E-SSRAN: Resolution Enhancement by Merging Multispectral and Hyperspectral Satellite Data

by

Karansinh Atulsinh Padhiar

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

University of Alberta

Abstract

Remote sensing is an effective tool to monitor and assess the dynamics across the Earth's surface. Despite significant technological advancements, there remains a constant demand for high-resolution remote sensing data in spatial, spectral, and temporal contexts. Such high-resolution data is crucial for accurate analysis in diverse fields including agriculture, environmental monitoring, disaster mapping, and many more. However, acquiring such high-resolution remote sensing data is often expensive, with limited geographic coverage and restricted accessibility for general researchers. To address these challenges, our study leverages the power of convolutional neural networks (CNNs) to enhance the resolution of open-source satellite images. We introduce E-SSRAN (Extended Spatial-Spectral Residual Attention Network), an algorithm that integrates the complementary spatial and spectral characteristics of multispectral images (MSIs) and hyperspectral images (HSIs). Our approach consists of two key modules: Spectral-Enhancement and Spatial-Enhancement. The Spectral-Enhancement module uses E-SSRAN to enhance the spectral resolution of MSI by learning the mapping between MSI and HSI, while the Spatial-Enhancement module uses E-SSRAN to improve the spatial resolution of MSI by mapping low-resolution MSI to high-resolution MSI. Additionally, we proposed a comprehensive pipeline to merge the outputs of the Spatial- and Spectral-Enhancement modules, resulting in satellite images with enhanced spatial and spectral resolution. We evaluated the performance of E-SSRAN using MSI and HSI data acquired from Sentinel-2B and DESIS (DLR Earth

Sensing Imaging Spectrometer), respectively, focusing on agricultural areas in Central Alberta. Both qualitative and quantitative assessments demonstrate the effectiveness of E-SSRAN in enhancing the spatial and spectral resolution of low-resolution satellite images. The Spectral-Enhancement module achieved a mean absolute error (MAE) of 0.014, a root mean squared error (RMSE) of 0.018, a spectral angle mapper (SAM) score of 0.144, a universal image quality index (UIQI) score of 0.96, a peak signal-to-noise ratio (PSNR) of 37 dB, and a structural similarity index measure (SSIM) of 0.95. The Spatial-Enhancement module successfully improved the spatial resolution of the 20 m and 60 m ground sampling distance (GSD) Sentinel-2B bands to 10 m GSD. Furthermore, the 10 m GSD MSI produced by the Spatial-Enhancement module was subsequently processed through the Spectral-Enhancement module, generating 10 m GSD HSI. This integrated approach produces satellite images with high spatial and spectral resolution, demonstrating that E-SSRAN significantly improves the quality of satellite images by optimizing the trade-off between spatial and spectral resolutions. Our method has the potential to add substantial value to various applications in agriculture and beyond, by providing more detailed and accurate insights for satellite data analysis.

Preface

Parts of this thesis have been submitted and are under review for publication in the Springer Advances in Science, Technology and Innovation (ASTI) series, managed by the IEREK Interdisciplinary Series for Sustainable Development.

Hard work, willpower and dedication.
For a person with these qualities, the sky is the limit.

– Milkha Singh, Indian track and field sprinter.

Acknowledgements

Completing this thesis marks a significant milestone in my master's studies at the University of Alberta. This achievement would not have been possible without the support and belief of many individuals, to whom I am deeply grateful.

I express my deepest gratitude to my supervisor, Dr. Irene Cheng, for her exemplary guidance and invaluable suggestions. Her trust and support were crucial in guiding me through pivotal moments and making the right decisions. It has been an honor to work under her mentorship.

I am also grateful for the financial support provided by the University of Alberta and the Mitacs Globalink Graduate Fellowship, which enabled me to pursue my research without interruption.

I extend my sincere thanks to my thesis examination committee, Dr. Anup Basu and Dr. Jun Jin for their insightful feedback and encouragement, which played a key role in the successful completion of my thesis.

A heartfelt thank you goes to my friends and colleagues for all the engaging and thought-provoking discussions we shared, which inspired me to think beyond the obvious.

I am profoundly grateful to my parents for their constant support, love, and encouragement throughout this journey.

Lastly, I would like to thank everyone who has, directly or indirectly, contributed to this endeavor. Your support and contributions have been invaluable.

Sincerely yours,
Karansinh Atulsinh Padhiar

Contents

List of Tables

List of Figures

Acronyms

2D 2-Dimensional

ALI Advanced Land Imager

CNN Convolutional Neural Network

DESIS DLR Earth Sensing Imaging Spectrometer

DLR German Aerospace Center

DN Digital Number

E-SSRAN Extended Spatial-Spectral Residual Attention Network

EM Electromagnetic

EO-1 Earth Observing One

EOS Earth Observing System

ESA European Space Agency

GAN Generative Adversarial Network

GB Gigabyte

GPU Graphics Processing Unit

GSD Ground Sampling Distance

HSI Hyperspectral Image

ISS International Space Station

LLE Locally Linear Embedding

LR Learning Rate

MAE Mean Absolute Error

MAP Maximum a Posteriori

MB Megabyte

ML Machine Learning

MODIS Moderate Resolution Imaging Spectroradiometer

MSE Mean Squared Error

MSI Multispectral Image

MTF Modulation Transfer Function

MUSES Multi-User System for Earth Sensing

OLI Operational Land Imager

PCA Principal Component Analysis

PSNR Peak Signal-to-Noise Ratio

ReLU Rectified Linear Unit

RGB Red-Green-Blue

RMSE Root Mean Squared Error

SAM Spectral Angle Mapper

SGD Stochastic Gradient Descent

SSIM Structural Similarity Index Measure

SSRAN Spatial-Spectral Residual Attention Network

SWIR Shortwave Infrared

TBE Teledyne Brown Engineering

TIR Thermal Infrared

TIRS Thermal Infrared Sensor

UIQI Universal Image Quality Index

VNIR Visible and Near-Infrared

Chapter 1

Introduction

Satellite-based remote sensing has become an indispensable tool for monitoring and assessing the Earth's surface and atmosphere on regional, continental, and global scales [62]. Traditional *in situ* techniques for monitoring the Earth's surface are often labor-intensive, time-consuming, expensive, and sometimes infeasible due to the poor accessibility of certain regions [48]. In contrast, remote sensing technology has garnered significant interest in recent decades due to its ability to provide detailed information about the Earth's surface across the Electromagnetic (EM) spectrum, covering large geographic areas in a cost-effective manner. This capability supports a wide range of applications, including environmental monitoring, agricultural planning, forestry management, water resources assessment, snow and glacier monitoring, urban development, and disaster monitoring and mitigation.

1.1 Multispectral and Hyperspectral Imaging

Multispectral and hyperspectral imaging are two prevalent types of remote sensing data. Multispectral sensors capture data in several discrete, broad wavelength bands, typically ranging from 3 to 36 bands, across the EM spectrum. Examples include the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments onboard `Landsat-8`, which collectively record 11 spectral bands in the Visible and Near-Infrared (VNIR), Thermal Infrared (TIR), and Shortwave Infrared (SWIR) portions of the EM spectrum. The `Sentinel-2 mission`, consisting of Sentinel-2A and Sentinel-2B

satellites, captures 13 spectral bands in the VNIR and SWIR spectral ranges. Other multispectral instruments include the `Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)` on the Terra platform of NASA's (National Aeronautics and Space Administration) Earth Observing System (EOS) that captures 14 spectral bands, and the `Moderate Resolution Imaging Spectroradiometer (MODIS)` instrument aboard the Terra and Aqua satellites capturing 36 spectral bands. Multispectral sensors are relatively cost-effective, making them suitable for a wide variety of applications, including land cover classification [56], [40], vegetation monitoring [1], precision agriculture [25], [81], water quality assessment [2], [77], soil moisture content monitoring [36] and many others.

Hyperspectral sensors, in contrast, record data in hundreds of narrow, contiguous wavelength bands across the EM spectrum [32]. This finer spectral resolution enables the identification of materials and substances based on their unique biochemical and biophysical characteristics, which otherwise are indistinguishable from multispectral sensors [58], [41]. For instance, the `Hyperion` instrument onboard the Earth Observing One (EO-1) satellite collects 220 spectral channels ranging from 357 nm to 2576 nm at a bandwidth of 10 nm. Other hyperspectral instruments include the DLR Earth Sensing Imaging Spectrometer (DESIS) capturing 235 spectral bands, the `Earth surface Mineral dust source InvesTigation (EMIT)` instrument onboard the International Space Station (ISS) that records 285 bands, and the `Environmental Mapping and Analysis Program (EnMAP)` instrument that captures 228 spectral bands. Due to the abundant spectral information that it offers, hyperspectral imaging is widely used for various applications including mineral mapping and exploration [6], environmental geology [50], vegetation monitoring and mapping [64], [65], [1] and environmental monitoring [52]. However, it often requires domain expertise and greater storage and processing capabilities.

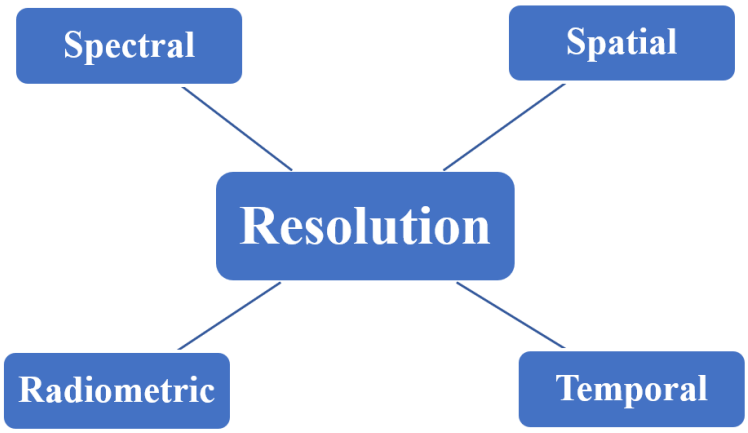

Figure 1.1: Types of resolution in remote sensing.

1.2 Types of Resolution in Remote Sensing

The resolution of satellite images plays a crucial role in determining their applications. It varies based on the satellite's orbit, primary objective, and sensor design. The four primary types of resolution (Figure 1.1) are: radiometric, spectral, spatial, and temporal resolutions [44].

1.2.1 Radiometric Resolution

Radiometric resolution refers to the amount of information each pixel can store, measured in bits. For example, an 8-bit resolution allows each pixel to store up to 256 intensity levels.

1.2.2 Spectral Resolution

Spectral resolution describes a sensor's ability to capture data at finer wavelength bandwidths across the EM spectrum. It determines the number of spectral bands and their width. The narrower the wavelength range (bandwidth) for each band, the higher the spectral resolution. Higher spectral resolution allows for the detection of minor spectral changes, such as vegetation stress or molecular absorption [51]. For instance, panchromatic images consist of a single, broad band of wavelength spanning across a wide range of the EM spectrum. Multispectral sensors typically record data in a limited number of

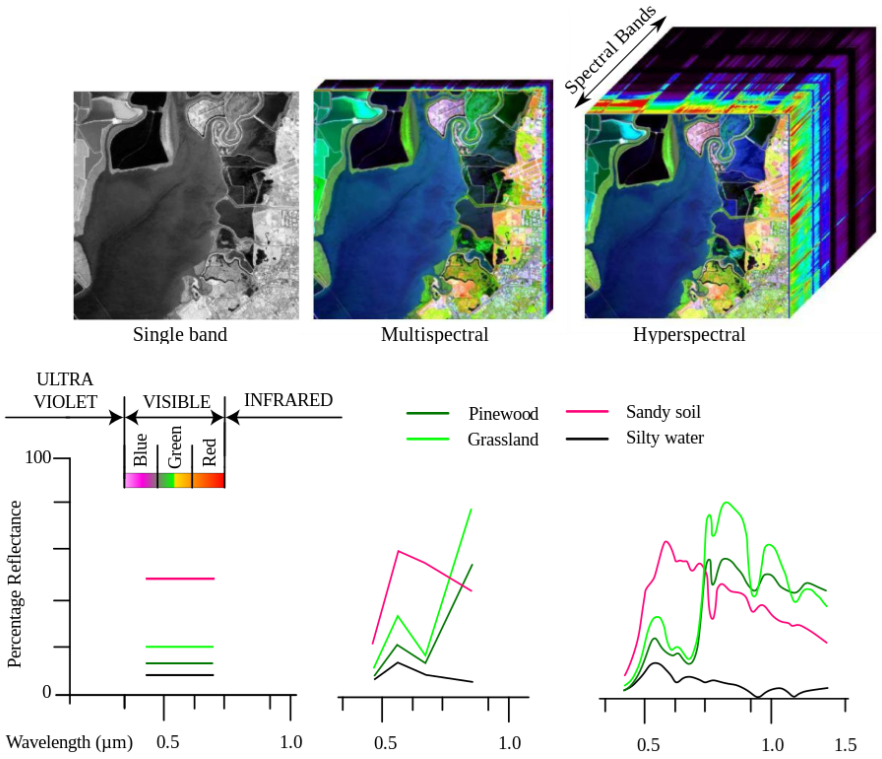

Figure 1.2: A panchromatic (single band) image, a multispectral image, a hyperspectral cube over the same region, and their corresponding spectral signatures (Source).

bands (3-36 bands), whereas hyperspectral sensors capture data across hundreds of spectral bands, together forming a hyperspectral cube, thus offering a higher spectral resolution compared to the panchromatic and multispectral images. Figure 1.2 illustrates a panchromatic image, a multispectral image, and a hyperspectral cube over the same region, along with their corresponding spectral signatures. Since panchromatic and multispectral images have a limited number of bands, their spectral signatures are discrete. In contrast, a hyperspectral cube consists of hundreds of spectral bands, therefore, its spectral signature is a continuous curve offering relatively more information about the observed region. This detailed spectral information can be used to distinguish between different entities including vegetation, water bodies, soil types, minerals, land cover, etc.

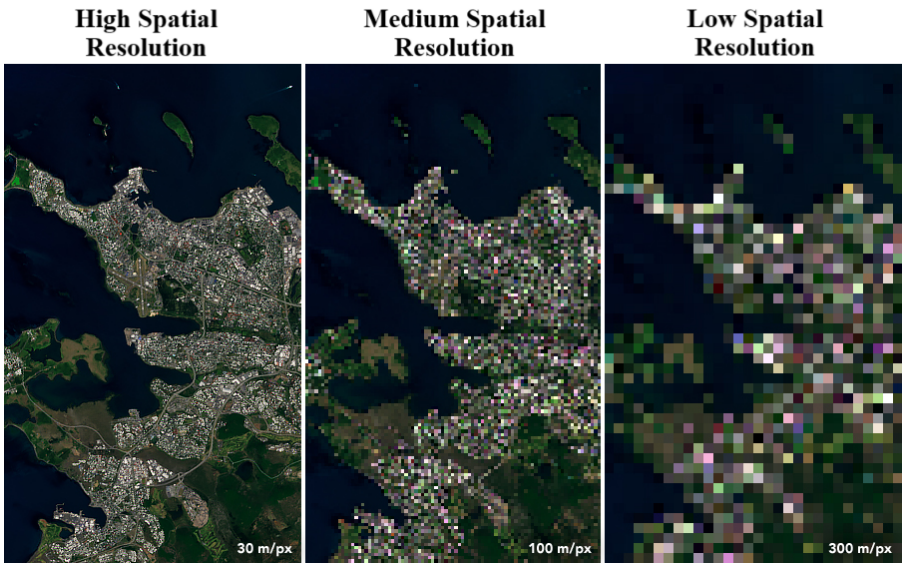

Figure 1.3: Images captured at different spatial resolutions (30 m, 100 m, 300 m). Landsat-8 image capturing Reykjavik, Iceland (date: July 7, 2019) (Credit: NASA Earth Observatory).

1.2.3 Spatial Resolution

Spatial resolution defines the geographic area represented by each pixel in an image. For example, Sentinel-2B and Landsat-8 capture images at a spatial resolution of 30 m, meaning each pixel corresponds to a 30 m $\times$ 30 m geographic area on the ground. Sensors with high spatial resolution can capture finer details of the Earth's surface such as roads, vehicles, small water bodies, etc. [51]. Figure 1.3 compares images captured at different spatial resolutions: 30 m, 100 m, and 300 m. It can be noticed that the image with high spatial resolution (30 m/pixel) contains more details compared to the other counterparts (100 m/pixel and 300 m/pixel).

1.2.4 Temporal Resolution

Temporal resolution refers to the frequency at which a satellite revisits the same location on the Earth's surface. The temporal resolution of a satellite can be influenced by its orbit, swath width, and sensor characteristics. Higher temporal resolution means more frequent revisits. Geostationary satellites, such as the Sentinel-2 series, have a high temporal resolution with a revisit

period of around five days, while polar-orbiting satellites like Landsat-8 have a revisit period of 16 days, resulting in a lower temporal resolution. High temporal resolution (frequent revisit period) is essential for monitoring dynamic changes and short-term phenomena, such as vegetation growth, weather patterns, and natural disasters.

1.3 Trade-offs in Resolution

While it would be preferable to have satellites that capture data with high radiometric, spatial, spectral, and temporal resolutions simultaneously, it is practically challenging due to several constraints. These include limitations in sensor technology, orbital parameters such as satellite altitude and inclination, data processing and transmission constraints, and application-specific requirements. Enhancing one type of resolution often compromises another, leading to various trade-offs.

1.3.1 Spatial-Temporal Trade-off

Acquiring data with high spatial resolution requires a narrower swath width. This, in turn, can increase the satellite's revisit time, resulting in lower temporal resolution. We refer to this as the spatial-temporal trade-off where an increase in the spatial resolution may lead to a decrease in the temporal resolution and vice-versa. For example, the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor captures images of the entire Earth every 1-2 days, but at coarser spatial resolutions (250 m to 1000 m Ground Sampling Distance (GSD)). In contrast, Landsat-8 records data at 30 m spatial resolution but with a revisit period of 16 days.

1.3.2 Spectral-Temporal Trade-off

Recording images at higher spectral resolution generates larger volumes of data, which requires more processing power and bandwidth for transmission, potentially reducing the frequency of data capture (temporal resolution). Hyperspectral sensors like DESIS and EO-1 Hyperion provide detailed spectral

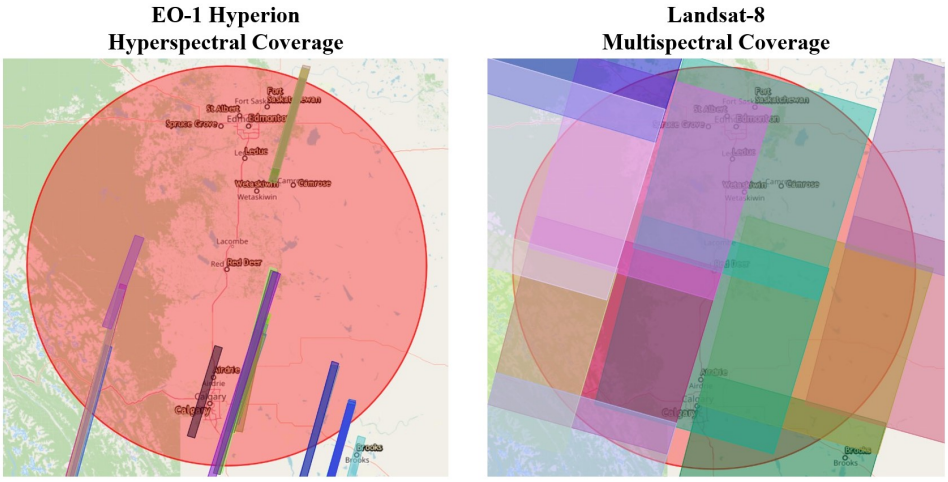

Figure 1.4: Comparing the temporal resolution of EO-1 Hyperion and Landsat-8 satellites over Central Alberta between January 01, 2005, and December 31, 2023 (Credit: USGS EarthExplorer).

information but typically have longer revisit times due to the complexity and volume of data they collect. Conversely, multispectral sensors such as those on Sentinel-2 or the Landsat series offer more frequent revisits (higher temporal resolution) but with fewer and broader spectral bands relative to hyperspectral sensors. This allows multispectral sensors to monitor changes more frequently but with less spectral detail. Figure 1.4 compares the number of Hyperspectral Image (HSI) and Multispectral Image (MSI) recorded by EO-1 Hyperion and Landsat-8 satellites, respectively, over Central Alberta, Canada, between January 01, 2005, and December 31, 2023. The EO-1 Hyperion sensor captured 57 HSI, while the Landsat-8 sensor captured 6,922 MSI during the same period over the same geographic region. This comparison highlights the spectral-temporal trade-off between hyperspectral and multispectral sensors.

1.3.3 Spatial-Spectral Trade-off

Sensors that capture data at high spatial resolution require larger apertures to gather more amount of light and record finer details, whereas the sensors capturing detailed spectral information (high spectral resolution) need complex optics capable of focusing light across a broad range of narrow wavelengths. Multispectral sensors can record data at higher spatial resolution but often suf-

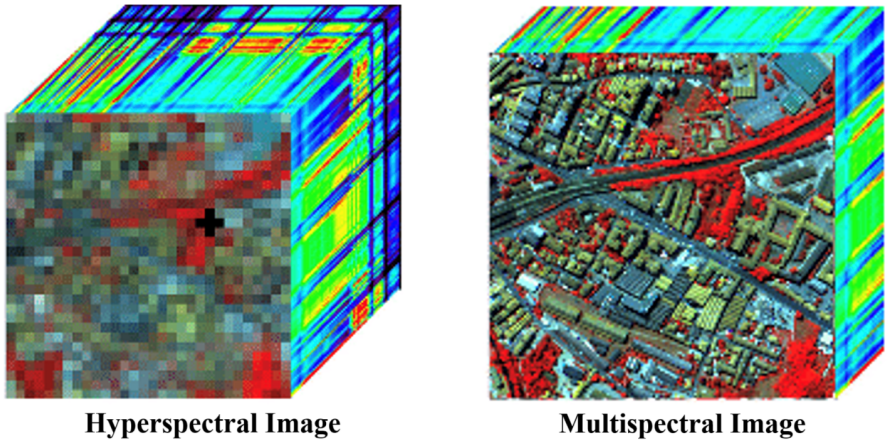

Figure 1.5: Spatial-spectral trade-off between HSI and MSI (Source).

fer from lower spectral resolution as they capture a limited number of spectral bands (3 – 36 bands). In contrast, hyperspectral sensors offer higher spectral resolution, capturing hundreds of spectral bands, but often at lower spatial resolution [72], [38]. This spatial-spectral trade-off between HSI and MSI is illustrated in Figure 1.5. Designing sensors that balance high spatial and spectral resolution is complex and expensive, leading to trade-offs between spatial and spectral resolution while balancing the sensor performance with cost-effectiveness and reliability.

1.4 Contributions

For decades, multispectral and hyperspectral sensors have provided valuable data in spectral, spatial, radiometric, and temporal contexts, significantly enhancing our understanding of surface and atmospheric changes [62]. However, there remains a need for data with enhanced spatial, spectral, and temporal resolutions to deepen our knowledge and explore new applications. A 2019 survey by Wu et al. [72] indicates that improving the spectral, spatial, and temporal capabilities of the Landsat sensor would greatly benefit numerous areas, including ecosystems, agriculture, forestry, disaster response, human health, climate, and water resources monitoring. The users expressed a desire for weekly to sub-weekly cloud-free observations, enhanced spatial resolution (around 10 m GSD), and additional narrow spectral bands at 10 nm band-

width in the VNIR and TIR regions (400–2500 nm), essential for applications in agriculture, forestry, mineral resources, and human health. Ideally, users seek sub-weekly to daily cloud-free observations, a higher spatial resolution of 5 m GSD, and a continuous 10 nm width spectral capability (hyperspectral imaging) over the solar reflective region (400–2500 nm) for applications in geoscience, natural hazards, vegetation characterization, water resource monitoring, cryosphere studies, and land cover mapping.

Moreover, for various applications, it would be beneficial to have remote sensing data at consistent resolutions. However, due to sensor limitations and application-specific requirements, several multispectral satellites record data at different spatial resolutions. For instance, Sentinel-2B records bands at 10 m, 20 m, and 60 m GSD. Similarly, the MODIS instruments capture data in three native spatial resolutions: 250 m, 500 m, and 1000 m GSD; Landsat-8 acquires data at 15 m, 30 m, and 100 m spatial resolutions.

The contributions of this thesis are summarized below.

- We introduce the Extended Spatial-Spectral Residual Attention Network (E-SSRAN) to address the challenges arising from the trade-off between spatial and spectral resolution in satellite images.

- The E-SSRAN architecture effectively integrates the spatial details from MSIs with the rich spectral information from HSIs, resulting in satellite images with both high spatial and spectral resolution.

- Our method consists of two modules:

    1. **Spectral-Enhancement module:** This module employs E-SSRAN to enhance the spectral resolution of MSI by mapping MSI to HSI, resulting in a spectrally enhanced MSI that is equivalent to an HSI.

    2. **Spatial-Enhancement module:** This module uses E-SSRAN to improve the spatial resolution of MSI by mapping low-resolution MSI to high-resolution MSI. Specifically, the network enhances the

20 m and 60 m GSD bands of Sentinel-2B to 10 m GSD, thus providing all 12 Sentinel-2B bands at a uniform 10 m GSD.

- **Integrated Processing Pipeline:** We propose a comprehensive pipeline to merge the outputs of the Spectral- and Spatial-Enhancement modules. In this process, the 10 m GSD MSI generated by the Spatial-Enhancement module is further processed through the Spectral-Enhancement module to produce 10 m GSD HSI. This pipeline ensures the creation of satellite images with both high spatial and spectral resolution.

- **Performance Evaluation:** The proposed E-SSRAN algorithm is evaluated using MSI and HSI data from the Sentinel-2B and DESIS satellites, with a focus on agricultural areas in Central Alberta. The evaluation process includes both qualitative and quantitative assessments:

  - **Quantitative evaluation:** Six widely recognized metrics are used to measure performance: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Spectral Angle Mapper (SAM), Universal Image Quality Index (UIQI), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM).
  - **Qualitative Evaluation:** This involves visually inspecting and comparing the spectral signatures of the predicted satellite images with original satellite images from Sentinel-2B and DESIS.

These contributions demonstrate the potential of E-SSRAN and the proposed pipeline to significantly enhance the spatial and spectral resolution of satellite images, thereby providing valuable improvements for various remote sensing applications.

1.5 Organization of the Thesis

The remainder of the thesis is organized as follows:

- Chapter 2: Literature Review – This chapter provides a review of previous studies, highlighting their advantages and limitations.

- Chapter 3: Dataset Description – This chapter describes the Sentinel-2B MSI and DESIS HSI datasets used for this study and provides an overview of the study area.

- Chapter 4: Spectral-Enhancement module – This chapter focuses on the Spectral-Enhancement module, detailing the data preprocessing steps and the proposed E-SSRAN architecture, along with evaluation metrics and the results.

- Chapter 5: Spatial-Enhancement module – This chapter describes the proposed Spatial-Enhancement module, including the preprocessing steps involved in the pipeline followed by the results.

- Chapter 6: Merging the Spatial- and Spectra-Enhancement modules – This chapter proposes a pipeline to merge the outputs of the Spectral-Enhancement and the Spatial-Enhancement modules to generate satellite images with enhanced spatial and spectral resolution.

- Chapter 7: Conclusion and Future Directions – This chapter presents the conclusions of the study and outlines future research directions.

- Chapter 8: References – This chapter lists the articles, research papers, books, and blogs referenced during the study.

Chapter 2

Literature Review

For understanding the semantics of real-world images, it is crucial to have high-resolution images characterized by sharp, clear boundaries and rich visual descriptions of the observed scene. However, acquiring such images through hardware-based techniques, such as reducing pixel size or increasing sensor size, is often impractical and expensive [46]. Consequently, most of the studies have focused on applying algorithm-based approaches to enhance the resolution of low-resolution images. These methods, commonly referred to as super-resolution techniques, aim to produce high-resolution images from one or more low-resolution observations of the same scene [33]. Prior studies [43], [47] have demonstrated that improving image resolution through super-resolution methods enhances the accuracy and robustness of vision-based applications across various domains, including medical imaging [24], [69], action recognition [45], [57], gait recognition [3], [84], and aerial imaging [82], [83].

Super-resolution methods are traditionally categorized based on the number of low-resolution images used: single-image super-resolution and image fusion techniques [33], [46]. Early single-image super-resolution methods relied on interpolation approaches such as linear, bicubic, and cubic splines interpolation. Although these methods are simple and quick, they tend to produce overly smooth and blurry images as they add little to no additional information. To address this, more advanced reconstruction-based and learning-based methods have been developed. Reconstruction-based approaches utilize certain priors or constraints, such as distributions or energy functions to improve

image quality [13], [4], [26]. On the other hand, learning-based methods use Machine Learning (ML) algorithms to establish relationships between low-resolution and high-resolution patches, which have shown promising results due to their ability to learn from large datasets. However, excessive data can introduce noise and blurring, therefore, it is necessary to maintain a balance between training data size and reconstruction quality.

With the advancements in ML, numerous learning-based models have emerged to tackle the super-resolution problem. These methods are further classified into five categories [33]: neighbor embedding methods [7], sparse coding methods [9], self-exemplar methods [23], locally linear regression methods [76], and deep learning methods [63], [28].

Neighbor embedding methods leverage the similarity in the local geometries between low-resolution and high-resolution patches, computing high-resolution patches as weighted averages of local neighbors. For example, Chang et al. [10] applied Locally Linear Embedding (LLE) [55] to weight learning, by assuming that each sample and its neighbors lie on or near a locally linear patch of the manifold. This idea significantly influenced early coding-based methods.

Sparse coding methods represent image patches as sparse linear combinations of elements from a pre-constructed sparse dictionary. Yang et al. [78], for instance, trained a joint dictionary to find a sparse, over-complete coefficient matrix that clearly described the relationship between low-resolution and high-resolution patches. However, these methods are generally slow due to increased memory usage.

Natural images often exhibit self-similarity, which inspired self-exemplar methods. Glasner et al. [16] proposed a scale space pyramid trained with internal data to map low-resolution pairs to high-resolution. Despite their potential, high computational costs limit the widespread use of these methods.

Locally linear regression methods, introduced by Timofte et al. [66], address the computational speed issues of sparse coding by using multiple lightweight dictionaries instead of a single complex one. Consequently, multiple studies have used supervised ML approaches [59], [49] to replace heavy

sparse coding dictionaries for effectively mapping low-resolution images to high-resolution and reducing computational costs.

Deep learning-based methods can be categorized into Convolutional Neural Network (CNN)-based and Generative Adversarial Network (GAN)-based approaches. These methods use deep neural network architectures to learn the mapping between low-resolution and high-resolution images. With the development of CNNs, super-resolution techniques have become very popular in signal and image processing. This has led to an increase in the development of deep learning-based single-image super-resolution techniques in the remote sensing domain, focusing on the spectral or spatial statistical distributions of a low-resolution image to generate a high-resolution image [26], [35], [68]. For instance, Lanaras et al. [31] proposed DSen2, a network architecture designed to effectively super-resolve the lower-resolution (20 m and 60 m GSD) Sentinel-2B bands to 10 m GSD. Zheng et al. [87] proposed a novel network architecture named Spatial-Spectral Residual Attention Network (SSRAN) that simultaneously explored the spatial and spectral characteristics of MSI to reconstruct an HSI.

Compared to single-image super-resolution techniques, the increased availability of satellite images from multiple Earth observation satellites over the last few decades has led to an increase in the development of image fusion techniques for image enhancement. For instance, HSIs have been widely used for various applications in the remote sensing domain for accurate identification and classification of ground-based objects owing to their ability to capture abundant spectral information. However, they often suffer from low spatial resolution compared to multispectral and panchromatic images. Therefore, various studies have focused on enhancing the spatial resolution of HSIs using image fusion techniques.

Image fusion has emerged as a valuable method for combining the distinct characteristics of high spectral and spatial resolutions of hyperspectral and multispectral/panchromatic images. Hyperspectral image super-resolution using an auxiliary image has gained significant interest in the remote sensing community [37]. One of the commonly addressed techniques, namely, hyper-

sharpening, involves the fusion of hyperspectral and high-resolution multispectral images to enhance the spatial resolution of HSI [60]. Maximum a Posteriori (MAP) estimation [19], one of the earliest hyper-sharpening techniques, is a Bayesian model-based algorithm that can work with any number of spectral bands in the primary hyperspectral and auxiliary (multispectral or panchromatic) images. For example, Yang et al. [79] propose a method to fuse HSI and MSI using a deep CNN with two branches for extracting spectral and spatial features. Dian et al. [14] introduced a deep CNN-based residual learning approach to learn the image priors essential for estimating a high-resolution HSI. Zheng et al. [86] presented an unsupervised CNN-based fusion approach with three coupled autoencoder networks for spectral unmixing. Qu et al. [54] and Xie et al. [73] performed hyperspectral-multispectral fusion using networks based on a spectral mixture model.

Compared to multispectral and hyperspectral images, panchromatic images offer much higher spatial resolution, capturing more details of the ground objects. For instance, panchromatic sensors like the Advanced Land Imager (ALI) and the Landsat Enhanced Thematic Mapper record panchromatic images at a pixel resolution of 15 m GSD, whereas WorldView and QuickBird achieve sub-meter spatial resolutions. As a result, the fusion of hyperspectral and panchromatic images, known as hyperspectral pan-sharpening, has shown significant performance for various remote sensing applications. Several recent studies have focused on hyperspectral pan-sharpening. For instance, Qu et al. [53] proposed a detail-injection network combined with a guided filter. Li et al. [34] proposed a hyperspectral pan-sharpening algorithm employing improved Principal Component Analysis (PCA) and an optimal weighted fusion strategy. He et al. conducted two notable studies related to hyperspectral pan-sharpening and produced remarkable results: (1) HyPNN [22], a spectrally predictive CNN, and (2) HSpeNet [21], a spectral fidelity CNN with multiple skip connections. Xie et al. [75] proposed a CNN performing deep residual mapping based on the idea of deep priors. Zheng et al. [89] adopted a deep prior and dual attention residual network for adaptively learning spectral and spatial features simultaneously.

Despite capturing images at higher spatial resolution, panchromatic sensors often contain limited spectral information, making it challenging to estimate the spectra of HSIs. Additionally, HSIs sometimes have spectral bands in wavelength ranges that are not commonly covered by panchromatic sensors. These issues have restricted the direct fusion of low-resolution HSIs with high-resolution panchromatic images. To overcome these limitations, recent studies employed a method that synthesizes an intermediate, latent MSI from the hyperspectral data cube rather than directly fusing hyperspectral and panchromatic images. For instance, Xie et al. [74] propose a 3-D generative adversarial network-based hyperspectral pan-sharpening framework. Lu et al. [38] introduced a novel technique named CCNNF, that combines hyper-sharpening and multispectral image pan-sharpening based on a two-stage cascaded CNN.

Overall, research in satellite image enhancement has largely focused on improving either the spectral or spatial resolution independently. For instance, previous work has often concentrated on improving the spatial resolution of HSIs or enhancing the spectral resolution of MSIs. However, attempts to simultaneously improve both spatial and spectral resolution have been limited, primarily due to the unavailability of suitable data.

Our study is motivated by the growing need for remote sensing data with both high spatial and spectral resolution. Such data can offer substantial value to various remote sensing applications by providing more detailed and accurate insights for satellite data analysis [17]. In this article, we aim to develop a method to effectively enhance both spatial and spectral resolution of satellite images.

Our work is inspired by the SSRAN architecture [87] because of its ability to exploit both spatial and spectral characteristics of satellite images. Although SSRAN was designed for enhancing the spectral resolution of MSIs, we extend its capabilities towards enhancing both spatial and spectral resolution of satellite images.

Chapter 3

Dataset Description

3.1 Multispectral Images: Sentinel-2B

The Sentinel-2 mission, managed by the European Space Agency (ESA) in collaboration with the European Commission's Copernicus Programme, encompasses two satellites: Sentinel-2A and Sentinel-2B, launched on June 23, 2015, and March 7, 2017, respectively. These satellites operate in a sun-synchronous orbit, each at a 180° phase offset from the other, resulting in a five-day revisit interval at the equator. Covering all terrestrial areas except Antarctica, the Sentinel-2 mission has rapidly gained prominence due to its superior spatial, spectral, radiometric, and temporal resolutions, revolutionizing Earth observation across multiple disciplines. Applications of Sentinel-2 data range from conventional land-cover mapping to advanced hydrological and environmental studies including hydrology, water resource management, and monitoring dynamic geophysical variables. The high quality, widespread accessibility, and global coverage of Sentinel-2 make it a pivotal tool for present and future Earth observation endeavors, serving as the driving force behind this research.

For our study, we leverage MSIs from Sentinel-2B, comprising 13 spectral bands, each characterized by distinct spectral properties and offering spatial resolutions of 10 m, 20 m, or 60 m GSD as detailed in Table 3.1. These spectral bands support various applications including land-cover monitoring, agriculture, forestry, biophysical variable mapping, water resource mapping, and disaster mapping. Notably, the 10 m and 20 m bands are crucial for high-resolution applications, while the 60 m bands serve specific purposes such as

atmospheric correction and cloud estimation. In practice, these bands are initially captured at 20 m GSD and then downsampled to 60 m to enhance the signal-to-noise ratio [31].

The registration errors between the Sentinel-2B bands are minimal and generally ignored in previous studies [12]. Moreover, Band B10, intended for cirrus cloud detection, often exhibits inferior radiometric quality and is prone to striping artifacts. As a result, it is omitted from various quality control processes, hence we exclude it.

Sentinel-2 data can be downloaded for free from the `Copernicus Services Data Hub`. For processing, we use the Level-2A product, which provides atmospherically corrected surface reflectance images. Atmospheric correction addresses the correction of Rayleigh scattering, the absorption and scattering effects of atmospheric gases (ozone, oxygen, and water vapor), and aerosol particles. Each Level-2A product consists of 110 $\times$ 110 km^2 size tiles in UTM/WGS84 projection. Further preprocessing details are discussed in Section 4.1.

Spectral Bands	**Spectral Range (nm)**	**Spatial Resolution (m)**
Band 1 – (Coastal/Aerosol)	433 – 453	60
Band 2 – (Blue)	458 – 523	10
Band 3 – (Green)	543 – 578	10
Band 4 – (Red)	650 – 680	10
Band 5 – (Red edge 1)	698 – 713	20
Band 6 – (Red edge 2)	733 – 748	20
Band 7 – (Red edge 3)	773 – 793	20
Band 8 – (NIR)	785 – 900	10
Band 8A – (NIRn)	855 – 875	20
Band 9 – (WV)	935 – 955	60
Band 11 – (SWIR-1)	1565 – 1655	20
Band 12 – (SWIR-2)	2100 – 2280	20

Table 3.1: Sentinel-2B bands used for this study.

3.2 Hyperspectral Images: DESIS

The DLR Earth Sensing Imaging Spectrometer (DESIS) [30] is a push broom hyperspectral imaging spectrometer stationed on the Multi-User System for Earth Sensing (MUSES) platform aboard the ISS. Developed collaboratively by the German Aerospace Center (DLR) and Teledyne Brown Engineering (TBE), DESIS was launched on June 29, 2018, to provide detailed images of the Earth's surface for various applications including fire detection, change detection, maritime domain awareness, and atmospheric research.

DESIS operates within the VNIR spectral range (400 nm to 1000 nm), capturing 235 channels with a minimum spectral resolution of approximately 2.55 nm and a signal-to-noise ratio of 195 dB. At the nadir view and a reference altitude of 400 km, DESIS acquires imagery at a GSD of $\sim$30 m, delivered in the form of tiles measuring around 1024×1024 pixels (30×30 km^2 geographic area). This translates to a swath width of 30 km. Operating within the ISS orbit inclination of 51.6°, DESIS scans approximately 90% of the inhabited Earth every 3-5 days.

DESIS data is available free of cost for scientific purposes and can be downloaded from the `EOWEB GeoPortal (EGP)` data hub. For our study, we use the Level-2A Surface Reflectance product that offers high-quality, analysis-ready atmospherically corrected surface reflectance images.

3.3 Area of Study

Our study area encompasses the town of Olds, located in the Mountain View County of central Alberta, Canada. Olds is a key agriculture hub, known for its fertile soils and rich agricultural heritage. The landscape includes open prairies, river valleys, and cultivated farmland, supporting a diverse range of crops.

Olds and its surrounding area feature a spectrum of soil types, including the nutrient-rich Chernozem and clay-rich Luvisol. These soil compositions offer optimal conditions for agriculture, facilitating the growth of various crops

vital to the region's agricultural output. Favorable weather conditions, including ample sunshine and adequate precipitation, further enhance agricultural productivity. As a result, a variety of staple crops are grown in the region. Wheat and barley serve as primary cereal crops, while canola, renowned for its oil content, occupies a prominent position in the region's oilseed production. Oats, valued for their versatility in both human consumption and livestock feed, also feature prominently in the agricultural landscape of Olds. This diverse agricultural setting provides an ideal context for our study, allowing us to leverage the capabilities of multispectral and hyperspectral imagery to monitor and analyze agricultural activities in this region.

Chapter 4

Spectral-Enhancement Module

The primary objective of the Spectral-Enhancement module is to improve the spectral resolution of MSI such that the output of this module is a spectrally enhanced MSI that is equivalent to an HSI. In Section 4.1, we briefly describe the data used in this study and outline the preprocessing steps involved in preparing the data for the Spectral-Enhancement module. In Section 4.2, we propose the network architecture used for enhancing the spectral resolution of MSIs. Finally, Section 4.3 explains the evaluation metrics used in the study, while Section 4.4 presents the results achieved by our approach.

4.1 Data Preprocessing

Data preprocessing is a very crucial step in any remote sensing application. It ensures the dataset is free from any kind of noise, anomalies or missing values. Preprocessing is particularly essential because satellite data often suffer from various challenges, such as atmospheric interference, sensor noise, and inconsistent lighting conditions, which can degrade image quality and compromise the accuracy of subsequent analysis. Moreover, raw satellite images may contain missing or corrupted data, which need to be addressed to avoid incorrect results. Therefore, preprocessing is necessary to standardize the data, eliminate inconsistencies, and enhance the overall quality of remote sensing data.

For enhancing the spectral resolution of MSI, we utilized HSI and MSI obtained from DESIS and Sentinel-2B satellites, respectively, captured on August 4, 2020, and covering the same geographic region centered on the town

of Olds. This alignment was crucial to minimize atmospheric discrepancies and ensure consistency in the dataset. The images contained a variety of landscapes including vegetation, bare soil, water bodies, and build-up areas.

Atmospheric correction is a crucial preprocessing step for remote sensing data, aimed at removing atmospheric interferences, such as scattering and absorption of light, to reveal surface features accurately. However, since both DESIS and Sentinel-2B images were Level-2A atmospherically corrected products, this step was omitted. Further, the DESIS HSI was converted from the Digital Number (DN) to surface reflectance values using Equation (4.1).

$$L_{i,j,B} = G_B * DN_{i,j,B} + O_B \tag{4.1}$$

where $L_{i,j,B}$ denotes the surface reflectance value for the pixel (i, j) in the band B; G_B is the band-specific multiplicative factor; $DN_{i,j,B}$ represents the DN value; and O_B is the band-specific offset value. The surface reflectance values represent the amount of light reflected by the Earth's surface without any interference from the atmosphere. Following this, a visual inspection was conducted to identify and exclude regions affected by noise, cloud cover, shadows, and snow, thus, ensuring data integrity and reliability. Furthermore, regions characterized by undefined pixels (black background) were excluded from the study to maintain analytical precision.

Since the main objective of the Spectral-Enhancement module is to enhance the spectral resolution of MSI, both MSI and HSI must have the same geographic extent and GSD. Therefore, we performed geographic alignment, region clipping, and spatial resampling on the images to ensure a pixel-to-pixel correspondence between the MSI and HSI. This was necessary as MSI and HSI were gathered from two different satellite systems. Geographic alignment ensured that both datasets contained the same geographic region and followed the same coordinate reference system. Region clipping retained only the overlapping areas between the MSI and HSI, and spatial resampling was performed to maintain a uniform spatial resolution across the datasets.

As discussed in Section 3.1, Sentinel-2B bands have a varying GSD – 10 m,

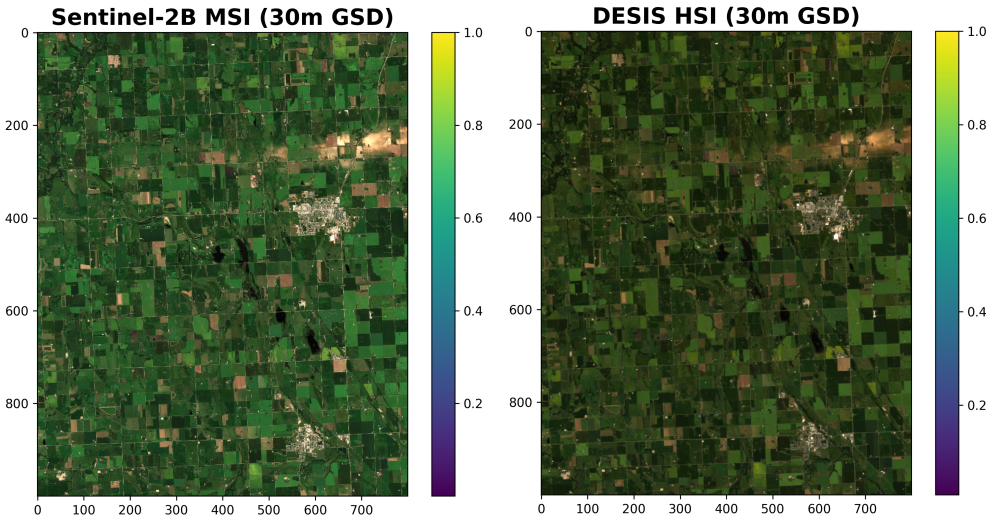

Figure 4.1: True color (RGB) image of Sentinel-2B MSI (R- Band4, G- Band3, B- Band2) and DESIS HSI (R- Band90, G- Band52, B- Band25) attained after data preprocessing.

20 m, and 60 m. To match with the 30 m GSD of DESIS data, the 10 m and 20 m GSD bands of Sentinel-2B were downsampled, and the 60 m GSD bands were upsampled to 30 m GSD using bilinear interpolation. The resultant HSI and MSI had a dimension of 1000 $\times$ 800 pixels each, containing 235 and 12 bands, respectively. The true color images of the resultant MSI and HSI are shown in Figure 4.1.

An entire satellite image is usually very large to fit into computer memory for training and testing purposes. Therefore, the HSI and MSI pair were divided into non-overlapping tiles of size 200 $\times$ 200 pixels. Then, the HSI-MSI tiles were split randomly into training and testing sets at a 70:30 ratio, with 10% of the training set designated for validation. Following this, the tiles in the training and the validation sets were separately subdivided into overlapping patches of size 50 $\times$ 50 pixels with a stride length of 25 $\times$ 25 pixels, resulting in a 50% overlap. The tiles in the testing set were subdivided into non-overlapping patches of size 50 $\times$ 50 pixels. This strategic approach ensured that there was no geographic overlap between the training, validation, and testing sets, thus enhancing the robustness and generalizability of the analysis. Further detailed explanations regarding the patch size rationale, loss

function, etc. are provided in Section 4.4.

4.2 Method

Inspired by Zheng et al. [87], we introduce E-SSRAN (Extended Spatial-Spectral Residual Attention Network) to effectively integrate the spatial and spectral characteristics of MSI and HSI. While the architecture design of E-SSRAN shares similarities with SSRAN [87], key modifications have been made to the loss function and the patch size during training (explained in Section 4.4). Additionally, although SSRAN was initially developed to enhance the spectral resolution of MSI to reconstruct an HSI, E-SSRAN extends these capabilities by enhancing both spatial and spectral resolution of satellite images.

In the Spectral-Enhancement module, E-SSRAN learns the mapping between MSI and HSI by exploiting the spatial and spectral details of the two image modalities and uses this information to reconstruct a spectrally enhanced MSI that is equivalent to an HSI.

Let $X_{patch} \in \mathbb{R}^{p \times p \times B_M}$ and $y_{patch} \in \mathbb{R}^{p \times p \times B_H}$, respectively, represent the MSI and HSI patches with the same GSD. $p \times p$ denotes the patch size; B_M and B_H are the number of bands in MSI and HSI, respectively, such that $B_H >> B_M$. The MSI patches, X_{patch}, are fed into the E-SSRAN architecture for training, while the HSI patches, y_{patch}, serve as the ground truth.

The E-SSRAN architecture consists of three main components – feature extraction, nonlinear mapping, and reconstruction – capable of exploiting both spatial and spectral information of MSI for reconstructing the HSI. During the training phase, E-SSRAN learns the mapping, $\Psi^{Spectral}$, between the MSI and HSI data, where

$$\Psi^{Spectral} : \mathbb{R}^{p \times p \times B_M} \rightarrow \mathbb{R}^{p \times p \times B_M}. \tag{4.2}$$

During the testing phase, the trained model uses the learned MSI-to-HSI mapping to generate a spectrally enhanced MSI that is equivalent to an HSI. Since the primary objective is to enhance the spectral resolution of MSI, the GSD

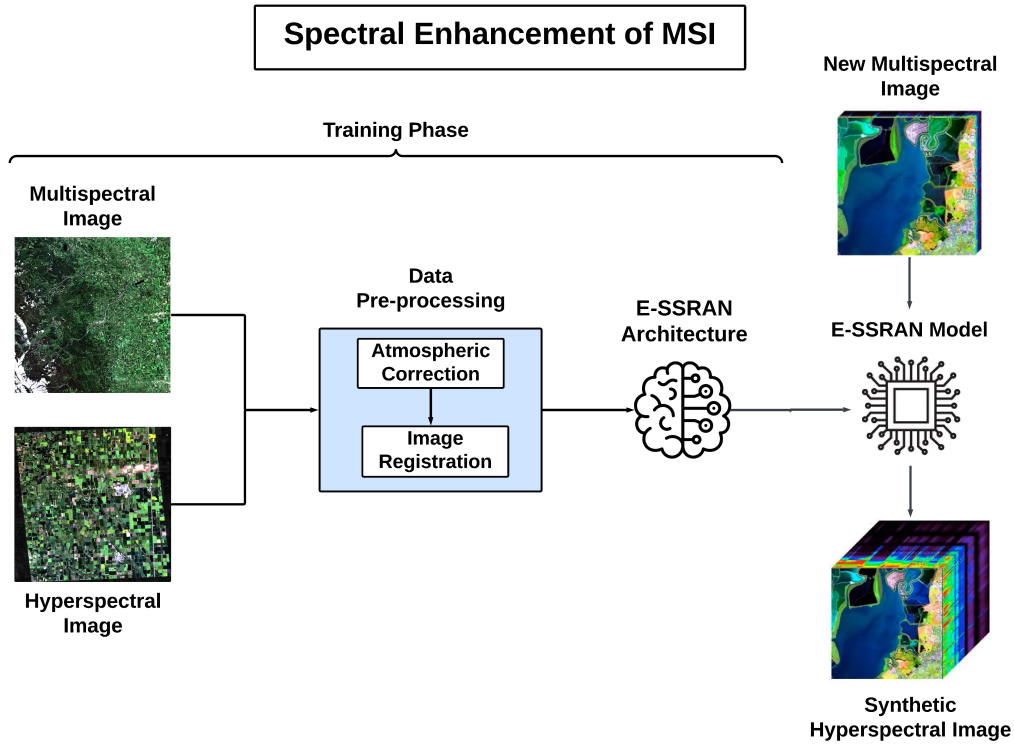

Figure 4.2: Overview of the steps involved in the Spectral-Enhancement module for enhancing the spectral resolution of MSI using E-SSRAN.

and the spatial size of the spectrally enhanced MSI predicted by the model are the same as that of the input MSI patches. Figure 4.2 provides an overview of the steps involved in the Spectral-Enhancement module.

The feature extraction component of the E-SSRAN architecture consists of an input layer that takes the MSI patches (X_{patch}) as input. This is followed by a 2-Dimensional (2D) convolution layer, and a Rectified Linear Unit (ReLU) layer used to extract multispectral features from the input MSI patches. The kernel size of the 2D convolution layer is set to 1×1 for specifically extracting the spectral features. Employing multiple convolutions in this phase, as suggested by Zheng et al. [87], was deemed unnecessary, given the numerous convolutions applied in the subsequent nonlinear mapping component. Therefore, the feature extraction component extracts multispectral features from the input MSI patches.

The nonlinear mapping component aims to map the multispectral features

of the input MSI to hyperspectral features. It consists of several spatial-spectral residual blocks for simultaneously exploiting the spatial and spectral information to learn the MSI-to-HSI feature mapping. Each spatial-spectral residual block consists of a spatial branch and a spectral branch attached in parallel, followed by a concatenation layer and a set of 2D convolution layers with 1×1 kernels and a ReLU layer. The spatial branch extracts spatial features using two 2D convolution layers with 3×3 kernels, each followed by a ReLU layer. The 2D convolution layers with 3×3 kernels offer computational efficiency and have demonstrated effective spatial feature extraction [5], [15]. The spectral branch, on the other hand, focuses exclusively on learning spectral features and consists of two 2D convolution layers with 1×1 kernels, each followed by a ReLU layer. Employing 1×1 kernels in the spectral branch ensures that the 2D convolution layers do not incorporate spatial details from the neighboring pixels. Next, a concatenation layer combines the spatial features from the spatial branch and the spectral features from the spectral branch across the channel dimension. An additional 2D convolution layer with 1×1 kernels and a ReLU layer is added following this to generate the spatial-spectral features. Moreover, a long, additive skip connection is introduced at this point from the beginning of the spatial-spectral residual block. Multiple such spatial-spectral residual blocks are joined in series within the nonlinear mapping component to enhance the nonlinearity of the model.

Lastly, the reconstruction component reconstructs the target HSI patch from the hyperspectral features learned in the nonlinear mapping component. It consists of a 2D convolution layer with 1×1 kernels; the number of kernels in this convolution layer is set to B_H, the desired number of bands in the target HSI output. However, solely relying on a 2D convolution layer for the HSI reconstruction may not entirely preserve the neighboring spectral correlation prior [18], [85] of the resulting HSI patch. The neighboring spectral correlation prior is defined as the high correlation between the adjacent spectral bands in an [39], [88]. Preserving the neighboring spectral correlation prior is essential for enhancing the quality of the reconstructed HSI patch. To address this, a neighboring spectral attention module [87] is introduced in the recon-

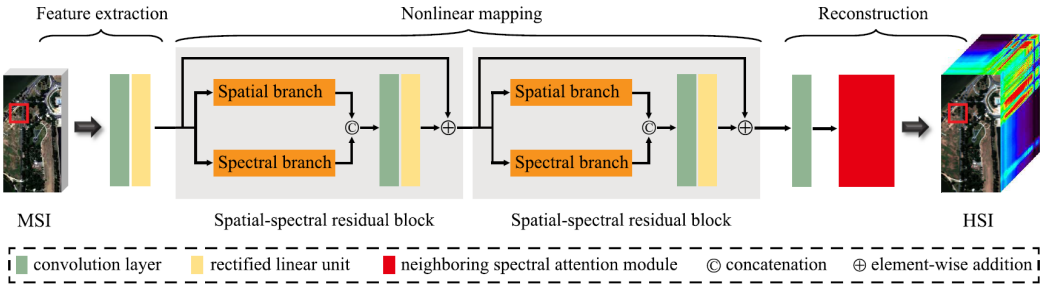

Figure 4.3: Visual representation of the E-SSRAN architecture.

struction component, consisting of a global average pooling operation followed by a 1D convolution layer and a sigmoid activation layer.

Additionally, a neighboring spectral attention module is also integrated into the spatial-spectral residual block of the nonlinear mapping component, after the last set of 1×1 kernel 2D convolution and ReLU layer, to further improve the neighboring spectral band correlation. The reconstruction component outputs spectrally enhanced MSI patches that are equivalent to HSI, $\hat{y}_{patch} \in \mathbb{R}^{p \times p \times B_H}$. y_{patch} has the same GSD as that of the input MSI patches, X_{patch}, but with a spectral resolution of B_H bands. These predicted patches, $\hat{y}_{patch}$, are then compared with the ground-truth HSI patches, y_{patch}, during model evaluation. Figure 4.3 provides a visual representation of the components of the E-SSRAN architecture, inspired by Zheng et al. [87].

4.3 Evaluation Metrics

We carried out an extensive evaluation of the Spectral-Enhancement module proposed for enhancing the spectral resolution of MSI. The quantitative assessment was performed using a combination of six widely used evaluation metrics [8]: MAE, RMSE, SAM [80], UIQI [70], PSNR [27], and SSIM [71]. MAE measures the average absolute differences between predicted and actual image patches. RMSE computes the square root of the average squared differences between predicted and actual patches. A lower value of MAE and RMSE signifies better performance. SAM quantifies the spectral similarity between the predicted and the actual spectra. Each element of the SAM score

is a spectral angle measured in radians within the range [0, 3.142], where a smaller SAM score indicates a strong match between the predicted and actual spectral signatures [42]. UIQI assesses the overall image quality on a scale of [0, 1], where a UIQI score of 1 indicates a perfect match between the predicted and actual images. PSNR evaluates the peak signal quality relative to the noise level with higher values indicating better quality. SSIM measures the structural similarity between images, ranging from 0 to 1, with 1 representing exact structural similarity.

MAE and RMSE are calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{4.4}$$

where n is the total number of bands in the image; Y_i represents the pixel value in the ground truth image Y for the band i; $\hat{Y}_i$ denotes the predicted image pixel value for the band i; and $|.|$ signifies the absolute difference between the ground truth and predicted pixel values.

SAM score (α) can be calculated as:

$$\alpha = cos^{-1}\left[\frac{\sum_{i=1}^{n} t_i r_i}{\sqrt{\sum_{i=1}^{n} t_i^2}\sqrt{\sum_{i=1}^{n} r_i^2}}\right] \tag{4.5}$$

where t and r are the test spectra and reflectance spectra, respectively, and n is the number of bands in the image.

UIQI score can be computed using [70].

PSNR score is calculated as:

$$PSNR = 10log_{10}\left(\frac{R^2}{MSE}\right) \tag{4.6}$$

where R is the maximum fluctuation in the input image data type. For example, if it is an 8-bit unsigned integer data type, R is 255. And Mean Squared Error (MSE) is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{4.7}$$

where n is the number of bands in the image; Y_i represents the pixel value in the ground truth Y for the band i; $\hat{Y}_i$ denotes the predicted image pixel value for the band i.

SSIM is calculated based on three characteristics of an image: luminance, contrast, and structure.

$$Luminance, l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4.8}$$

$$Contrast, c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4.9}$$

$$Structure, s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4.10}$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} are the local means, standard deviations, and cross-covariance, respectively, for the images x, y; C_1, C_2, and C_3 are the regularization constants for luminance, contrast, and structure terms, respectively.

Then, the SSIM score is calculated as:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \tag{4.11}$$

Furthermore, qualitative evaluation is performed by comparing the spectral signatures of the reconstructed HSI ($\hat{y}_{patch}$) with the ground-truth HSI (y_{patch}), and by visual examination of the images.

4.4 Results and Description

This section provides a comprehensive analysis of the parameters used to train the E-SSRAN architecture and evaluates their impact on the spectral enhancement of MSI.

We set the number of kernels to 64 for all 2D convolution layers, except in the reconstruction component where B_H kernels were used in the final 2D convolution layer to reconstruct the spectrally enhanced MSI, transforming it into an HSI. The network weights were initialized using the Xavier weight initialization technique [20] with $L2$ kernel regularization. MAE ($L1$ norm) was used as the loss function, and the Adam optimizer [29] was employed to

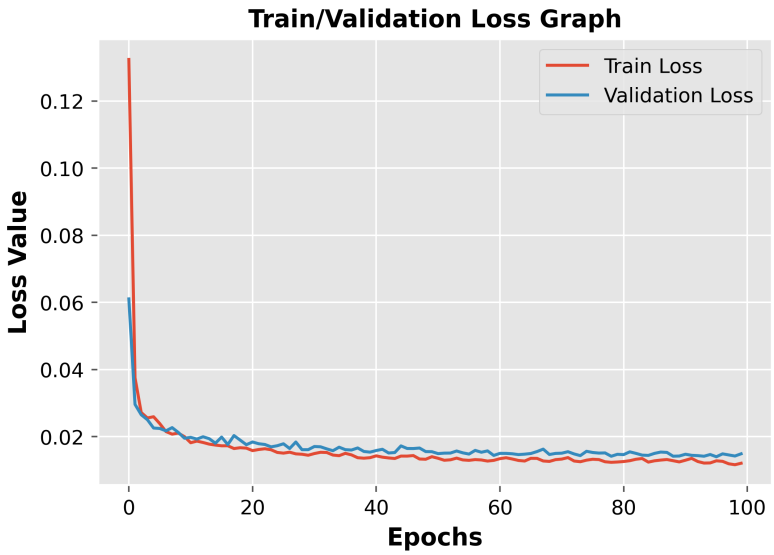

Figure 4.4: Variation in training and validation losses during the training of the E-SSRAN architecture for spectral enhancement of MSI using network $\Psi^{Spectral}$.

minimize the loss. The batch size was set to 64. Both training and validation losses were monitored throughout the training process. Figure 4.4 shows the variation in training and validation losses during the training of the E-SSRAN architecture for spectral enhancement of MSI using network $\Psi^{Spectral}$. The minimal gap between training and validation losses suggested that the model did not overfit.

The base learning rate was initially set to 0.01. If the validation loss did not decrease over eight consecutive epochs, the learning rate was reduced by a factor of 0.1 to aid model convergence. The maximum number of epochs was set to 200, with an early stopping callback to terminate training if no improvement in validation loss was observed over fifteen consecutive epochs. All analyses were conducted using Python 3 [67] and the Keras deep learning API [11]. The model was trained on an NVIDIA GeForce RTX 2080 SUPER Graphics Processing Unit (GPU) with 8 Gigabyte (GB) dedicated memory.

We analyzed various hyperparameters to understand their effects on the performance of the Spectral-Enhancement module. This included different loss functions, patch sizes, the number of spatial-spectral residual blocks in the nonlinear mapping component, the number of neighboring spectral bands

in the neighboring spectral attention module of the reconstruction component, and different base learning rates.

4.4.1 Loss Function

We used MAE ($L1$ norm) as the loss function:

$$Loss_{MAE} = \frac{1}{n} \sum_{i=1}^{n} | Y_i - M_i(X, \theta) | \tag{4.12}$$

where n is the total number of bands in the HSI, Y_i represents the pixel value in the ground truth image Y for the band i, $M_i(X, \theta)$ denotes the predicted HSI pixel value for the band i, and $|.|$ signifies the absolute difference between the ground truth and predicted pixel values.

We compared the performance of MAE with the MSE loss function ($L2$ norm) used in previous studies [87]. Results (Table 4.1) indicated that the $L1$ norm converged faster and yielded superior results compared to the $L2$ norm (MSE), likely because the $L1$ norm is more robust to outliers. Hence, MAE ($L1$ norm) was chosen as the loss function for all the subsequent analyses.

Loss Function	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
MAE	**0.014**	**0.018**	**0.144**	**0.959**	**37.00**	**0.955**
MSE	0.022	0.028	0.260	0.906	32.44	0.908

Table 4.1: Quantitative comparison of E-SSRAN in terms of MAE and MSE loss functions for training the Spectral-Enhancement model.

4.4.2 Patch Size

The spatial dimensions of the MSI and HSI patches were set to 50 × 50 pixels, corresponding to a receptive field of several hundred meters on the ground. This size was sufficient for capturing local low-level texture features and small semantic structures, such as individual farms or small water bodies. In contrast, Zheng et al. [87] proposed a patch size of 4 × 4 pixels for training the SSRAN model, which was considered too small for our study as it limited the CNN's ability to learn spatial details from neighboring pixels.

We conducted a quantitative comparison of various patch sizes: 4 × 4, 20 × 20, 40 × 40, 50 × 50, and 100 × 100 pixels. Figure 4.5 provides a visual

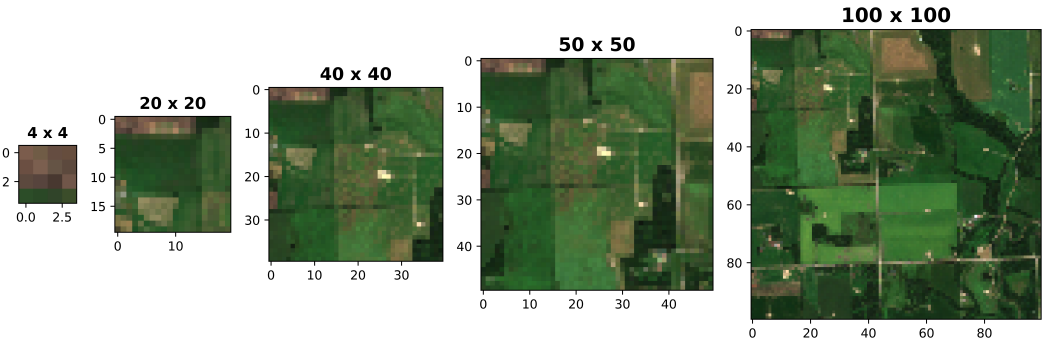

Figure 4.5: Visual representation of different patch sizes: 4 × 4, 20 × 20, 40 × 40, 50 × 50, and 100 × 100 pixels.

representation of different patch sizes. Patch sizes of 4 × 4, 20 × 20 pixels provide very blurry images as compared to the other larger patch sizes. Table 4.2 presents the performance metrics of training the Spectral-Enhancement module for each patch size keeping other parameter choices consistent. The model performance improved with an increase in patch size up to a certain limit, with 50 × 50 pixels demonstrating the best performance for learning spatial and spectral details for spectral enhancement, followed by the 100 × 100 pixels patch size. Thus, our further analysis was conducted using a patch size of 50 × 50 pixels.

Patch Size	**Stride**	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
4 × 4	4	0.016	0.019	0.095	-	38.35	-
20 × 20	10	0.016	0.021	0.151	0.952	36.61	0.938
40 × 40	20	0.016	0.022	0.173	0.952	35.92	0.938
50 × 50	25	**0.014**	0.018	**0.144**	**0.959**	**37.00**	**0.955**
100 × 100	50	0.013	0.018	0.151	0.956	36.82	0.952

Table 4.2: Quantitative comparison of different patch sizes for training the E-SSRAN architecture.

4.4.3 Number of Spatial-Spectral Residual Blocks

Spatial-spectral residual blocks in the nonlinear mapping component of E-SSRAN are used to introduce nonlinearity in the MSI-to-HSI mapping. Too many blocks can increase the number of weight parameters, while too few can result in underfitting. Multiple experiments (reported in Table 4.3) with

different numbers of spatial-spectral residual blocks suggested that five spatial-spectral residual blocks provided the best performance for spectral enhancement. Consequently, we set the number of spatial-spectral residual blocks to five in subsequent experiments.

Num. of Blocks	MAE	RMSE	SAM	UIQI	PSNR	SSIM
1	0.016	0.021	0.168	0.953	35.76	0.940
2	0.016	0.021	0.164	0.953	35.85	0.942
3	0.016	0.022	0.176	0.952	35.56	0.941
4	0.015	0.021	0.163	0.956	36.36	0.944
5	**0.014**	0.018	**0.144**	**0.959**	**37.00**	**0.955**
6	0.015	0.021	0.172	0.953	36.21	0.944
7	0.015	0.020	0.159	0.955	36.45	0.946

Table 4.3: Quantitative comparison of different numbers of spatial-spectral residual blocks in the nonlinear mapping component of E-SSRAN architecture.

4.4.4 Number of Neighboring Spectral Bands

Adjacent bands of an HSI are highly correlated. The primary objective of the neighboring spectral attention module in the reconstruction component of the E-SSRAN architecture is to maintain this high spectral correlation between adjacent bands in the reconstructed HSI. Table 4.4 compares the performance with different number of neighboring spectral bands in the neighboring spectral attention module, indicating that five neighboring bands provided optimal results.

Num. of Bands	MAE	RMSE	SAM	UIQI	PSNR	SSIM
1	0.017	0.022	0.173	0.953	35.62	0.935
3	0.015	0.020	0.160	0.955	36.48	0.947
5	**0.014**	**0.018**	**0.144**	**0.959**	**37.00**	**0.955**
7	0.015	0.020	0.157	0.954	36.43	0.946
9	0.014	0.020	0.155	0.957	36.76	0.948

Table 4.4: Quantitative comparison of different number of spatial-spectral residual blocks in the nonlinear mapping component of E-SSRAN architecture.

4.4.5 Learning Rate

The Learning Rate (LR) is a critical hyperparameter in training neural networks. A low learning rate may result in a very slow training process, while a high learning rate could prevent convergence. Therefore, choosing an optimal learning rate value is very important. For training the E-SSRAN architecture in the Spectral-Enhancement module, we experimented with various base learning rates and listed the quantitative results in Table 4.5. A base learning rate of 0.01 provided the best performance in terms of enhancing the spectral resolution of MSI. Consequently, all subsequent analyses were performed using this base learning rate.

Base LR	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
0.1	0.018	0.025	0.184	0.946	34.60	0.927
0.07	0.017	0.023	0.181	0.950	35.20	0.932
0.01	**0.014**	**0.018**	**0.144**	**0.959**	37.00	**0.955**
0.007	0.015	0.020	0.155	0.957	36.72	0.948
0.001	0.014	0.018	0.156	0.958	37.04	0.952

Table 4.5: Quantitative comparison of different base learning rates for training the E-SSRAN architecture in the Spectral-Enhancement module.

4.4.6 Comparing the performance of E-SSRAN with SSRAN

The E-SSRAN model extends the capabilities of the original SSRAN model [87] by addressing both spectral and spatial enhancement of MSI. Table 4.6 provides a quantitative performance comparison between E-SSRAN and SSRAN.

For training the SSRAN model, we followed the exact configuration described in the original study, including a patch size of 4×4 pixels and three spatial-spectral residual blocks in the nonlinear mapping component of the architecture. The model was trained using the MSE loss function and the Stochastic Gradient Descent (SGD) optimizer. The initial learning rate was set to 0.01 with a decay factor of 0.1 after every fifty epochs. A batch size of 100 was used, and the model was trained for a total of 200 epochs.

In contrast, the E-SSRAN model was trained using a larger patch size of 50 × 50 pixels and incorporated five spatial-spectral residual blocks in the nonlinear mapping component of the architecture. Increasing the patch size allowed the model to learn features at a larger spatial extent, improving its ability to capture spatial dependencies. Additionally, increasing the number of spatial-spectral residual blocks in the nonlinear mapping component of the architecture enhanced the model's complexity and its ability to learn nonlinear dependencies. The MAE loss function, along with the Adam optimizer, was used for training. Compared to MSE, MAE is less sensitive to outliers and typically converges faster. The initial learning rate was set to 0.01, with a decay factor of 0.1 applied when the validation loss plateaued. A batch size of 64 was used, and the model was trained for 100 epochs.

The SSRAN model required a total training time of 48.48 minutes, averaging to 14.5 seconds per epoch. Conversely, the E-SSRAN model required only 5.47 minutes, averaging less than 5 seconds per epoch. The SSRAN architecture consisted of approximately 0.29 million trainable parameters (∼1.10 Megabyte (MB) size), whereas the E-SSRAN architecture consisted of approximately 0.47 million trainable parameters (1.79 MB size). Despite the smaller model size, SSRAN took longer to train compared to E-SSRAN, likely due to the smaller patch size, which increased the number of training samples.

Algorithm	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
SSRAN [87]	0.016	0.018	0.123	-	37.44	-
E-SSRAN	**0.014**	**0.018**	0.144	**0.959**	37.00	0.955

Table 4.6: Comparing the quantitative performance of SSRAN [87] and E-SSRAN.

Furthermore, Figure 4.6 provides a qualitative evaluation between SSRAN and E-SSRAN by comparing the spectral signatures of the predicted HSI generated by both models. The spectral signatures of the HSI predicted by the SSRAN model were non-uniform and noisy compared to those predicted by the E-SSRAN model. This discrepancy is likely due to the small patch size used for training the SSRAN model, which can disturb the spectral correlation between neighboring bands, resulting in non-uniform and noisy spectral signa-

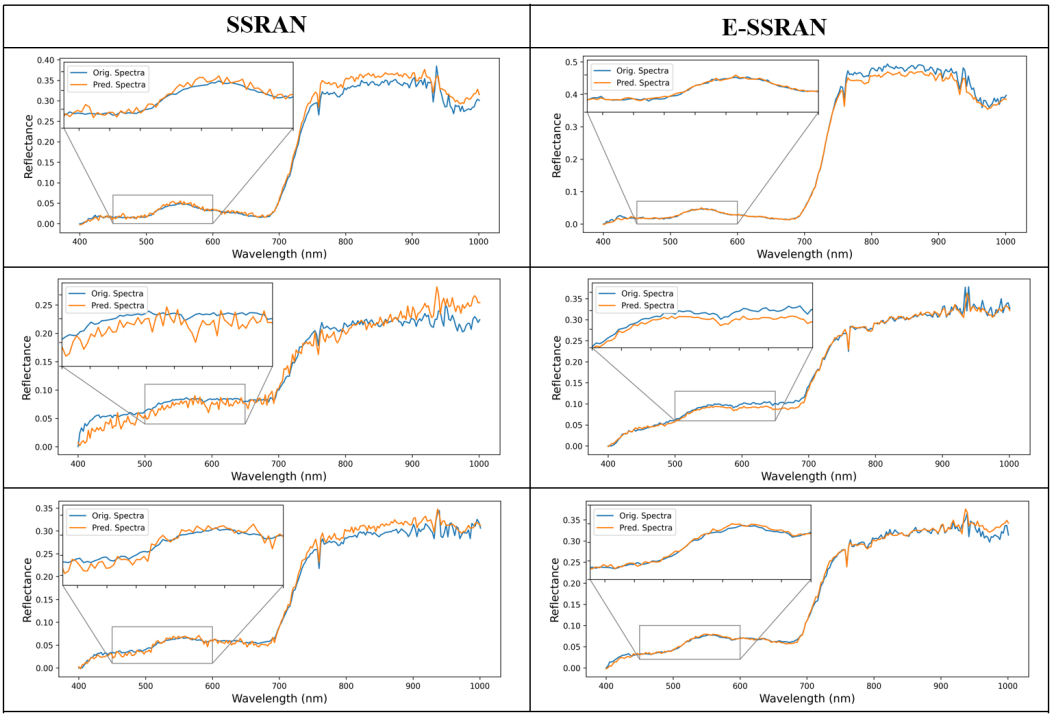

Figure 4.6: Comparing spectral signatures of the predicted HSI using SSRAN and E-SSRAN.

tures. The larger patch size (50 × 50 pixels) used for training the E-SSRAN model ensured better preservation of the spectral correlation.

Figure 4.7 provides a visual comparison of the original DESIS HSI and the HSI predicted by the E-SSRAN model in the Spectral-Enhancement module over different spectral bands (Band-50, Band-100, Band-150, Band-200, and Band-250). It can be observed that the predicted HSI is very close to the original HSI, further justifying the performance of E-SSRAN for enhancing the spectral resolution of MSI.

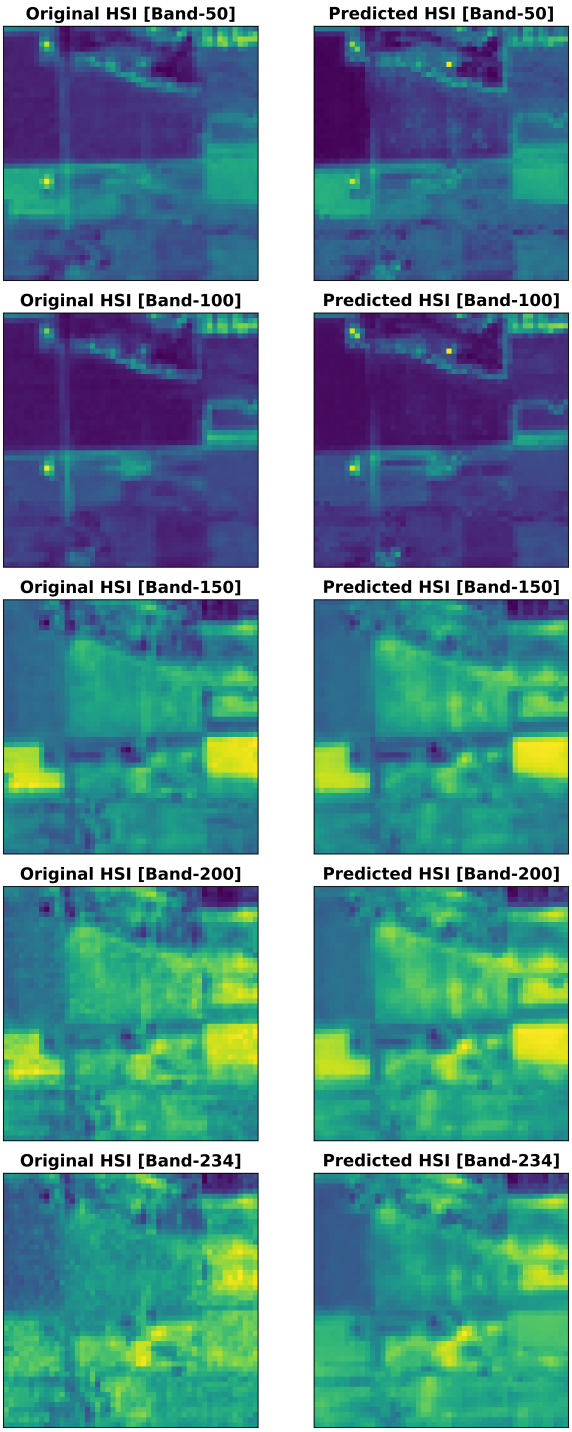

Figure 4.7: Visual comparison of original DESIS HSI and the HSI predicted in the Spectral-Enhancement module over different spectral bands (Band-50, Band-100, Band-150, Band-200, and Band-250).

Chapter 5

Spatial-Enhancement module

As discussed in Section 3.1, the Sentinel-2B satellite provides imagery in bands with varying GSDs of 10 m, 20 m, and 60 m. These differences arise due to limitations in storage and transmission bandwidth, enhanced signal-to-noise ratios, and other application-specific requirements. Figure 5.1 illustrates three sample images captured by the Sentinel-2B satellite at 10 m (left column), 20 m (middle column), and 60 m (right column) GSD. The images with lower resolutions (20 m and 60 m) show significantly fewer spatial details compared to the 10 m band, with the 60 m band being the most affected. This reduction in spatial resolution can restrict the utility of satellite images across various domains. Therefore, it is often desirable to have a uniform high resolution (10 m GSD) across all bands.

Basic methods such as bilinear or bicubic interpolation can increase the spatial resolution of low-resolution images quickly, but they often tend to produce blurry outputs with minimal additional information. In contrast, advanced techniques utilizing CNNs [31] aim to provide superior results by smartly upsampling the images. These methods use the available high-resolution bands to recover spatial details in low-resolution bands while preserving spectral information. The Spatial-Enhancement module seeks to enhance the spatial resolution of the 20 m and 60 m bands to 10 m GSD, maintaining spectral integrity across the bands.

CNNs, as a supervised learning technique, require abundant training data where both multi-resolution inputs and corresponding high-resolution outputs

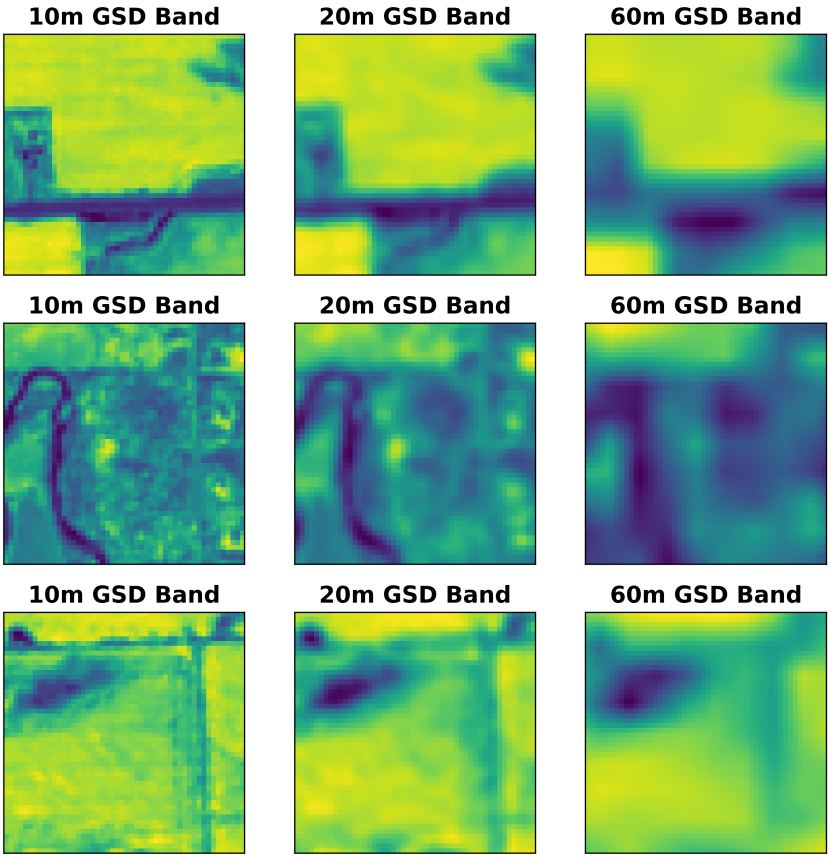

Figure 5.1: Sample images from Sentinel-2B satellite at 10 m (left column), 20 m (middle column), and 60 m (right column) GSD.

are available. A major challenge in CNN-based spatial enhancement is constructing training, validation, and test sets due to the unavailability of ground truth data at 10 m GSD for the 20 m and 60 m bands. Despite considerable efforts, such as using aerial hyperspectral data and advanced simulation technology, synthesizing realistic high-resolution data remains challenging. Consequently, prior studies [31] have relied on the assumption that the transfer of spatial details from high-resolution to low-resolution bands is scale-invariant, relying solely on the relative resolution difference rather than the absolute GSD of the images. This implies that the relationships between bands of different resolutions exhibit self-similarity within a relevant scale range, a concept supported by existing literature on self-similarity in image analysis [61], [16].

For our purposes, this assumption needs to hold only within a limited range of up to a 6× resolution difference, which is less than one order of magnitude. Scale-invariance implies that the relationships between, for example, 20 m → 10 m and 40 m → 20 m spatial enhancements are approximately similar. Therefore, a CNN-based algorithm can be trained for 40 m → 20 m spatial enhancement and then used for 20 m → 10 m enhancement. The same logic applies to 60 m → 10 m and 360 m → 60 m enhancements – a CNN-based algorithm trained for 360 m → 60 m spatial enhancement and then used for 60 m → 10 m enhancement. Due to the absence of 10 m ground truth data for the 20 m and 60 m bands, quantitative assessment must be performed at reduced resolutions, specifically 40 m → 20 m and 360 m → 60 m. This way, the 2× spatial enhancement from 20 m to 10 m GSD can be learned from ground truth images at lower resolutions: 40 m and 20 m GSD. Similarly, the 6× spatial enhancement from 60 m to 10 m GSD can be learned from ground truth images at lower resolutions: 360 m and 60 m GSD. If this invariance holds, the learned spatial-spectral correlations will be accurate. Following this premise, virtually limitless training data can be generated by synthetically downsampling raw Sentinel-2 images.

5.1 Method

To enhance the spatial resolution of Sentinel-2B images, we used the MSI captured on August 4, 2020, centered on the town of Olds. Similar to Section 4.1, the preprocessing steps in the Spatial-Enhancement module included atmospheric correction and visual inspection to exclude noise, cloud cover, shadows, and snow regions. Regions with undefined pixels (black background) were removed from the images. This resulted in a total of 12 Sentinel-2B bands, further categorized based on their GSD into three sets as detailed in Table 5.1.

Set	**GSD**	**Sentinel-2B Bands**
A	10 m	B2, B3, B4, B8
B	20 m	B5, B6, B7, B8a, B11, B12
C	60 m	B1, B9

Table 5.1: Sentinel-2B bands categorized into three sets based on their GSD.

To generate training data, we downsampled the original Sentinel-2B bands at a desired scale factor f. This was achieved by blurring the bands with a Gaussian filter at a standard deviation of $\sigma = 1/f$ pixels, mimicking the Modulation Transfer Function (MTF) of Sentinel-2. Subsequently, the bands were downsampled by averaging over $f \times f$ windows, with $f = 2$ and $f = 6$. Thus, two distinct datasets were generated for 20 m $\rightarrow$ 10 m and 60 m $\rightarrow$ 10 m spatial enhancement. Following this, the MSIs were divided into non-overlapping tiles of size 200 $\times$ 200 pixels and split randomly into training and testing sets at a ratio of 70:30, with an additional 10% of the training set allocated for validation. To ensure geographically distinct regions, the tiles in each of the training, validation, and testing sets were separately subdivided into overlapping patches of size 50 $\times$ 50 pixels, with an overlap of 50%.

The first dataset was derived by downsampling the original 10 m and 20 m bands at a scale factor $f = 2$, resulting in "high-resolution" images at 20 m GSD and "low-resolution" images at 40 m GSD (Figure 5.2). Subsequently, the 40 m bands were upsampled using bilinear interpolation to match the target resolution (20 m) before being passed into the network for training. This

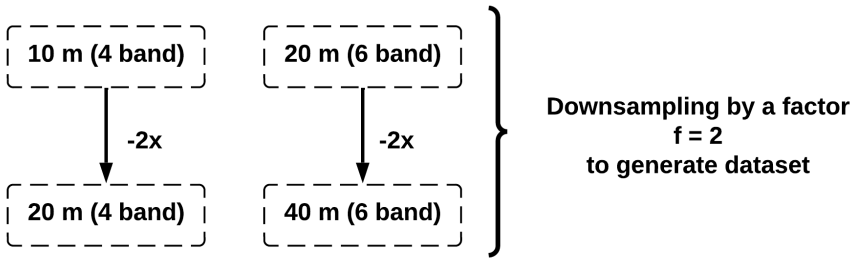

Figure 5.2: Downsampling the original 10 m and 20 m Sentinel-2B bands by a scale factor of $f = 2$

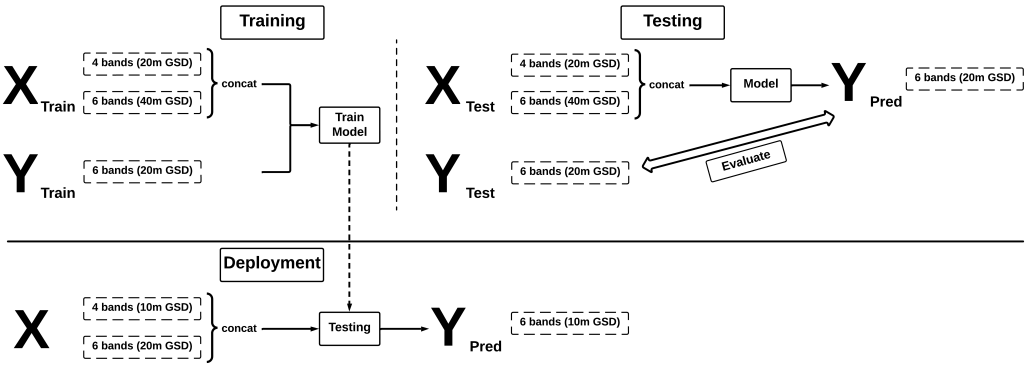

Figure 5.3: An overview of the method used for training the network $\Psi_{2\times}$

dataset, consisting of 20 m and 40 m bands, was used to train the network,

$$\Psi_{2\times} : \mathbb{R}^{W \times H \times 4} \times \mathbb{R}^{W/2 \times H/2 \times 6} \rightarrow \mathbb{R}^{W \times H \times 6}, \tag{5.1}$$

for $2\times$ spatial enhancement (40 m $\rightarrow$ 20 m). Once trained, the network $\Psi_{2\times}$ was applied to the original 20 m bands (set B) to obtain spatially enhanced bands with 10 m GSD (20 m $\rightarrow$ 10 m). Figure 5.3 provides a pictorial representation of this approach.

Similarly, the second dataset included images at 60 m, 120 m, and 360 m GSD, generated by downsampling the original 10 m, 20 m, and 60 m bands, respectively, at a scale factor $f = 6$. The 120 m and 360 m bands were then upsampled using bilinear interpolation to match the target resolution (60 m) before being fed into the network for training. Thus, the 60 m, 120 m, and 360 m bands were used to train the network,

$$\Psi_{6\times} : \mathbb{R}^{W \times H \times 4} \times \mathbb{R}^{W/2 \times H/2 \times 6} \times \mathbb{R}^{W/6 \times H/6 \times 2} \rightarrow \mathbb{R}^{W \times H \times 2}, \tag{5.2}$$

for 6× spatial enhancement (360 m → 60 m). Once trained, the network $\Psi_{6\times}$ was applied to the original 60 m bands (set C) to generate spatially enhanced data with 10 m GSD (60 m → 10 m).

5.2 E-SSRAN Architecture Updates for Spatial Enhancement

To enhance the spatial resolution of Sentinel-2B MSI, we utilized the same E-SSRAN architecture described in Section 4.2. This architecture is effective in mapping low spatial resolution bands to high spatial resolution, making it suitable for both the Spectral-Enhancement and Spatial-Enhancement modules. The E-SSRAN architecture features separate spatial and spectral branches in the nonlinear mapping component, allowing it to learn the mapping between low-resolution and high-resolution MSI. Additionally, the neighboring spectral attention module in the reconstruction component of the E-SSRAN architecture helps preserve the spectral correlation between bands. However, for the Spatial-Enhancement module, minor modifications were made to the input layer and the final 2D convolution layer, while the rest of the architecture remained unchanged.

Let the spatial dimensions of the bands in the set A be denoted by $W \times H$, and the intensities of the bands in sets A, B, and C be represented as $X_A \in \mathbb{R}^{W \times H \times 4}$, $X_B \in \mathbb{R}^{W/2 \times H/2 \times 6}$, and $X_C \in \mathbb{R}^{W/6 \times H/6 \times 2}$, respectively. The network $\Psi_{2\times}$ takes two inputs, X_A and $\bar{X}_B$, where $\bar{X}_B \in \mathbb{R}^{W \times H \times 6}$ is the bilinearly interpolated version of X_B to match the spatial size of X_A. The final 2D convolution layer in the reconstruction component of $\Psi_{2\times}$ contains six kernels to match the number of bands in the set B. The output of the network is the predicted "high-resolution" MSI patches $\bar{y}_B \in \mathbb{R}^{W \times H \times 6}$, with a spatial resolution of 20 m GSD. These are compared with the ground truth MSI patches $y_B \in \mathbb{R}^{W \times H \times 6}$ during model evaluation.

Similarly, the network $\Psi_{6\times}$ takes three inputs, X_A, $\bar{X}_B$, and $\bar{X}_C$. Here, $\bar{X}_B \in \mathbb{R}^{W \times H \times 6}$ and $\bar{X}_C \in \mathbb{R}^{W \times H \times 2}$ are the bilinearly interpolated versions of X_B and X_C, respectively, to match the spatial size of X_A. The final

2D convolution layer in the reconstruction component of $\Psi_{6\times}$ contains two kernels to match the number of bands in the set C. The output is the predicted "high-resolution" MSI patches $\bar{y}_C \in \mathbb{R}^{W \times H \times 2}$, with a spatial resolution of 60 m GSD, which are further compared with the ground truth MSI patches $y_C \in \mathbb{R}^{W \times H \times 2}$ during evaluation.

Once trained and evaluated on the spatially downsampled counterparts, the network $\Psi_{2\times}$ is used for 20 m $\rightarrow$ 10 m spatial enhancement, while the network $\Psi_{6\times}$ is used for 60 m $\rightarrow$ 10 m spatial enhancement, producing spatially enhanced bands at 10 m GSD. Therefore, the Spatial-Enhancement module employs two networks: $\Psi_{2\times}$ for enhancing spatial resolution of the 20 m bands to 10 m GSD, and $\Psi_{6\times}$ for improving the 60 m bands to 10 m GSD. The resultant 10 m bands can only be assessed through visual examination.

5.3 Results and Discussion

This section presents the quantitative and qualitative results of the Spatial-Enhancement module designed to improve the spatial resolution of Sentinel-2B MSI. The results are divided based on the two networks used in the Spatial-Enhancement module: 1) network $\Psi_{2\times}$, used for the 20 m $\rightarrow$ 10 m enhancement, and 2) network $\Psi_{6\times}$, used for the 60 m $\rightarrow$ 10 m spatial enhancement. The model configurations defined in Section 4.4 for the Spectral-Enhancement module remain consistent for training the two networks in the Spatial-Enhancement module. We used the MAE (Equation 4.12) as the loss function along with the Adam optimizer [29]. The base learning rate was set to 0.01, with a batch size of 64. The training was conducted for a maximum of 200 epochs, with early stopping if the validation loss did not decrease for fifteen consecutive epochs.

5.3.1 Network $\Psi_{2\times}$: 20 m $\rightarrow$ 10 m Spatial Enhancement

Due to the absence of 10 m ground truth data for the 20 m Sentinel-2B bands, network $\Psi_{2\times}$ was trained at a lower resolution, specifically 40 m $\rightarrow$ 20 m. Figure 5.4 illustrates the variation in training and validation losses for the E-SSRAN architecture during this spatial enhancement. The minimal gap

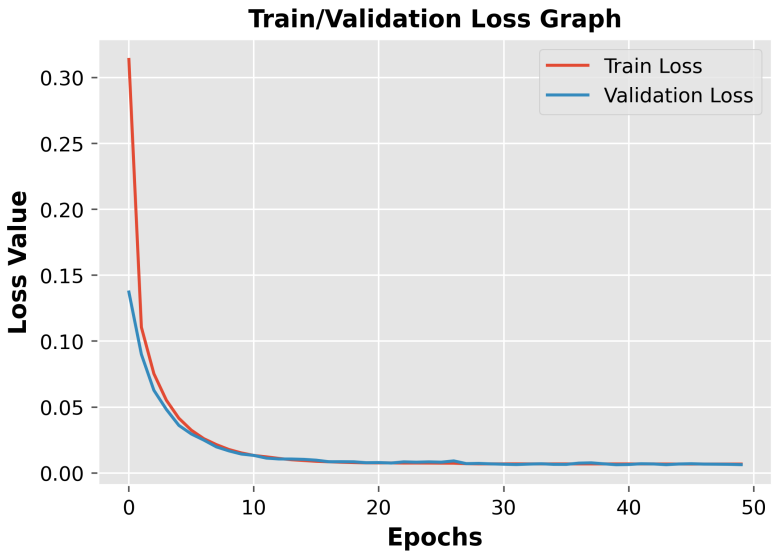

Figure 5.4: Variation in training and validation losses while training network $\Psi_{2\times}$ for 40 m $\rightarrow$ 20 m spatial enhancement.

between training and validation losses indicated that the model did not overfit. The network $\Psi_{2\times}$ successfully learned the mapping between the "low-resolution" 40 m bands and the "high-resolution" 20 m bands for 2$\times$ spatial enhancement.

Table 5.2 presents the quantitative results of network $\Psi_{2\times}$, comparing its performance with the DSen2 network architecture proposed by Lanaras et al. [31]. While DSen2 performs similarly to E-SSRAN in terms of quantitative metrics, E-SSRAN shows slightly better performance. Additionally, E-SSRAN is effective for both Spectral-Enhancement and Spatial-Enhancement modules, unlike DSen2.

Algorithm	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
DSen2 [31]	0.005	0.007	0.038	0.982	43.89	0.953
E-SSRAN	**0.004**	**0.006**	**0.033**	**0.998**	**44.67**	**0.964**

Table 5.2: Comparing the quantitative performance of DSen2 [31] and E-SSRAN for the 40 m $\rightarrow$ 20 m spatial enhancement using network $\Psi_{2\times}$.

Figure 5.5 compares the 20 m bands predicted by network $\Psi_{2\times}$ for the 40 m $\rightarrow$ 20 m spatial enhancement. The left column shows the 40 m "low-resolution" band input, the middle column shows the 20 m "high-resolution"

band predicted by the network (the 40 m $\rightarrow$ 20 m spatial enhancement), and the right column displays the original 20 m band from Sentinel-2B. The 20 m band predicted by the network contains more spatial details compared to the 40 m GSD input and is very similar to the original 20 m Sentinel-2B band. This demonstrates the network's ($\Psi_{2\times}$) ability to enhance the spatial resolution of low-resolution Sentinel-2B bands.

Once trained, the network $\Psi_{2\times}$ can be used for 20 m $\rightarrow$ 10 m spatial enhancement, generating spatially enhanced bands at 10 m GSD that can be evaluated visually. The results are shown in Figure 5.6 compares the spatially enhanced 10 m bands predicted by network $\Psi_{2\times}$ for the 20 m $\rightarrow$ 10 m spatial enhancement. The left column shows the original 20 m band from Sentinel-2B input to the network, and the right column shows its corresponding spatially enhanced 10 m band predicted by the network. The 10 m band predicted by the network contains more spatial details compared to the original 20 m Sentinel-2B band, proving the network's capability to enhance the spatial resolution from 20 m to 10 m GSD.

5.3.2 Network $\Psi_{6\times}$: 60 m $\rightarrow$ 10 m spatial enhancement

Due to the lack of 10 m ground truth data for the 60 m Sentinel-2B bands, network $\Psi_{6\times}$ was trained at lower resolutions, specifically 360 m $\rightarrow$ 60 m. Figure 5.7 shows the variation in training and validation losses during the training process for the E-SSRAN architecture aimed at the 360 m $\rightarrow$ 60 m spatial enhancement using network $\Psi_{6\times}$. The network effectively learns the mapping between the "high-resolution" 60 m bands and the "low-resolution" 360 m bands for 6$\times$ spatial enhancement.

Table 5.3 presents the quantitative performance of network $\Psi_{6\times}$ for the 360 m $\rightarrow$ 60 m spatial enhancement. Additionally, Figure 5.8 provides a visual comparison of the 60 m bands predicted by network $\Psi_{6\times}$ for the 360 m $\rightarrow$ 60 m spatial enhancement. The left column shows the 360 m "low-resolution" band input to the network, the middle column displays the 60 m "high-resolution" band predicted by the network, and the right column shows the original 60 m band from Sentinel-2B (set C). The 60 m band predicted by the network

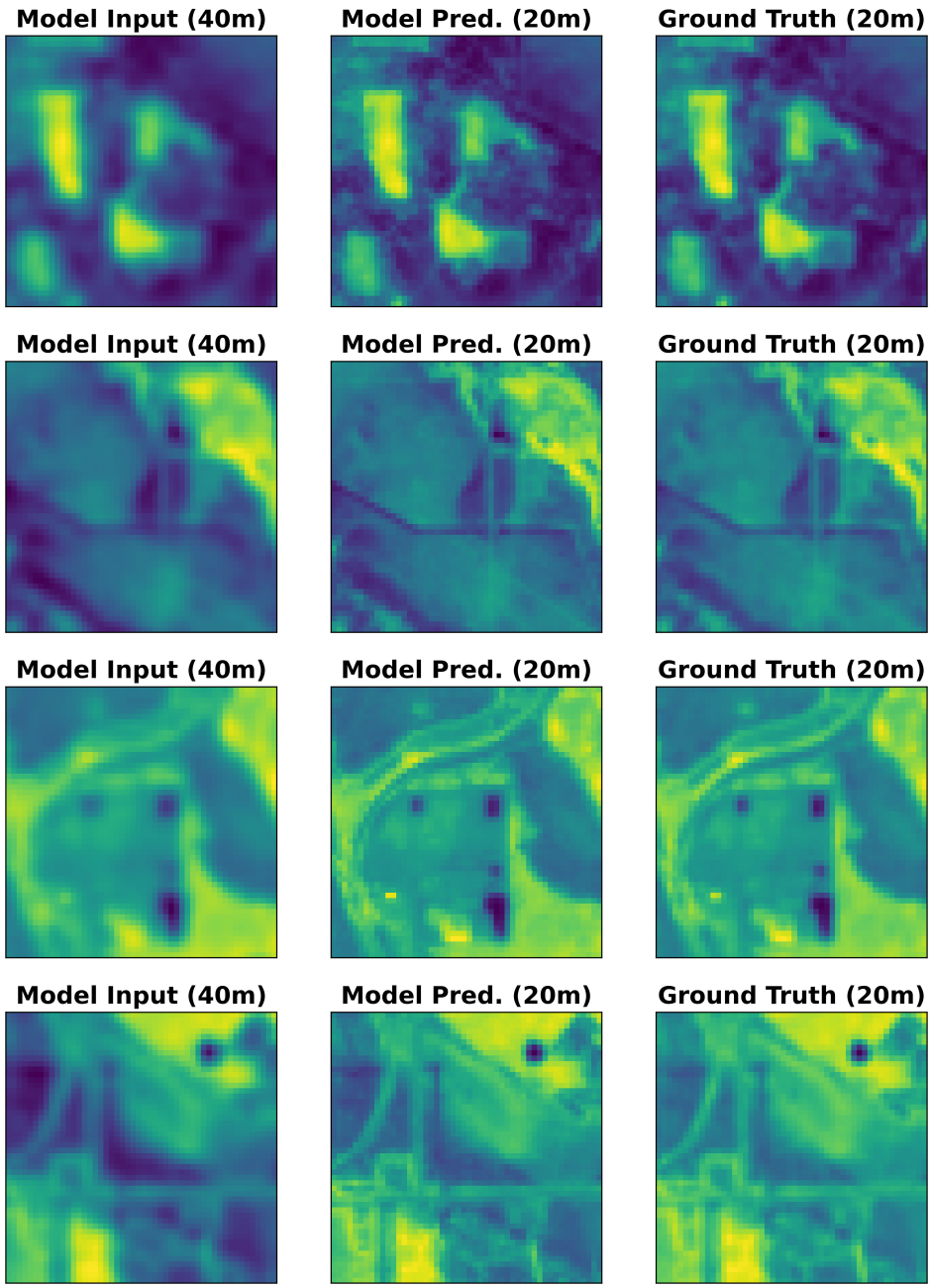

Figure 5.5: Qualitative comparison of the 40 m "low-resolution" band (Band-B8a) input (left), the 20 m "high-resolution" band (Band-B8a) predicted by network ($\Psi_{2\times}$) (middle), and the original 20 m band (Band-B8a) from Sentinet-2B (right) for the 40 m $\rightarrow$ 20 m spatial enhancement.

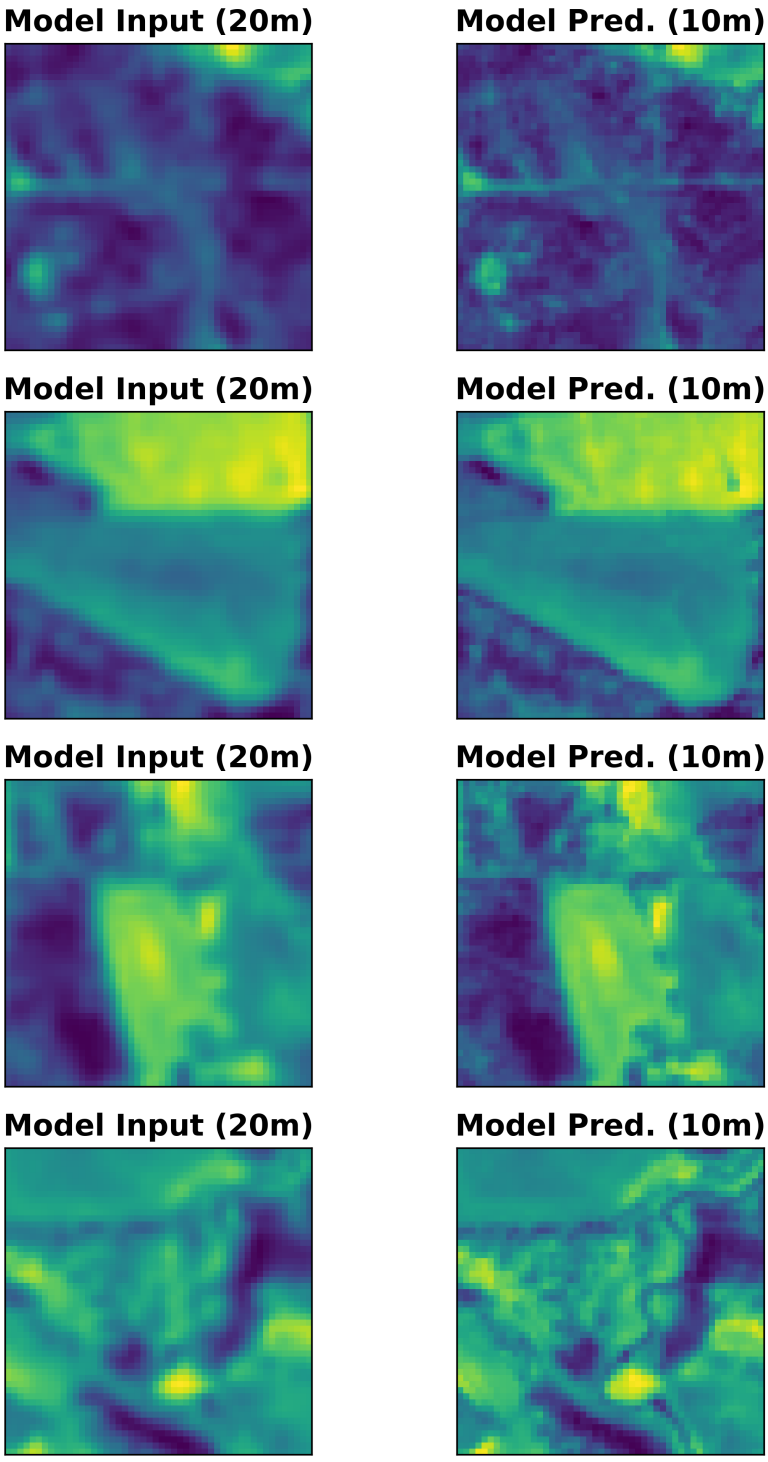

Figure 5.6: Qualitative comparison of the original 20 m Sentinel-2B band (Band-B8a) input (left) and its corresponding spatially enhanced 10 m band (Band-B8a) predicted by network $\Psi_{2\times}$ (right) for the 20 m $\rightarrow$ 10 m spatial enhancement.

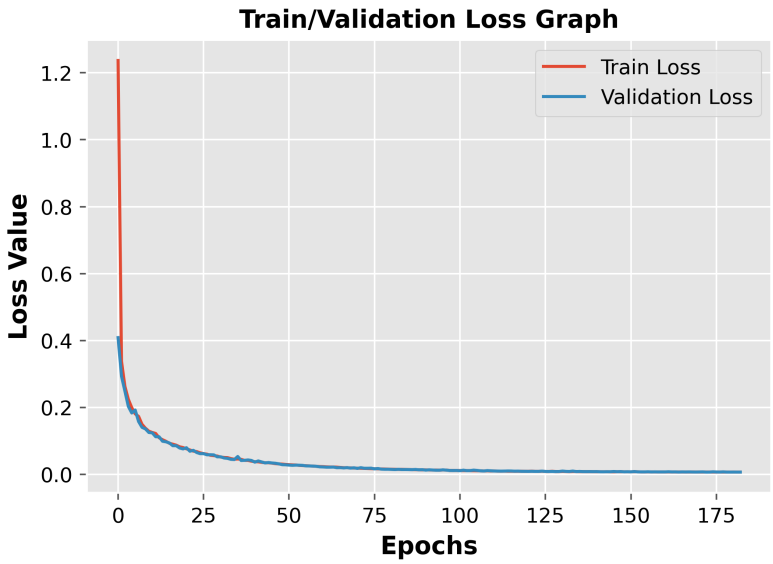

Figure 5.7: Variation in training and validation losses while training network $\Psi_{6\times}$ for the 360 m $\rightarrow$ 60 m spatial enhancement.

contains more spatial information compared to the 360 m GSD input and closely resembles the original 60 m Sentinel-2B band. This demonstrates that network $\Psi_{6\times}$ can effectively enhance the spatial resolution of low-resolution Sentinel-2B bands.

Algorithm	**MAE**	**RMSE**	**SAM**	**UIQI**	**PSNR**	**SSIM**
E-SSRAN	0.004	0.005	0.074	0.989	47.01	0.935

Table 5.3: Quantitative performance of network $\Psi_{6\times}$ for the 360 m $\rightarrow$ 60 m spatial enhancement.

Once trained, network $\Psi_{6\times}$ can be used for 60 m $\rightarrow$ 10 m spatial enhancement, generating spatially enhanced bands at 10 m GSD that can be evaluated visually. The results are shown in Figure 5.9 compares the spatially enhanced 10 m bands predicted by network $\Psi_{6\times}$ for the 60 m $\rightarrow$ 10 m spatial enhancement. The left column shows the original 60 m band from Sentinel-2B (set C) input to the network, and the right column displays the corresponding spatially enhanced 10 m band predicted by the network. The 10 m band predicted by the network contains more spatial details compared to the original 60 m Sentinel-2B band demonstrating the network's ability to enhance the spatial resolution from 60 m to 10 m GSD.

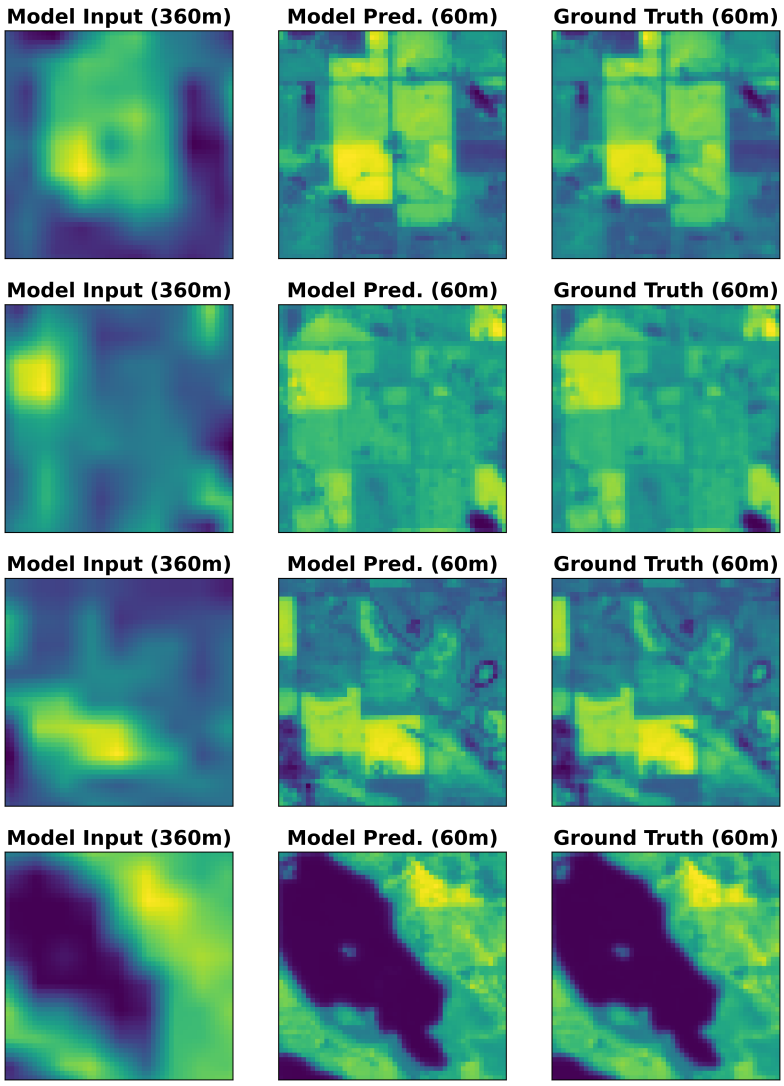

Figure 5.8: Qualitative comparison of the 360 m "low-resolution" band (Band-B9) input to the network (left), the 60 m "high-resolution" band (Band-B9) predicted by the network $\Psi_{6\times}$ (middle), and the original 60 m band (Band-B9) from Sentinet-2B (right) for the 360 m $\rightarrow$ 60 m spatial enhancement.

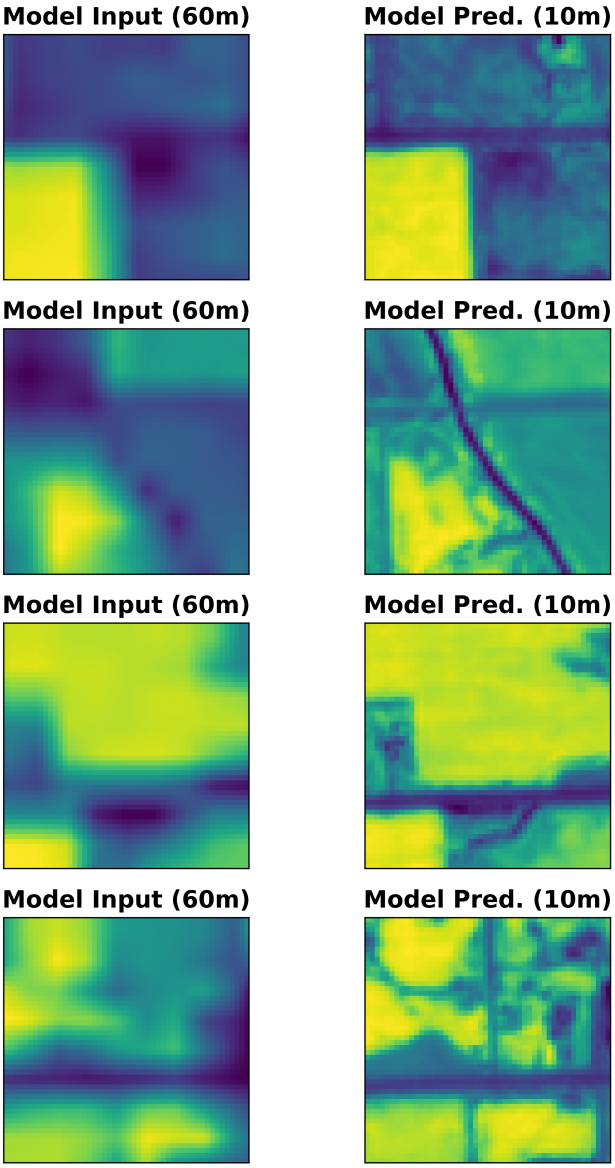

Figure 5.9: Qualitative comparison of the original 60 m Sentinel-2B band (Band-B9) input to the network (left) and the corresponding spatially enhanced 10 m band (Band-B9) predicted by network $\Psi_{6\times}$ (right) for the 60 m $\rightarrow$ 10 m spatial enhancement.

Chapter 6

Merging the Spatial- and Spectral-Enhancement Modules

In this section, we propose a pipeline to integrate the outputs of the Spatial- and Spectral-Enhancement modules to obtain a satellite image with improved spatial and spectral resolution (shown in Figure 6.1). Initially, we train the model $\Psi^{Spectral}$ using the E-SSRAN architecture to enhance the spectral resolution of MSI. For training $\Psi^{Spectral}$, we use MSI patches with a 30 m GSD, enabling the model to learn the mapping between MSI and HSI. Once trained, $\Psi^{Spectral}$ generates a spectrally enhanced MSI (that is spectrally equivalent to an HSI) at the same GSD as input MSI.

Following this, in the Spatial-Enhancement module, we train two separate E-SSRAN models to improve the spatial resolution of Sentinel-2B MSI: The first model, $\Psi_{2\times}$, is designed to enhance the spatial resolution of 20 m Sentinel-2B bands to 10 m GSD, while the second model, $\Psi_{6\times}$, focuses on improving the spatial resolution of 60 m Sentinel-2B bands to 10 m GSD. This results in a spatially enhanced MSI with all 12 Sentinel-2B bands at 10 m GSD.

Finally, the spatially enhanced 10 m GSD Sentinel-2B MSI is fed into the $\Psi^{Spectral}$ model in the Spectral-Enhancement module, producing an HSI at 10 m GSD. Thus, the proposed pipeline can: 1) enhance the spatial resolution of all 12 Sentinel-2B bands to 10 m GSD using models $\Psi_{2\times}$ and $\Psi_{6\times}$, 2) enhance the spectral resolution of MSI using the $\Psi^{Spectral}$ model, and 3) improve the spatial resolution of HSI from 30 m to 10 m GSD. In other words, the pipeline enhances the spectral resolution of MSI by learning the MSI-to-

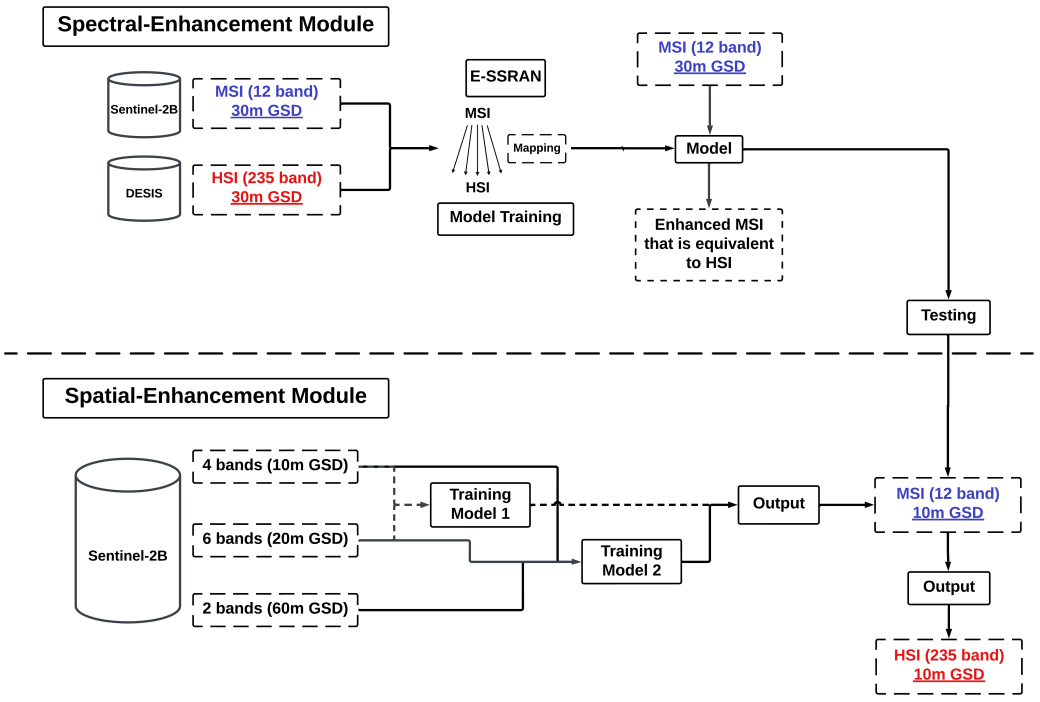

Figure 6.1: Merging the outputs of the Spatial- and Spectral-Enhancement modules

HSI mapping (as detailed in the Spectral-Enhancement module), improves the spatial resolution of MSI by mapping low-resolution MSI to high-resolution MSI (as explained in the Spatial-Enhancement module), and enhances the spatial resolution of HSI by merging the outputs of the Spatial- and Spectral-Enhancement modules. All this comprehensive enhancement is achieved using the E-SSRAN architecture.

The 10 m GSD HSI generated by our pipeline can only be evaluated visually since the original HSI is available only at 30 m GSD. Figure 6.2 provides true-color images of the spatially enhanced 10 m GSD HSI generated by our pipeline. The left column shows the original 30 m GSD HSI captured by DESIS, and the right column displays the corresponding spatially enhanced 10 m GSD HSI predicted by the $\Psi^{Spectral}$ network. It can be observed that the 10 m GSD HSI predicted by the network contains more spatial details compared to the original 30 m DESIS HSI, demonstrating the network's ability to enhance the spatial resolution of HSI from 30 m to 10 m GSD by merging the outputs of the Spatial- and Spectral-Enhancement modules. Additionally,

the spectral signatures of the 10 m GSD HSI predicted by the network are very similar to that of the original 30 m HSI (Figure 6.3), indicating that our algorithm preserves spectral integrity among the bands while enhancing spatial resolution.

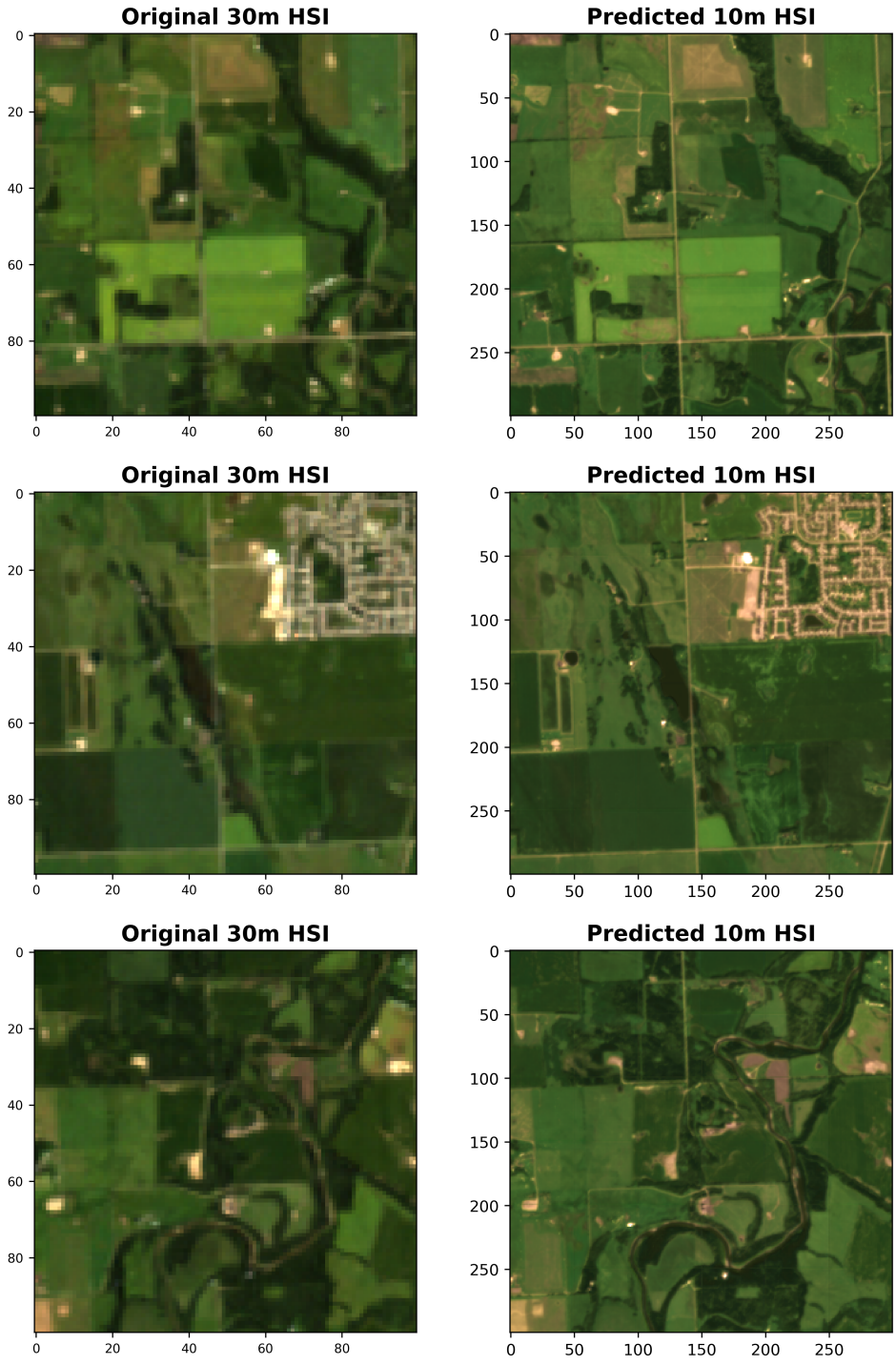

Figure 6.2: True color (RGB) images comparing the original 30 m GSD HSI obtained from DESIS (left) and the spatially enhanced 10 m GSD HSI predicted by the $\Psi^{Spectral}$ network (right).

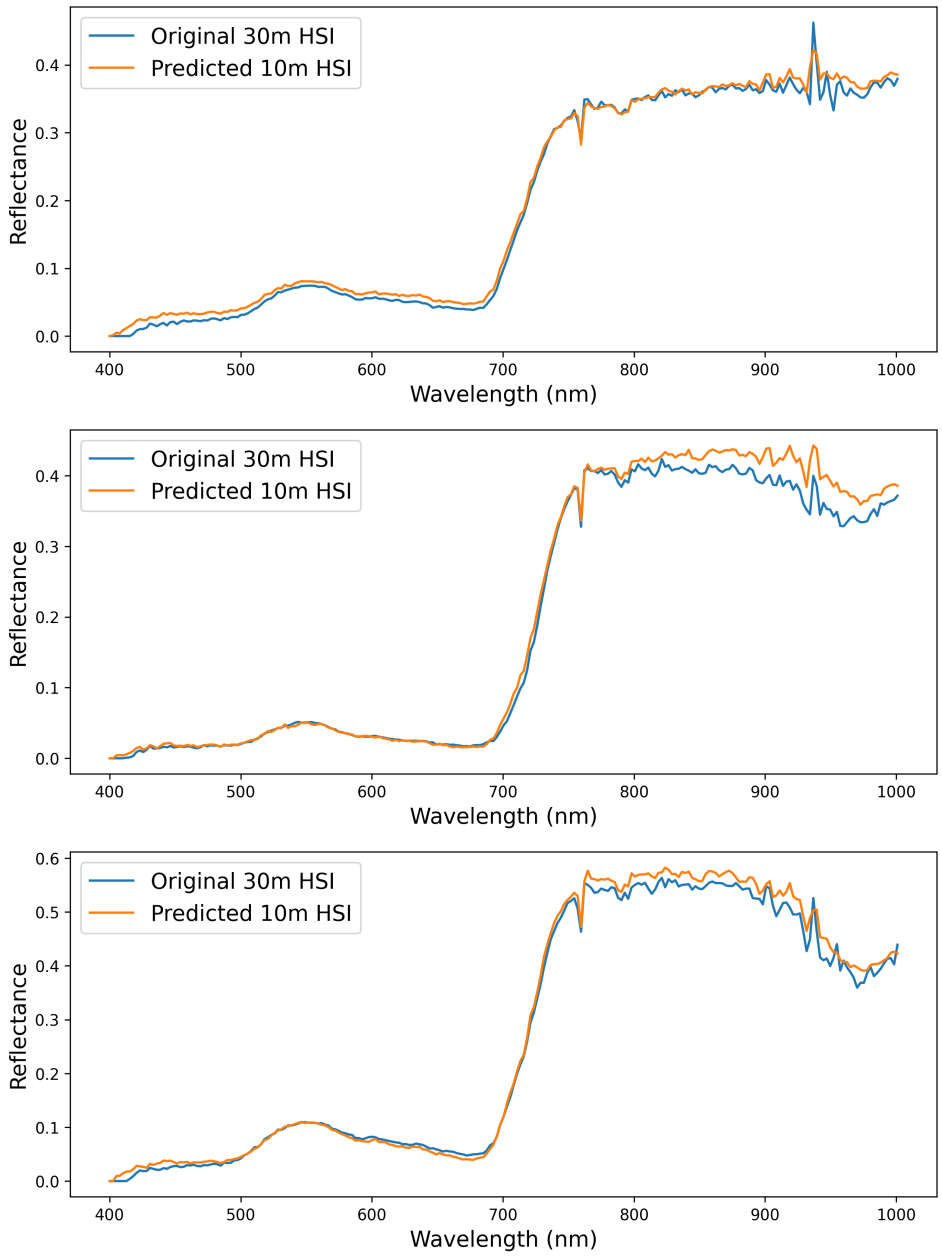

Figure 6.3: Comparing the spectral signatures of the original 30 m GSD HSI obtained from DESIS (blue color) and the spatially enhanced 10 m GSD HSI predicted by the $\Psi^{Spectral}$ network (orange color).

Chapter 7

Conclusion and Future Directions

In this study, we presented Extended Spatial-Spectral Residual Attention Network (E-SSRAN) to effectively integrate the spatial and spectral characteristics of MSIs and HSIs, respectively, and address the challenges associated with the trade-off between spatial and spectral resolution of satellite images. Our method comprises two main modules: 1) the Spectral-Enhancement module, which leverages E-SSRAN to enhance the spectral resolution of MSI by learning the MSI-to-HSI mapping, resulting in a spectrally enhanced MSI that is equivalent to an HSI, and 2) the Spatial-Enhancement module, which uses E-SSRAN to improve the spatial resolution of MSIs by mapping low-resolution MSIs to high-resolution MSIs. Specifically, it enhances the 20 m and 60 m GSD bands of Sentinel-2B to 10 m GSD, resulting in all 12 Sentinel-2B bands at a uniform 10 m GSD. Furthermore, we developed a comprehensive pipeline to merge the outputs from both modules. The 10 m GSD MSI generated by the Spatial-Enhancement module is further processed through the Spectral-Enhancement module, producing 10 m GSD HSI. This integrated approach yields satellite images that exhibit high spatial and spectral resolution.

The performance of our proposed algorithm was evaluated using MSI and HSI data from the Sentinel-2B and DESIS satellites, respectively, focusing on agricultural regions in Central Alberta. The results were promising both qualitatively and quantitatively. The Spectral-Enhancement module achieved an RMSE of 0.018, a UIQI score of 0.96, a PSNR value of 37 dB, and an SSIM

of 0.95. The Spatial-Enhancement module successfully improved the spatial resolution of the 20 m and 60 m Sentinel-2B bands to 10 m GSD. These findings indicate that E-SSRAN significantly enhances the quality of satellite images by optimizing the balance between spatial and spectral resolutions.

Our method has the potential to offer substantial value to various applications in agriculture and beyond, by providing more detailed and accurate insights for satellite data analysis. Future work will involve applying the proposed algorithm to datasets from other satellites, such as Landsat-8 and EO-1 Hyperion, to further assess its performance. Furthermore, the high-resolution data generated by our algorithm can be utilized in various applications, including land cover classification, mineral mapping, etc. to determine how effectively the enhanced data addresses real-world problems compared to existing low-resolution data. Additionally, future efforts will focus on reducing noise in the last hundred bands of the HSI and enhancing spectral correlation among the bands to further refine the performance of the algorithm.

References

[1] E. Adam, O. Mutanga, and D. Rugege, *Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review*, 2010. DOI: 10.1007/s11273-009-9169-z.

[2] G. E. Adjovu, H. Stephen, D. James, and S. Ahmad, *Overview of the application of remote sensing in effective monitoring of water quality parameters*, 2023. DOI: 10.3390/rs15071938.

[3] N. Akae, Y. Makihara, and Y. Yagi, "Gait recognition using periodic temporal super resolution for low frame-rate videos," in *2011 International Joint Conference on Biometrics, IJCB 2011*, 2011. DOI: 10.1109/IJCB.2011.6117530.

[4] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Transactions on Image Processing*, vol. 14, 10 2005, ISSN: 10577149. DOI: 10.1109/TIP.2005.851684.

[5] B. Arad, R. Timofte, O. Ben-Shahar, *et al.*, "Ntire 2020 challenge on spectral reconstruction from an rgb image," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, 2020. DOI: 10.1109/CVPRW50498.2020.00231.

[6] E. Bedini, "The use of hyperspectral remote sensing for mineral exploration: A review," *Journal of Hyperspectral Remote Sensing*, vol. 7, 4 2017, ISSN: 2237-2202. DOI: 10.29150/jhrs.v7.4.p189-211.

[7] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *BMVC 2012 - Electronic Proceedings of the British Machine Vision Conference 2012*, 2012. DOI: 10.5244/C.26.135.

[8] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)? -arguments against avoiding rmse in the literature," *Geoscientific Model Development*, vol. 7, 3 2014, ISSN: 19919603. DOI: 10.5194/gmd-7-1247-2014.

[9] Y. Chai, J. Ren, H. Zhao, Y. Li, J. Ren, and P. Murray, "Hierarchical and multi-featured fusion for effective gait recognition under variable scenarios," *Pattern Analysis and Applications*, vol. 19, 4 2016, ISSN: 14337541. DOI: 10.1007/s10044-015-0471-5.

[10] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004. DOI: 10.1109/cvpr.2004.1315043.

[11] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[12] S. Clerc and M. Team, *S2 mpc - data quality report*, 2018. [Online]. Available: http://earth.esa.int/documents/247904/685211/Sentinel-2-Data-Quality-Report.

[13] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007. DOI: 10.1109/CVPR.2007.383028.

[14] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, 11 2018, ISSN: 21622388. DOI: 10.1109/TNNLS.2018.2798162.

[15] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Joint camera spectral response selection and hyperspectral image recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, 1 2022, ISSN: 19393539. DOI: 10.1109/TPAMI.2020.3009999.

[16] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009. DOI: 10.1109/ICCV.2009.5459271.

[17] Y. Gu, Y. Zhang, and J. Zhang, "Integration of spatial-spectral information for resolution enhancement in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, 5 2008, ISSN: 01962892. DOI: 10.1109/TGRS.2008.917270.

[18] R. Hang, Z. Li, Q. Liu, and S. S. Bhattacharyya, "Prinet: A prior driven spectral super-resolution network," in *Proceedings - IEEE International Conference on Multimedia and Expo*, vol. 2020-July, 2020. DOI: 10.1109/ICME46284.2020.9102772.

[19] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Transactions on Image Processing*, vol. 13, 9 2004, ISSN: 10577149. DOI: 10.1109/TIP.2004.829779.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Dec. 2015, pp. 1026–1034, ISBN: 978-1-4673-8391-2. DOI: 10.1109/ICCV.2015.123.

[21] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, 2020, ISSN: 21511535. DOI: 10.1109/JSTARS.2020.3025040.

[22] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "Hyperpnn: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, 8 2019, ISSN: 21511535. DOI: 10.1109/JSTARS.2019.2917584.

[23] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, 2015. DOI: 10.1109/CVPR.2015.7299156.

[24] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, 2017. DOI: 10.1109/CVPR.2017.613.

[25] E. R. Hunt, W. D. Hively, S. J. Fujikawa, D. S. Linden, C. S. Daughtry, and G. W. McCarty, "Acquisition of nir-green-blue digital photographs from unmanned aircraft for crop monitoring," *Remote Sensing*, vol. 2, 1 2010, ISSN: 20724292. DOI: 10.3390/rs2010290.

[26] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Transactions on Computational Imaging*, vol. 6, 2020, ISSN: 23339403. DOI: 10.1109/TCI.2020.2996075.

[27] D. Johnson, "Signal-to-noise ratio," *Scholarpedia*, vol. 1, p. 2088, Jan. 2006. DOI: 10.4249/scholarpedia.2088.

[28] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, 2016. DOI: 10.1109/CVPR.2016.182.

[29] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[30] D. Krutz, R. Müller, U. Knodt, *et al.*, "The instrument design of the dlr earth sensing imaging spectrometer (desis)," *Sensors (Switzerland)*, vol. 19, 7 2019, ISSN: 14248220. DOI: 10.3390/s19071622.

[31] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, 2018, ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2018.09.018.

[32] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, 1 2002, ISSN: 10535888. DOI: 10.1109/79.974718.

[33] X. Li, Y. Wu, W. Zhang, R. Wang, and F. Hou, "Deep learning methods in real-time image super-resolution: A survey," in *Journal of Real-Time Image Processing*, vol. 17, 2020. DOI: 10.1007/s11554-019-00925-3.

[34] Y. Li, J. Qu, W. Dong, and Y. Zheng, "Hyperspectral pansharpening via improved pca approach and optimal weighted fusion strategy," *Neurocomputing*, vol. 315, 2018, ISSN: 18728286. DOI: 10.1016/j.neucom.2018.07.030.

[35] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 9 2021, ISSN: 15580644. DOI: 10.1109/TGRS.2021.3049875.

[36] M. Lobato, W. R. Norris, R. Nagi, A. Soylemezoglu, and D. Nottage, "Machine learning for soil moisture prediction using hyperspectral and multispectral data," in *Proceedings of 2021 IEEE 24th International Conference on Information Fusion, FUSION 2021*, 2021. DOI: 10.23919/fusion49465.2021.9627067.

[37] L. Loncan, L. B. D. Almeida, J. M. Bioucas-Dias, *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, 3 2015, ISSN: 21686831. DOI: 10.1109/MGRS.2015.2440094.

[38] X. Lu, J. Zhang, D. Yang, L. Xu, and F. Jia, "Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image," *IEEE Transactions on Image Processing*, vol. 30, pp. 6815–6828, 2021, ISSN: 1057-7149. DOI: 10.1109/TIP.2021.3098246.

[39] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, 8 2020, ISSN: 15580644. DOI: 10.1109/TGRS.2020.2963848.

[40] L. S. Macarringue, É. L. Bolfe, and P. R. M. Pereira, "Developments in land use and land cover classification techniques in remote sensing: A review," *Journal of Geographic Information System*, vol. 14, 01 2022, ISSN: 2151-1950. DOI: 10.4236/jgis.2022.141001.

[41] D. G. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, 1 2002, ISSN: 10535888. DOI: 10.1109/79.974724.

[42] MathWorks, *Sam*, https://www.mathworks.com/help/images/ref/sam.html#mw_5dd49a76-8bc3-44f9-ade3-b59a3944ca4e_sep_mw_dc72bab6-b469-451e-ba87-f21ac3e3fe05, Accessed: 2024-06-05.

[43] P. Milanfar, "Super-resolution imaging (1st ed.)," *CRC Press*, 2011, ISSN: 9781315217994. DOI: https://doi.org/10.1201/9781439819319.

[44] NASA, *Introduction to remote sensing*, https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing, Accessed: 2024-06-01.

[45] K. Nasrollahi, S. Escalera, P. Rasti, *et al.*, "Deep learning based super-resolution for improved action recognition," in *5th International Conference on Image Processing, Theory, Tools and Applications 2015, IPTA 2015*, 2015. DOI: 10.1109/IPTA.2015.7367098.

[46] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Machine Vision and Applications*, vol. 25, 6 2014, ISSN: 14321769. DOI: 10.1007/s00138-014-0623-4.

[47] K. Nguyen, C. Fookes, S. Sridharan, M. Tistarelli, and M. Nixon, "Super-resolution for biometrics: A comprehensive survey," *Pattern Recognition*, vol. 78, 2018, ISSN: 00313203. DOI: 10.1016/j.patcog.2018.01.002.

[48] S. L. Ozesmi and M. E. Bauer, "Satellite remote sensing of wetlands," *Wetlands Ecology and Management*, vol. 10, 5 2002, ISSN: 09234861. DOI: 10.1023/A:1020908432489.

[49] E. Perez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn, "Psyco: Manifold span reduction for super resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, 2016. DOI: 10.1109/CVPR.2016.203.

[50] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review," *Journal of Applied Remote Sensing*, vol. 15, 03 2021, ISSN: 19313195. DOI: 10.1117/1.jrs.15.031501.

[51] C. Pohl and J. L. V. Genderen, *Review article multisensor image fusion in remote sensing: Concepts, methods and applications*, 1998. DOI: 10.1080/014311698215748.

[52] A. B. Pour, B. Zoheir, B. Pradhan, and M. Hashim, *Editorial for the special issue: Multispectral and hyperspectral remote sensing data for mineral exploration and environmental monitoring of mined areas*, 2021. DOI: 10.3390/rs13030519.

[53] J. Qu, Y. Li, and W. Dong, "Hyperspectral pansharpening with guided filter," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, 11 2017, ISSN: 1545598X. DOI: 10.1109/LGRS.2017.2755679.

[54] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018. DOI: 10.1109/CVPR.2018.00266.

[55] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, 5500 2000, ISSN: 00368075. DOI: 10.1126/science.290.5500.2323.

[56] S. S. Rwanga and J. M. Ndambuki, "Accuracy assessment of land use/land cover classification using remote sensing and gis," *International Journal of Geosciences*, vol. 08, 04 2017, ISSN: 2156-8359. DOI: 10.4236/ijg.2017.84033.

[57] M. S. Ryoo, B. Rothrock, C. Fleming, and H. J. Yang, "Privacy-preserving human activity recognition from extreme low resolution," in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017. DOI: 10.1609/aaai.v31i1.11233.

[58] R. N. Sahoo, S. S. Ray, and K. R. Manjunath, "Hyperspectral remote sensing of agriculture," *Current Science*, vol. 108, 5 2015, ISSN: 00113891.

[59] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, 2015. DOI: 10.1109/CVPR.2015.7299003.

[60] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on sim-ga data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, 6 2015, ISSN: 21511535. DOI: 10.1109/JSTARS.2015.2440092.

[61] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007. DOI: 10.1109/CVPR.2007.383198.

[62] M. Shrestha, D. Helder, and J. Christopherson, "Dlr earth sensing imaging spectrometer (desis) level 1 product evaluation using radcalnet measurements," *Remote Sensing*, vol. 13, 12 2021, ISSN: 20724292. DOI: 10.3390/rs13122420.

[63] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, 2017. DOI: 10.1109/CVPR.2017.298.

[64] P. S. Thenkabail, M. K. Gumma, P. Teluguntla, and M. I. Ahmed, "Hyperspectral remote sensing of vegetation and agricultural crops," *Photogrammetric Engineering and Remote Sensing*, vol. 79, 9 2013, ISSN: 00991112. DOI: 10.1201/b11222-41.

[65] P. S. Thenkabail, J. G. Lyon, and A. Huete, "Advances in hyperspectral remote sensing of vegetation and agricultural crops," in Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation, 2019. DOI: 10.1201/9781315164151-1.

[66] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013. DOI: 10.1109/ICCV.2013.241.

[67] G. Van Rossum and F. L. Drake Jr, *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[68] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Transactions on Industrial Electronics*, vol. 68, 11 2021, ISSN: 15579948. DOI: 10.1109/TIE.2020.3038096.

[69] Y. H. Wang, J. Qiao, J. B. Li, P. Fu, S. C. Chu, and J. F. Roddick, "Sparse representation-based mri super-resolution reconstruction," *Measurement: Journal of the International Measurement Confederation*, vol. 47, 1 2014, ISSN: 02632241. DOI: 10.1016/j.measurement.2013.10.026.

[70] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, 3 2002, ISSN: 10709908. DOI: 10.1109/97.995823.

[71] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, 4 2004, ISSN: 10577149. DOI: 10.1109/TIP.2003.819861.

[72] Z. Wu, G. Snyder, C. Vadnais, *et al.*, "User needs for future landsat missions," *Remote Sensing of Environment*, vol. 231, 2019, ISSN: 00344257. DOI: 10.1016/j.rse.2019.111214.

[73] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, 2019. DOI: 10.1109/CVPR.2019.00168.

[74] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "Hpgan: Hyperspectral pansharpening using 3-d generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 1 2021, ISSN: 15580644. DOI: 10.1109/TGRS.2020.2994238.

[75] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, 5 2020, ISSN: 21622388. DOI: 10.1109/TNNLS.2019.2920857.

[76] C. Y. Yang and M. H. Yang, "Fast direct super-resolution by simple functions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013. DOI: 10.1109/ICCV.2013.75.

[77] H. Yang, J. Kong, H. Hu, Y. Du, M. Gao, and F. Chen, "A review of remote sensing for water quality retrieval: Progress and challenges," *Remote Sensing*, vol. 14, 8 2022, ISSN: 20724292. DOI: 10.3390/rs14081770.

[78] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, 11 2010, ISSN: 10577149. DOI: 10.1109/TIP.2010.2050625.

[79] J. Yang, Y. Q. Zhao, and J. C. W. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sensing*, vol. 10, 5 2018, ISSN: 20724292. DOI: 10.3390/rs10050800.

[80] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *Summaries of the 4th JPL Airborne Earth Science Workshop, JPL Publication 92-41*, 1992.

[81] P. J. Zarco-Tejada, J. A. Berni, L. Suárez, G. Sepulcre-Cantó, F. Morales, and J. R. Miller, "Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection," *Remote Sensing of Environment*, vol. 113, 6 2009, ISSN: 00344257. DOI: 10.1016/j.rse.2009.02.016.

[82] H. Zhang, Z. Yang, L. Zhang, and H. Shen, "Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences," *Remote Sensing*, vol. 6, 1 2013, ISSN: 20724292. DOI: 10.3390/rs6010637.

[83] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Processing*, vol. 92, 9 2012, ISSN: 01651684. DOI: 10.1016/j.sigpro.2012.01.020.

[84] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-resolution gait recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 986–996, 2010. DOI: 10.1109/TSMCB.2010.2042166.

[85] W. Zhang, X. Li, and L. Zhao, "A fast hyperspectral feature selection method based on band correlation analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, 11 2018, ISSN: 15580571. DOI: 10.1109/LGRS.2018.2853805.

[86] K. Zheng, L. Gao, W. Liao, *et al.*, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 3 2021, ISSN: 15580644. DOI: 10.1109/TGRS.2020.3006534.

[87] X. Zheng, W. Chen, and X. Lu, "Spectral super-resolution of multispectral images using spatial–spectral residual attention network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022, ISSN: 0196-2892. DOI: 10.1109/TGRS.2021.3104476.

[88] X. Zheng, Y. Yuan, and X. Lu, "Hyperspectral image denoising by fusing the selected related bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, 5 2019, ISSN: 01962892. DOI: 10.1109/TGRS.2018.2875304.

[89] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, 11 2020, ISSN: 15580644. DOI: 10.1109/TGRS.2020.2986313.