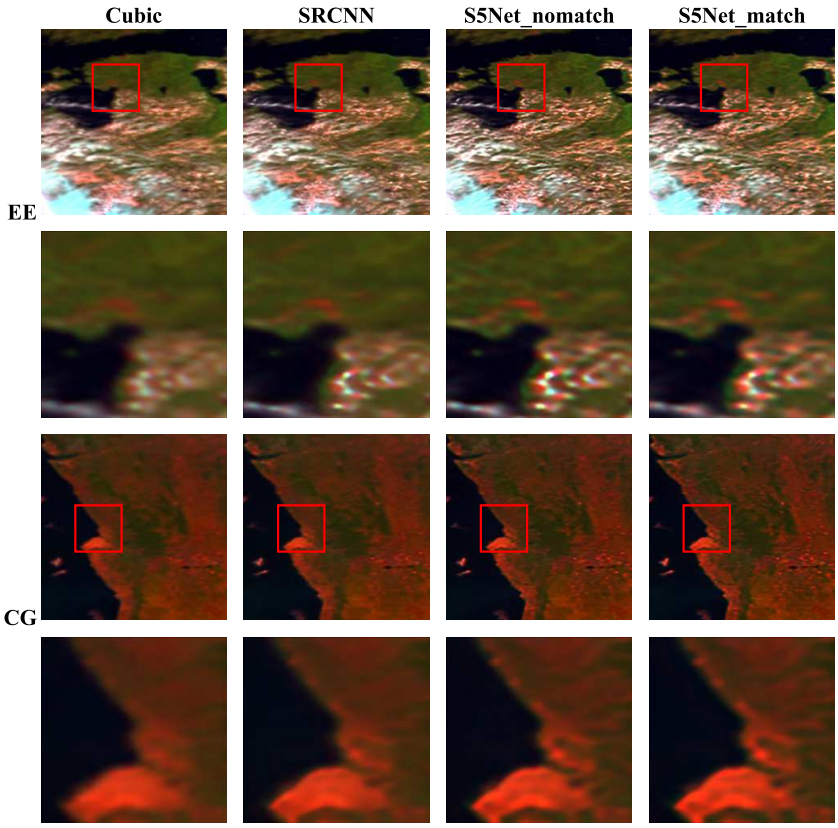

Figure 4.19: False-colour visualisation of super-resolved images for EE and CG orbits for scaling ratio 4 at FR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 512 × 512 pixel crop and a 128 × 128 pixel crop are shown in the first and second rows, respectively. From left to right: the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

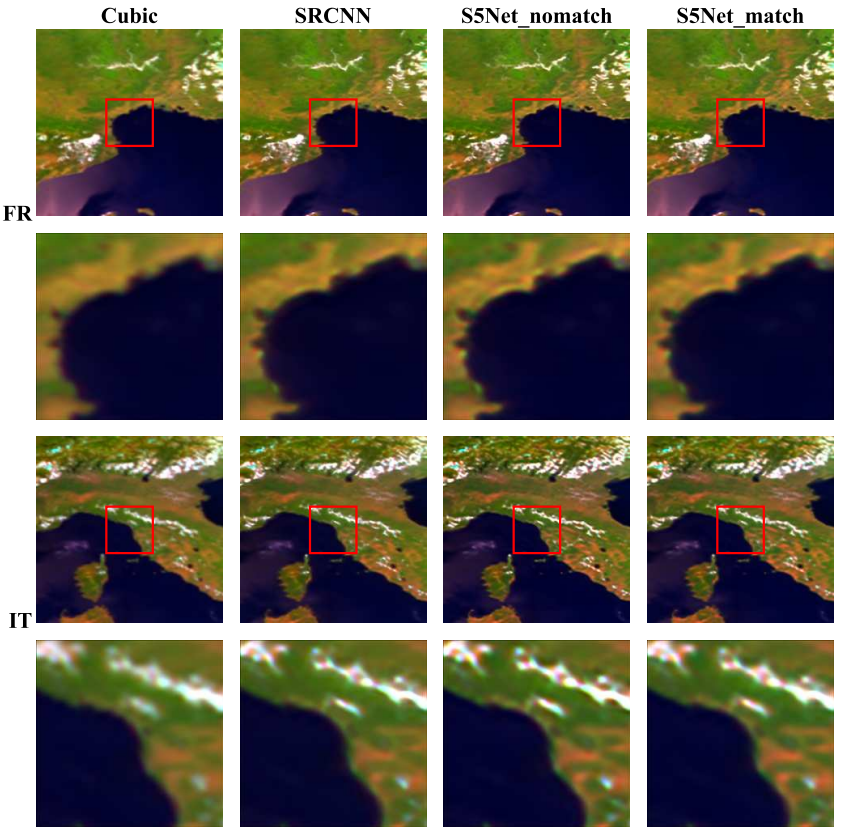

Figure 4.20: False-colour visualisation of super-resolved images for FR and IT orbits for scaling ratio 4 at FR. The red, green, and blue channels correspond to the SWIR, NIR, and UVIS monochromatic bands, respectively. For each orbit, a 512 × 512 pixel crop and a 128 × 128 pixel crop are shown in the first and second rows, respectively. From left to right: the output of bicubic interpolation, SRCNN (Dong et al., 2016a), PSF-unaware S5Net (S5Net_nomatch), and PSF-aware S5Net (S5Net_match).

confirm that visual inspection is crucial in the FR protocol, and demonstrate that S5Net — especially in its PSF-aware version (S5Net_match) — better preserves high-frequency details and structures while maintaining perceptual quality.

These observations highlight the importance of the FR protocol: unlike the RR protocol, it directly reflects the actual target resolution, making it a more realistic and application-relevant setting for evaluating SR performance.

4.2.3 Evaluation of the super-resolution method on the whole S5P imageries

This section presents the evaluation of the proposed PSF-aware SISR approach on the complete S5P images, composed by all channels acquired by the sensor. It includes a comparative analysis of the proposed dynamic multi-directional cascade fine-tuning algorithm against benchmark methods, once optimised.

4.2.3.1 Optimization of benchmark fine-tuning algorithms

To optimise all benchmark fine-tuning strategies employed for S5Net (Section 4.1.4.3), we selected the entire preprocessed orbit (Section 4.1.1). To limit computational complexity, only a single representative orbit was chosen (Table 4.1), namely IN.

As regards S5Net_PCA, the number k of principal components to be super-resolved — both when considering the entire image (GSR) and the image split by DEM (DSR) — is optimised based on the ratio of the explained variance. Figure 4.21 illustrates the trend of the explained variance ratio with respect to the number of principal components in the GSR case,

whereas Figure 4.22 shows the corresponding trends in the DSR scenario.

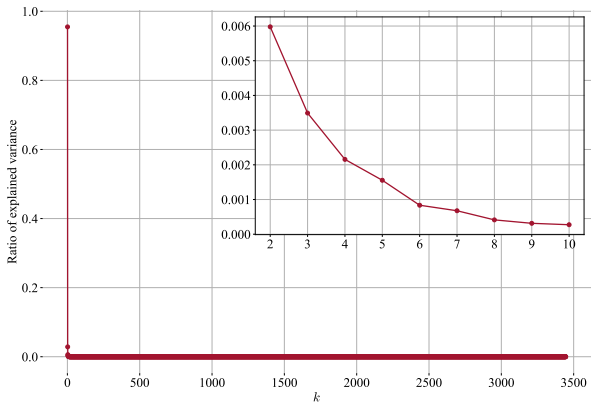

Figure 4.21: Explained variance ratio as a function of the number of principal components k for the entire IN image (GSR case). An inset at the top right shows a zoom on components 2 to 10.

As shown in both Figure 4.21 and Figure 4.22, the ratio of the explained variance displays a clear saturation trend, with approximately 95% of the total variance captured by the first principal component in both cases. This justifies the use of the inset plots, which provide a focused view on the first nine components after the first one.

In particular, the marginal gain beyond the ninth component is minimal across all cases. This confirms that selecting $k = 9$ principal components to be super-resolved by S5Net in the S5Net_PCA framework — both under the GSR and DSR scenarios — is sufficient to retain the informative content while avoiding unnecessary computational cost.

As regards the S5Net_static state-of-the-art approach, the number of iterations i_c was optimised over the interval $\{1, 3, 7, 15, 30, 75, 150, 300, 750, 1500, 3000\}$ by evaluating the RMSE value, which is the square root of the MSE value (Equation 3.5). To avoid excessive computational demands,

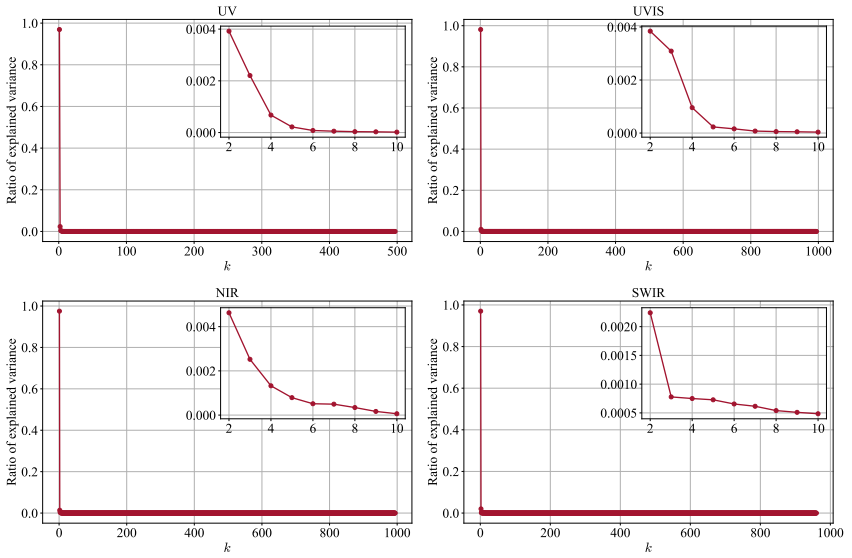

Figure 4.22: Explained variance ratio as a function of the number of principal components k for the IN image, separately for each DEM (DSR case). An inset at the top right shows a zoom on components 2 to 10.

only 100 consecutive spectral channels are sampled from the entire preprocessed IN orbit within the UVIS DEM scope. Consecutive channels are selected to simulate the accumulation of reconstruction error across adjacent wavelengths that occurs both in the GSR and DSR scenarios. The analysis was conducted following the RR protocol with a scaling ratio of 4.

The trend of the RMSE as a function of the number of fine-tuning iterations i_c is shown in Figure 4.23.

The best performance is achieved at $i_c = 75$, where the RMSE reaches its minimum. Beyond $i_c = 300$, the error increases consistently, indicating that excessive training iterations are detrimental. This behaviour is likely due to error accumulation during sequential fine-tuning, as each channel is initialised from the weights of the preceding one. When the number of epochs becomes too high, transferring overfitted weights leads

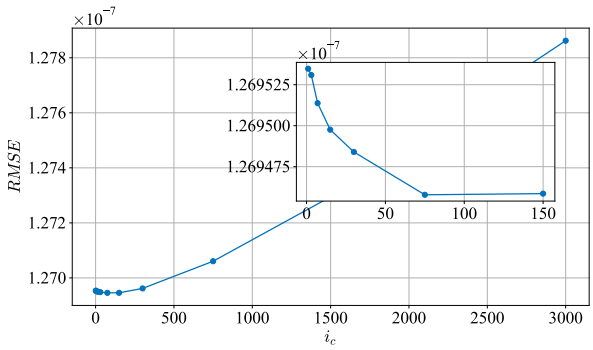

Figure 4.23: Evolution of the RMSE value as a function of the number of training iterations i_c for the S5Net_static approach. A zoomed-in view for $i_c \leq 150$ highlights the early-stage behaviour of the plot. The plot refers to 100 consecutive spectral channels of the IN image under the RR protocol with a scaling ratio of 4.

to performance degradation — such that a single epoch may yield better results than more than 300. For this reason, the number of iterations in S5Net_static was fixed to $i_c = 75$ per channel for both the DSR and GSR cases.

4.2.3.2 Assessment of the S5Net fine-tuning strategy

To ensure a fair and computationally efficient comparison of the proposed dynamic multi-directional cascade fine-tuning strategy for S5Net (S5Net_dynamic) against benchmark approaches, all methods were evaluated on a entire orbit. To avoid excessive computational load, a single representative preprocessed orbit — namely the IN image (Table 4.1) — was used. The corresponding results are shown in Figure 4.24, which reports the PSNR values (Equation 4.1) achieved by the proposed S5Net_dynamic, the baseline approach (S5Net_all), and the two state-of-the-art methods (S5Net_PCA and S5Net_static), once they have all been optimised, for both the GSR and DSR scenarios. All results are obtained using the

RR protocol with a scaling ratio of 4.

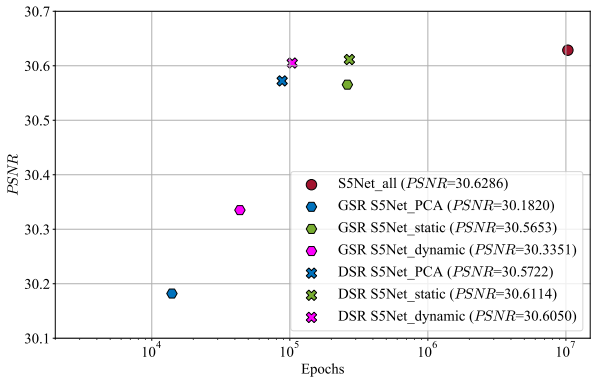

Figure 4.24: The plot shows the PSNR values as a function of the number of training iterations (logarithmic scale) for the proposed fine-tuning approach (S5Net_dynamic), the baseline method (S5Net_all), and the two state-of-the-art strategies (S5Net_PCA and S5Net_static). For clarity, the corresponding numerical PSNR values are also reported.

Among the evaluated approaches, S5Net_all (a red circle) achieves the highest PSNR score. However, this performance comes at the cost of an exceedingly high number of training iterations (on the order of 10^7), which makes it impractical for large-scale SR applications. The DSR setting (indicated by the cross marker), in which the image is split according to the DEM, consistently outperforms the GSR setting (indicated by the hexagon marker), where the image is processed as a whole. This improvement can be attributed to the stronger spectral correlation observed between channels belonging to the same DEM, as opposed to the weaker inter-DEM correlation.

S5Net_PCA (in blue) is the least effective approach in both cases. Although it benefits from a lower computational cost, even when optimised, it fails to reach competitive performance levels. As expected, S5Net_static (in green) achieves better PSNR than the proposed S5Net_dynamic (in magenta), thanks to the higher number of training epochs. Nevertheless, the

proposed S5Net_dynamic offers a favourable trade-off in the DSR case, which is the most relevant one due to its superior performance. Specifically, it requires only 104,092 iterations compared to the 270,075 needed by S5Net_static, while losing merely 0.006 dB in terms of PSNR.

Therefore, in the subsequent full-framework evaluation, we consider only the most promising strategies: S5Net_all as a baseline, S5Net_static as benchmark and S5Net_dynamic in both GSR and DSR configurations. S5Net_PCA is excluded due to its poor performance.

4.2.3.3 Final results on super-resolution of S5P radiance imageries

The full evaluation of the proposed PSF-aware SISR framework was carried out using the entire preprocessed dataset of S5P Level-1b radiance images (Section 4.1.1), which includes ten selected orbits (see Table 4.1 for details on the considered orbits) across different predefined areas and 3445 spectral channels.

As discussed in the previous subsection, under the FR protocol, consistency with the LR image is correlated with the results observed in the RR setting. Similarly, performance remains stable across different scaling ratios, indicating that evaluation at a single, representative ratio is sufficient. To reduce computational load, we therefore assess the complete framework under the RR protocol using the scaling ratio most commonly adopted in remote sensing applications, namely 4.

Table 4.14 reports all full-reference metrics introduced in Section 4.1.2, namely UIQI (Equation 4.3), ERGAS (Equation 4.2), SAM (Equation 4.4), and PSNR (Equation 4.1). The table includes results for the baseline

	Method	$Q \uparrow$	$ERGAS \downarrow$	$SAM \downarrow$	$PSNR \uparrow$
	Cubic	0.8715	3.3480	**1.7428**	32.2591
	CGA	0.8901	3.5982	2.6987	32.1371
	SRCNN	0.8897	*3.1093*	1.9326	33.0021
	VDSR	*0.8978*	3.1215	1.6252	*33.0432*
	EDSR	0.8822	3.5016	1.8280	32.1042
	PAN	0.8807	3.4041	<u>1.7795</u>	32.2358
	S5Net_all	<u>0.9033</u>	**2.8766**	2.0807	**33.5674**
GSR	S5Net_st	*0.9021*	*2.9483*	2.0659	*33.4235*
GSR	S5Net_dy	0.8917	3.0131	*2.0635*	33.2006
DSR	S5Net_st	**0.9044**	<u>2.9206</u>	2.0859	<u>33.5618</u>
DSR	S5Net_dy	0.9030	2.9280	*2.0742*	33.5435

Table 4.14: Quantitative evaluation of the proposed full PSF-aware SISR method, S5Net, conducted using different fine-tuning algorithms: one baseline (S5Net_all), one state-of-the-art method (S5Net_st, which stands for S5Net_static), and our proposed dynamic multi-directional cascade fine-tuning algorithm (S5Net_dy, which stands for S5Net_dynamic), evaluated under both GSR and DSR scenarios. The results are obtained under the RR protocol with a scaling factor of 4. All full-reference metrics are reported as averages across the orbits listed in Table 4.1. The overall best result for each image quality metric is shown in bold, while the second-best is underlined. When relevant, the best result within each sub-block is italicised.

approach — bicubic interpolation (Section 2.3.1.1) — and the performance of five state-of-the-art approaches, the deconvolution-based method solved via CGA (Section 2.3.1.2) — evaluated after optimization — and four DL-based SISR methods, including SRCNN (Dong et al., 2016a), VDSR (Kim et al., 2016a), EDSR (Lim et al., 2017), and PAN (Zhao et al., 2020).

Our proposed PSF-aware SISR approach, here simply indicated as S5Net, is reported in all promising fine-tuning configurations identified in the preceding subsection, namely S5Net_all, S5Net_static, and S5Net_dynamic, for both the GSR and DSR scenarios.

First, when S5Net is trained separately on all spectral channels (S5Net_all), it achieves the best performance in terms of PSNR and ERGAS. These two metrics, both dependent on the RMSE metric, indicate that full per-channel fine-tuning leads to higher-fidelity reconstructions. However, this improvement comes with a clear trade-off: S5Net_all performs worse in terms of UIQI and SAM, which reflect perceptual image quality and spectral consistency, respectively. This suggests that while the spatial details are better preserved, the structural similarity and spectral fidelity are negatively impacted when treating channels independently.

In contrast, S5Net_static achieves the highest UIQI score under the DSR setting, highlighting its effectiveness in preserving overall image quality from a structural perspective. The fact that this result occurs in the DSR scenario reinforces the broader trend that emerges from the table: DSR consistently outperforms GSR across nearly all metrics and methods (+0.14 dB for S5Net_static and +0.34 dB for S5Net_dynamic on PSNR metric). This observation suggests that processing images split by DEM — rather than entire orbits — is beneficial, likely due

to the reduced intra-region variability.

Interestingly, S5Net_static also achieves PSNR values very close to those of S5Net_all, particularly in the DSR setting, with minimal performance degradation (-0.005 dB on PSNR). Likewise, S5Net_dynamic shows similar behaviour: in the DSR configuration, its results are nearly indistinguishable from the full-channel baseline across all metrics (-0.024 dB on PSNR). This is a promising result, as it indicates that our proposed dynamic fine-tuning strategy can approach the performance of fully separated training, but at a significantly lower computational cost — an aspect that will be further discussed in the following.

Bicubic interpolation yields the lowest SAM values across all methods, indicating the highest spectral consistency. However, this result is not surprising, as bicubic interpolation inherently avoids introducing spectral distortions. At the same time, its performance in terms of PSNR is low, highlighting the fundamental trade-off between spectral preservation and pixel-based reconstruction.

Among the fine-tuning methods, S5Net_dynamic achieves better spectral consistency than S5Net_static, particularly in the GSR setting (2.0635 for S5Net_dynamic against 2.0659 for S5Net_static, with 2.0807 for S5Net_all). This result suggests that our dynamic fine-tuning mechanism is more effective at adapting to spectral variations across channels. Additionally, the GSR version of S5Net_dynamic shows a better SAM value than its DSR counterpart (2.0635 for GSR and 2.0742 for DSR), confirming that, although DSR is generally superior in terms of reconstruction accuracy and perceptual quality, GSR may offer slight advantages in maintaining spectral coherence.

With regard to computational complexity, and in line with the observations made for the IN orbit in the detailed

	Method	Iterations
	S5Net_all	10,335,000
GSR	S5Net_static	261,300
GSR	S5Net_dynamic	83,487
DSR	S5Net_static	270,075
DSR	S5Net_dynamic	96,100

Table 4.15: The table reports the average number of iterations required by each of the fine-tuning approaches employed in S5Net to perform the full evaluation presented in Table 4.14.

computational load analysis presented in the preceding subsection, Table 4.15 reports the average number of iterations required by each of the fine-tuning approaches introduced in Table 4.14. Specifically, we compare S5Net_all, S5Net_static, and S5Net_dynamic under both the GSR and DSR scenarios. The number of iterations for both S5Net_all and S5Net_static is fixed across all orbits, whereas for S5Net_dynamic it is averaged over the orbits.

From a computational perspective, the results in Table 4.15 confirm the expected trade-offs between performance and efficiency. As anticipated from the previous analysis, the fully channel-wise fine-tuning approach, S5Net_all, is by far the most computationally expensive, requiring a total of 10,335,000 iterations — $i_s = 3000$ for each of the 3445 spectral channels. Although this approach yields the best results for RMSE-based metrics such as PSNR and ERGAS, the cost is prohibitive, particularly when scalability to large-scale or real-time applications is considered.

Comparatively, the DSR scenario exhibits a higher computational load than its GSR counterpart for both S5Net_static and S5Net_dynamic. This increase is mainly due to the presence of four channels fine-tuned with $i_s = 3000$

iterations. Nonetheless, the added computational burden in DSR is moderate, and the improved qualitative performance observed in Table 4.14 justifies the cost.

Most notably, the S5Net_dynamic approach proves to be the most computationally efficient overall. Its adaptive fine-tuning strategy significantly reduces the number of required iterations while maintaining performance levels close to those of S5Net_all. This confirms the practical value of the proposed dynamic multi-directional cascade fine-tuning strategy, as it offers a favourable compromise between reconstruction quality and computational cost.

All super-resolved images for five different orbits — CS, IT, IN, RS and US (see Table 4.1) — are shown in Figures 4.25–4.29, respectively, for the RR protocol at scaling ratio 4 and for an area of 128 × 128 pixels of the original image (Figure 4.2). In each figure, the images are displayed for the GT, all benchmark approaches listed in Table 4.14, as well as for two fine-tuning variants of S5Net, namely S5Net_all and S5Net_dynamic in the DSR scenario. All images are displayed using channels 3050, 2250, and 1700 for red, green, and blue, respectively.

The visual comparison across the different SR methods reveals clear differences in reconstruction quality. The bicubic interpolation results appear overly smooth, lacking fine detail. Conversely, the deconvolution-based approach produces images with a somewhat blocky and pixelated texture. DL-based state-of-the-art methods, such as VDSR, EDSR, and PAN, generate images with a soft, blended appearance, but tend to introduce some blur.

Among all the methods evaluated, S5Net_all consistently produces the most visually convincing and detailed reconstructions, particularly noticeable along complex edge structures such as coastlines — for example, the

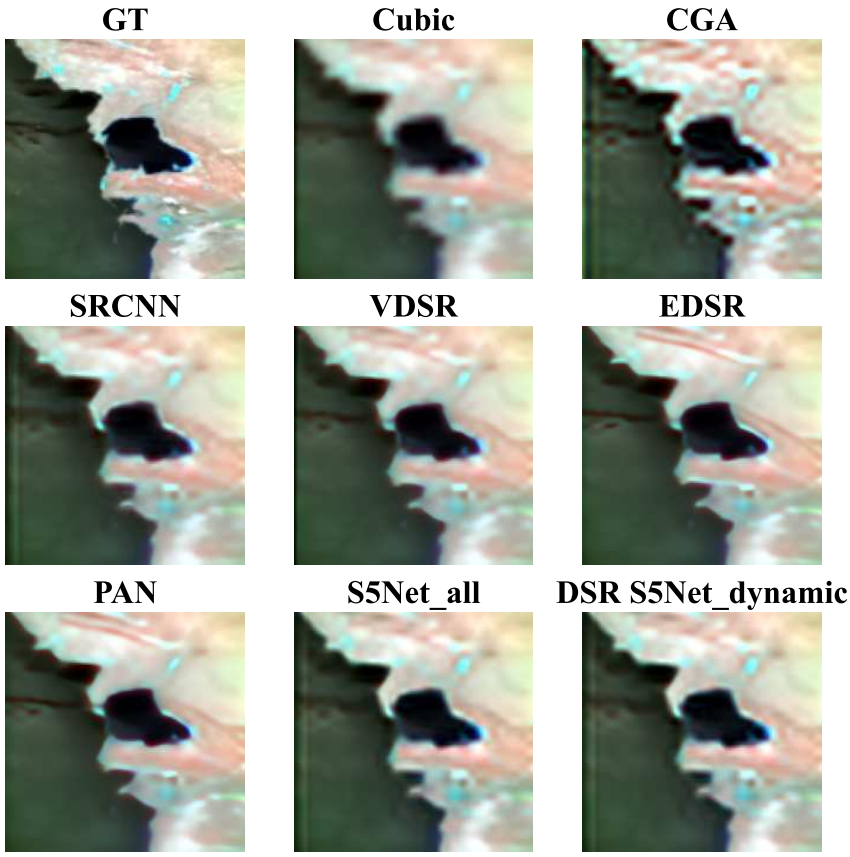

Figure 4.25: Super-resolved outputs for the full CS orbit for the RR protocol at scaling ratio 4. The images compare the GT, all benchmark methods and two S5Net fine-tuning variants, S5Net_all and S5Net_dynamic in the DSR configuration. All images are rendered using channels 3050 (red), 2250 (green), and 1700 (blue).

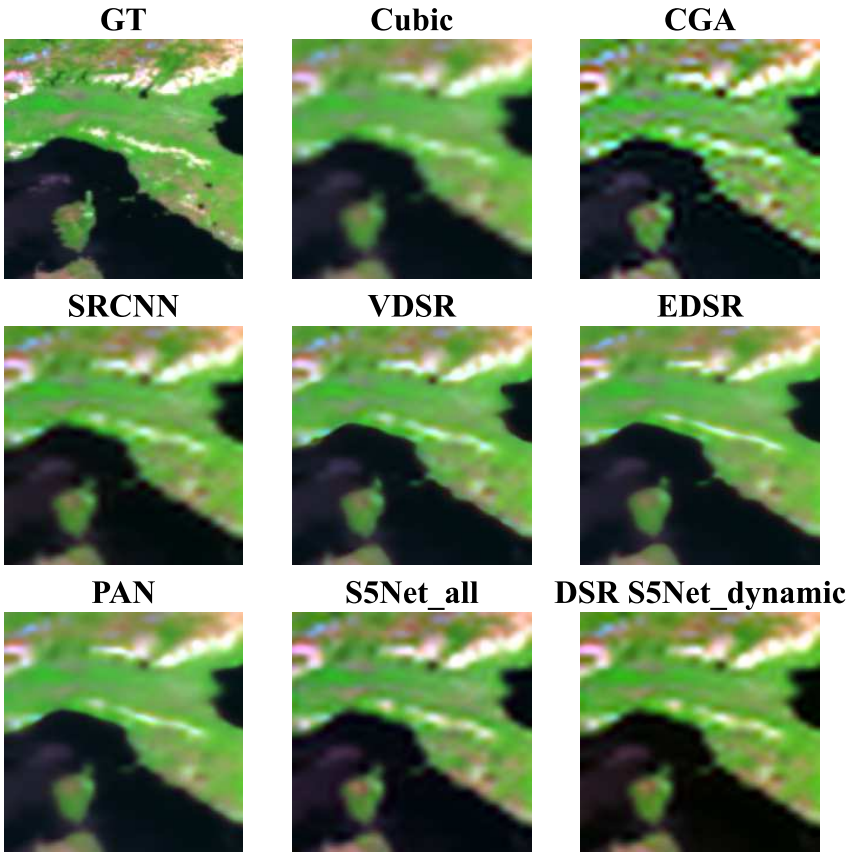

Figure 4.26: Super-resolved outputs for the full IT orbit for the RR protocol at scaling ratio 4. The images compare the GT, all benchmark methods and two S5Net fine-tuning variants, S5Net_all and S5Net_dynamic in the DSR configuration. All images are rendered using channels 3050 (red), 2250 (green), and 1700 (blue).

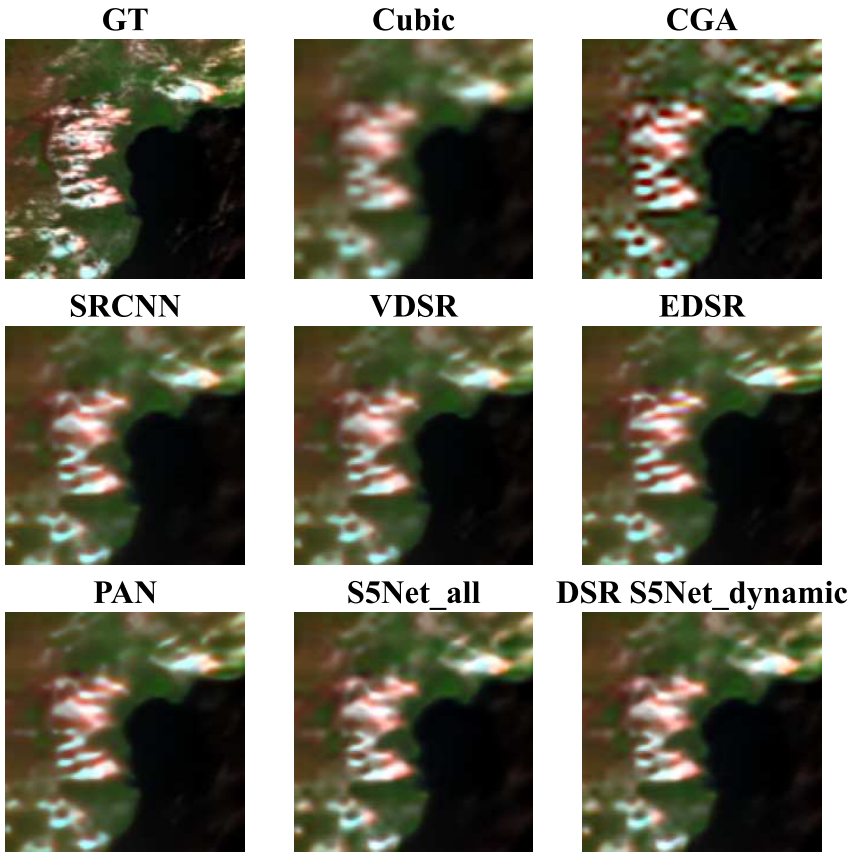

Figure 4.27: Super-resolved outputs for the full IN orbit for the RR protocol at scaling ratio 4. The images compare the GT, all benchmark methods and two S5Net fine-tuning variants, S5Net_all and S5Net_dynamic in the DSR configuration. All images are rendered using channels 3050 (red), 2250 (green), and 1700 (blue).

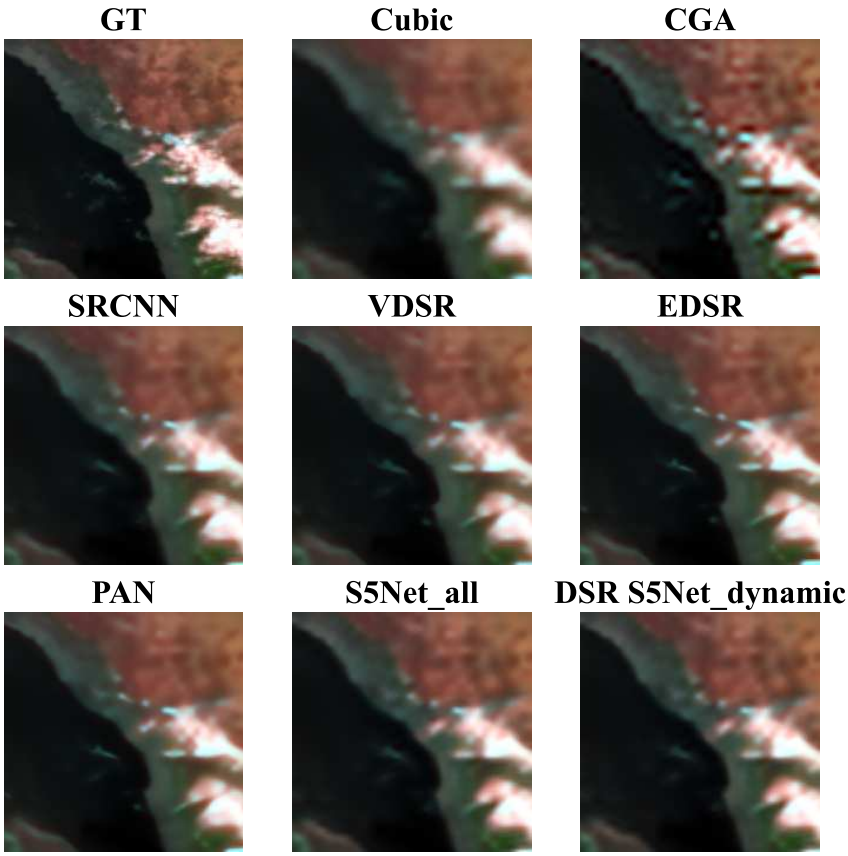

Figure 4.28: Super-resolved outputs for the full RS orbit for the RR protocol at scaling ratio 4. The images compare the GT, all benchmark methods and two S5Net fine-tuning variants, S5Net_all and S5Net_dynamic in the DSR configuration. All images are rendered using channels 3050 (red), 2250 (green), and 1700 (blue).

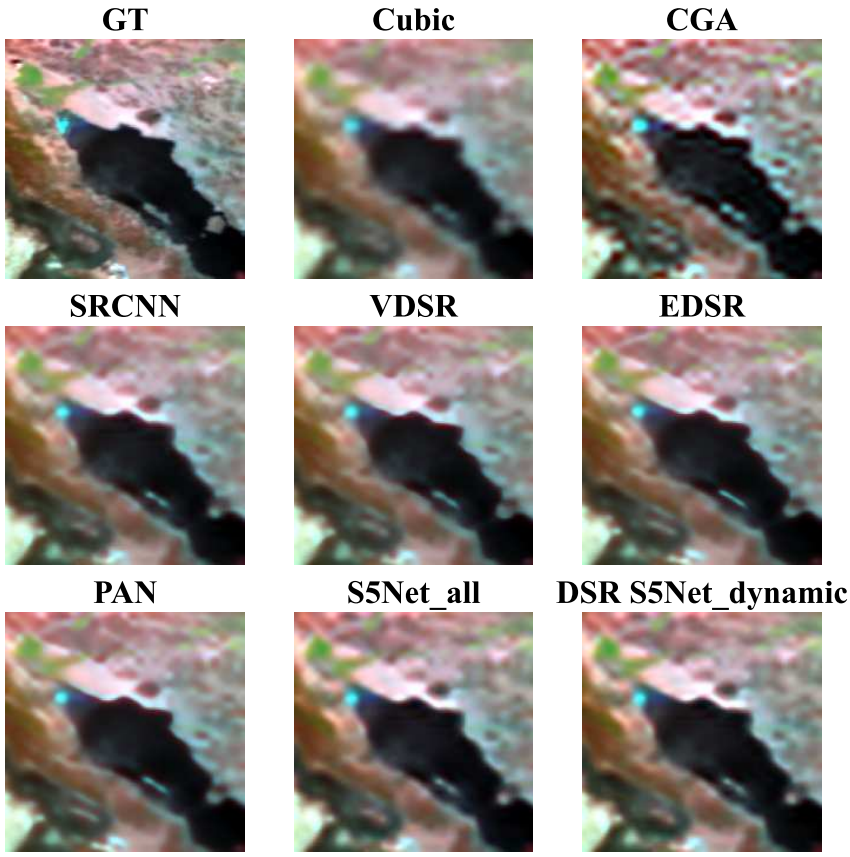

Figure 4.29: Super-resolved outputs for the full US orbit for the RR protocol at scaling ratio 4. The images compare the GT, all benchmark methods and two S5Net fine-tuning variants, S5Net_all and S5Net_dynamic in the DSR configuration. All images are rendered using channels 3050 (red), 2250 (green), and 1700 (blue).

Caspian Sea and the Red Sea in the CS and RS orbits, respectively, the Italian coast in IT, and the Gulf of Mexico in US. The precise delineation of these edges highlights the method's superior ability to reconstruct spatial details that are challenging for other approaches. It is also important to note that S5Net_dynamic in the DSR scenario, the computationally more efficient variant, yields results that are visually indistinguishable from those of S5Net_all, indicating that the reduced computational cost does not compromise image quality.

It should be emphasised that these images represent only three selected channels out of the full 3445 spectral bands available, and the primary focus of this research is on quantitative evaluation rather than visual inspection. Nevertheless, the high visual quality observed in the super-resolved images, despite the limited spectral subset, demonstrates the effectiveness of the proposed method.

4.3 Satellite-based prediction of oxidative potential

To illustrate the potential of the proposed SISR approach in a real-world setting, we consider an exemplary case study in air quality monitoring. Specifically, we investigate whether the PSF-aware method we propose can enhance the prediction of oxidative potential (OP) of particulate matter (PM), a health-relevant indicator introduced in Section 2.1.1.

4.3.1 Case study details

This section provides a description of the problem and the materials used in this analysis.

4.3.1.1 Background and motivation

OP is regarded as a promising health indicator (Gao et al., 2020). However, most epidemiological studies rely on data from a single monitoring site, leading to potential exposure misclassification and underestimation — or even omission — of associated health effects (Zeger et al., 2000). Since OP is more strongly influenced by PM chemical composition than by its concentration (Weber et al., 2021), land surface features visible in satellite imagery serve as indirect proxies, being closely tied to emission sources and factors shaping PM composition (Shangguan et al., 2022). In Carbone et al. (2026), this idea is operationalised in OPNet, a DL-based model designed to predict the OP of PM_{10} over Grenoble, France, using satellite-based surface features and auxiliary variables. Here, we adopt OPNet as a testbed to evaluate the contribution of our SISR framework.

4.3.1.2 Study area

Grenoble, whose location and satellite view is shown in Figure 4.30, is the most populous metropolitan area in the Alps. The French city lies on a valley floor, surrounded by mountains reaching nearly 3,000 m. This topography limits air circulation, trapping pollutants near the ground and leading to elevated PM levels (Borlaza et al., 2021).

Dataset: OP data were derived from 24-hour PM_{10} samples collected at three different ground-stations in Grenoble — UC (urban centre), UB (urban background) and SU (suburban), whose locations are shown in Figure 4.30 — between 2017 and 2021, following the protocols described in (Calas et al., 2018, 2017). The locations of these stations are shown in Figure 4.30. To account for meteorological and temporal variability

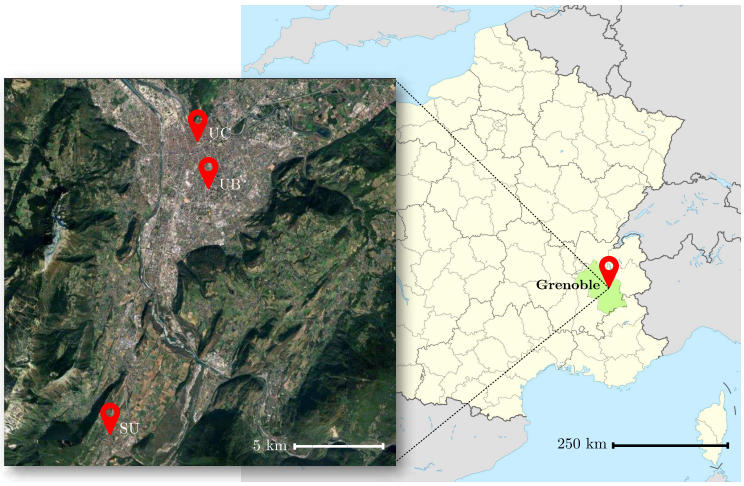

Figure 4.30: Study area in southeastern France (right) and locations of air quality monitoring stations in Grenoble (left). In particular, UC stands for urban centre, UB for urban background, and SU for suburban.

affecting PM composition (World Health Organization, 2021), auxiliary inputs included nine meteorological variables from ERA5/ERA5-Land (Hersbach et al., 2020; Muñoz-Sabater et al., 2021) and two temporal indicators. Satellite inputs consisted of S5P Level-1b radiances cropped into 50×50-pixel patches centred on Grenoble, using only three channels near the red, green, and blue ranges. To highlight the role of SR, we generated three dataset:

- Original LR images;
- Images enhanced via bicubic interpolation (Section 2.3.1.1) with scaling ratios of 2 and 4;
- Images enhanced using our PSF-aware SISR method, S5Net, also with scaling ratios of 2 and 4.

Each image in these datasets was paired with the corresponding daily averages of the 11 auxiliary variables, PM_{10}, and OP values.

4.3.2 Methodology

OPNet (Carbone et al., 2026) is a joint DL-based model composed of (1) a deep residual CNN with 50 layers (He et al., 2016) as the backbone, which extracts N_i surface features from a single satellite image, and (2) a 3-layer feed-forward neural network (Rumelhart et al., 1986) with n neurones as the predictor, which estimates the OP for a given day based on the concatenation of the N_i image features and the $N_a = 11$ auxiliary variables, resulting in $N_i + 11$ input features. A graphical representation of OPNet is depicted in Figure 4.31.

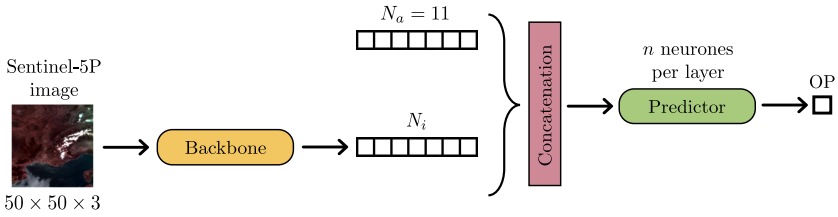

Figure 4.31: A graphical representation of the employed OPNet. In the figure, N_a and N_i represent the number of additional variables and the number of surface features extracted from the image by the backbone, respectively, and n is the number of neurones in each layer of the predictor.

The optimization of OPNet is split into two stages: a domain-adaptive task using PM_{10} data, followed by a domain-specific task on the target OP measurements. This design leverages PM_{10} as a proxy for OP.

4.3.3 Performance analysis

In this section, we first outline the experimental setup, followed by a detailed presentation of the results obtained from the analysis.

4.3.3.1 Experimental setup

We conduct experiments to evaluate the potential for estimating OP from satellite-based surface features extracted from the three datasets combined with the 11 auxiliary variables. Performance is assessed using three standard regression metrics: R-squared (Equation 3.3); RMSE, defined as the square root of MSE (Equation 3.5); and normalised mean absolute error (NMAE) (Equation 4.6):

$$NMAE = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N \cdot \bar{\mathbf{y}}}, \tag{4.6}$$

where $\bar{\mathbf{y}}$ is the mean of the reference values $\mathbf{y} = [y_1, \ldots, y_N]$, with N denoting the total number of values; $\hat{y}_i$ is the i-th predicted value; and $|\cdot|$ denotes the absolute value operator. Lower NMAE values indicate improved prediction accuracy.

To assess computational efficiency, we additionally report: the number of floating-point operations (FLOPs), representing the number of multiply-accumulate operations (MAC); the number of trainable parameters, reflecting model capacity; and the total time (in seconds), indicating the duration of model optimization.

To evaluate the contribution of each architectural component we compare three OPNet configurations. Model A removes the backbone and trains only the predictor on the 11 auxiliary variables. Model $\mathrm{B_{PM}}$ excludes the auxiliary variables and predicts OP solely from the N_i image surface features. Model $\mathrm{A{+}B_{PM}}$ corresponds to the complete OPNet architecture.

All experiments were run on an Nvidia RTX6000 GPU using Python 3.11.2, PyTorch 2.1.0+cu121, PyTorch Lightning 2.1.0, and Torchvision 0.16.0.

4.3.3.2 Experimental results

Table 4.16 and Table 4.17 report the results for OP prediction across all evaluation metrics, OPNet configurations, and image datasets, for scaling ratios 2 and 4 respectively.

Method	**Dataset**	**Performance**		
		$RMSE \downarrow$	$MAE \downarrow$	$R^2 \uparrow$
A	-	0.3399	0.2343	0.8352
B_{PM}	LR	0.4364	0.3855	0.7284
	Cubic	0.3184	0.2722	0.8554
	S5Net	0.3122	0.2390	0.8609
$A+B_{PM}$	LR	0.2897	0.2230	0.8803
	Cubic	<u>0.2815</u>	<u>0.2207</u>	<u>0.8870</u>
	S5Net	**0.2469**	**0.2032**	**0.9130**

Table 4.16: Results of OP prediction for the different OPNet configurations using three satellite image datasets: original LR images (LR), images enhanced via bicubic interpolation (Cubic), and images super-resolved with S5Net (S5Net) with scaling ratio 2.

As expected, increasing the scaling ratio generally leads to a decrease in prediction performance, highlighting the inherent difficulty of accurately super-resolving images at higher magnification factors. The LR images consistently yield the poorest results across all OPNet configurations, confirming that applying SR techniques reliably improves the predictive accuracy for OP. In particular, S5Net consistently outperforms bicubic interpolation, demonstrating the clear advantage of using our PSF-aware SR method over a generic interpolation approach.

The comparison between using only satellite images and combining them with auxiliary variables highlights the importance of multimodal inputs. As the backbone is pre-trained on PM_{10} data, even when auxiliary variables are

Method	**Dataset**	**Performance**		
		$RMSE \downarrow$	$MAE \downarrow$	$R^2 \uparrow$
A	-	0.3399	0.2343	0.8352
B_{PM}	LR	0.4364	0.3855	0.7284
	Cubic	0.3441	0.2960	0.8311
	S5Net	0.3320	0.2592	0.8428
A+B_{PM}	LR	0.2897	<u>0.2230</u>	0.8803
	Cubic	<u>0.2839</u>	0.2410	<u>0.8851</u>
	S5Net	**0.2515**	**0.2134**	**0.9098**

Table 4.17: Results of OP prediction for the different OPNet configurations using three satellite image datasets: original LR images (LR), images enhanced via bicubic interpolation (Cubic), and images super-resolved with S5Net (S5Net) with scaling ratio 4.

omitted, the model can still capture relevant information, although the combination of both inputs remains the most effective strategy. Notably, while LR images alone perform worse than using only the auxiliary variables, applying SISR allows satellite images alone to outperform the A case, emphasising the benefit of enhanced spatial detail. Overall, these results demonstrate that SISR is a crucial step for enhancing predictive performance and that S5Net provides a clear improvement over standard strategies.

Table 4.18 reports the complexity of OPNet across all tested configurations and datasets, for both scaling ratios 2 and 4.

As expected, model A, which uses only the auxiliary variables, is the lightest, but also excludes image information entirely. In contrast, configurations including the image backbone (B_{PM} and A+B_{PM}) are substantially more demanding, with FLOPs in the range of 4.29–4.30 GM and over 16M parameters, regardless of the dataset or scaling ratio. The small variation in FLOPs can be attributable to differences in input resolution,

Method	Dataset	Ratio	Complexity		
			FLOPs	**Params**	**Time**
A	-	-	2.76 kM	5.14k	78.8
B_{PM}	LR	-	4.29 GM	16.2M	183.5
	Cubic	2	4.30 GM	16.3M	42.6
		4	4.30 GM	16.3M	42.5
	S5Net	2	4.29 GM	16.2M	131.8
		4	4.29 GM	16.2M	154.5
$A+B_{PM}$	LR	-	4.29 GM	16.2M	225.4
	Cubic	2	4.29 GM	16.2M	47.8
		4	4.30 GM	16.3M	63.9
	S5Net	2	4.29 GM	16.2M	47.8
		4	4.29 GM	16.2M	132.2

Table 4.18: Complexity analysis of OP prediction for the different OPNet configurations using three satellite image datasets: original LR images (LR), bicubic-interpolated images (Cubic), and S5Net-super-resolved images (S5Net) at scaling ratios 2 and 4. In the table, kM and GM denote kMACs and GMACs, respectively, while k and M indicate multiplications by 10^3 and 10^6.

while the number of parameters remains nearly constant across datasets. Increasing the scaling ratio from 2 to 4 generally leads to slightly longer times, particularly when using S5Net. However, the difference remains modest, and does not significantly impact the overall complexity of the models. These results confirm that incorporating satellite images increases the computational load compared to using only auxiliary variables. Nonetheless, the added complexity is manageable and justified by the substantial performance gains reported in Table 4.16 and Table 4.17, especially when using super-resolved images via S5Net.

Conclusions

This thesis tackled the persistent challenge posed by the limited nominal spatial resolution of atmospheric monitoring satellites, with particular emphasis on Sentinel-5 Precursor (S5P). While ground-based monitoring stations offer precise, high-quality observations, their spatial coverage is inherently limited. Satellites, on the other hand, provide unique, consistent, and global measurements of atmospheric composition, yet this comes at the expense of spatial detail due to the trade-off between spectral coverage and spatial sampling.

The primary objective of this research was thus to enhance the spatial resolution of S5P radiance data without compromising its spectral fidelity, thereby enabling novel opportunities for both scientific analysis and practical applications in Earth observation. The central contribution is a comprehensive framework for single-image super-resolution (SISR) specifically designed for S5P radiance imagery. While maintaining a lightweight architecture, the framework integrates several innovative and impactful components:

- Sensor-aware degradation modelling. A formal characterization of the TROPOspheric Monitoring Instrument (TROPOMI)'s point spread function (PSF) was developed to generate realistic low- and high-resolution training pairs. By explicitly modelling the sensor's imaging physics, this step addresses a

critical gap in the literature, where most approaches assume sensor-independent degradations that fail to capture the actual acquisition characteristics.

- Spectral-scalable dynamic multi-directional cascade fine-tuning. To manage the unprecedented spectral dimensionality of S5P (approximately 4000 channels), the framework introduces a dynamic fine-tuning strategy. Unlike independent per-channel training, static global approaches, or other dimensionality-reduction methods such as principal component analysis (PCA), this approach dynamically adapts the number of training iterations according to the correlations between channels. This design preserves spectral fidelity while optimising computational efficiency and spatial reconstruction quality.

- Dedicated preprocessing and dataset generation. A complete preprocessing pipeline was implemented to prepare S5P radiance data for learning-based SISR. Given the complexity, volume, and characteristics of hyperspectral radiance data, this step is essential and has been largely overlooked in previous studies. The resulting dataset constitutes a novel resource for the community.

Collectively, these components establish a framework that achieves a balance between computational efficiency, spatial enhancement, and spectral consistency — an accomplishment not previously demonstrated in existing approaches.

Experimental results consistently demonstrate the clear advantage of explicitly incorporating sensor-specific PSF information into the SISR framework. Across all evaluated scaling ratios, the PSF-aware configuration outperforms PSF-unaware variants and state-of-the-art methods in all

standard full-reference metrics, while maintaining comparable spectral consistency. This highlights the critical importance of accurate sensor modelling for reconstructing high-frequency spatial details, edges, and complex patterns in remote sensing data, where each pixel carries physically meaningful information.

A detailed examination of the results reveals several nuanced observations.

First, a sensitivity analysis was conducted to evaluate the robustness of S5Net to moderate mismatches between the assumed and true PSFs. The results indicate the presence of a range of fine-tuning gains around the optimal value where the reconstruction PSNR remains mostly unchanged, demonstrating that the proposed model is robust to imperfect knowledge of the sensor-specific PSF.

A preliminary generalization capability evaluation confirms the effectiveness of S5Net beyond the image-specific training setting. Across all tested scaling ratios and for all tested spectral channels, S5Net consistently outperforms all benchmark approaches. These results demonstrate that the network architecture contributes significantly to the accurate reconstruction of fine-scale spatial structures, even under conventional supervised training conditions. While this analysis represents a controlled, idealised scenario that cannot be replicated operationally due to the lack of true HR references, it provides an important validation of S5Net's design principles. The network's ability to recover fine spatial details in this context highlights the generalization potential of the proposed method.

In contrast, the subsequent analyses consider the performance of S5Net under the image-specific framework. As true high-resolution references are unavailable, this evaluation

provides a practical assessment of the network's ability to reconstruct fine spatial structures under realistic conditions and highlights key aspects of S5Net's performance.

First, the performance gain from PSF-aware modelling varies across spectral domains. For ultraviolet (UV) and short-wave infrared (SWIR) bands, where the actual and assumed PSFs are similar, improvements are modest. In contrast, for ultraviolet and visible (UVIS) and near-infrared (NIR) bands, which display larger deviations from typical assumptions, the PSF-aware approach yields substantial reductions in reconstruction error. This indicates that the benefits of accurate PSF integration are directly proportional to the degree of mismatch between actual and assumed acquisition characteristics, emphasizing the need for sensor-specific calibration to achieve optimal results.

Second, although the PSF-aware SISR consistently improves pixel-level fidelity with the available image in the absence of ground-truths (GTs), visual inspection remains essential. In these cases, PSF-aware outputs demonstrate superior edge preservation and finer structural details compared to PSF-unaware or benchmark methods, confirming the benefit of incorporating PSF knowledge into the reconstruction process.

Third, the results illustrate the inherent trade-off between reconstruction accuracy and the scaling factor. As expected, performance declines with increasing upscaling ratios due to the ill-posed nature of the inverse problem. Nevertheless, the PSF-aware framework maintains a consistent advantage over benchmark methods even at the highest tested ratio, demonstrating robustness and generalization capability. This suggests that embedding physics-informed priors can partially alleviate the limitations inherent in super-resolution.

The comprehensive evaluation of the full framework, including

the dynamic fine-tuning approach, provides additional insights. Considering the trade-off between error-based metrics and perceptual/spectral metrics, it is evident that full per-channel fine-tuning maximises pixel-level accuracy but may slightly compromise structural and spectral consistency. Independent channel-wise optimization enhances high-frequency spatial detail reconstruction but can introduce subtle spectral distortions or reduce overall structural similarity. Hence, this strategy is most suitable for applications demanding absolute accuracy, albeit at a very high computational cost. In contrast, static multi-channel fine-tuning achieves superior structural and perceptual quality, particularly when processing is performed band-wise, highlighting the advantages of reducing intra-region variability for hyperspectral SISR. The dynamic fine-tuning approach introduced in this thesis represents a favourable compromise, maintaining results close to the fully independent baseline while drastically reducing the number of training iterations even when compared to the static approach. Additionally, it exhibits slightly improved spectral consistency in the global scenario, indicating that dynamic fine-tuning effectively accommodates channel-specific variations without sacrificing spatial reconstruction quality. This capability is particularly relevant when scaling to the full 3445-channel dataset or when near-real-time processing is required.

Visual inspection of the final super-resolved outputs supports these quantitative trends. Both the fully independent baseline and the dynamic band-wise approach provide crisp edge delineation and faithful reconstruction across coastlines and other complex features. The near-indistinguishability of the two methods confirms that the dynamic strategy preserves high perceptual quality while significantly reducing computational

cost, a crucial aspect for operational deployment.

Beyond algorithmic performance, the findings also illustrate the practical value of super-resolution in real-world scenarios. For instance, in predicting the oxidative potential (OP) of particulate matter over urban areas using satellite-based surface features, low-resolution inputs consistently underperform compared to their super-resolved counterparts. This confirms that current sensor resolutions are insufficient to capture the fine-scale spatial patterns critical for air quality assessment in urban areas. By enhancing spatial detail, SISR substantially improves model accuracy, both as a standalone input and in combination with auxiliary variables.

While direct validation with independent high-resolution sources or in-situ monitoring stations is not feasible at present due to the lack of true HR references and limited availability of urban surface measurements, preliminary efforts using OPNet provide a first indication of the potential applicability of the framework for urban air-quality modelling. In future work, extending such validation with additional ground-based networks for selected pollutants could further demonstrate practical utility and robustness.

Similarly, while some performance improvements reported in this thesis are modest in absolute terms, they remain meaningful for end-user applications that rely on accurate fine-scale spatial information. Future studies could extend operational validation and application-specific performance assessments — including, but not limited to, the estimation of OP — to more comprehensively quantify the impact of super-resolved data on predictive modelling and decision-making in real-world scenarios.

Overall, these findings reinforce that super-resolution should not be treated merely as a preprocessing step. Rather, it acts

as a critical enabler for predictive modelling in remote sensing contexts where high-resolution GTs are unavailable and spatial precision is paramount. By increasing the effective resolution of S5P radiance data, the framework can support a wide array of downstream applications, from atmospheric chemistry studies to operational air quality monitoring. Importantly, the approach is generalizable: the principles of sensor-specific degradation modelling and dynamic spectral fine-tuning can be applied to other hyperspectral missions, providing a versatile tool for the Earth observation community and opening multiple avenues for further research.

Acknowledgements

I would like to express my sincere gratitude to all the people who supported me during the development of this thesis.

First and foremost, I would like to thank my supervisors, Prof. Rocco Restaino and Dr. Gemine Vivone, for their constant guidance, support, valuable advice and continuous feedback throughout the development of this thesis. Their expertise, encouragement, patience and insightful comments have been fundamental to the completion of this research and have greatly contributed to my personal and academic growth.

I would also like to thank all the professors I had the opportunity to learn from throughout my academic journey, whose teaching has played an important role in shaping my academic path.

I would also like to express my sincere thanks to Prof. Jocelyn Chanussot and Dr. Gaëlle Uzu for their support and helpful discussions during my research secondment at the Institut des Géosciences de l'Environnement (IGE) in Grenoble, France. I am also grateful to the many scientists there who contributed to the collection of the particulate matter (PM) and oxidative potential (OP) samples used in the application part of this work.

I would also like to thank my colleagues for the stimulating discussions during this journey.

Finally, I would like to express my heartfelt gratitude to my family and friends for their constant encouragement and support, as well as to all those who, directly or indirectly, contributed to the completion of this thesis.

Bibliography

Abrams, J.Y., Weber, R.J., Klein, M., Sarnat, S.E., Chang, H.H., Strickland, M.J., Verma, V., Fang, T., Bates, J.T., Mulholland, J.A., et al., 2017. Associations between ambient fine particulate oxidative potential and cardiorespiratory emergency department visits. Environmental Health Perspectives 125, 107008. doi:10.1289/EHP1545.

Aburaed, N., Alkhatib, M.Q., Marshall, S., Zabalza, J., Al Ahmad, H., 2023. A review of spatial enhancement of hyperspectral remote sensing imaging techniques. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16, 2275–2300. doi:10.1109/JSTARS.2023.3242048.

Achakulwisut, P., Brauer, M., Hystad, P., Anenberg, S.C., 2019. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: estimates from global datasets. The Lancet. Planetary Health 3, e166–e178. doi:10.1016/S2542-5196(19)30046-4.

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54, 4311–4322.

Ahn, N., Kang, B., Sohn, K.A., 2018. Fast, accurate, and lightweight super-resolution with cascading residual network, in: Computer Vision — ECCV 2018: 15th European Conference, Proceedings, Part X, pp. 256–272. doi:10.1007/978-3-030-01249-6_16.

Aiazzi, B., Baronti, S., Selva, M., Alparone, L., 2013. Bi-cubic interpolation for shift-free pan-sharpening. ISPRS Journal of Photogrammetry and Remote Sensing 86, 65–76. doi:10.1016/j.isprsjprs.2013.09.007.

Akgun, T., Altunbasak, Y., Mersereau, R., 2005. Super-resolution reconstruction of hyperspectral images. IEEE Transactions on Image Processing 14, 1860–1875. doi:10.1109/TIP.2005.854479.

Anwar, S., Barnes, N., 2022. Densely residual Laplacian super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 1192–1204. doi:10.1109/TPAMI.2020.3021088.

Arun, P.V., Buddhiraju, K.M., Porwal, A., Chanussot, J., 2020. CNN-based super-resolution of hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing 58, 6106–6121. doi:10.1109/TGRS.2020.2973370.

Bates, J.T., Weber, R.J., Abrams, J., Verma, V., Fang, T., Klein, M., Strickland, M.J., Sarnat, S.E., Chang, H.H., Mulholland, J.A., et al., 2015. Reactive oxygen species generation linked to sources of atmospheric particulate matter and cardiorespiratory effects. Environmental Science & Technology 49, 13605–13612. doi:10.1021/acs.est.5b02967.

Blau, Y., Michaeli, T., 2018. The perception-distortion tradeoff, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6228–6237. doi:10.1109/CVPR.2018.00652.

Borlaza, L.J.S., Weber, S., Uzu, G., Jacob, V., Cañete, T., Micallef, S., Trébuchon, C., Slama, R., Favez, O., Jaffrezo, J.L., 2021. Disparities in particulate matter (PM_{10}) origins and oxidative potential at a city scale (Grenoble, France) – Part 1: Source apportionment at three neighbouring sites.

Atmospheric Chemistry and Physics 21, 5415–5437. doi:10.5194/acp-21-5415-2021.

Borsdorff, T., Aan de Brugh, J., Hu, H., Aben, I., Hasekamp, O., Landgraf, J., 2018. Measuring carbon monoxide with TROPOMI: First results and a comparison with ECMWF-IFS analysis data. Geophysical Research Letters 45, 2826–2832. doi:10.1002/2018GL077045.

Bovensmann, H., Burrows, J.P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V.V., Chance, K.V., Goede, A.P.H., 1999. SCIAMACHY: Mission objectives and measurement modes. Journal of the Atmospheric Sciences 56, 127 – 150. doi:10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenthal, S., et al., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter, in: Proceedings of the National Academy of Sciences of the United States of America, pp. 9592–9597. doi:10.1073/pnas.1803222115.

Burrows, J.P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.U., et al., 1999. The global ozone monitoring experiment (GOME): Mission concept and first scientific results. Journal of the Atmospheric Sciences 56, 151–175. doi:10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2.

Calas, A., Uzu, G., Kelly, F.J., Houdier, S., Martins, J.M.F., Thomas, F., Molton, F., Charron, A., Dunster, C., Oliete, A., et al., 2018. Comparison between five acellular oxidative potential measurement assays performed with detailed chemistry on PM_{10} samples from the city of Chamonix (France).

Atmospheric Chemistry and Physics 18, 7863–7875. doi:10.5194/acp-18-7863-2018.

Calas, A., Uzu, G., Martins, J.M.F., Voisin, D., Spadini, L., Lacroix, T., Jaffrezo, J.L., 2017. The importance of simulated lung fluid (SLF) extractions for a more relevant evaluation of the oxidative potential of particulate matter. Scientific Reports 7, 11617. doi:10.1038/s41598-017-11979-3.

Campisi, P., Egiazarian, K., 2007. Blind image deconvolution: theory and applications. CRC Press, Boca Raton.

Carbone, A., Hough, I., Vivone, G., Chanussot, J., Restaino, R., Dupont, H., Jaffrezo, J.L., Uzu, G., 2026. OPNet: A deep-learning approach for estimating particulate matter's oxidative potential from satellite imagery. Atmospheric Environment: X 29, 100406. doi:https://doi.org/10.1016/j.aeaoa.2025.100406.

Carbone, A., Restaino, R., Vivone, G., 2024a. Efficient hyperspectral super-resolution of Sentinel-5P data via dynamic multidirectional cascade fine-tuning. IEEE Geoscience and Remote Sensing Letters 21, 1–5. doi:10.1109/LGRS.2024.3454155.

Carbone, A., Restaino, R., Vivone, G., Chanussot, J., 2024b. Model-based super-resolution for Sentinel-5P data. IEEE Transactions on Geoscience and Remote Sensing 62, 1–16. doi:10.1109/TGRS.2024.3387877.

Chance, K., et al., 2019. TEMPO green paper: Chemistry, physics, and meteorology experiments with the tropospheric emissions: Monitoring of pollution instrument, in: Sensors, Systems, and Next-Generation Satellites XXIII. doi:10.1117/12.2534883.

Chang, H., Yeung, D.Y., Xiong, Y., 2004. Super-resolution through neighbor embedding, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern

Recognition, 2004. CVPR 2004., pp. I–I. doi:10.1109/CVPR.2004.1315043.

Chang, K.L., Cooper, O.R., West, J.J., Serre, M.L., Schultz, M.G., Lin, M., Marécal, V., Josse, B., Deushi, M., Sudo, K., Liu, J., Keller, C.A., 2019. A new method (M^3Fusion v1) for combining observations and multiple model output for an improved estimate of the global surface ozone distribution. Geoscientific Model Development 12, 955–978. doi:10.5194/gmd-12-955-2019.

Chen, C., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., 2023a. A review of hyperspectral image super-resolution based on deep learning. Remote Sensing 15, 2853. doi:10.3390/rs15112853.

Chen, J., Hoek, G., 2020. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. Environment International 143, 105974. doi:10.1016/j.envint.2020.105974.

Chen, W., Zheng, X., Lu, X., 2021. Hyperspectral image super-resolution with self-supervised spectral-spatial residual network. Remote Sensing 13. doi:10.3390/rs13071260.

Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C., 2023b. Activating more pixels in image super-resolution transformer, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22367–22377. doi:10.1109/CVPR52729.2023.02142.

Cheng, X., Li, X., Yang, J., Tai, Y., 2018. SESR: Single image super resolution with recursive squeeze and excitation networks, in: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 147–152. doi:10.1109/ICPR.2018.8546130.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., et al., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution:

an analysis of data from the Global Burden of Diseases Study 2015. The Lancet 389, 1907–1918. doi:10.1016/S0140-6736(17)30505-6.

Courrèges-Lacoste, G.B., Arcioni, M., Meijer, Y., Bèzy, J.L., Bensi, P., Langen, J., 2008. Sentinel-4: The geostationary component of the GMES atmospheric monitoring mission, in: Proceedings of the 7th International Conference on Space Optics (ICSO).

Cremers, D., Reid, I., Saito, H., Yang, M.H., 2015. A+: Adjusted anchored neighborhood regression for fast super-resolution, in: Computer Vision – ACCV 2014, pp. 111–126. doi:10.1007/978-3-319-16817-3_8.

Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L., 2019. Second-order attention network for single image super-resolution, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11057–11066. doi:10.1109/CVPR.2019.01132.

Dominutti, P., Mudway, I., Neses, A., Bougiatioti, K., Mihalopoulos, N., Hoek, G., Harrison, R., Uzu, G., 2023. Harmonization of oxidative potential of PM monitoring for application in pilots. Technical Report. CNRS. URL: https://riurbans.eu/wp-content/uploads/2023/03/RI-URBANS_D11_D2_3.pdf.

Dong, C., Loy, C.C., He, K., Tang, X., 2016a. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 38, 295–307. doi:10.1109/TPAMI.2015.2439281.

Dong, C., Loy, C.C., Tang, X., 2016b. Accelerating the super-resolution convolutional neural network, in: Computer Vision – ECCV 2016, Springer International Publishing. pp. 391–407.

Duchon, C.E., 1979. Lanczos filtering in one and two dimensions. Journal of Applied Meteorology and Climatology 18, 1016 – 1022. doi:10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.

Eilers, P.H.C., Ruckebusch, C., 2022. Fast and simple super-resolution with single images. Scientific Reports 12, 11241. doi:10.1038/s41598-022-14874-8.

European Space Agency, 2013. Algorithm theoretical basis document for the TROPOMI L01b data processor. Technical Report. URL: https://sentiwiki.copernicus.eu/web/s5p-documents.

European Space Agency, 2016. Readme for TROPOMI pixel response function. Technical Report. URL: https://sentiwiki.copernicus.eu/web/s5p-documents.

European Space Agency, 2022. Input/output data specification for the TROPOMI L01b data processor. Technical Report. URL: https://sentiwiki.copernicus.eu/web/s5p-documents.

Fan, C., Wu, C., Li, G., Ma, J., 2017. Projections onto convex sets super-resolution reconstruction based on point spread function estimation of low-resolution remote sensing images. Sensors 17, 362. doi:10.3390/s17020362.

Fernandez-Beltran, R., Latorre-Carmona, P., Pla, F., 2017. Single-frame super-resolution in remote sensing: a practical overview. International Journal of Remote Sensing 38, 314–354. doi:10.1080/01431161.2016.1264027.

Freeman, W.T., Pasztor, E.C., Carmichael, O.T., 2000. Learning low-level vision. International Journal of Computer Vision 40, 25–47. doi:10.1023/A:1026501619075.

Gao, D., Ripley, S., Weichenthal, S., Godri Pollitt, K.J., 2020. Ambient particulate matter oxidative potential: Chemical determinants, associated health effects, and strategies for risk

management. Free Radical Biology and Medicine 151, 7–25. doi:10.1016/j.freeradbiomed.2020.04.028.

GBD 2019 Risk Factors Collaborators, 2020. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet (London, England) 396, 1223–1249. doi:10.1016/S0140-6736(20)30752-2.

Goldberg, D.L., Anenberg, S.C., Kerr, G.H., Mohegh, A., Lu, Z., Streets, D.G., 2021. TROPOMI NO_2 in the United States: A detailed look at the annual averages, weekly cycles, effects of temperature, and correlation with surface NO_2 concentrations. Earth's Future 9, e2020EF001665. doi:10.1029/2020EF001665.

Goodkind, A.L., Tessum, C.W., Coggins, J.S., Hill, J.D., Marshall, J.D., 2019. Fine-scale damage estimates of particulate matter air pollution reveal opportunities for location-specific mitigation of emissions. Proceedings of the National Academy of Sciences 116, 8775–8780. doi:10.1073/pnas.1816102116.

Guarino, G., Ciotola, M., Vivone, G., Scarpa, G., 2024. Band-wise hyperspectral image pansharpening using CNN model propagation. IEEE Transactions on Geoscience and Remote Sensing 62, 1–18. doi:10.1109/TGRS.2023.3339337.

Han, W., Chang, S., Liu, D., Yu, M., Witbrock, M., Huang, T.S., 2018. Image super-resolution via dual-state recurrent networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1654–1663. doi:10.1109/CVPR.2018.00178.

He, H., Siu, W.C., 2011. Single image super-resolution using Gaussian process regression, in: CVPR 2011, pp. 449–456. doi:10.1109/CVPR.2011.5995713.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.

He, Z., Dai, T., Lu, J., Jiang, Y., Xia, S.T., 2020. Fakd: Feature-affinity based knowledge distillation for efficient image super-resolution, in: 2020 IEEE International Conference on Image Processing (ICIP), pp. 518–522. doi:10.1109/ICIP40778.2020.9190917.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146. doi:10.1002/qj.3803.

Hestenes, M.R., Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems. Journal of research of the National Bureau of Standards 49, 409–435.

Hodnebrog, Ø., Stordal, F., Berntsen, T.K., 2011. Does the resolution of megacity emissions impact large scale ozone? Atmospheric Environment 45, 6852–6862. doi:10.1016/j.atmosenv.2011.01.012.

Hosseiny, B., Mahdianpari, M., Hemati, M., Radman, A., Mohammadimanesh, F., Chanussot, J., 2024. Beyond supervised learning in remote Sensing: A systematic review of deep learning approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17, 1035–1052. doi:10.1109/JSTARS.2023.3316733.

Hu, J., Li, Y., Xie, W., 2017. Hyperspectral image super-resolution by spectral difference learning and spatial error correction. IEEE Geoscience and Remote Sensing Letters 14, 1825–1829. doi:10.1109/LGRS.2017.2737637.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. doi:10.1109/CVPR.2018.00745.

Hu, J.F., Huang, T.Z., Deng, L.J., Dou, H.X., Hong, D., Vivone, G., 2022a. Fusformer: A Transformer-based fusion network for hyperspectral image super-resolution. IEEE Geoscience and Remote Sensing Letters 19, 1–5. doi:10.1109/LGRS.2022.3194257.

Hu, J.F., Huang, T.Z., Deng, L.J., Jiang, T.X., Vivone, G., Chanussot, J., 2022b. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. IEEE Transactions on Neural Networks and Learning Systems 33, 7251–7265. doi:10.1109/TNNLS.2021.3084682.

Hu, Y., Li, J., Huang, Y., Gao, X., 2020. Channel-wise and spatial feature modulation network for single image super-resolution. IEEE Transactions on Circuits and Systems for Video Technology 30, 3911–3927. doi:10.1109/TCSVT.2019.2915238.

Huang, B., He, B., Wu, L., Guo, Z., 2021. Deep residual dual-attention network for super-resolution reconstruction of remote sensing images. Remote Sensing 13, 2784. doi:10.3390/rs13142784.

Hui, Z., Wang, X., Gao, X., 2018. Fast and accurate single image super-resolution via information distillation network, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 723–731. doi:10.1109/CVPR.2018.00082.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.M., Dominguez, J.J., Engelen, R., Eskes, H., Flemming, J., et al., 2019. The CAMS reanalysis of atmospheric composition. Atmospheric Chemistry and Physics 19, 3515–3556. doi:10.5194/acp-19-3515-2019.

Irani, M., Peleg, S., 1991. Improving resolution by image registration. CVGIP: Graphical Models and Image Processing 53, 231–239. doi:10.1016/1049-9652(91)90045-L.

Jiang, K., Wang, Z., Yi, P., Jiang, J., Xiao, J., Yao, Y., 2018. Deep distillation recursive network for remote sensing imagery super-resolution. Remote Sensing 10, 1700. doi:10.3390/rs10111700.

Johnson, N.M., Hoffmann, A.R., Behlen, J.C., Lau, C., Pendleton, D., Harvey, N., Shore, R., Li, Y., Chen, J., Tian, Y., Zhang, R., 2021. Air pollution and children's health – a review of adverse effects associated with prenatal exposure from fine to ultrafine particulate matter. Environmental Health and Preventive Medicine 26, 72. doi:10.1186/s12199-021-00995-5.

Karner, A.A., Eisinger, D.S., Niemeier, D.A., 2010. Near-roadway air quality: synthesizing the findings from real-world data. Environmental Science & Technology 44, 5334–5344. doi:10.1021/es100008x.

Keshava, N., Mustard, J., 2002. Spectral unmixing. IEEE Signal Processing Magazine 19, 44–57. doi:10.1109/79.974727.

Keys, R., 1981. Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing 29, 1153–1160. doi:10.1109/TASSP.1981.1163711.

Kim, J., Jeong, U., Ahn, M.H., Kim, J.H., Park, R.J., Lee, H., Song, C.H., Choi, Y.S., Lee, K.H., Yoo, J.M., et al., 2020. New era of air quality monitoring from space: Geostationary environment monitoring spectrometer (GEMS). Bulletin of the American Meteorological Society 101, E1–E22. doi:10.1175/BAMS-D-18-0013.1.

Kim, J., Lee, J.K., Lee, K.M., 2016a. Accurate image super-resolution using very deep convolutional networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1646–1654. doi:10.1109/CVPR.2016.182.

Kim, J., Lee, J.K., Lee, K.M., 2016b. Deeply-recursive convolutional network for image super-resolution, in: 2016

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1637–1645. doi:10.1109/CVPR.2016.181.

Kim, K.I., Kwon, Y., 2010. Single-image super-resolution using sparse regression and natural image prior. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1127–1133. doi:10.1109/TPAMI.2010.25.

Larkin, A., Geddes, J.A., Martin, R.V., Xiao, Q., Liu, Y., Marshall, J.D., Brauer, M., Hystad, P., 2017. Global land use regression model for nitrogen dioxide air pollution. Environmental Science & Technology 51, 6957–6964. URL: https://doi.org/10.1021/acs.est.7b01148, doi:10.1021/acs.est.7b01148.

Layton, W., Sussman, M., 2017. Numerical Linear Algebra. Glasstree.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105–114. doi:10.1109/CVPR.2017.19.

Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R.T., Haines, A., Ramanathan, V., 2019. Effects of fossil fuel and total anthropogenic emission removal on public health and climate. Proceedings of the National Academy of Sciences 116, 7192–7197. doi:10.1073/pnas.1819989116.

Levelt, P., van den Oord, G., Dobber, M., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., Saari, H., 2006. The ozone monitoring instrument. IEEE Transactions on Geoscience and Remote Sensing 44, 1093–1101. doi:10.1109/TGRS.2006.872333.

Li, J., Cui, R., Li, B., Song, R., Li, Y., Dai, Y., Du, Q., 2020. Hyperspectral image super-resolution by band attention through adversarial learning. IEEE Transactions on Geoscience and Remote Sensing 58, 4304–4318. doi:10.1109/TGRS.2019.2962713.

Li, J., Fang, F., Mei, K., Zhang, G., 2018. Multi-scale residual network for image super-resolution, in: Computer Vision — ECCV 2018, pp. 527–542. doi:10.1007/978-3-030-01237-3_32.

Li, X., Orchard, M., 2001. New edge-directed interpolation. IEEE Transactions on Image Processing 10, 1521–1527. doi:10.1109/83.951537.

Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., 2017. Enhanced deep residual networks for single image super-resolution, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1132–1140. doi:10.48550/arXiv.1707.02921.

Liu, C., Fan, Z., Zhang, G., 2022a. GJTD-LR: A trainable grouped joint tensor dictionary with low-rank prior for single hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing 60, 1–17. doi:10.1109/TGRS.2022.3204049.

Liu, J., Tang, J., Wu, G., 2020. Residual feature distillation network for lightweight image super-resolution, in: Computer Vision — ECCV 2020 Workshops, pp. 41–55. doi:10.1007/978-3-030-67070-2_2.

Liu, J., Wu, Z., Xiao, L., Wu, X.J., 2022b. Model inspired autoencoder for unsupervised hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing 60, 1–12. doi:10.1109/TGRS.2022.3143156.

Lu, T., Wang, J., Zhang, Y., Wang, Z., Jiang, J., 2019. Satellite image super-resolution via multi-scale residual deep neural network. Remote Sensing 11, 1588. doi:10.3390/rs11131588.

Mathew, A., Shekar, P.R., Nair, A.T., Mallick, J., Rathod, C., Bindajam, A.A., Alharbi, M.M., Abdo, H.G., 2024. Unveiling urban air quality dynamics during COVID-19: A Sentinel-5P TROPOMI hotspot analysis. Scientific Reports 14, 21624. doi:10.1038/s41598-024-72276-4.

Mei, S., Yuan, X., Ji, J., Zhang, Y., Wan, S., Du, Q., 2017. Hyperspectral image spatial super-resolution via 3D full convolutional neural network. Remote Sensing 9, 1139. doi:10.3390/rs9111139.

Michaeli, T., Irani, M., 2013. Nonparametric blind super-resolution, in: 2013 IEEE International Conference on Computer Vision, pp. 945–952. doi:10.1109/ICCV.2013.121.

Moser, B., Raue, F., Frolov, S., Hees, J., Palacio, S., Dengel, A., 2023. Hitchhiker's guide to super-resolution: Introduction and recent advances. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 9862–9882. doi:10.1109/TPAMI.2023.3243794.

Moser, B.B., Shanbhag, A.S., Raue, F., Frolov, S., Palacio, S.M., Dengel, A., 2024. Diffusion models, image super-resolution, and everything: A survey. IEEE Transactions on Neural Networks and Learning Systems 36, 11793–11813.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., et al., 2016. The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview. Atmospheric Measurement Techniques 9, 1279–1301. doi:10.5194/amt-9-1279-2016.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13, 4349–4383. URL: https://essd.copernicus.org/articles/13/4349/2021/, doi:10.5194/essd-13-4349-2021.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, pp. 807–814.

Neal, L.S., Dalvi, M., Folberth, G., McInnes, R.N., Agnew, P., O'Connor, F.M., Savage, N.H., Tilbee, M., 2017. A description and evaluation of an air quality model nested within global and regional composition-climate models using MetUM. Geoscientific Model Development 10, 3941–3962. doi:10.5194/gmd-10-3941-2017.

Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, H., 2020. Single image super-resolution via a holistic attention network, in: Computer Vision — ECCV 2020, pp. 191–207. doi:10.1007/978-3-030-58610-2_12.

Peters, R., Ee, N., Peters, J., Booth, A., Mudway, I., Anstey, K.J., 2019. Air pollution and dementia: A systematic review. Journal of Alzheimer's disease: JAD 70, S145–S163. doi:10.3233/JAD-180631.

Qu, Y., Qi, H., Kwan, C., 2018. Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition , 2511–2520doi:10.1109/CVPR.2018.00266.

Rashid, C.S., Bansal, A., Simmons, R.A., 2018. Oxidative stress, intrauterine growth restriction, and developmental programming

of type 2 diabetes. Physiology 33, 348–359. doi:10.1152/physiol.00023.2018.

Rew, R., Davis, G., 1990. NetCDF: an interface for scientific data access. IEEE Computer Graphics and Applications 10, 76–82. doi:10.1109/38.56302.

Riess, T.C.V.W., Boersma, K.F., van Vliet, J., Peters, W., Sneep, M., Eskes, H., van Geffen, J., 2022. Improved monitoring of shipping NO_2 with TROPOMI: decreasing NO_x emissions in European seas during the COVID-19 pandemic. Atmospheric Measurement Techniques 15, 1415–1438. doi:10.5194/amt-15-1415-2022.

Robbins, H., Monro, S., 1951. A stochastic approximation method. The Annals of Mathematical Statistics 22, 400–407. doi:10.1214/aoms/1177729586.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. doi:10.1038/323533a0.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M., 2023. Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 4713–4726. doi:10.1109/TPAMI.2022.3204461.

Scales, L.E., 1985. Non-linear least squares, in: Introduction to Non-Linear Optimization. Macmillan Education UK, London, pp. 110–138. doi:10.1007/978-1-349-17741-7_5.

Schowengerdt, R.A., 2007. Remote sensing, models, and methods for image processing. 3rd ed., Academic Press, Burlington, MA.

Schultz, R., Stevenson, R., 1996. Extraction of high-resolution frames from video sequences. IEEE Transactions on Image Processing 5, 996–1011. doi:10.1109/83.503915.

Sdraka, M., Papoutsis, I., Psomas, B., Vlachos, K., Ioannidis, K., Karantzalos, K., Gialampoukidis, I., Vrochidis, S., 2022. Deep learning for downscaling remote sensing images: Fusion and super-resolution. IEEE Geoscience and Remote Sensing Magazine 10, 202–255. doi:10.1109/MGRS.2022.3171836.

Shaddick, G., Thomas, M.L., Amini, H., Broday, D., Cohen, A., Frostad, J., Green, A., Gumy, S., Liu, Y., Martin, R.V., et al., 2018. Data integration for the assessment of population exposure to ambient air pollution for global burden of disease assessment. Environmental Science & Technology 52, 9069–9078. doi:10.1021/acs.est.8b02864.

Shangguan, Y., Zhuang, X., Querol, X., Li, B., Moreno, N., Trechera, P., Sola, P.C., Uzu, G., Li, J., 2022. Characterization of deposited dust and its respirable fractions in underground coal mines: Implications for oxidative potential-driving species and source apportionment. International Journal of Coal Geology 258, 104017. doi:10.1016/j.coal.2022.104017.

Shankar, K., Mehendale, H.M., 2014. Oxidative stress, in: Encyclopedia of Toxicology (Third Edition). Academic Press, Oxford, pp. 735–737. doi:10.1016/B978-0-12-386454-3.00345-6.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Shaw, G., Manolakis, D., 2002. Signal processing for hyperspectral image exploitation. IEEE Signal Processing Magazine 19, 12–16. doi:10.1109/79.974715.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: 2016 IEEE Conference on Computer Vision

and Pattern Recognition (CVPR), pp. 1874–1883. doi:10.1109/CVPR.2016.207.

Shocher, A., Cohen, N., Irani, M., 2018. Zero-shot super-resolution using deep internal learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2018.00329.

Siddique, M.A., Naseer, E., Usama, M., Basit, A., 2024. Estimation of surface-level NO_2 using satellite remote sensing and machine learning: A review. IEEE Geoscience and Remote Sensing Magazine 12, 8–34. doi:10.1109/MGRS.2024.3398434.

Sies, H., 2018. On the history of oxidative stress: Concept and some aspects of current development. Current Opinion in Toxicology 7, 122–126. doi:10.1016/j.cotox.2018.01.002.

Sun, J., Xu, Z., Shum, H.Y., 2008. Image super-resolution using gradient profile prior, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. doi:10.1109/CVPR.2008.4587659.

Tai, Y., Yang, J., Liu, X., 2017. Image super-resolution via deep recursive residual network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2790–2798. doi:10.1109/CVPR.2017.298.

Thenkabail, P.S., Lyon, J.G., 2016. Hyperspectral Remote Sensing of Vegetation. CRC Press.

Timofte, R., De, V., Gool, L.V., 2013. Anchored neighborhood regression for fast example-based super-resolution, in: 2013 IEEE International Conference on Computer Vision, pp. 1920–1927. doi:10.1109/ICCV.2013.241.

Tong, T., Li, G., Liu, X., Gao, Q., 2017. Image super-resolution using dense skip connections, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 4809–4817. doi:10.1109/ICCV.2017.514.

United Nations, 1948. Universal Declaration of Human Rights. URL: https://www.un.org/en/about-us/universal-declaration-of-human-rights.

United States Congress, 1970. Clean air act. URL: https://www.epa.gov/clean-air-act-overview.

US Environmental Protection Agency, 2016. Particulate matter (PM) basics. URL: https://www.epa.gov/pm-pollution/particulate-matter-pm-basics.

Veefkind, J.P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H.J., de Haan, J.F., Kleipool, Q., et al., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sensing of Environment 120, 70–83. doi:10.1016/j.rse.2011.09.027.

Vivone, G., Dalla Mura, M., Garzelli, A., Pacifici, F., 2021. A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 6102–6118. doi:10.1109/JSTARS.2021.3086877.

Vohra, K., Vodonos, A., Schwartz, J., Marais, E.A., Sulprizio, M.P., Mickley, L.J., 2021. Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. Environmental Research 195, 110754. doi:10.1016/j.envres.2021.110754.

Wald, L., 2002. Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions. Presses des MINES.

Wald, L., Ranchin, T., Mangolini, M., 1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. Photogrammetric Engineering and Remote Sensing 63, 691–699.

Wang, H., Cheng, S., Li, Y., Du, A., 2023. Lightweight remote-sensing image super-resolution via attention-based multilevel feature fusion network. IEEE Transactions on Geoscience and Remote Sensing 61, 1–15. doi:10.1109/TGRS.2023.3336524.

Wang, Q., Li, Q., Li, X., 2021a. Hyperspectral image super-resolution using spectrum and feature context. IEEE Transactions on Industrial Electronics 68, 11276–11285. doi:10.1109/TIE.2020.3038096.

Wang, S., Wang, B., 2014. Super-resolution restoration of multispectral images based on principal component analysis, in: 2014 12th International Conference on Signal Processing (ICSP), pp. 841–846. doi:10.1109/ICOSP.2014.7015122.

Wang, X., Chen, J., Wei, Q., Richard, C., 2022a. Hyperspectral image super-resolution via deep prior regularization with parameter estimation. IEEE Transactions on Circuits and Systems for Video Technology 32, 1708–1723. doi:10.1109/TCSVT.2021.3078559.

Wang, X., Wu, Y., Ming, Y., Lv, H., 2020. Remote sensing imagery super resolution based on adaptive multi-scale feature fusion network. Sensors 20, 1142. doi:10.3390/s20041142.

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., Min, H., 2022b. A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sensing 14, 5423. doi:10.3390/rs14215423.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C., 2019. ESRGAN: Enhanced super-resolution generative adversarial networks, in: Computer Vision — ECCV 2018 Workshops, pp. 63–79. doi:10.1007/978-3-030-11021-5_5.

Wang, Z., Bovik, A., 2002. A universal image quality index. IEEE Signal Processing Letters 9, 81–84. doi:10.1109/97.995823.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612. doi:10.1109/TIP.2003.819861.

Wang, Z., Chen, J., Hoi, S.C.H., 2021b. Deep learning for image super-resolution: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 3365–3387. doi:10.1109/TPAMI.2020.2982166.

Weber, S., Uzu, G., Favez, O., Borlaza, L.J.S., Calas, A., Salameh, D., Chevrier, F., Allard, J., Besombes, J.L., Albinet, A., et al., 2021. Source apportionment of atmospheric PM_{10} oxidative potential: synthesis of 15 year-round urban datasets in France. Atmospheric Chemistry and Physics 21, 11353–11378. doi:10.5194/acp-21-11353-2021.

Wei, W., Nie, J., Zhang, L., Zhang, Y., 2022. Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement. IEEE Transactions on Geoscience and Remote Sensing 60, 1–15. doi:10.1109/TGRS.2020.3039534.

World Bank, 2016. The cost of air pollution: strengthening the economic case for action. Technical Report. Washington (DC). URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/781521473177013155.

World Bank, 2021. The world by income and region. Technical Report. Washington (DC). URL: https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.

World Health Organization, 2021. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Technical Report. URL: https://www.who.int/publications/i/item/9789240034228.

Wu, S., Huang, B., Wang, J., He, L., Wang, Z., Yan, Z., Lao, X., Zhang, F., Liu, R., Du, Z., 2021. Spatiotemporal mapping and assessment of daily ground NO_2 concentrations in China using high-resolution TROPOMI retrievals. Environmental Pollution 273, 116456. doi:10.1016/j.envpol.2021.116456.

Xie, Q., Zhou, M., Zhao, Q., Meng, D., Zuo, W., Xu, Z., 2019. Multispectral and hyperspectral image fusion by MS/HS fusion net, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1585–1594. doi:10.1109/CVPR.2019.00168.

Yang, A., Janssen, N.A.H., Brunekreef, B., Cassee, F.R., Hoek, G., Gehring, U., 2016. Children's respiratory health and oxidative potential of $PM_{2.5}$: the PIAMA birth cohort study. Occupational and Environmental Medicine 73, 154–160. doi:10.1136/oemed-2015-103175.

Yang, C.Y., Yang, M.H., 2013. Fast direct super-resolution by simple functions, in: 2013 IEEE International Conference on Computer Vision, pp. 561–568. doi:10.1109/ICCV.2013.75.

Yang, J., Wright, J., Huang, T., Ma, Y., 2008. Image super-resolution as sparse representation of raw image patches, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. doi:10.1109/CVPR.2008.4587647.

Ye, S., Zhao, S., Hu, Y., Xie, C., 2023. Single-image super-resolution challenges: A brief review. Electronics 12, 2975. doi:10.3390/electronics12132975.

Yuhas, R.H., Goetz, A.F.H., Boardman, J.W., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm, in: JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop.

Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., Cohen, A., 2000. Exposure measurement error in

time-series studies of air pollution: concepts and consequences. Environmental Health Perspectives 108, 419–426. doi:10.1289/ehp.00108419.

Zhang, A., Lipton, Z.C., Li, M., Smola, A.J., 2023. Transposed convolution, in: Dive into Deep Learning. Cambridge University Press. URL: https://d2l.ai/chapter_computer-vision/transposed-conv.html.

Zhang, Q., Jiang, X., Tong, D., Davis, S.J., Zhao, H., Geng, G., Feng, T., Zheng, B., Lu, Z., Streets, D.G., et al., 2017. Transboundary health impacts of transported global air pollution and international trade. Nature 543, 705–709. doi:10.1038/nature21712.

Zhang, W., Liu, Y., Dong, C., Qiao, Y., 2022. RankSRGAN: Super resolution generative adversarial networks with learning to rank. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 7149–7166. doi:10.1109/TPAMI.2021.3096327.

Zhang, Y., Chen, H., Chen, X., Deng, Y., Xu, C., Wang, Y., 2021. Data-free knowledge distillation for image super-resolution, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7848–7857. doi:10.1109/CVPR46437.2021.00776.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks, in: Computer Vision — ECCV 2018: 15th European Conference, Proceedings, Part VII, pp. 294–310. doi:10.1007/978-3-030-01234-2_18.

Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y., 2019. Residual non-local attention networks for image restoration, in: 7th International Conference on Learning Representations, ICLR.

Zhao, H., Kong, X., He, J., Qiao, Y., Dong, C., 2020. Efficient image super-resolution using pixel attention, in: 2020 European

Conference on Computer Vision (ECCV), pp. 56–72. doi:10.1007/978-3-030-67070-2_3.

Zhao, N., Wei, Q., Basarab, A., Dobigeon, N., Kouame, D., Tourneret, J.Y., 2016. Fast single image super-resolution using a new analytical solution for L2 - L2 problems. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 25, 3683–3697. doi:10.1109/TIP.2016.2567075.

Zheng, K., Gao, L., Liao, W., Hong, D., Zhang, B., Cui, X., Chanussot, J., 2021. Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution. IEEE Transactions on Geoscience and Remote Sensing 59, 2487–2502. doi:10.1109/TGRS.2020.3006534.

Zhu, F., Zhao, Q., 2019. Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 2453–2460. doi:10.1109/ICCVW.2019.00300.

Zhu, Z., Hou, J., Chen, J., Zeng, H., Zhou, J., 2021. Hyperspectral image super-resolution via deep progressive zero-centric residual learning. IEEE Transactions on Image Processing 30, 1423–1438. doi:10.1109/TIP.2020.3044214.