Politecnico di Torino

Computer Engineering - Artificial Intelligence and Data Analytics

A.Y. 2023/2024
Graduation Session December 2024

Enhancing Spatial Resolution in Sentinel-3 Data - A Landsat 8 Supervised Approach

Supervisor
Prof. Giovanni Squillero

Co-Supervisor
Dott. Giacomo Blanco
Dott. Lorenzo Innocenti
Dott. Luca Barco

Candidate
Filippo Greco

Abstract

The temperature of Earth's surface has become in recent years critical for various research areas including weather forecasting, environmental monitoring, vegetation health analysis, urban planning, human well-being, and agricultural management. A major issue arising from urbanization is the development of urban heat islands — a phenomenon analyzed within microclimate studies, which focuses on localized areas that exhibit significantly higher temperatures than their surroundings. This challenge has increased the demand for finer spatial resolution of Land Surface Temperature (LST), enabling stakeholders and policymakers to make informed decisions and implement effective strategies within their areas of responsibility. Remote sensing instruments on satellites have been a main tool for estimating LST. In this case study datasets of temperature measurements come from Sentinel-3 and Landsat 8, which are two satellite missions that use spectral bands to calculate LST, each with its own limitations: Sentinel-3 provides multiple readings per day but with a spatial resolution of 1 km, whereas Landsat 8 offers a finer spatial resolution of 100 m but revisits the same area only every eight days. Moreover the two satellites' dataset have different types of data content and these discrepancies are addressed to make them as similar as possible. The goal of this thesis is to prepare data from the aforementioned satellites — along with auxiliary data including Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM), Land Cover (LC), and weather data that act as additional information — and propose new deep neural network architectures that lay their foundation on established models such as SR-CNN, SE-SRCNN, EDSR, SRGAN. These models are well-known for the task of Super Resolution (SR), a technique proven in recent years to be effective in generating high-resolution images from low-resolution counterparts. The proposed models perform 10x upscaling of the low-resolution Sentinel-3 images, using Landsat 8 data as the target, leveraging the strength of Sentinel-3's temporal resolution and Landsat 8's spatial resolution. The performance of the models is evaluated based on metrics including PSNR, MAE, and SSIM and a comparison between the results is provided.

Contents

List of Figures

List of Tables

Acronyms

BT Brightness Temperature. The temperature at which a black body would emit the same amount of radiation, often used in remote sensing to measure the radiation emitted by the Earth's surface. v, vi, 11, 15, 16, 27, 29, 34–36, 44, 46–53, 55–60

BT2LST Brightness Temperature to Land Surface Temperature. A conversion method proposed in this thesis that transforms brightness temperature (BT) measurements, typically obtained from thermal infrared satellite data, into land surface temperature (LST) estimates. v, vi, 36, 42–61

CNN Convolutional Neural Network. A deep learning algorithm primarily used in image processing, which automatically detects patterns like edges and textures. 5, 6

DEM Digital Elevation Model. A 3D representation of terrain elevations, used in geographic and remote sensing applications. iii, 3, 11–13, 21, 34–36, 43, 45–61, 63

doy Day of the Year. The sequential day number of a given year, from 1 to 365 or 366 in leap years. 12, 13, 35, 43, 46–61, 63

DR-SE-SRCNN Deep Recursive Super-Resolution Convolutional Neural Network. A novel architecture proposed in this thesis that enhances image resolution by integrating a deep recursive layer into the feature extractor of the SRCNN model. iv, vi, 38, 59–61

DRCNN Deep Recursive Convolutional Neural Network. A type of neural network that performs super-resolution tasks by recursively applying the same set of filters to enhance image resolution. 6

EDSR Enhanced Deep Super-Resolution. An advanced deep learning model designed to improve image super-resolution accuracy and quality. iii, v, 6, 7, 38–40, 50–55, 57, 61, 63

ESA European Space Agency. An intergovernmental organization dedicated to the exploration of space, responsible for satellite missions and research in Earth observation, planetary science, and astronomy. 14

ESRGAN Enhanced Super-Resolution Generative Adversarial Network. A deep learning model that improves the quality of images through super-resolution techniques by employing a GAN framework. iii, 7, 8

EUMETSAT European Organisation for the Exploitation of Meteorological Satellites. An intergovernmental organization that operates weather and climate-monitoring satellites, providing data to support weather forecasting, environmental monitoring, and climate research. 14

GAN Generative Adversarial Network. A deep learning framework composed of two neural networks, a generator and a discriminator, which compete against each other to create data (such as images) that is indistinguishable from real-world data. 7

GIS Geographic Information System. A system designed to capture, store, analyze, manage, and present spatial or geographic data, commonly used for mapping, environmental monitoring, urban planning, and other location-based analyses. 12, 41

HR High Resolution. A term used to describe images or data with fine spatial details, typically characterized by a small pixel size. 2, 3, 29, 35, 36, 45, 48

IR Infrared. A region of the electromagnetic spectrum with wavelengths longer than visible light, typically used in remote sensing to measure thermal radiation and detect heat patterns on Earth's surface, such as land and sea temperatures. iii, 17

LC Land Cover. Refers to the physical material at the surface of the Earth, such as urban areas, forests, water bodies, and agricultural land. 3, 8, 9, 11–13, 34–36, 43, 45–61, 63

LR Low Resolution. A term used to describe images or data with less spatial detail, characterized by larger pixel sizes. iv–vi, 2, 3, 35, 36, 38, 40, 43–60

LST Land Surface Temperature. The temperature of the Earth's surface, as observed via remote sensing techniques. iii–vi, 1, 3, 8, 11, 13, 16, 18, 19, 27, 29, 30, 33–36, 44–61

LU/LC Land Use/Land Cover. A classification system used in remote sensing and environmental science to categorize the Earth's surface based on how land is utilized (land use) and the physical characteristics of the land cover (such as forest, water bodies, urban areas, and agricultural land). iii, 22

MAE Mean Absolute Error. A measure of errors between paired observations, commonly used to evaluate prediction accuracy. 3, 41–61, 63

MLP Multilayer Perceptron. A type of artificial neural network that consists of multiple layers of neurons, typically including an input layer, one or more hidden layers, and an output layer. 35

MODIS Moderate Resolution Imaging Spectroradiometer. A key instrument aboard NASA's Terra and Aqua satellites, used for large-scale global observations, including land cover, LST, and NDVI. 1, 20

MSE Mean Squared Error. A measure used to evaluate the accuracy of a prediction model by calculating the average of the squares of the errors, which are the differences between predicted and actual values. 34, 35, 42

NaN Not a Number. A symbolic representation used in computing to denote undefined or unrepresentable numerical values, often resulting from operations like division by zero or the square root of a negative number. 12, 29

NASA National Aeronautics and Space Administration. The United States government agency responsible for the civilian space program, as well as aeronautics and aerospace research. 18

NDVI Normalized Difference Vegetation Index. A measure of vegetation health and density, calculated from the difference between near-infrared and red light reflected by vegetation. iii, 3, 8, 9, 11–13, 16, 20, 34, 35, 43, 45–61, 63

NIR Near-Infrared. A region of the electromagnetic spectrum with wavelengths just longer than visible light, typically ranging from 0.7 to 1.5 micrometers, often used in remote sensing to study vegetation, land cover, and water bodies. 8, 20

OLCI Ocean and Land Colour Instrument. An instrument aboard Sentinel-3 used for ocean and land observation, monitoring biological activity and environmental changes. 14

OLI Operational Land Imager. A sensor on Landsat 8 that captures imagery of the Earth's surface, used for monitoring land cover and land use changes. iii, 18, 19

PSNR Peak Signal-to-Noise Ratio. A measure used to evaluate the quality of reconstructed images or video, typically used in super-resolution tasks. 3, 41, 44–61, 63

RRDB Residual-in-Residual Dense Block. A block used in deep learning networks like ESRGAN, to stabilize the training process and improve performance in image super-resolution tasks. 7

SE-EDSR Squeeze and Extract Enhanced Deep Super-Resolution. A novel model proposed in this thesis that combines the principles of the Squeeze-and-Excitation (SE) mechanism with the Enhanced Deep Super-Resolution (EDSR) architecture to improve image resolution. iv, vi, 40, 57, 58, 61, 63

SE-SRCNN Super-Resolution Convolutional Neural Network with Squeeze and Extract. A deep learning model for image super-resolution that incorporates a squeeze-and-extract mechanism to efficiently capture and enhance features, improving the resolution of low-resolution images. iii, v, 9, 38, 40, 48, 49, 53, 59, 61

SLSTR Sea and Land Surface Temperature Radiometer. An instrument aboard Sentinel-3 designed for the measurement of sea and land surface temperatures, using thermal infrared channels. iii, 13–16

SR Super Resolution. A technique used in image processing that enhances the spatial resolution of images, allowing for the generation of high-resolution images from low-resolution inputs through various algorithms, including deep learning methods. 2, 3, 5, 8, 11, 15, 18, 22, 27, 29, 33, 35–37, 45, 48, 53, 57

SRCNN Super-Resolution Convolutional Neural Network. One of the earliest deep learning models developed for enhancing image resolution. iii, v, 5, 6, 8, 37, 38, 45–48, 53, 55

SRGAN Super-Resolution Generative Adversarial Network. A deep learning-based approach to generating high-resolution images from low-resolution input, using a GAN architecture. iii, v, vi, 7, 35, 39, 55, 56

SSIM Structural Similarity Index Measure. A method for measuring the similarity between two images, commonly used in image quality assessments. 3, 41, 44–61, 63

SWIR Short-Wave Infrared. A region of the electromagnetic spectrum with wavelengths ranging from 1.5 to 3 micrometers, used in remote sensing for applications such as detecting moisture, mapping vegetation health, and identifying minerals. 8

TIFF Tagged Image File Format. A flexible raster graphics file format commonly used for storing high-quality images, supporting various color depths and multiple layers or channels, and often used in scientific and medical imaging. 12

TIR Thermal Infrared. The portion of the electromagnetic spectrum used for measuring thermal radiation emitted by surfaces, typically from 3 to 15 micrometers. iii, 18, 19

TOA Top of Atmosphere. Refers to the outermost layer of the Earth's atmosphere, where satellites measure radiance emitted from the Earth's surface. 11, 13, 15, 16

UHI Urban Heat Island. A phenomenon where urban areas experience significantly higher temperatures than their surrounding rural areas due to human activities, dense infrastructure, and reduced vegetation. 1

USGS United States Geological Survey. A scientific agency of the U.S. government responsible for studying and monitoring natural resources, hazards, and landscapes, including mapping, remote sensing, and providing data on earth science topics such as geology, hydrology, and biology. 18

VDSR Very Deep Super-Resolution. A convolutional neural network designed for image super-resolution tasks, which utilizes a very deep architecture to learn residuals between low-resolution and high-resolution images, enabling high-quality reconstruction. 6

1 Introduction

High-resolution images of Earth's surface temperature are increasingly needed across various research areas, primarily due to the demand for microclimate predictions, which are made in a small area with high variance with respect to the surroundings. An example of which are Urban Heat Islands (UHI), a well-known term for areas within cities or densely populated urban zones that experience significantly higher temperatures than surrounding rural areas. This temperature difference is mainly caused by human activities and the urban environment itself.

The objective of the proposed thesis is explore new deep learning neural network architectures able to enhance the spatial resolution of the remote sensed Land Surface Temperature (LST) measured by Sentinel-3 using Landsat 8 as target. The proposed models can also process auxiliary data that can help give ulterior context to each sample and improve the results.

This introductory chapter provides a brief overview of the key concepts and the frequently used terms, to support the readers in their task.

1.1 Land Surface Temperature

LST is the measurement of the temperature of the Earth's surface, typically obtained through remote sensing instruments on satellites. These instruments capture thermal infrared radiation emitted by the surface, which is then used to calculate the temperature [1, 2] . LST measurements can vary due to factors like surface type (e.g., vegetation, water, or urban areas), atmospheric conditions, and surface emissivity (the efficiency of different surfaces to emit heat). Satellite missions such as Landsat [3], MODIS [4], and Sentinel-3 [5] provide LST data at varying spatial and temporal resolutions (i.e. 1 km, twice a day for Sentinel-3 and 100 m, once each 8 days for Landsat 8), allowing for the monitoring of surface temperature patterns across the globe. LST and related measurements such as air temperature and near-surface temperature are pivotal for numerous applications, including weather forecasting, environmental monitoring, vegetation health analysis, urban planning, human well-being, and agricultural management [1, 6–12].

1.2 Spatial Resolution and Super Resolution

Spatial resolution refers to the scale or size of the smallest unit of an image capable of distinguishing objects [13]. In this thesis, the images under consideration are the product of remote sensing from the Sentinel-3 and Landsat 8 missions. Spatial resolution in satellite images is the size of one pixel — the smallest detail visible to the sensor. The ensemble of multiple pixels covering a specific geographic area is also called raster, of which a representation is shown in Figure 1.1. A higher density of points in the area, or equivalently, a higher number of pixels in the raster, corresponds to better resolution, enabling more precise observations.

Enhancing spatial resolution involves generating a high-resolution (HR) representation of the data from its low-resolution (LR) counterpart, a problem commonly referred to as Super Resolution (SR). Resolution enhancement is especially necessary in fields such as video surveillance, medical diagnosis, and remote sensing applications [14], including satellite imagery — the focus of this thesis.

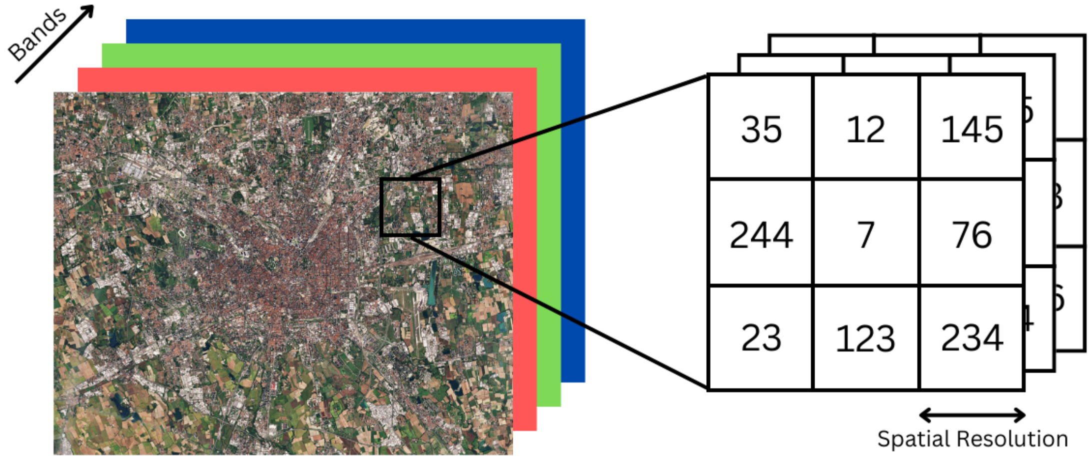


Figure 1.1: A visual representation for a multi channel raster on a specific region over the city of Milan

Machine learning approaches, particularly Artificial Neural Networks, have recently been the subject of study for their use as a SR technique to address the aforementioned problem. Nowadays, they are considered to provide the best capability in enhancing image quality and detail.

With a clear understanding of spatial resolution and its implications in remote sensing, the following section describes the problem undertaken on this thesis offering possible solutions and showing their limitations.

1.3 Formulation of the Problem

In the field of remote sensing, one of the main challenges is the trade-off between spatial resolution and temporal frequency of satellite imagery. While Sentinel-3 offers frequent revisit times, its spatial resolution of 1 km for thermal channels limits its effectiveness for applications requiring fine detail. Conversely, Landsat 8 provides a higher spatial resolution of 100 m, making it more suitable for detailed analysis; however, it revisits any given location approximately every 8 days. This disparity presents a unique problem for researchers and practitioners who aim to leverage the strengths of both satellite missions. Integrating the temporal frequency of Sentinel-3 with the spatial detail of Landsat 8 becomes critical for enhancing decision-making processes in various domains.

This thesis revisits the State Of The Art of SR and proposes deep-learning models designed to enhance the spatial resolution of Sentinel-3 LST data by a factor of 10x, from 1 km up to 100 m reached by Landsat 8, that is used as target. The models aim to produce high-resolution imagery that retains the temporal advantages of Sentinel-3. In pursuit of this goal, low-resolution images are supplemented with auxiliary features, such as other remotely sensed data, topological data, and derived data including Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM), Land Cover (LC) and weather data. These additional features aid the model in learning the intricate relationships between LR and HR images.

In SR task LR images are usually the exact counterpart of the worsened HR version, but in this particular case study data come from different satellites and thus there is not the strict correspondence between pairs of LR-HR. This is due mainly to the different level of preprocessing the two dataset have and also by the difference in temperature and cloud position due to the time shift between the two satellites' passage. This issue is addressed in this work by proposing a neural network, called BT2LST, able to shorten the disparity between LR-HR pairs.

The performance of this model will be evaluated using metrics such as Mean Absolute Value (MAE), Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) which are essential for assessing the quality of enhanced imagery. A comparison is also made with respect to other common techniques for image resizing such as bicubic interpolation to evaluate the effectiveness and quality of the results.

The thesis addresses the problem by breaking the discussion into six chapters, each of which is described in detail below:

1. **Chapter 1 - Introduction.** The current chapter describes the content of the thesis, the proposed problem and the corresponding solutions, along with some introductory description of land surface temperature and super resolution.

2. **Chapter 2 - Related Works.** This chapter reviews existing literature and research relevant to the thesis topic. It discusses previous studies, methodologies, and findings that have contributed to the understanding of the problem addressed.

3. **Chapter 3 - Datasets.** It describes the collection of data utilized to train the deep neural networks presented in this work, including how the data was acquired and any preprocessing steps applied.

4. **Chapter 4 - Methodology.** It outlines the methods and techniques used to conduct the research. It includes details on the supporting approaches to address secondary problems and limitations and the deep neural networks trained for the main objective — super resolution.

5. **Chapter 5 - Experimental Results** It presents the findings of the research, including quantitative and qualitative results. It includes visualizations, tables, and comparisons between the proposed approach and existing methods, discussing successes, limitations, and interpretations of the results.

6. **Chapter 6 - Conclusions** It summarizes the key findings of the research and reflects on its contributions to the field. It also includes suggestions for future work.

2 Related Works

In this chapter, an overview of the literature on the topic of SR will be provided, starting from traditional SR methods applied to optical images and progressing to the state of the art techniques in the field of remote sensing. The discussion will emphasize the significant advancements in SR technologies and how these techniques have evolved to meet the specific challenges of remote sensing applications. Moreover, particular attention will be given to the impact of microclimate analysis, which involves studying localized atmospheric variables that can influence environmental phenomenons such as urban heat island. The purpose of this analysis is to highlight how SR can be a powerful tool for improving the spatial resolution of climate data and enhancing the precision of microclimate monitoring.

2.1 Super Resolution for Optical Images

The foundation of SR can be traced back to the image restoration techniques developed in the 1960s. One of the earliest examples of this concept emerged from the restoration of images transmitted by remote spacecraft from the U.S. National Aeronautics and Space Administration (NASA). Although the specifics of these image restoration procedures were never widely disclosed, they significantly enhanced the quality of the images received from these spacecraft. Following this early work, a burst of research activity in image restoration emerged, evolving into what is today known as SR, a relevant area of research in Computer Vision [15].

Common methods and algorithms that preceded Neural Networks are based on dictionaries as collections of basis elements to represent an image as a sparse linear combination of these elements [16, 17]. Additionally, mathematical methods such as bicubic interpolation are widely used for image resizing and are often employed as a benchmark for comparison against newer techniques, including those explored in this study.

Dong et al. (2014) [18] introduced a pivotal deep learning approach to SR represented as a deep CNN (SRCNN). The method optimizes all layers jointly, demonstrating state-of-the-art restoration quality with a lightweight structure and achieving fast speed. This work marked a significant shift from traditional methods to utilizing CNNs for SR. Figure 2.1, adapted from the original paper [18], illustrates the overall architecture of SRCNN and the purpose of each layer.

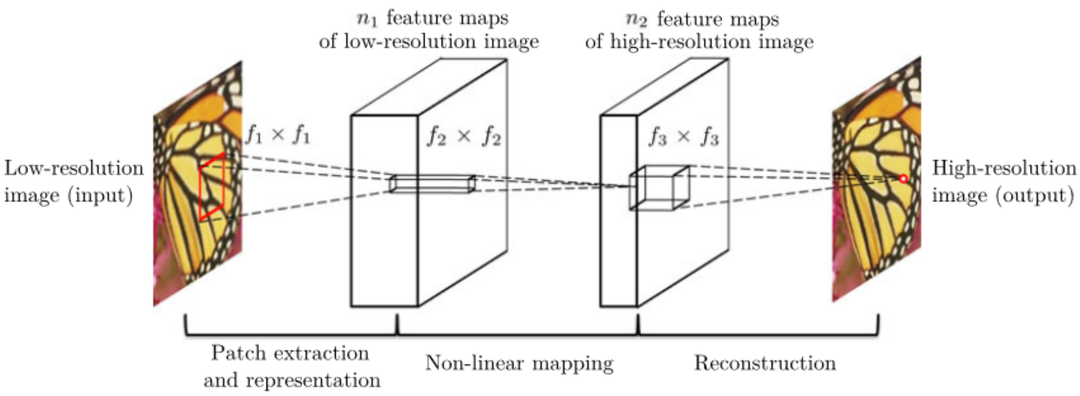
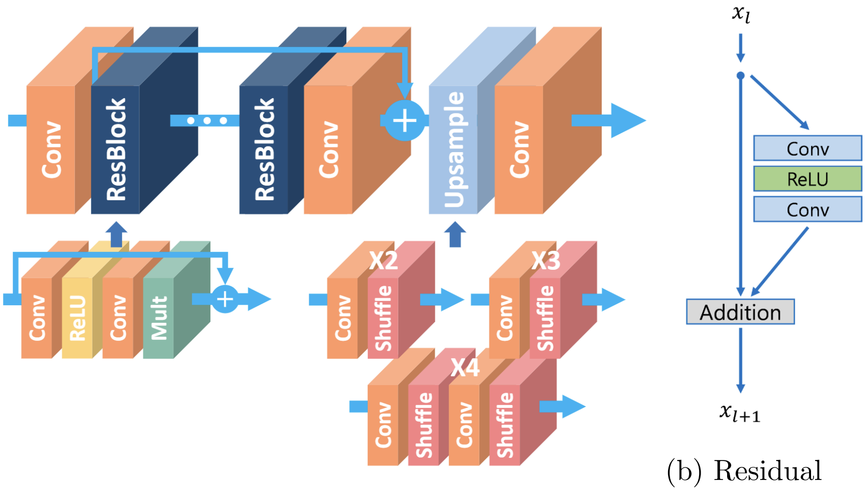


Figure 2.1: SRCNN architecture and the purpose of each layer [18]

On the foundation of this work multiple recent studies have proposed new CNN-based super resolution models, that result to be faster, like the network proposed by *Dong et al. (2016)* [19], and/or more accurate with deeper architectures (VDSR) or similarly multiple recursive layers (DRCNN) as proposed by *Kim et al.* (2016) [20, 21]. Some other architectures, from which the new proposed architectures take inspiration from, are more recent and utilize the concept of residual blocks. *Lim et al.* [22] propose an Enhanced Deep Super Resolution Network (EDSR) that optimizes residual networks by removing some unnecessary components, i.e. batch normalizations, as shown in Figure 2.2.



(a) EDSR architecture

(b) Residual Block

Figure 2.2: EDSR architecture for 2x, 3x and 4x scaling factors [22]

Ledig et al. [23], shown in Figure 2.3 implement a Super Resolution GAN (SRGAN), in which the generator is similar to the EDSR described above.

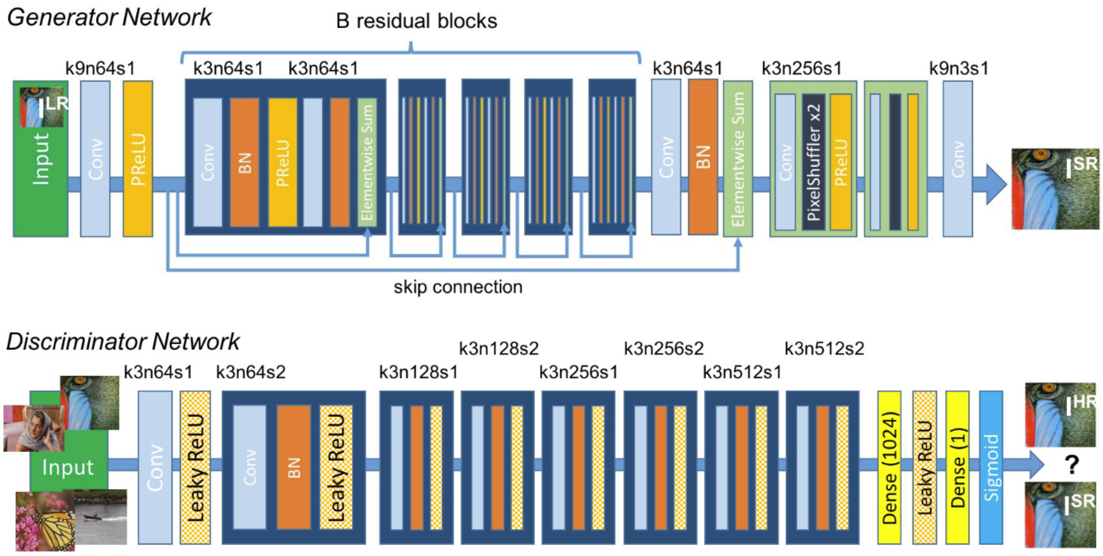


Figure 2.3: SRGAN architecture [23]

Wang et al. [24] propose an Enhanced Super Resolution Generative Adversarial Network (ESRGAN) that improves upon the SRGAN by introducing the Residual-in-Residual Dense Block (RRDB) [25], in which each layer's output is connected to every subsequent layer in the block. The changes between SRGAN and ESRGAN are represented in Figure 2.4

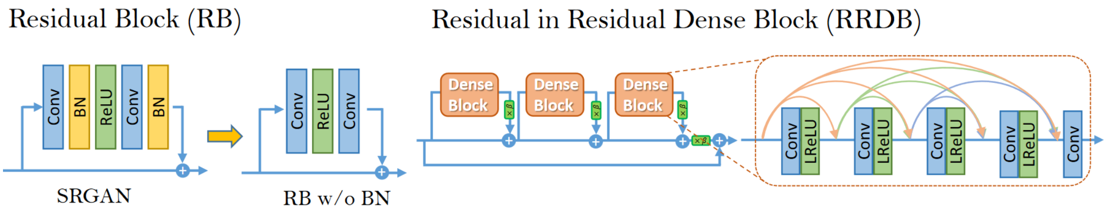


Figure 2.4: Differences between ESRGAN and SRGAN [24]

2.2 Super Resolution for Remote Sensing

While this study primarily focuses on using SR Neural Networks to enhance LST data, numerous studies have explored mapping air temperature and other remote sensing data with various machine learning models, such as regression and random forests. Although not directly aimed at LST SR, these works offer valuable insights into mapping remote sensing variables, which can be generalized to LST estimation. The task of air temperature mapping for example can be seen as potentially similar to LST estimation since both involve complex relationships between temperature and environmental variables such as vegetation, surface cover, and urban heat island effects. This conceptual similarity justifies an analysis of these related works, as the methodologies and findings in air temperature mapping can be adapted and extended to the estimation of LST.

Some examples of works that don't only make use of deep neural networks but some other ML tools such as support vector machines and random forests are the ones by *Ho et al.* [9] and *Vulova et al.* [7]. These implementations give insight about which data could be of potential usefulness when mapping air temperature, such NDVI and LC that will be used in this thesis. Besides the aforementioned methods, recent studies have demonstrated the potential of common SR neural networks, typically utilized for optical images, when applied to remote sensing data. Some of them have utilized SRCNN with different scaling factors and for different data, including *Liebel et al.* [26] with 2x on Sentinel-2 multi spectral bands and *Nomura et al.* [27] with 25x on NDVI. *Singh et al.* [28] investigated 30x scaling of precipitation data using SRCNN iteratively multiple times at small 2x steps instead of a big 30x leap. *Pouliot et al.* [29] proposed DCR_SRCNN, an improved version of SRCNN with residual connections and a deeper architecture, achieving a 2.3 scaling ratio for red, NIR, and SWIR bands of Sentinel-2 and Landsat data. *Izumi et al.* [30] tested SRCNN, RRDBNet (multiple residual and dense blocks), and ESRGAN for 4x scaling on ocean temperature data. *Nguyen et al.* [31] proposed a new U-Net for 4x scaling of LST data without leveraging other auxiliary data other than LST itself. Finally *Yasuda et al.* [10] proposed one of the first SRCNN networks that employ auxiliary data for remote sensing SR by implementing SE-SRCNN, a network that proposes an attention mechanism able to grasp the importance of each individual channel. As shown in Figure 2.5 takes inspiration from the SRCNN architecture and its main innovations are a skip connection and the squeeze-and-excitation (SE) module that serves as an attention mechanism. The skip connection makes the training process more stable and faster, and the SE block improves the accuracy of SR.

Most of the previously mentioned works don't take into consideration auxiliary data for the SR task, which is exactly what we want to experiment in the thesis. For this purpose here are described some pertinent studies, including those

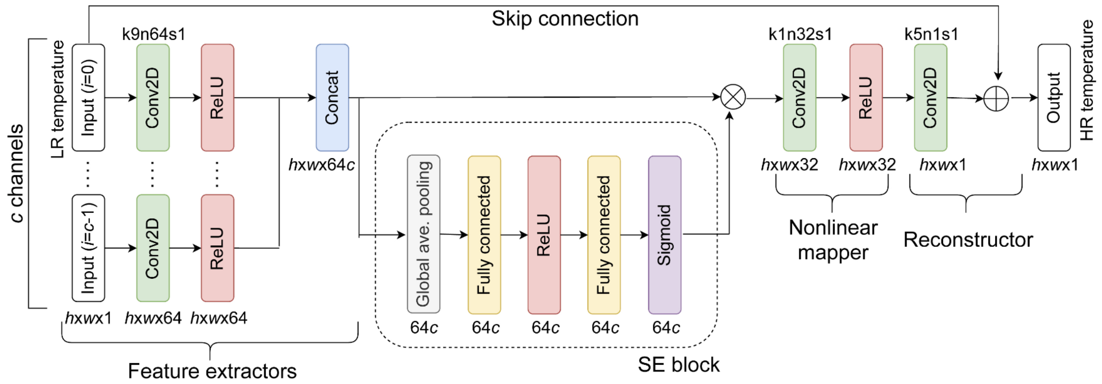
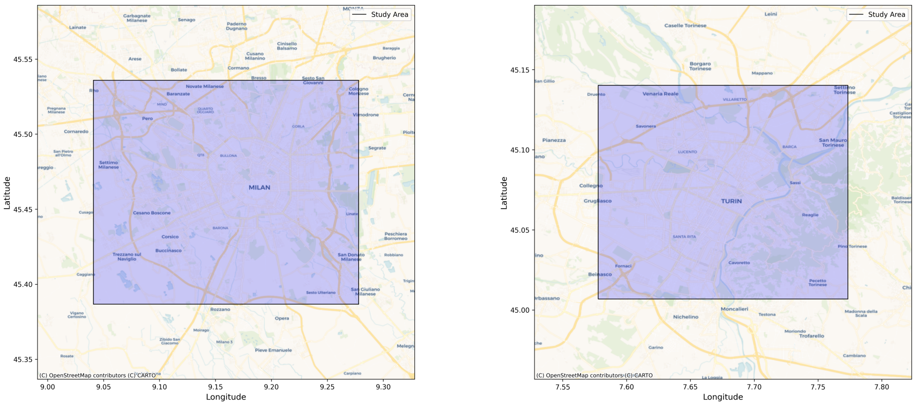


Figure 2.5: SE-SRCNN architecture proposed by *Yasuda et al.* [10]

by *Shen et al.* [8], *Yang et al.* [11], and *Han et al.* [12], that illustrate how auxiliary data—encompassing remote sensing, sociological, and geographical information—can significantly improve air temperature predictions in microclimate environments around weather stations. Their research demonstrates that incorporating a diverse range of attributes, such as NDVI, elevation, land cover (LC), population density, road density, wind speed, soil moisture content, albedo, humidity, pressure, solar radiation, and temporal factors like day of the year, as well as geographic coordinates (latitude and longitude), provides valuable insights that enhance predictive accuracy.

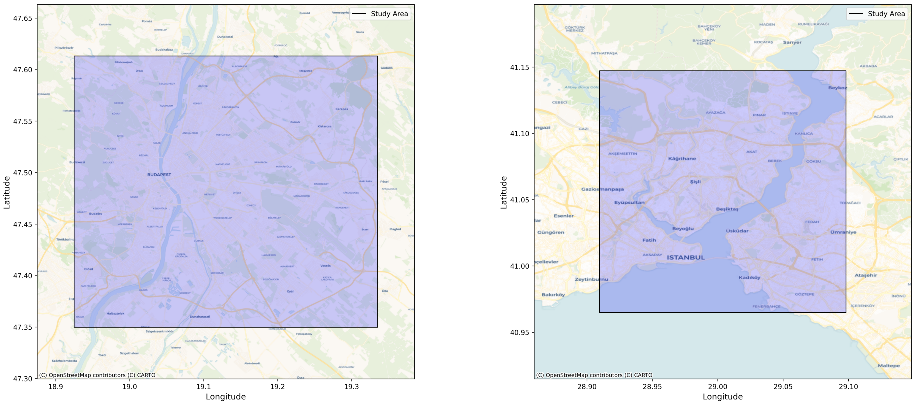
3 Datasets

The study utilizes various remote sensing datasets that drive the training of the SR models. The two main data sources are Sentinel-3 and Landsat 8 that provide LST directly or surrogate measures such as TOA BT. As far as the auxiliary data this research includes NDVI, DEM, LC and weather data from meteorological stations. All the data are available for the cities of Milan, Turin and Budapest from January 2018 to December 2023 and for Istanbul for the year 2023. The specific areas of study are shown in Figure 3.1.



(a) Milan

(b) Turin



(c) Budapest

(d) Istanbul

Figure 3.1: Areas of study for each city

Table 3.1 provides an overview of all the sources, including their bands/attributes, spatial and temporal resolution.

When combining data from different sources that may cover varying date ranges, an intersection is applied to retain only the dates for which all datasets are available. Moreover missing values (i.e. NaN values) do not account for the loss and consequently for weights update during backpropagation. Before the input is fed to the network NaN values are filled with the mean for each specific channel.

Dataset are loaded from different file types: Sentinel-3, Landsat 8, NDVI, DEM and LC are all opened from a single or multiple GeoTIFF file/s. GeoTIFF is a TIFF (Tagged Image File Format) file format widely used for satellite imagery that contains geographic metadata that enables the image to be tied to a specific location on Earth's surface, allowing it to be used in geographic information systems (GIS). The weather data is stored in CSV format, as it lacks spatial information; the daily values remain consistent across the entire city.

Additionally we can classify data sources into four distinct types derived from the combination of two discrete classes: static versus dynamic and raster versus scalar. Sentinel-3, Landsat 8 and NDVI are dynamic rasters, since their measurements change in time and space. DEM and LC are static rasters that can vary only spatially but they are almost static through time (changes happen in the scale of years). Weather data include dynamic scalars, measurements that change through time but are fixed spatially for each city (the value is propagated on all rasters points).

Finally, from both Sentinel-3 and Landsat 8 rasters, it is possible to extract the day of the year (doy) for each observation, as well as calculate the time difference in seconds (time_shift) between the two satellite measurements. These two additional information are taken into consideration to give ulterior context to the input. They both are dynamic scalars.

Data Source	**Description**	**Bands/ Attributes**	**Unit of Measurement**	**Spatial Resolution**	**Time Resolution**
Sentinel-3 SLSTR SL_1_RBT	TOA Radiance and Brightness Temperature	S6	K	1 km	0.5 days
		S7	K		
		S8	K		
		F1	K		
		F2	K		
Landsat 8 Collection 2 LST	Land Surface Temperature	LST	K	100 m	8 days
DEM, Copernicus	Digital Elevation Model	DEM	m	10 m	-
NDVI, Sentinel-2	Normalized Difference Vegetation Index	NDVI	-	10 m	1 month
LC Urban Atlas CLMS	Land Cover	LC	-	10 m	-
Weather Visual Crossing	Air temperature	temp	°C	-	1 day
	Dew point	dew	°C		
	Humidity percentage	humidity	%		
	Amount of rain	precip	mm		
	Snowfall measure	snow	cm		
	Accumulated snow depth	snowdepth	cm		
	Wind speed	windspeed	km/h		
	Wind direction	winddir	°		
	Atmospheric pressure	pressure	hPa		
	Cloud coverage percentage	cloudcover	%		
	Solar radiation	solarradiation	W/m^2		
	Emitted solar energy	solarenergy	MJ/m^2		
	Ultraviolet radiation index	uvindex	-		
Calculated	Day of the Year	doy	-	-	1 day
	Difference between time observation	time_shift	s		

Table 3.1: Data sources with their bands, unit of measurement, time and spatial resolution

3.1 Sentinel-3

Sentinel-3 is a mission of Earth observation satellites developed and operated by the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) as part of the European Union's Copernicus Programme. The mission began with the launch of Sentinel-3A on February 16, 2016, followed by Sentinel-3B on April 25, 2018. Sentinel-3C and Sentinel-3D will eventually replace the two already in orbit and their launch is currently still in the planning stages [32]. Providing accurate, timely and easily accessible information to improve the management of the environment, understand and mitigate the effects of climate change and ensure civil security, Copernicus is the most ambitious Earth observation programme to date.

Sentinel-3 satellites are equipped with a suite of advanced sensors that capture data across multiple spectral bands (each one refers to a specific range of wavelengths in the electromagnetic spectrum that a sensor can capture), and at various spatial resolutions. The combination of the sensors on a satellite is also called payload, shown in Figure 3.2. These sensors include instruments for measuring ocean color, sea surface temperature, land surface temperature, vegetation health and atmospheric composition. In particular its instrument packages include a Sea and Land Surface Temperature Radiometer (SLSTR), an Ocean and Land Colour Instrument (OLCI) and the Altimetry Surface Topography Mission (STM) payload.

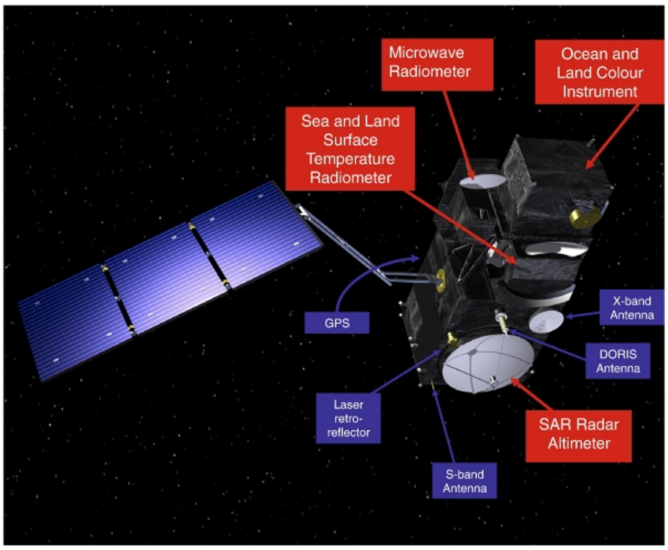


Figure 3.2: Sentinel-3 satellite and payloads [5]

The instrument that provides the data for the thesis is SLSTR, which measures in nine spectral channels and two additional fire bands that have their radiometric measurements expressed in Top Of Atmosphere (TOA) brightness temperatures (BT). Sentinel-3 delivers measurements at a spatial resolution of 500 m for visible/near-infrared and short-wavelength infrared channels and at 1 km for the thermal infrared channels.

The Sentinel-3 satellites operate in a near-polar, sun-synchronous orbit, which ensures that they pass near the poles on each orbit and cross any given point on Earth's surface at the same local solar time each day. This synchronization allows for uniform solar illumination, ensuring consistent lighting conditions for each image or data sample, which is crucial for accurate temporal comparisons. Such consistency is vital for monitoring environmental changes, including deforestation, urban expansion, and seasonal vegetation dynamics. The two operational Sentinel-3 satellites, Sentinel-3A and Sentinel-3B, are phased +/-140° apart along their identical orbital paths, significantly reducing revisit times. For example, SLSTR achieves a revisit time of less than one day at the equator. With only one spacecraft, the mean global coverage revisit time for dual-view SLSTR observations at the equator is 1.9 days, which reduces to 0.9 days with both spacecraft in operation. Due to orbital convergence, these revisit times decrease at higher latitudes, providing more frequent coverage, such as twice-daily observations for latitudes similar to those of European cities included in this study. This frequent coverage is particularly advantageous for monitoring dynamic environmental changes in these regions [5].

Although the Sentinel-3 satellites offer a short revisit time, their spatial resolution does not match the finer detail provided by the Landsat 8 mission. This limitation in spatial resolution affects the precision of certain applications that require high-detail imagery. To address this challenge, this thesis explores the use of SR techniques to enhance the spatial resolution of Sentinel-3 data. The goal is to bridge the gap between the temporal frequency of Sentinel-3 and the spatial quality of Landsat 8, thereby improving the accuracy of environmental monitoring.

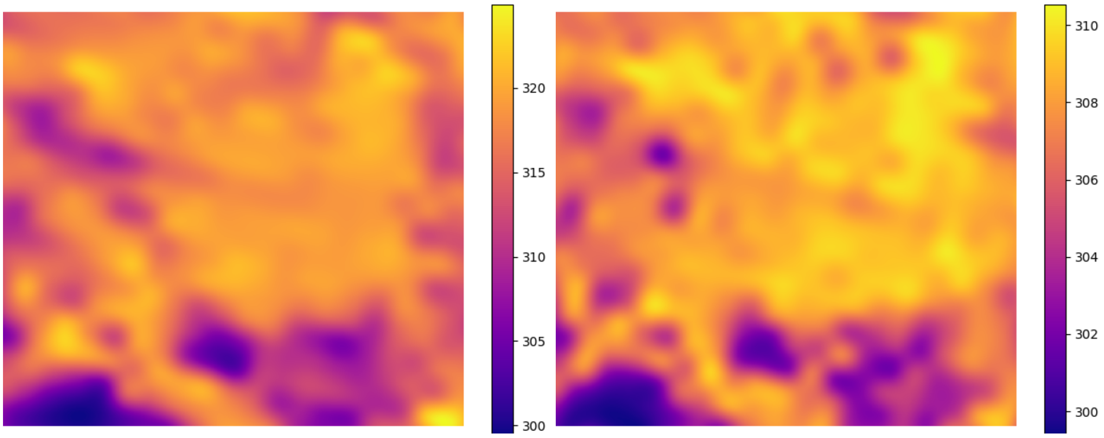
Operationally, the processing of SLSTR data is divided into three major levels, each producing distinct data products. Level-0 processing organizes and checks raw data from the instrument's source packets, creating internal products (SL_0_SLT) that are not distributed to users but serve as the foundation for further processing. Level-1 processing generates the SL_1_RBT product, which provides radiance and brightness temperature data for each pixel, along with annotations like pixel location and surface properties. This product includes data from bands S1-S6 at 500 m resolution, though this study focuses solely on the 1 km resolution bands S7, S8, S9 (thermal infrared), and F1, F2 (visible/near-infrared fire bands).

A list of the 11 SLSTR SL_1_RBT bands is shown in Figure 3.3, along with their wavelength, resolution and brief description. At Level-2, data is further refined to generate specialized products such as land surface temperature (SL_2_LST), fire radiative power (SL_2_FRP), or aerosol optical depth (SL_2_AOD). However, this study bypasses the LST product and focuses instead on the direct use of Level-1B radiance and brightness temperature data from bands S7, S8, S9, F1, and F2, examples of which are shown in Figure 3.4.

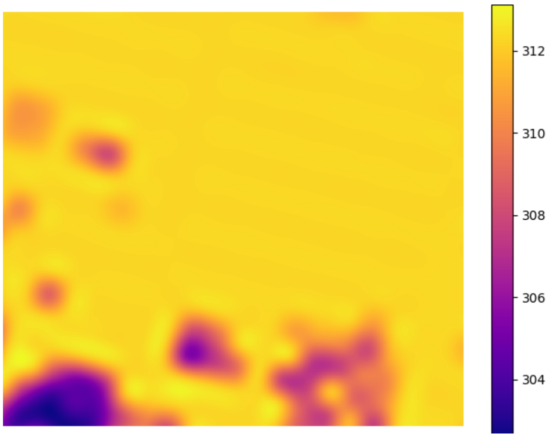
Although using directly the LST product from the original dataset would be the best approach for our case study, this was not achievable given the limitations inherent in the working context, thus creating the need of retrieving it from TOA BT and potentially with auxiliary information such as NDVI. However, this problem and its solution is discussed later in Subsection 4.1.3 [5].

Band	Central Wavelength (nm)	Bandwidth (nm)	Function	Comments		Resolution (m)
S1	554.27	19.26	>Cloud screening, vegetation monitoring, aerosol	VNIR	Solar Reflectance Bands	500
S2	659.47	19.25	NDVI, vegetation monitoring, aerosol			
S3	868.00	20.60	NDVI, cloud flagging,Pixel co-registration			
S4	1374.80	20.80	Cirrus detection over land	SWIR		
S5	1613.40	60.68	loud clearing, ice, snow,vegetation monitoring			
S6	2255.70	50.15	Vegetation state and cloud clearing			
S7	3742.00	398.00	SST, LST, Active fire	Thermal IR Ambient bands (200 K -320 K)		1000
S8	10854.00	776.00	SST, LST, Active fire			
S9	12022.50	905.00	SST, LST			
F1	3742.00	398.00	Active fire	Thermal IR fire emission bands		
F2	10854.00	776.00	Active fire			

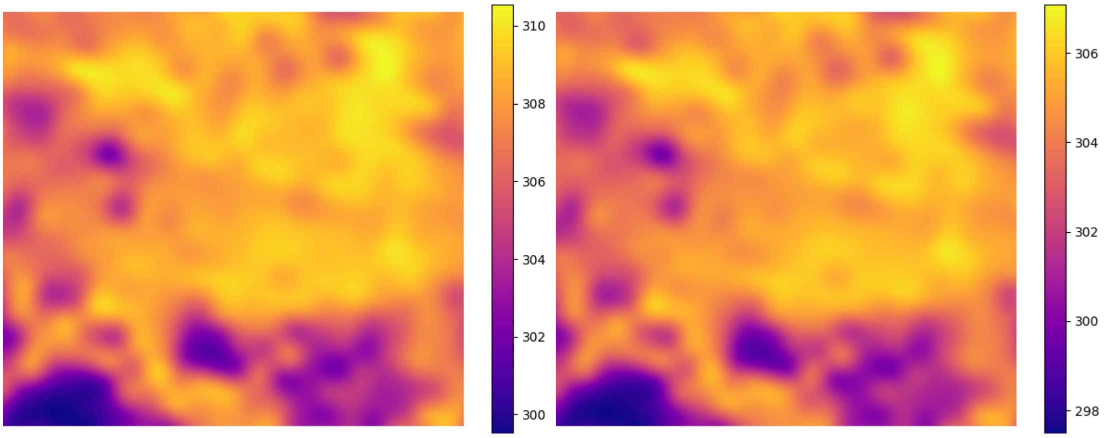
Figure 3.3: Sentinel-3 SLSTR bands along with wavelength and resolution [5]



(a) S7 (b) S8



(c) S9



(d) F1 (e) F2

Figure 3.4: Examples of Sentinel-3 rasters. a, b and c represent the three thermal IR ambient bands, d and e instead the thermal IR fire emission bands. The rasters where taken at the same time on the city of Milan

3.2 Landsat 8

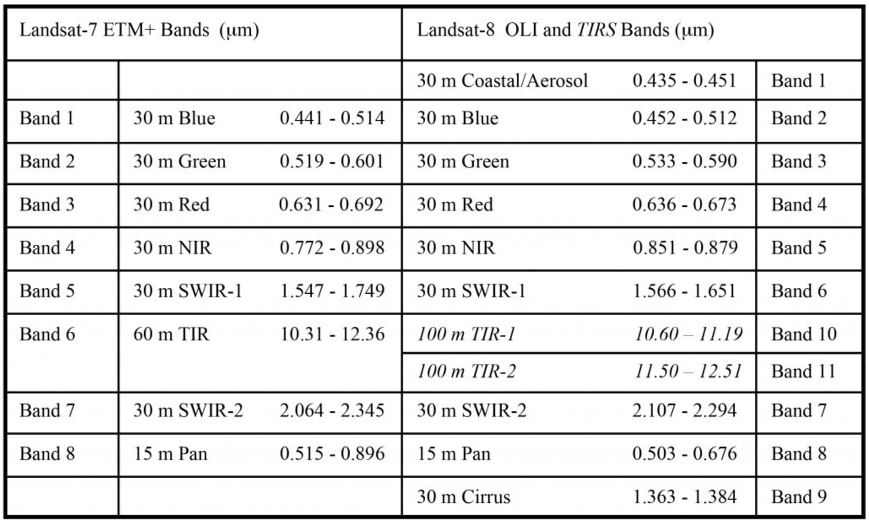
The Landsat program has been a cornerstone of Earth observation since its inception in 1972, providing invaluable data for monitoring land use, climate change, and natural resource management. Landsat 8 (formerly the Landsat Data Continuity Mission, LDCM), launched on an Atlas-V rocket from Vandenberg Air Force Base, California, on February 11, 2013, is a collaboration between NASA and the U.S. Geological Survey. It continues the legacy of its predecessors in delivering high-quality imagery for various applications, with future missions such as Landsat 9, launched in September 2021, aiming to further enhance this capability.

Similar to Sentinel-3, Landsat 8 operates in a sun-synchronous, near-polar orbit, providing moderate-resolution measurements ranging from 15 m to 100 m, depending on the spectral frequency. The two operational satellite missions, Sentinel-3 and Landsat 8, are temporally separated by approximately 8 days, resulting in measurements that are each 8 days apart. This configuration facilitates more frequent observations of the Earth's surface, which is essential for effectively monitoring environmental changes over time.

The satellite's payload consists of two instruments: the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS), which together encompass a total of 12 spectral bands. The OLI provides high-resolution optical imagery for land use and land cover classification, while the TIRS measures thermal infrared radiation to assess surface temperature by converting the thermal radiance captured by the sensor into temperature values, accounting for surface emissivity and atmospheric conditions. The Landsat 8 LST product will serve as a target for SR [33].

The Landsat archive is organized into a tiered collection management structure, with two main collections: Landsat Collection 1 encompasses all Level-1 data acquired from Landsat satellites 1 through 8 since 1972. This collection ensures a consistent archive of known data quality, facilitating continuous improvement of the archive and access to data as it is acquired. It represents a significant change in the management of the Landsat archive, providing users with a reliable source of data for various applications. However, as of December 30, 2022, Landsat Collection 1 data and products are no longer available for download from the United States Geological Survey (USGS). Landsat Collection 2 marks the second major reprocessing event of the USGS Landsat Level-1 archive and includes all Landsat satellites from 1 to 9. Collection 2 introduces several product improvements that leverage advancements in data processing and algorithm development. A notable feature of Collection 2 is the enhanced absolute geolocation accuracy of the global ground reference dataset used in the Landsat Level-1 processing flow.

Additionally, Collection 2 incorporates updated global digital elevation models, calibration and validation updates, and provides global Level-2 products such as surface reflectance and LST (used in this study), covering the period from 1982 to the present [3]. An example of LST raster is shown in Figure 3.6.



Landsat-7 ETM+ Bands (μm)			Landsat-8 OLI and *TIRS* Bands (μm)		
			30 m Coastal/Aerosol	0.435 - 0.451	Band 1
Band 1	30 m Blue	0.441 - 0.514	30 m Blue	0.452 - 0.512	Band 2
Band 2	30 m Green	0.519 - 0.601	30 m Green	0.533 - 0.590	Band 3
Band 3	30 m Red	0.631 - 0.692	30 m Red	0.636 - 0.673	Band 4
Band 4	30 m NIR	0.772 - 0.898	30 m NIR	0.851 - 0.879	Band 5
Band 5	30 m SWIR-1	1.547 - 1.749	30 m SWIR-1	1.566 - 1.651	Band 6
Band 6	60 m TIR	10.31 - 12.36	*100 m TIR-1*	*10.60 – 11.19*	Band 10
			100 m TIR-2	*11.50 – 12.51*	Band 11
Band 7	30 m SWIR-2	2.064 - 2.345	30 m SWIR-2	2.107 - 2.294	Band 7
Band 8	15 m Pan	0.515 - 0.896	15 m Pan	0.503 - 0.676	Band 8
			30 m Cirrus	1.363 - 1.384	Band 9

Figure 3.5: Landsat 7 ETM+ bands and Landsat 8 OLI and TIRS bands with their respective wavelength [33]

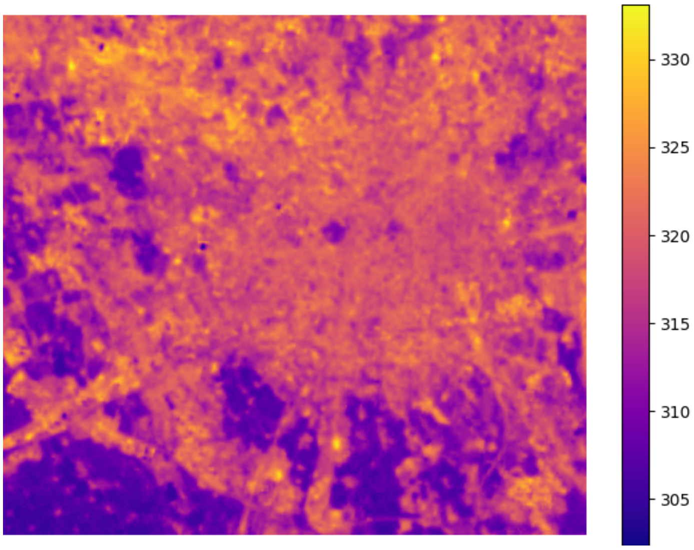


Figure 3.6: Example of Landsat 8 raster representing LST on the city of Milan

3.3 Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is a widely used vegetation index calculated from the reflectance values of different spectral bands. It can be derived from various datasets, including Landsat missions and MODIS (Moderate Resolution Imaging Spectroradiometer). However, in this study, we specifically utilize products from the Sentinel-2 satellites, which operate with the MultiSpectral Instrument (MSI) that captures information across 13 spectral bands, ranging from the visible to the shortwave infrared spectrum. NDVI specifically utilizes Band 4 (B4), corresponding to the red light region (RED), and Band 8 (B8), which captures light in the near-infrared region (NIR). An example of NDVI raster is provided in Figure 3.7 for the city of Milan.

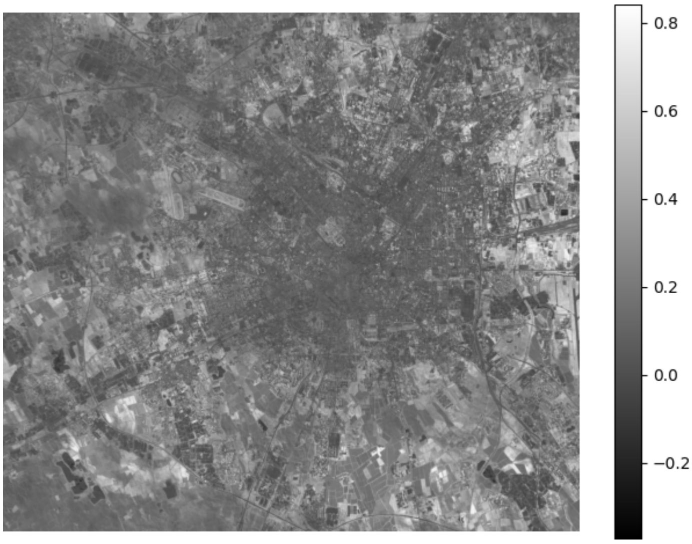


Figure 3.7: Example of NDVI raster on the city of Milan

The Normalized Difference Vegetation Index is defined as:

$$NDVI := \text{Index}(NIR, RED) = \frac{NIR - RED}{NIR + RED}$$

For Sentinel-2 in particular:

$$NDVI := \text{Index}(B8, B4) = \frac{B8 - B4}{B8 + B4}$$

NDVI is a crucial tool in climate change and environmental studies since it provides valuable insights into plant health and biomass. This index helps scientists monitor vegetation changes over time, assess the impact of climate change on ecosystems, and guide sustainable land management practices. NDVI's ability to offer real-time data makes it indispensable for tracking environmental changes and implementing adaptive strategies to mitigate the adverse effects of climate change [34].

3.4 Digital Elevation Model

The Digital Elevation Model (DEM) dataset [35] used in this work is derived from the Copernicus mission. Specifically, the Copernicus DEM is a Digital Surface Model (DSM) that captures the Earth's surface, including buildings, infrastructure, and vegetation. It is available in three different resolutions: EEA-10, GLO-30, and GLO-90, with the EEA-10 version, utilized in this work, providing European coverage at a 10 m resolution in GeoTIFF format. Data was acquired through the TanDEM-X mission between 2011 and 2015 [36]. An example of DEM raster is reported in Figure 3.8

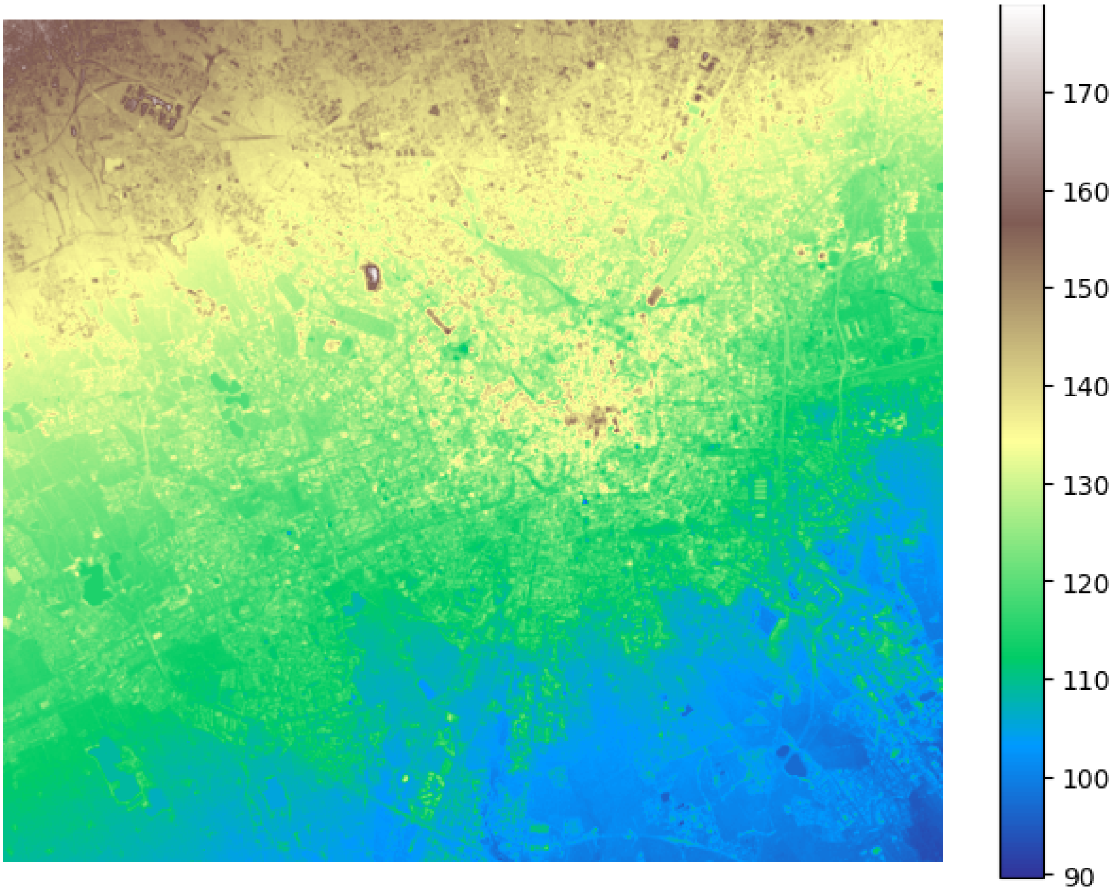
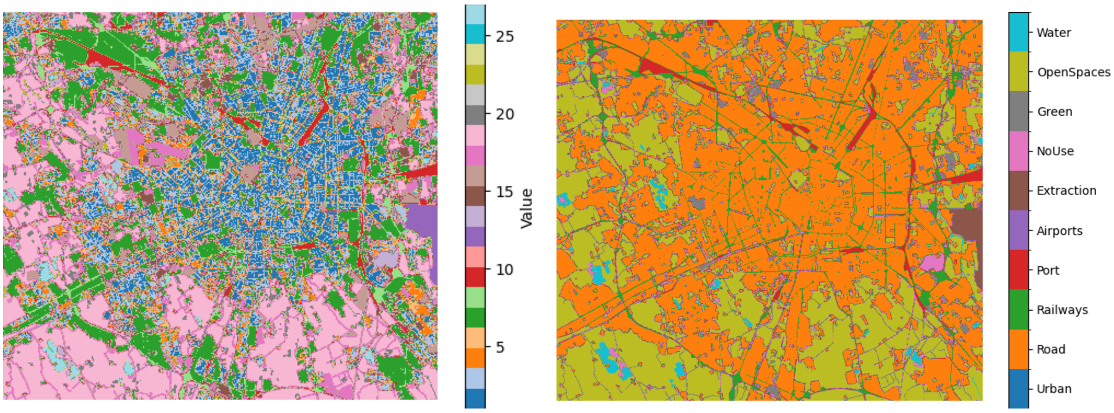


Figure 3.8: DEM raster for the city of Milan

3.5 Land Cover

Land Cover data is provided by Copernicus which provides a dataset each 6 years. The most recent in their collection is Urban Atlas Land Cover/Land Use (LU/LC) 2018 [37] that offers detailed land cover and land use information, covering the 38 EEA (European Economic Area) countries and the United Kingdom [38]. It classifies land cover and land use into two categories: 17 urban classes with the Minimum Mapping Unit (MMU) of 0.25 ha and 10 rural classes with the MMU of 1 ha for a grand total of 27 classes which are grouped into 10 more generic classes providing a simpler representation. LU/LC is provided as vector data but for simplicity it has been rasterized into a GeoTIFF format with a spatial resolution of 10 m, which will then be transformed into a $10 \times H \times W$ raster with a one hot encoding on each channel for each LU/LC grouped class. The native classes and their grouped product are listed in Table 3.2. Moreover a representation of the Land Cover of the city of Milan is shown in Figures 3.9a and 3.9b respectively before and after the grouping.

In this study, the hypothesis is that urban areas exhibit higher temperatures compared to rural and green/open space areas due to the urban heat island effect, where dense, impervious surfaces like asphalt and concrete absorb and retain more heat. In contrast, natural areas, such as forests, parks, and open spaces, tend to be cooler due to higher vegetation cover. Additionally, water bodies are expected to be cooler. Given that the land cover data used in this analysis has a high spatial resolution of 10 meters, it enables more detailed mapping of these temperature variations. This finer resolution should provide valuable input for the super-resolution (SR) model enhancing the accuracy of temperature predictions.



(a) Before (b) After

Figure 3.9: Urban Atlas LU/LC for the city of Milan before and after grouping into 10 classes

Class	Code	Land Use	Mapped Land Use
0	11100	Continuous Urban Fabric	Urban
1	11210	Discontinuous Dense Urban Fabric	Urban
2	11220	Discontinuous Medium Density Urban Fabric	Urban
3	11230	Discontinuous Low Density Urban Fabric	Urban
4	11240	Discontinuous Very Low Density Urban Fabric	Urban
5	11300	Isolated Structures	Urban
6	12100	Industrial, commercial, public, military and private units	Urban
7	12210	Fast transit roads and associated land	Road
8	12220	Other roads and associated land	Road
9	12230	Railways and associated land	Railways
10	12300	Port areas	Port
11	12400	Airports	Airports
12	13100	Mineral extraction and dump sites	Extraction
13	13300	Construction sites	Urban
14	13400	Land without current use	No Use
15	14100	Green urban areas	Green
16	14200	Sports and leisure facilities	Urban
17	21000	Arable land (annual crops)	Green
18	22000	Permanent crops (vineyards, fruit trees, olive groves)	Green
19	23000	Pastures	Green
20	24000	Complex and mixed cultivation patterns	Green
21	25000	Orchards at the fringe of urban classes	Green
22	31000	Forests	Green
23	32000	Herbaceous vegetation associations	Green
24	33000	Open spaces with little or no vegetation	Open Spaces
25	40000	Wetland	Water
26	50000	Water bodies	Water

Table 3.2: Land Use Classes, Codes, and Corresponding Mapped Categories

3.6 Weather data

The weather dataset is built upon Visual Crossing API [39]. Visual Crossing has been a leading provider of weather data and enterprise analysis tools from 2003. Historical weather data sets within Visual Crossing Weather Data are created by analyzing the weather stations in the proximity of the requested location. The maximum distance that the system will search is 50 miles. The weather records of the different stations within the specified range will be statistically combined up and generate a single value. Records for weather stations closer to the requested location will be given more weight than the records for weather stations that are further away.

The list of the stations and their relative distance from the request point, also referred in this text as "Selected Point", is provided in Table 3.3. Moreover a map with the position of the stations is shown in Figure 3.10, along with the Selected Point which for each city represent the city center.

For the purpose of this work, it is useful to analyze the correlation between each weather attribute and land surface temperature. Therefore, a correlation heatmap was calculated and reported in Figure 3.11. A high absolute value of correlation with air temperature may indicate that the attribute is a strong indicator of heat or cold in a specific area. However, a limitation arises from the fact that, for each city, there is only one unique daily value that represents every pixel, which could reduce the relevance of weather data for improving our task.

Station Name	**Distance (m)**
LIML (Linate Airport)	8085.0
LINATE, IT	7841.0
DW6875 Besate IT	23956.0
CAMERI, IT	40577.0
LIMC (Milan Malpensa Airport)	39076.0
LIME (Milan Bergamo Airport)	45543.0
CW7783 Como IT	28723.0
MALPENSA, IT	39729.0
STABIO, SZ	46711.0
LUGANO, SZ	63310.0
MONTE GENEROSO, SZ	53364.0
BERGAMO ORIO AL SERIO, IT	46717.0
PIACENZA, IT	74971.0

Table 3.3: Distances of Weather Stations from the Selected Point

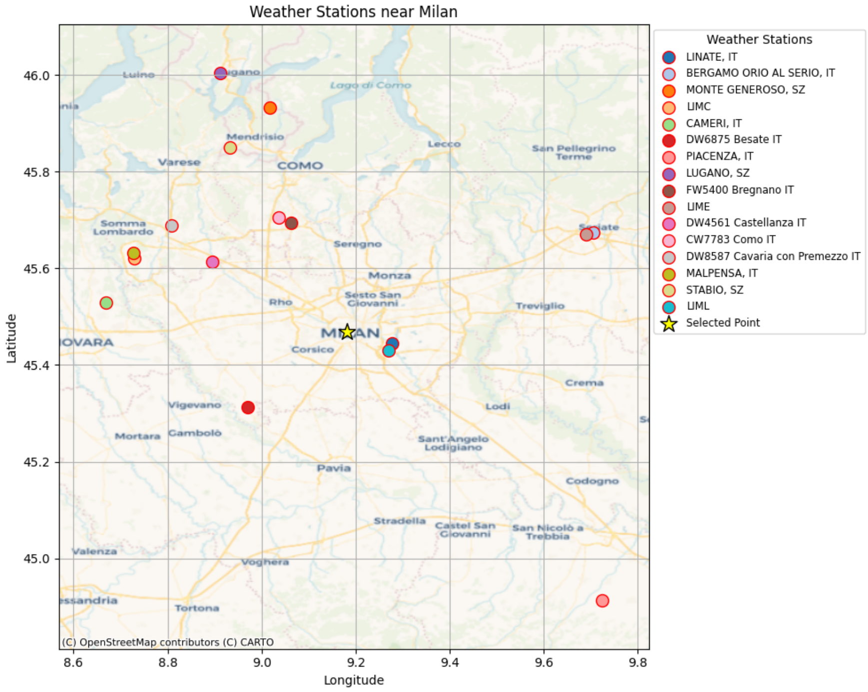
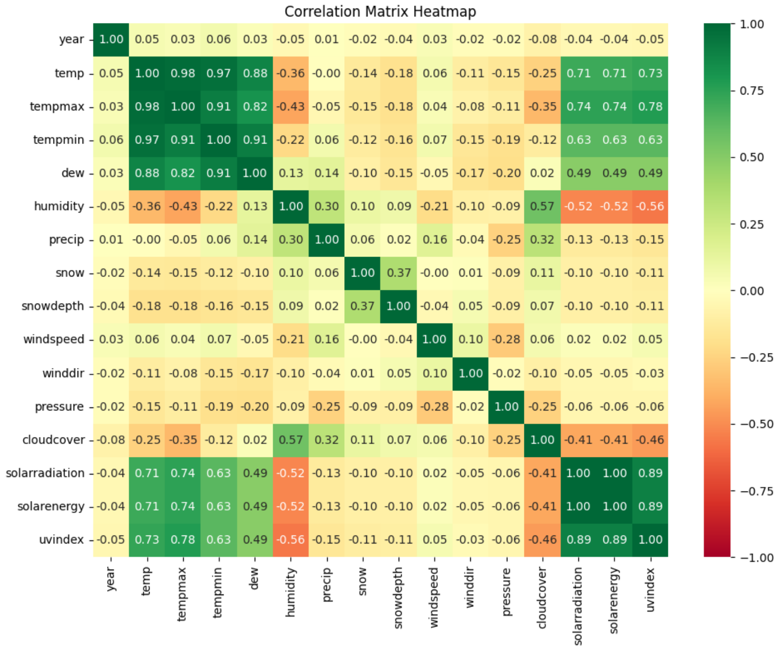


Figure 3.10: Weather stations near Milan used by Visual Crossing to provide weather data


Correlation Matrix Heatmap

	year	temp	tempmax	tempmin	dew	humidity	precip	snow	snowdepth	windspeed	winddir	pressure	cloudcover	solarradiation	solarenergy	uvindex
year	1.00	0.05	0.03	0.06	0.03	-0.05	0.01	-0.02	-0.04	0.03	-0.02	-0.02	-0.08	-0.04	-0.04	-0.05
temp	0.05	1.00	0.98	0.97	0.88	-0.36	-0.00	-0.14	-0.18	0.06	-0.11	-0.15	-0.25	0.71	0.71	0.73
tempmax	0.03	0.98	1.00	0.91	0.82	-0.43	-0.05	-0.15	-0.18	0.04	-0.08	-0.11	-0.35	0.74	0.74	0.78
tempmin	0.06	0.97	0.91	1.00	0.91	-0.22	0.06	-0.12	-0.16	0.07	-0.15	-0.19	-0.12	0.63	0.63	0.63
dew	0.03	0.88	0.82	0.91	1.00	0.13	0.14	-0.10	-0.15	-0.05	-0.17	-0.20	0.02	0.49	0.49	0.49
humidity	-0.05	-0.36	-0.43	-0.22	0.13	1.00	0.30	0.10	0.09	-0.21	-0.10	-0.09	0.57	-0.52	-0.52	-0.56
precip	0.01	-0.00	-0.05	0.06	0.14	0.30	1.00	0.06	0.02	0.16	-0.04	-0.25	0.32	-0.13	-0.13	-0.15
snow	-0.02	-0.14	-0.15	-0.12	-0.10	0.10	0.06	1.00	0.37	-0.00	0.01	-0.09	0.11	-0.10	-0.10	-0.11
snowdepth	-0.04	-0.18	-0.18	-0.16	-0.15	0.09	0.02	0.37	1.00	-0.04	0.05	-0.09	0.07	-0.10	-0.10	-0.11
windspeed	0.03	0.06	0.04	0.07	-0.05	-0.21	0.16	-0.00	-0.04	1.00	0.10	-0.28	0.06	0.02	0.02	0.05
winddir	-0.02	-0.11	-0.08	-0.15	-0.17	-0.10	-0.04	0.01	0.05	0.10	1.00	-0.02	-0.10	-0.05	-0.05	-0.03
pressure	-0.02	-0.15	-0.11	-0.19	-0.20	-0.09	-0.25	-0.09	-0.09	-0.28	-0.02	1.00	-0.25	-0.06	-0.06	-0.06
cloudcover	-0.08	-0.25	-0.35	-0.12	0.02	0.57	0.32	0.11	0.07	0.06	-0.10	-0.25	1.00	-0.41	-0.41	-0.46
solarradiation	-0.04	0.71	0.74	0.63	0.49	-0.52	-0.13	-0.10	-0.10	0.02	-0.05	-0.06	-0.41	1.00	1.00	0.89
solarenergy	-0.04	0.71	0.74	0.63	0.49	-0.52	-0.13	-0.10	-0.10	0.02	-0.05	-0.06	-0.41	1.00	1.00	0.89
uvindex	-0.05	0.73	0.78	0.63	0.49	-0.56	-0.15	-0.11	-0.11	0.05	-0.03	-0.06	-0.46	0.89	0.89	1.00

Figure 3.11: Correlation heatmap between each weather feature

As shown in Figure 3.11 dew, solar radiation, solar energy and UV index are the features that have the strongest correlation with respect to temperature measurements. We will thus consider these four to be of relevant information for our models. Moreover some features are strongly correlated (i.e. solar radiation, solar energy and UV index) and it is reasonable to pick only one of these (in our case solar energy) and exclude the others, mitigating redundant information.

Based on this analysis the two most relevant features are dew and solar energy and in Figure 3.12 it is demonstrated how combined they can well discriminate and map the temperature value.

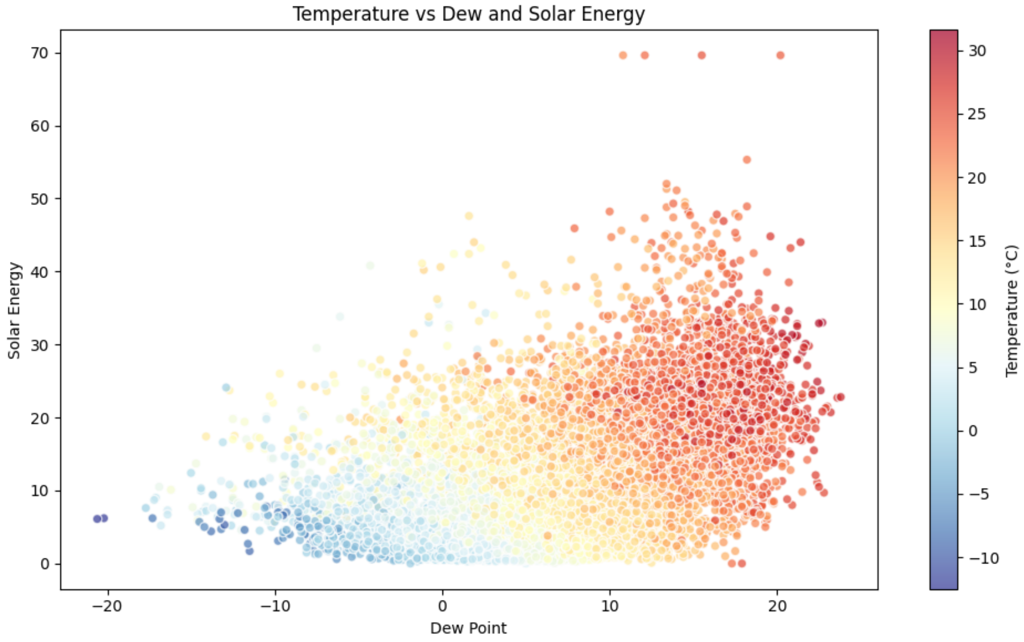


Figure 3.12: Land Surface Temperature in function of dew and solar radiation

3.7 Limitations

In any research endeavor, acknowledging limitations is crucial to provide a comprehensive understanding of the study's findings. Some major limitations related to the datasets of this study are reported below.

1. **Cloud Cover.** The presence of clouds can significantly impact the quality of remote sensing data. Clouds obscure the surface, preventing accurate measurements of BT and LST. While cloud masks are typically utilized to mitigate this issue, they are not provided in the starting dataset. Therefore, simple algorithms will be developed to minimize the effects of clouds as much as possible.

2. **Temporal Shift.** Temporal shifts refer to the discrepancies in timing between the collection of Sentinel-3 data and Landsat 8 measurements. Such shifts are caused by the differences in the orbits of these two satellites that pass over the same region in different periods of time. These inconsistencies can affect the accuracy of the SR model. One possible way of addressing this issue, that is tested in this work, is to provide the model with the temporal shift information to give context to the pair of input-target.

3. **Map of BT to LST.** In this study, the Sentinel-3 dataset consists of bands that represent BT, while the Landsat 8 product provides LST. The process of mapping BT to LST is complex and involves several assumptions and corrections, particularly atmospheric corrections, since BT measurements are influenced by atmospheric conditions. An alternative method proposed in this work is to use a neural network for the mapping, allowing it to learn the mapping function autonomously. However, the primary limitation arises from the inherent discrepancies between these two different measurements, which may affect the overall results.

Each of these limitations will be addressed in the next chapter, in which different solutions that can help mitigate their effect are proposed.

4 Methodology

This chapter describes the deep learning networks proposed to solve the SR task and the techniques adopted to address the dataset limitations described in the previous chapter, specifically the neural network utilized to mitigate the difference between Sentinel-3 BT measurements and Landsat 8 LST product.

4.1 Addressing dataset limitations

In this section, the text delves deeper into the problems mentioned in Chapter 3, proposing different solutions to mitigate their effects. For the sake of clarity, here are reported the three major problems in the utilized datasets:

1. **Cloud Cover.** This refers to the presence of clouds in the region of interest, which obstructs the measurement of BT and LST.
2. **Temporal Shift.** This refers to the difference in time between the measurements from Sentinel-3 and Landsat 8.
3. **Mapping BT to LST.** This refers to the discrepancies between BT and LST measurements from Sentinel-3 and Landsat 8, respectively.

4.1.1 Cloud Cover

Clouds can be obstacles for Sentinel-3 and Landsat 8, since their presence can prevent the correct reading of LST or more generally the spectral bands measures from which the latter is derived. The objective of the thesis imposes the comparison between Sentinel-3 and Landsat 8 LST HR images, this can cause a misleading comparison of instances of rasters that do not match between each other due to the fast dynamic movement of clouds that can alter the scene and thus lead to incorrect evaluation during testing.

For this reason cloud masks are needed to invalidate pixels that do not carry correct information on LST but rather a measurement of temperature of the clouds. The idea is to supplement the mask of null values, such as NaN instances, with the cloud masks of both Sentinel-3 and Landsat 8 in this fashion:

$$\text{mask}_{\text{tot}} = \text{mask}_{\text{NaN}} \vee \text{mask}_{\text{S,clouds}} \vee \text{mask}_{\text{L,clouds}}$$

where $\text{mask}_{\text{S,clouds}}$ and $\text{mask}_{\text{L,clouds}}$ are the cloud masks of Sentinel-3 and Landsat 8 respectively.

In this subsection we propose two methods for generating cloud masks:

1. **Method I - Temperature Threshold.** Clouds generally have lower temperatures compared to the ground, so a simple, fast, and effective method is to apply a threshold of 270 K (approximately 0 °C) to the temperature of each pixel. While this may lead to the exclusion of some valid LST pixels below 270 K, the number of correctly excluded cloud pixels far exceeds these, providing a significant benefit overall.

2. **Method II - Gaussian Distributions.** This method was proposed to extend Method I, addressing the instances for which clouds result in a temperature greater than 270 K. The technique tries to spot patterns in the LST rasters that do not resemble a normal behavior for that specific area. This can be done by defining a cloud-free raster $\boldsymbol{\mathcal{R}}_{\text{CF}}$ and then use it as reference for others LST rasters $\boldsymbol{\mathcal{R}}$ measuring the pixel-wise difference between the two:

   $$\boldsymbol{\mathcal{D}} := d(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{R}}_{\text{CF}}) = \boldsymbol{\mathcal{R}} - \boldsymbol{\mathcal{R}}_{\text{CF}}$$

   The raster (matrix) of distances $\boldsymbol{\mathcal{D}}$ that measures how similar the pixels are between $\boldsymbol{\mathcal{R}}$ and $\boldsymbol{\mathcal{R}}_{\text{CF}}$ can be flattened into a vector that can be interpreted as a distribution, that for simplicity we assume follows a Gaussian distribution.

   $$\boldsymbol{d} = \text{Flatten}(\boldsymbol{\mathcal{D}})$$

   $$\boldsymbol{d} \sim \mathcal{N}(\mu, \sigma^2)$$

   By calculating the mean μ_d and the standard variation σ_d of the distribution we can then detect points of $\boldsymbol{\mathcal{D}}$ with high deviations, the absolute Z-score is greater than a standard deviation threshold σ_T that is set to 3 (i.e. Points more than 3 standard deviations from the mean are excluded); data points that exceed this threshold are considered highly unusual and are typically extreme outliers. To additionally prevent incorrect clouds detection a difference threshold for each month δ_T^i is set to account for temperature rise in the summers and fall in the winters. The cloud masks for Method II for a month i are thus defined as:

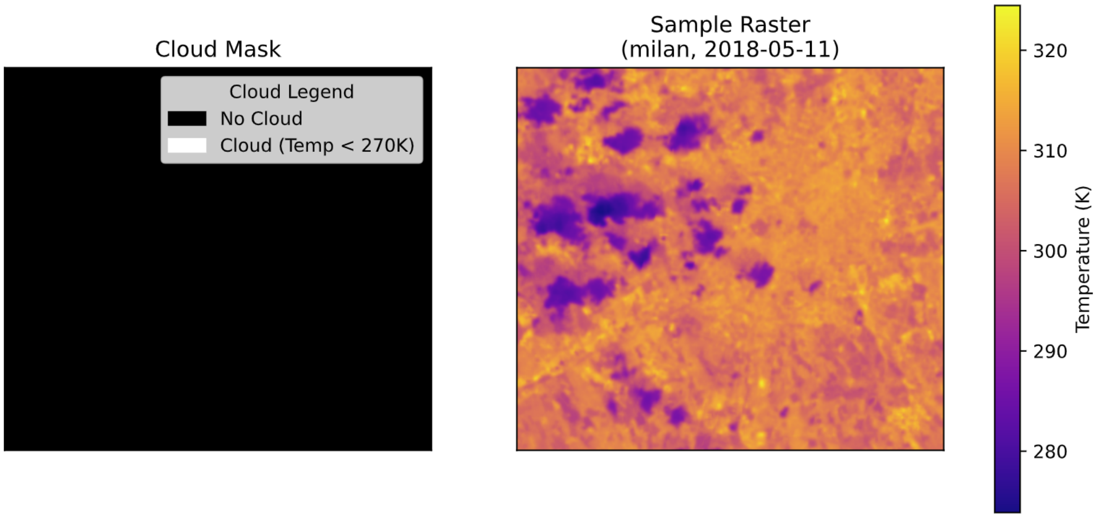
   $$\text{mask}_{\text{clouds}} := \boldsymbol{\mathcal{D}} < 270 \vee \left(\left| \frac{\boldsymbol{\mathcal{D}} - \mu_d}{\sigma_d} \right| > \sigma_T \wedge |\boldsymbol{\mathcal{D}}| > \delta_T^i \right)$$

   The values of $\delta_T^i \ \forall i \in [0, 11]$, equal for all cities, are listed in Table 4.1, and are chosen with respect to the reference raster, that for each city is selected in the summer.

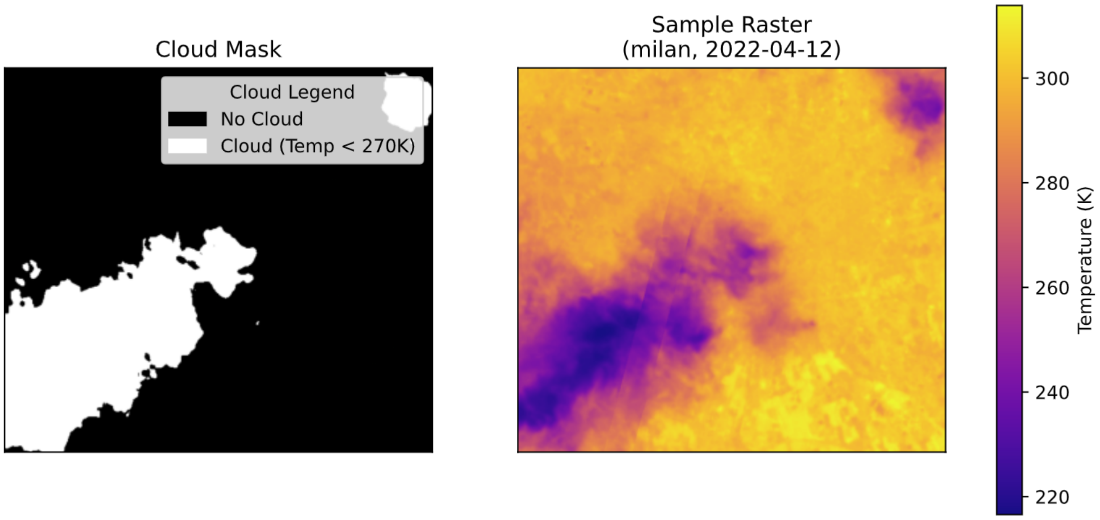
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
δ_T^i	45	40	30	30	25	20	15	20	25	30	40	45

Table 4.1

Figures 4.1 and 4.2 show examples of cloud masks for Landsat 8 rasters created using Method I and Method II respectively.



(a) Clouds above 270 K are not detected



(b) Clouds below 270 K are detected

Figure 4.1: Example of cloud mask generated with Method I

Figure 4.1a is an instance of Method I that fails detecting what they seem to be clouds over 270 K, although a ground truth mask is unavailable for verification, it is clear that the "cold stains" in the image are unusual as they are not present in the cloud-free reference raster shown in Figure 4.2. On the other hand Figure 4.1b demonstrates the method's efficiency in detecting clouds below 270 K.

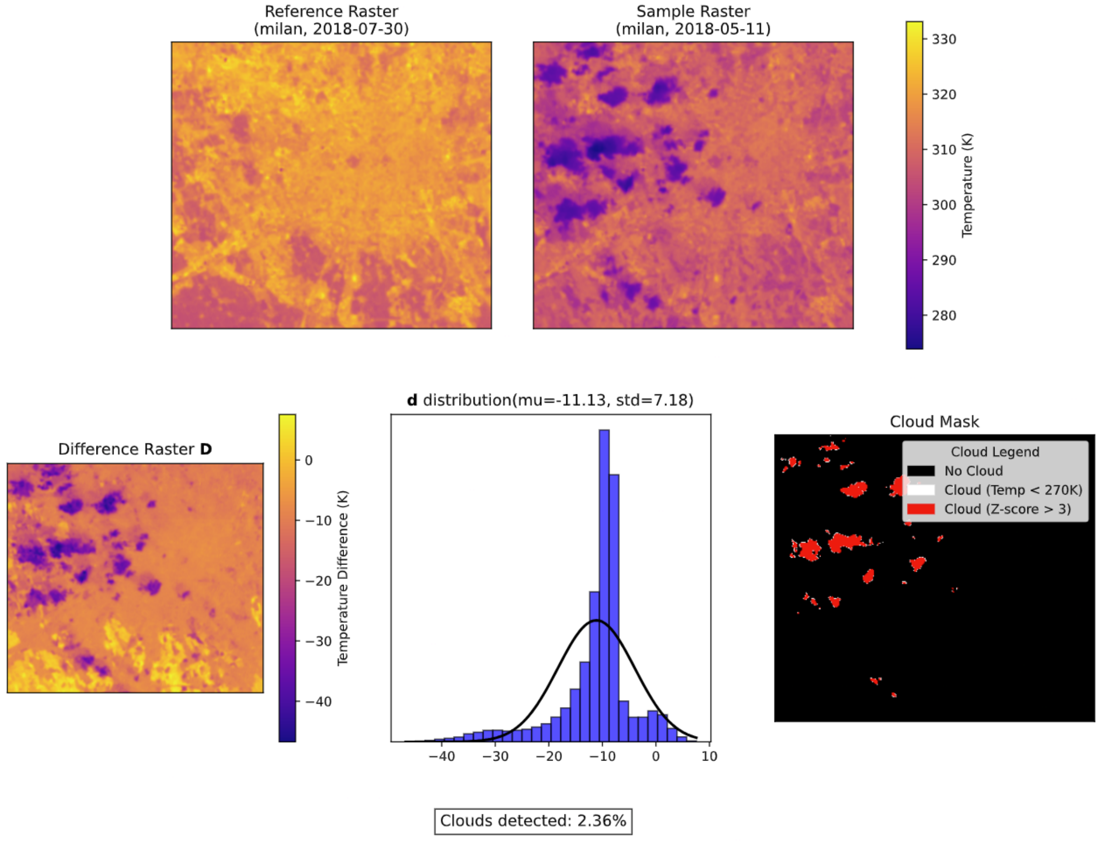


Figure 4.2: Example of cloud mask generated with Method II where clouds are above 270 K

The example in Figure 4.2, taken on the same date as Figure 4.1a (May 11, 2018), illustrates how Method II can be effective in detecting clouds above 270 K, by masking the points with high variance.

4.1.2 Temporal Shift between Sentinel-3 and Landsat 8

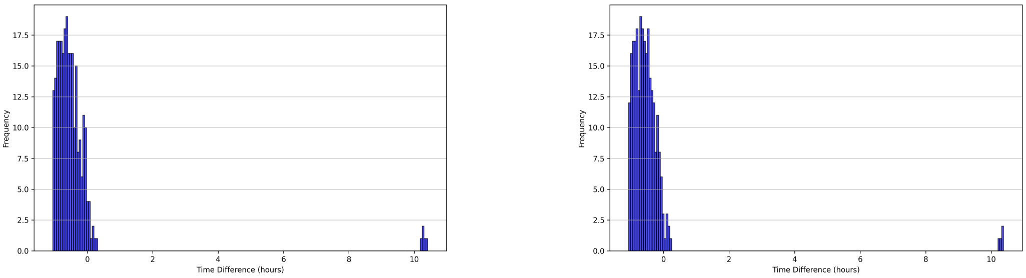
In this work, we refer as temporal shift as the amount of time that takes between Landsat 8 and Sentinel-3 to overpass on the same region, specifically:

$$\Delta_t := \mathcal{S}_t - \mathcal{L}_t$$

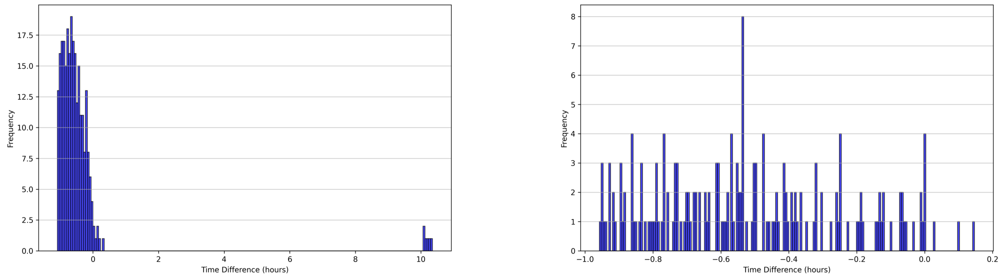
where $\mathcal{S}_t$ and $\mathcal{L}_t$ are the time expressed in hours of the passage of Sentinel-3 and Landsat-8 respectively. Time shift can be addressed with two major solutions:

1. By feeding this information into the networks along with the auxiliary data. This should give more information when applying SR from Sentinel-3 to Landsat 8.
2. By excluding pairs of (Sentinel-3 LST, Landsat 8 LST) that have Δ_t above a certain threshold—set to a 2-hour gap. This will exclude instances such as those represented in Figure 4.3, which represent pairs that differ by around 10 to 12 hours. Significant temporal shifts could introduce discrepancies in environmental conditions, such as changes in weather or surface temperature, producing outliers during training and poisoning the evaluation metrics during testing.

Both of these solutions will be applied to mitigate the temporal shift issue.



(a) Milan (b) Turin



(c) Budapest (d) Istanbul

Figure 4.3: Time difference in hours between Sentinel-3 and Landsat 8 satellite overpass for each city

4.1.3 Translation from Brightness to Land Surface Temperature

In this section it is described the methods adopted to translate BT of Sentinel-3 bands into LST. An example that shows the difference between BT and LST is reported in Figure 4.4.

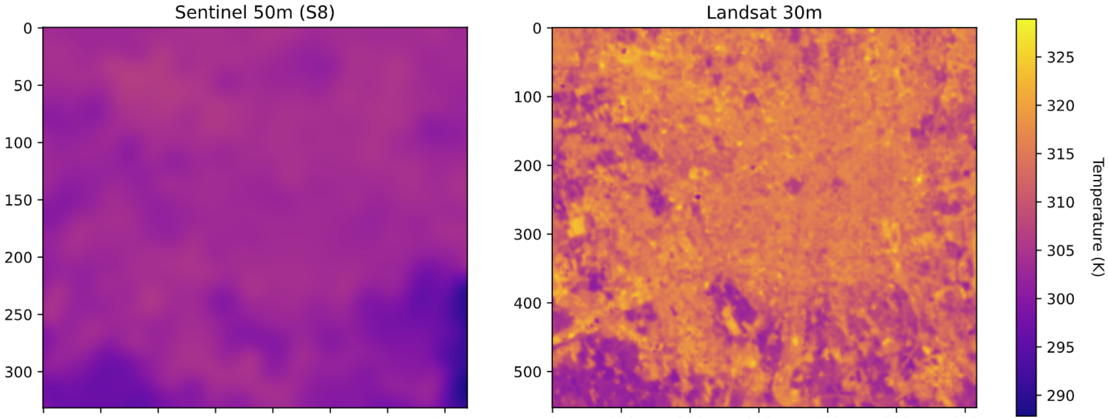


Figure 4.4: Gap between BT of S8 band of Sentinel-3 and LST of Landsat 8 for the city of Milan on June 6, 2018

This thesis proposes two different techniques:

1. **Regression.** The LST product is translated by linear transformation of Sentinel-3 BT bands and eventually additional data.

$$\widehat{\mathbf{LST}} = \mathbf{X}^T\boldsymbol{\beta} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1c} \\ 1 & x_{21} & \cdots & x_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nc} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_c \end{bmatrix}$$

   where n are the number of samples, i.e. pixels of the training set and c are the number of features per each pixel, accounting for one or more bands of Sentinel-3 and additional data such as NDVI, DEM, LC and weather data. The points are fit using Mean Squared Error (MSE):

$$\boldsymbol{\mathcal{L}} = \frac{1}{n}\left(\widehat{\mathbf{LST}} - \mathbf{LST}\right)^2$$

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \boldsymbol{\mathcal{L}}$$

   where **LST** are the Landsat 8 LST ground truth values.

2. **Neural Network.** Train a Multilayer Perceptron (MLP) to map Sentinel-3 BT to Landsat 8 LST. The MLP is composed of 4 layers with a leaky ReLU activation function for each one of them. The hidden sizes are respectively 64, 32, 16, 1. The model is trained using MSE as loss function. The input is represented by the set of flattened pixels of the images, where each pixel $\mathbf{X}$ is composed of C channels, representing a combination of the five bands of Sentinel-3 (S7, S8, S9, F1, F2) and the additional sources described in the dataset chapter (NDVI, DEM, LC, weather data, doy, time shift). The target $\mathbf{Y}$ instead is always Landsat-8 LST.

$$\mathbf{X} \in \mathbb{R}^C, \mathbf{Y} \in \mathbb{R}$$

4.2 Deep Neural Network Models

This section provides a detailed description of the shared preprocessing techniques applied to the data, as well as the deep neural networks employed for the task of SR. While the architecture changes from network to network their input $\mathbf{X}$ will be always a multi-channel raster.

$$\mathbf{X} \in \mathbb{R}^{C \times H \times W}$$

where C is the number of channels, H is the height and W is the width. The first channel of the input represents the LR LST value (or similarly BT value), meanwhile the remaining ones are the auxiliary data. The HR target raster $\mathbf{Y}$ has a similar representation with respect to the input with the difference that the number of channels is exactly 1, which represents HR Landsat 8 LST.

$$\mathbf{Y} \in \mathbb{R}^{1 \times H' \times W'}$$

The training of each of these networks is conducted using pairs of LR and HR images. For each training iteration, a LR image $\mathbf{X}$ is fed into the network, which try to learn the mapping

$$\hat{\mathbf{Y}} = f(\mathbf{X})$$

where $\hat{\mathbf{Y}}$ is the generated HR output from the network. This output is compared against the target HR image $\mathbf{Y}$ to calculate the loss, which in all cases, except for SRGAN and other adversarial models, will be MSE.

The tensors shapes of $\mathbf{X}$ and $\mathbf{Y}$ depend on the network, but we can distinguish two different scenarios:

1. **Case I - Isometric.** This is the case where $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ and $\mathbf{Y} \in \mathbb{R}^{1 \times H \times W}$. In this scenario, $\mathbf{X}$ was upscaled to match the dimensions of $\mathbf{Y}$ using bicubic interpolation.

2. **Case II - Non Isometric.** This is the scenario in which $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ and $\mathbf{Y} \in \mathbb{R}^{1\times H'\times W'}$ with $(H', W') = (rH, rW)$, where r represents the scaling factor of the SR network (the objective of the thesis sets $r = 10$). In this case the network will handle the upscaling process to produce an output that matches the dimensions of the target data.

Each sample target value $\mathbf{Y}$ is cropped so that H and W are the powers of 2 closest to the original dimensions of the raster. For example if a city's raster has HR dimensions $H \times W = 523 \times 615$ then the sample will be center-cropped to 512×512. The cropping operation is necessary to avoid complications with layers that perform operations such as convolution, upscaling, and downscaling. Additionally, it ensures consistency in input size, contributing to a more stable and effective training process.

LR images can be prepared in different ways:

1. **From HR counterparts.** The LR image is the product of HR: $\mathrm{LR} = f(\mathrm{HR})$, more precisely they are the HR images downscaled by a factor of 10x.
2. **Use Sentinel-3 BT S8.** LR images correspond to Sentinel-3 S8 band that represents the BT measurement closest to LST.
3. **Use Sentinel-3 LST.** LST of Sentinel-3 is calculated using BT2LST network that will translate Senintel-3 BT bands into the LST product.

During training HR images $\mathbf{Y}$ correspond to the Landsat 8 rasters, meanwhile the LR counterparts $\mathbf{X}$ are generated using one of the aforementioned methods, additionally stacked with the auxiliary data.

All inputs and target data are normalized using the training split statistics (rasters for the dates between 2018 and 2021) to improve training convergence speed and prevent numerical issues for large and small data values.

$$\mathbf{X}' = \frac{\mathbf{X} - \mu_{\mathbf{X},\mathrm{TR}}}{\sigma_{\mathbf{X},\mathrm{TR}}} \qquad \mathbf{Y}' = \frac{\mathbf{Y} - \mu_{\mathbf{Y},\mathrm{TR}}}{\sigma_{\mathbf{Y},\mathrm{TR}}}$$

LC channels are not normalized, since its content represents probabilities $p \in [0, 1]$. On the other hand, DEM is not normalized but rather scaled between -1 and 1. As far as the outputs of the networks $\hat{\mathbf{Y}}$, they are de-normalized by using the training statistics of the target data.

$$\hat{\mathbf{Y}}' = \hat{\mathbf{Y}} \cdot \sigma_{\mathbf{Y},\mathrm{TR}} + \mu_{\mathbf{Y},\mathrm{TR}}$$

They are then confronted with their respective target $\mathbf{Y}$. Moreover the batch of images fed to the network is randomly rotated and/or flipped to better generalize the models.

4.2.1 SRCNN

As mentioned in the previous chapter, SRCNN was one of the first deep neural networks developed for SR tasks, originally designed for three-channel RGB images. However, this work aims to evaluate the performance of SRCNN when fed with input images that have a different number of channels, denoted as $C \neq 3$. For this reason the first convolutional layer has been modified to accept a number of channels equal to the number of input features, that correspond to the temperature and the auxiliary data. The architecture utilized in this work for SRCNN consists of three convolutional layers, as described in Table 4.2. The first two layers incorporate ReLU activation functions. Notably, Convolutional Layer 2 has been modified from using a 1×1 kernel to a 5×5 kernel to ensure that the output dimensions match those of the target data.

Layer	**Filters**	**Kernel**	**ReLU**
Conv1	64	9x9	✓
Conv2	32	5x5	✓
Conv3	1	5x5	✗

Table 4.2: Convolutional Layers of SRCNN

Additionally padding of 10 pixels per each side has been added to the images before they are fed to the network to avoid undesired behaviors at the edges. The padding pixels are then removed once the network produces the output. Without the padding the output images had halos on the edges that did not reflect a normal behavior. The final architecture of the modified SRCNN is shown in Figure 4.5

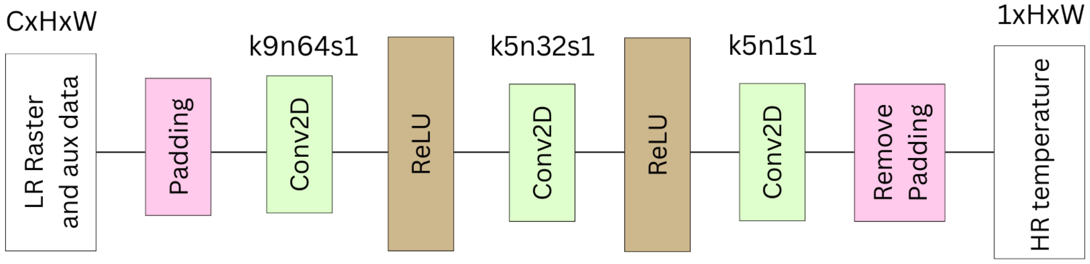


Figure 4.5: SRCNN proposed architecture

4.2.2 SE-SRCNN

SE-SRCNN is a network built to handle directly multiple input channels, and in the original paper was tested on simulated measurements for near the surface air temperature. In this work SE-SRCNN is utilized to test its performance on our datasets and it is for the most part the original architecture version, except for one modification: as previously done for SRCNN a padding of 10 pixels is added on the side of the images before they are fed to the network. This helps remove halos on the predicted images that could have been caused by the extremely small size of LR images.

We also explore a network architecture ablation by using recursive layers in a similar fashion to *Kim et al.* [21] (DR-SE-SRCNN). In Figure 4.6 it is shown the modifications applied to the feature extractors of the original SE-SRCNN architecture, which is represented in chapter 2, Figure 2.5.

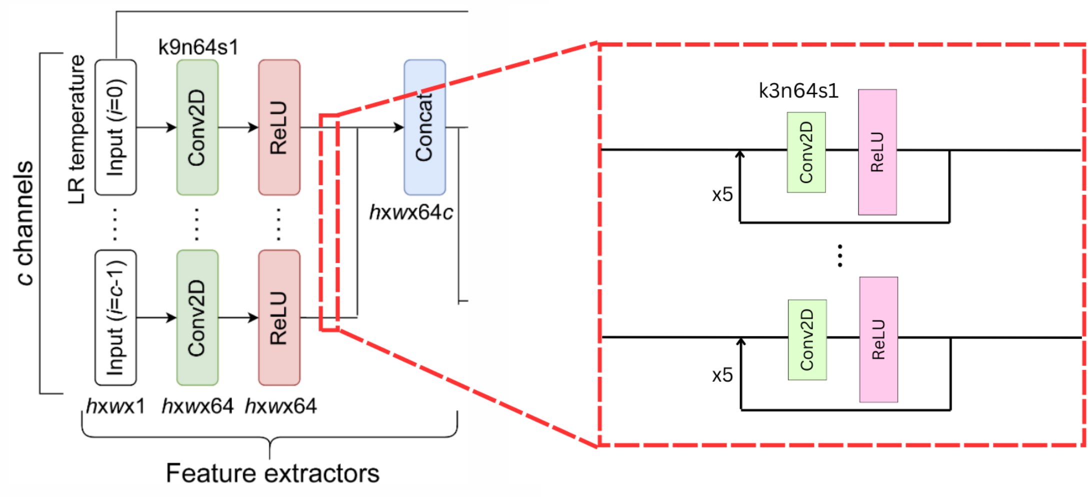


Figure 4.6: DR-SE-SRCNN feature extractor modification as an additional recursive layer, composed of a convolutional layer with a 3x3 kernel, 64 features and stride 1

4.2.3 EDSR

EDSR original architecture [22] was proposed for RGB images utilizing 2x, 3x, 4x as scaling factors, this is the reason why it is relevant to analyze its performances on multi channels remote sensing data and propose a way to upscale the LR image up to 10x and reach the goal of the thesis. Each convolutional layer, including the ones inside the 16 residual blocks have a fixed number of filters of 64 and a kernel of 3×3. The upsample block contain a convolutional layer that increase the features numbers by a factor of r and a pixel shuffle layer that rearranges elements from a tensor of shape $(*, C \times r^2, H, W)$ to a tensor of shape $(*, C, H \times r, W \times r)$, where r is an upscale factor. This block is different for each scaling factor and a description for each one of the latter is given in Table 4.3.

Scaling Factor	Layer	Description
2x	1	Convolutional - 128 filters, kernel size 3x3
	2	Pixel Shuffle - $r = 2$
3x	1	Convolutional - 192 filters, kernel size 3x3
	2	Pixel Shuffle - Pixel shuffle - $r = 3$
4x	1	Convolutional - 128 filters, kernel size 3x3
	2	Pixel Shuffle - $r = 2$
	3	Convolutional - 128 filters, kernel size 3x3
	4	Pixel Shuffle - $r = 2$
10x	1	Convolutional - 128 filters, kernel size 3x3
	2	Pixel Shuffle - $r = 2$
	3	Convolutional - 320 filters, kernel size 3x3
	4	Pixel Shuffle - $r = 5$

Table 4.3: Layers of the upscale module for different scaling factors

The upscaling of an images by a factor of 10x using EDSR can be achieved in two different ways: the first is to directly upscale the image using EDSR with an upsample block of 10x; the second method is to use an upsample block of 2x/3x/4x and then leverage resizing functions such as bicubic interpolation to scale the image to reach the target resolution. The order in which to apply EDSR and bicubic interpolation can also vary.

To implement the second method the network architecture was modified to perform 10x scaling operations by utilizing bicubic interpolation either before or after EDSR.

4.2.4 SRGAN

Original SRGAN architecture is modified to make the generator equal to EDSR network, specifically removing batch normalization layers and modifying first and last convolutional layers from 9x9 to 3x3 kernel size. This model uses the loss described in *Ledig et al.* paper [23]. Additionally the generator is a pretrained version of 400 epochs.

4.2.5 SE-EDSR

SE-EDSR is the enhancement of EDSR, that is supplemented with the SE attention mechanism block proposed in SE-SRCNN paper. The SE block helps the network by calculating weights for each channel that describe the magnitude of importance for each one of them. As shown in Figure 4.7 each channel is fed firstly to a convolutional layer and then to a series of 16 ResNet blocks. The mid-level features are then concatenated and fed to the SE block which will give the corresponding weights. After a weighted multiplication the LR feature maps are upsampled by a factor of 2x, 3x, 4x or 10x.

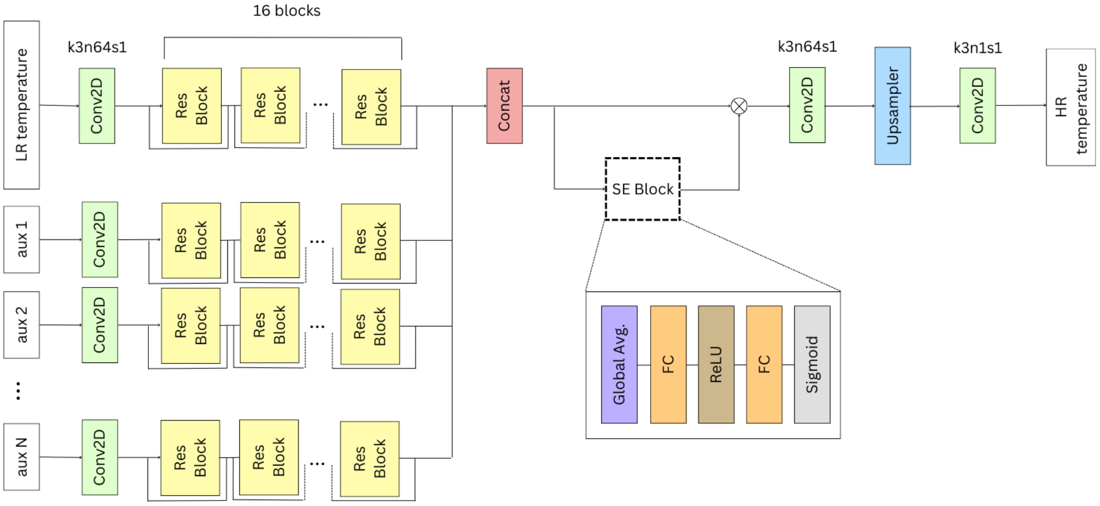


Figure 4.7: SE-EDSR proposed architecture

5 Experimental Results

This section presents the experimental results, comparing different methods using various metrics and parameter sets. It also describes the environment in which the experiments were conducted, including the frameworks, programming languages, hardware, and other relevant tools. The performance of the following models are all evaluated on three different metrics: MAE to measure the mean error in temperature difference, SSIM to evaluate the structural similarity for a visual point of view and PSNR to analyze the quality of the image.

5.1 Development tools, Programming Languages and Libraries

In this study, several technologies and frameworks were employed to facilitate efficient data processing, model development, and geospatial analysis.

For geospatial analysis, QGIS (Quantum GIS) was used to visualize and interpret raster datasets. It provided an interface for working with satellite imagery and other geospatial data, aiding in the understanding of spatial patterns and relationships. Alongside QGIS, Rasterio was employed to read, write, and manipulate the geospatial raster data within the Python environment.

The core machine learning and deep learning tasks were carried out using PyTorch and Lightning (PyTorch Lightning). Lightning provides a deep learning framework that automatizes the training loop and manages model validation and checkpointing. Among the most important libraries supporting these tasks were Pydantic, NumPy, Pandas and Matplotlib, which played a role in validating inputs, managing data and creating visualizations throughout the study.

To process imagery, OpenCV was used mainly for image resizing applying different interpolation techniques, including bicubic interpolation. The computational aspects of the project were executed on a high-performance cluster machine, in particular the jobs were run on a single NVIDIA GeForce RTX 2080 GPU with 12 GB of graphic memory along with 64 GB of RAM, scheduled via slurm. This hardware setup enabled faster processing and training of models, which was crucial for handling the neural network workloads efficiently.

5.2 BT2LST

BT2LST experiments are conducted using Adam optimizer, MSE loss and a starting learning rate of 1e-3, that is reduced with a plateau policy by a factor of 10 when the loss did not have any improvement in 50 epochs. The model is trained for 400 epochs, with batch size of 2048. The datasets comprehend rasters of the city of Milan, Turin, and Budapest from January 2018 to December 2023. Table 5.1 shows train, validation, and test split date ranges. A further analysis on unseen data is reported later in the text, that tests the performances of the models for the city of Istanbul, that was excluded during training.

Split	**Start**	**End**
Train	January 1, 2018	December 31, 2021
Validation	January 1, 2022	December 31, 2022
Test	January 1, 2023	December 31, 2023

Table 5.1: Dataset splits and their date range

BT2LST model is tested for different combinations of auxiliary data, as described in Table 5.2. Every test uses only S8 and S9 bands of the source satellite Sentinel-3, which have demonstrated to bring the best results. Additionally the time difference between satellite overpasses is fed as an additional channel.

As shown in Table 5.2, the information content of additional datasets generally lead to better results and the best performance is achieved by the combination of S8, S9, time shift, and solar energy, that achieve a MAE of 3.816 K.

BT2LST pre-trained network is used to translate all Sentinel-3 images into new versions that should resemble instances of Landsat 8 LST. These new images are then utilized as LR inputs to train different SR neural networks, with the hypothesis that they may lead to better results compared to Sentinel-3 dataset without translation.

Auxiliary Data	**MAE ↓**
-	5.786
NDVI	3.937
DEM	4.130
LC	4.124
doy	4.162
dew	4.116
solar energy	3.816
NDVI, DEM	3.941
NDVI, LC	3.894
DEM, LC	4.264
NDVI, DEM, LC	3.963
NDVI, DEM, LC, doy	3.930
NDVI, DEM, LC, dew	4.195
NDVI, DEM, LC, solar energy	4.241

Table 5.2: Results of BT2LST for different combinations of auxiliary data

5.3 Baseline, Bicubic Interpolation

Bicubic interpolation is a widely used resampling technique in image processing, particularly for increasing resolution of images. It improves upon the simpler nearest-neighbor and bilinear interpolation methods by considering the values of 16 neighboring pixels (4x4 grid) using cubic polynomials to interpolate the pixel values, resulting in smoother and more visually appealing images. In this section, we establish bicubic interpolation as the baseline method for evaluating the performance of our proposed super-resolution techniques. The choice of bicubic interpolation is based on its balance between computational efficiency and image quality, making it a standard reference for comparison. The results of bicubic interpolation are shown in Table 5.3, for different LR inputs and on different cities. All the input rasters have 1 km of spatial resolution, which is increased with bicubic interpolation by a factor of 10x and compared to Landsat 8 target rasters.

LR Input	City	MAE ↓	SSIM ↑	PSNR ↑
Landsat LST resampled at 1 km	Milan	1.802	0.970	44.319
	Turin	2.023	0.982	43.010
	Budapest	1.902	0.970	43.239
	Average	1.910	0.974	43.513
Sentinel-3 S8 BT	Milan	7.975	0.968	34.134
	Turin	7.839	0.980	33.586
	Budapest	7.221	0.968	33.490
	Average	7.669	0.972	33.738
Sentinel-3 LST after BT2LST	Milan	5.322	0.965	35.729
	Turin	6.265	0.979	34.674
	Budapest	7.254	0.966	33.786
	Average	6.290	0.970	34.721

Table 5.3: Results of bicubic interpolation with 10x scaling factor

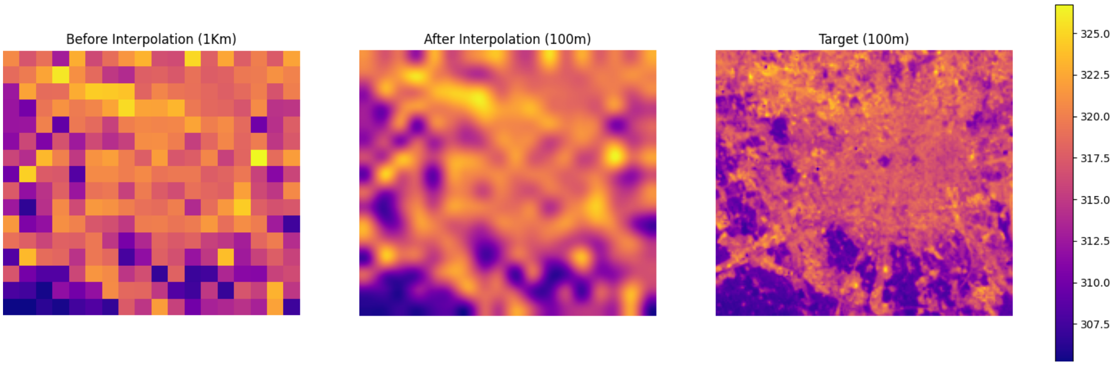


Figure 5.1: Visual results of bicubic interpolation with input LR Landsat LST for July 30th 2018 on the city of Milan

5.4 Deep Neural Network Models

All the models are run for 400 epochs, with a starting learning rate of 1e-3 that is decreased with a plateau policy by a factor of 10 each 50 epochs if the validation loss has not seen improvements. The models use Adam optimizer and a batch size of 16. Similarly to BT2LST, the dataset comprehending rasters from January 2018 to December 2023 for the cities of Milan, Turin, and Budapest is divided as described in Table 5.1.

For every model it is reported the training performances with the three different input type: Landsat 8, Sentinel-3 S8 bands and LST product retrieved with BT2LST, while the target is always Landsat 8. This is not to be confused with the LR input column, such the one in Table 5.4, which indicates the input utilized to test the model, that can be different from what has been used during training. In other words tables like 5.4 refer to models trained with LR Landsat 8 images, which are then tested additionally with other LR input types, i.e. Sentinel-3 S8 and BT2LST product. This approach highlights that Landsat 8 data will not be available at inference time. Conversely, models trained on Sentinel-3 S8 and BT2LST data are tested exclusively with that source, as this setup can be replicated during inference.

Training the network using LR Landsat 8 images degraded from HR counterparts is relevant to test the performances of the models on the usual SR task, which does not use different datasets between input and target. On the other hand, training with the remaining two input types, namely Sentinel-3 S8 band and BT2LST product, is vital to measure the actual metrics for the thesis objective.

SRCNN

From Table 5.4 it is possible to notice the effectiveness of auxiliary datasets in lowering MAE and increasing SSIM and PSNR. Notably, NDVI appears to provide the most valuable information content, as its presence almost always results in a reduction of approximately a 0.1 K. Specifically best results are carried by the combination of NDVI, DEM, and LC, which in comparison to standard bicubic interpolation reduces MAE from 1.910 to 1.613 K.

A further analysis on the performances of the model trained with Landsat 8 LR input, but tested on both Sentinel-3 S8 and BT2LST product, shows that the proposed translation network make predictions drastically worse, from 5.636 to 8.063 K. Nevertheless, a comparison between Tables 5.5 and 5.6 demonstrates that training the models directly with BT2LST LR product brings results similar to the training with Sentinel-3 S8. Although for SRCNN, BT2LST does not show improvements, the following neural networks will produce different outcomes.

Finally, Tables 5.5 and 5.6 indicate again a potential benefit for additional datasets, and more precisely best results are obtained using NDVI, DEM, LC, and solar energy all together, registering a MAE of 4.811 K that is far better than bicubic interpolation with 6.290 K at best.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.772	0.975	44.435
	NDVI	1.639	**0.980**	45.000
	DEM	1.741	0.976	44.612
	LC	1.765	0.976	44.454
	doy	1.778	0.975	44.398
	dew	1.793	0.975	44.285
	solar energy	1.823	0.975	44.201
	NDVI, DEM	1.619	**0.980**	45.127
	NDVI, LC	1.645	**0.980**	44.981
	DEM, LC	1.73	0.977	44.625
	NDVI, DEM, LC	**1.613**	**0.980**	**45.157**
	NDVI, DEM, LC, doy	1.636	**0.980**	45.078
	NDVI, DEM, LC, dew	1.696	**0.980**	44.706
	NDVI, DEM, LC, solar energy	1.738	0.979	44.446
Sentinel-3 S8 BT	-	5.685	0.971	36.208
	NDVI, DEM, LC	**5.636**	**0.976**	**36.402**
Sentinel-3 LST after BT2LST	-	**8.063**	0.969	**32.970**
	NDVI, DEM, LC	8.228	**0.974**	32.740

Table 5.4: Results of SRCNN trained with Landsat 8 LR images tested on different input types

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.201	0.971	37.172
NDVI	4.997	0.974	37.552
DEM	5.232	0.971	36.897
LC	5.154	0.971	37.093
doy	5.147	0.971	37.304
dew	5.237	0.971	37.388
solar energy	5.216	0.972	37.383
NDVI, DEM	5.212	**0.975**	37.277
NDVI, LC	5.368	0.973	37.131
DEM, LC	5.320	0.971	36.900
NDVI, DEM, LC	5.334	0.974	37.034
NDVI, DEM, LC, doy	5.027	0.973	37.578
NDVI, DEM, LC, dew	5.111	0.973	37.690
NDVI, DEM, LC, solar energy	**4.811**	0.974	**38.034**

Table 5.5: Results of SRCNN trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.210	0.971	36.611
NDVI	5.120	0.974	36.839
DEM	5.195	0.971	36.632
LC	5.486	0.971	36.362
doy	5.450	0.971	36.344
dew	5.234	0.971	36.824
solar energy	5.262	0.971	36.474
NDVI, DEM	5.032	**0.975**	37.091
NDVI, LC	5.286	0.974	36.729
DEM, LC	5.229	0.971	36.639
NDVI, DEM, LC	5.302	0.974	36.713
NDVI, DEM, LC, doy	**4.941**	0.973	37.238
NDVI, DEM, LC, dew	5.204	0.974	36.845
NDVI, DEM, LC, solar energy	4.945	0.973	**37.259**

Table 5.6: Results of SRCNN trained with Sentinel-3 LST after BT2LST LR images

SE-SRCNN

The observations on SE-SRCNN tests are aligned with what has already been said for SRCNN in the previous subsection. Remarkably, SE-SRCNN achieves better results with respect to SRCNN, applying NDVI, DEM, and LC produces HR reconstructions with a MAE of 1.597 K for the standard SR task, with the training on pairs of Landsat 8 LR-HR images.

Findings of Tables 5.7 and 5.8 suggest again that NDVI plays a key role, and together with DEM and LC decrease reconstruction error. Table 5.9 instead, highlights the importance of day of the year as a crucial variable to obtain 4.606 K MAE, that together with BT2LST marks the best results for mapping LR Sentinel-3 images into HR Landsat 8 ones.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.767	0.975	44.450
	NDVI	1.645	0.980	44.941
	DEM	1.703	0.978	44.895
	LC	1.714	0.976	44.768
	doy	1.794	0.975	44.269
	dew	1.787	0.975	44.369
	solar energy	1.797	0.975	44.332
	NDVI, DEM	1.621	0.980	45.147
	NDVI, LC	1.618	0.980	45.178
	DEM, LC	1.680	0.978	44.895
	NDVI, DEM, LC	**1.597**	**0.981**	**45.265**
	NDVI, DEM, LC, doy	1.611	**0.981**	45.225
	NDVI, DEM, LC, dew	1.634	0.980	45.133
	NDVI, DEM, LC, solar energy	1.650	0.980	45.064
Sentinel-3 S8 BT	-	5.692	0.970	36.198
	NDVI, DEM, LC	**5.678**	**0.976**	**36.365**
Sentinel-3 LST after BT2LST	-	**8.105**	0.969	**32.943**
	NDVI, DEM, LC	8.259	**0.975**	32.609

Table 5.7: Results of SE-SRCNN trained with Landsat 8 LR images tested on different input types

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.058	0.971	37.581
NDVI	4.858	0.974	37.973
DEM	5.034	0.971	37.354
LC	5.373	0.971	36.889
doy	4.993	0.971	37.612
dew	4.756	0.971	37.690
solar energy	5.038	0.971	37.297
NDVI, DEM	5.159	**0.974**	37.750
NDVI, LC	5.035	**0.974**	37.833
DEM, LC	5.255	0.972	37.456
NDVI, DEM, LC	5.209	0.973	37.483
NDVI, DEM, LC, doy	5.166	0.973	37.516
NDVI, DEM, LC, dew	**4.723**	0.973	**38.093**
NDVI, DEM, LC, solar energy	5.081	0.973	37.381

Table 5.8: Results of SE-SRCNN trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.375	0.971	36.148
NDVI	4.998	0.975	37.187
DEM	5.680	0.971	35.521
LC	5.261	0.971	36.784
doy	**4.606**	0.971	37.760
dew	5.129	0.971	36.771
solar energy	4.999	0.972	36.942
NDVI, DEM	4.951	0.975	37.406
NDVI, LC	4.964	0.973	37.250
DEM, LC	4.821	0.971	37.546
NDVI, DEM, LC	4.919	0.975	37.580
NDVI, DEM, LC, doy	5.092	**0.976**	37.412
NDVI, DEM, LC, dew	4.615	0.973	**37.768**
NDVI, DEM, LC, solar energy	4.840	0.975	37.297

Table 5.9: Results of SE-SRCNN trained with Sentinel-3 LST after BT2LST LR images

EDSR

EDSR test results are presented for various combinations of data, LR input types used during training and testing, and the ablation of bicubic interpolation applied at different magnitudes and stages of the network. Specifically, in contrast with the previous models, Tables 5.10, 5.11, and 5.12 outline metrics for the model trained with LR Landsat as input, respectively with a full 10x EDSR, a 4x and a 2x EDSR with the help of an additional bicubic interpolation, before or after the neural network.

To avoid extreme verbosity and the introduction of additional tables, experiments for the training with LR inputs as Sentinel-3 S8 and BT2LST product are only conducted for a specific configuration, with 2x EDSR using bicubic interpolation before the inputs are fed to the network.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.587	0.980	**45.704**
	NDVI	1.700	0.978	44.928
	DEM	1.579	0.981	45.671
	LC	1.712	0.977	44.765
	doy	1.662	0.980	45.229
	dew	1.788	0.977	44.399
	solar energy	1.878	0.975	44.083
	NDVI, DEM	1.736	0.978	44.688
	NDVI, LC	1.630	0.980	45.317
	DEM, LC	**1.570**	**0.982**	45.626
	NDVI, DEM, LC	1.618	0.980	45.393
	NDVI, DEM, LC, doy	1.716	0.978	44.790
	NDVI, DEM, LC, dew	1.725	0.978	44.749
	NDVI, DEM, LC, solar energy	1.839	0.975	44.241
Sentinel-3 S8 BT	-	5.561	0.970	36.123
	DEM, LC	**5.451**	**0.974**	**36.402**
Sentinel-3 LST after BT2LST	-	8.384	0.967	32.351
	DEM, LC	**8.268**	**0.972**	**32.514**

Table 5.10: Results of EDSR 10x trained with Landsat 8 LR images tested on different input types

Before				
LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.618	0.978	45.235
	NDVI	1.642	0.978	45.036
	DEM	1.568	0.981	45.532
	LC	**1.451**	**0.983**	**46.211**
	doy	1.602	0.979	45.305
	dew	1.808	0.975	44.226
	solar energy	1.834	0.975	44.084
	NDVI, DEM	1.635	0.979	45.089
	NDVI, LC	1.605	0.979	45.296
	DEM, LC	1.558	0.980	45.482
	NDVI, DEM, LC	1.650	0.978	45.012
	NDVI, DEM, LC, doy	1.643	0.978	45.071
	NDVI, DEM, LC, dew	1.681	0.977	44.811
	NDVI, DEM, LC, solar energy	1.751	0.976	44.481
Sentinel-3 S8 BT	-	5.702	0.970	36.149
	LC	**5.642**	**0.977**	**36.423**
Sentinel-3 LST after BT2LST	-	**8.316**	0.968	**32.573**
	LC	8.374	**0.975**	32.487
After				
Landsat LST resampled at 1 km	-	1.620	0.979	45.237
	NDVI	1.638	0.979	45.099
	DEM	1.534	**0.981**	45.685
	LC	**1.523**	**0.981**	**45.719**
	doy	1.755	0.976	44.453
	dew	1.810	0.975	44.181
	solar energy	1.802	0.975	44.322
	NDVI, DEM	1.617	0.979	45.198
	NDVI, LC	1.620	0.979	45.203
	DEM, LC	1.580	0.980	45.335
	NDVI, DEM, LC	1.632	0.979	45.115
	NDVI, DEM, LC, doy	1.623	0.979	45.159
	NDVI, DEM, LC, dew	1.642	0.979	45.040
	NDVI, DEM, LC, solar energy	1.662	0.978	44.952
Sentinel-3 S8 BT	-	5.634	0.970	36.230
	LC	**5.594**	**0.976**	**36.420**
Sentinel-3 LST after BT2LST	-	8.335	0.968	32.497
	LC	**8.163**	**0.974**	**32.726**

Table 5.11: Results of EDSR 4x plus bicubic interpolation before and after trained with Landsat 8 LR images tested on different input types

Before				
LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.655	0.978	45.013
	NDVI	1.613	0.980	45.104
	DEM	1.486	0.982	45.938
	LC	1.428	0.977	46.248
	doy	1.699	0.977	44.742
	dew	1.794	0.975	44.265
	solar energy	1.839	0.975	44.105
	NDVI, DEM	1.592	0.980	45.276
	NDVI, LC	1.557	0.980	45.476
	DEM, LC	**1.410**	**0.983**	**46.380**
	NDVI, DEM, LC	1.612	0.980	45.185
	NDVI, DEM, LC, doy	1.572	0.980	45.497
	NDVI, DEM, LC, dew	1.576	0.980	45.372
	NDVI, DEM, LC, solar energy	1.638	0.979	45.038
Sentinel-3 S8 BT	-	5.675	0.970	36.144
	DEM, LC	**5.606**	**0.978**	**36.429**
Sentinel-3 LST after BT2LST	-	8.383	0.968	32.472
	DEM, LC	**8.331**	**0.976**	**32.602**
After				
Landsat LST resampled at 1 km	-	1.677	0.977	44.929
	NDVI	1.616	0.979	45.140
	DEM	1.436	0.982	46.234
	LC	**1.418**	**0.983**	**46.308**
	doy	1.766	0.976	44.480
	dew	1.805	0.975	44.178
	solar energy	1.842	0.975	44.078
	NDVI, DEM	1.634	0.980	45.034
	NDVI, LC	1.561	0.980	45.478
	DEM, LC	1.435	**0.983**	46.214
	NDVI, DEM, LC	1.553	0.981	45.502
	NDVI, DEM, LC, doy	1.585	0.980	45.421
	NDVI, DEM, LC, dew	1.630	0.979	45.074
	NDVI, DEM, LC, solar energy	1.818	0.976	44.065
Sentinel-3 S8 BT	-	5.706	0.970	36.120
	LC	**5.572**	**0.978**	**36.430**
Sentinel-3 LST after BT2LST	-	8.418	0.968	32.491
	LC	**8.298**	**0.975**	**32.559**

Table 5.12: Results of EDSR 2x plus bicubic interpolation before and after trained with Landsat 8 LR images tested on different input types

EDSR achieves better results than SRCNN and SE-SRCNN for the standard task of SR, with 1.410 K of MAE, 0.983 of SSIM, and 46.380 of PSNR. The latter results are highlighted in Table 5.12, and indicate the utilization of DEM and LC as auxiliary dataset.

As shown in Table 5.13, the average MAE for different auxiliary data is clearly better for EDSR 2x with bicubic interpolation applied before; the test of the models trained with Sentinel-3 S8 BT and Senintel-3 LST calculated with BT2LST, reported in the following tables, will be only conducted using this configuration.

		Bicubic Interpolation Position	
		Before	After
Scaling Factor	2x	**1.493**	1.627
	4x	1.646	1.647

Table 5.13: Comparison between bicubic interpolation position before and after for with different EDSR scaling factor

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.175	0.971	37.283
NDVI	4.999	**0.974**	37.651
DEM	5.180	0.971	37.066
LC	5.637	0.971	36.777
doy	4.939	0.971	**37.736**
dew	5.615	0.970	36.749
solar energy	5.659	0.970	36.545
NDVI, DEM	5.119	**0.974**	37.638
NDVI, LC	5.177	**0.974**	37.114
DEM, LC	5.316	0.971	36.690
NDVI, DEM, LC	5.122	**0.974**	37.227
NDVI, DEM, LC, doy	**4.915**	0.973	37.419
NDVI, DEM, LC, dew	5.520	0.971	36.775
NDVI, DEM, LC, solar energy	5.124	0.971	37.169

Table 5.14: Results of EDSR 2x with bicubic interpolation applied before, trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE** ↓	**SSIM** ↑	**PSNR** ↑
-	5.141	0.971	36.738
NDVI	5.412	0.974	36.321
DEM	5.439	0.971	36.229
LC	5.026	0.971	37.026
doy	5.308	0.971	36.709
dew	5.093	0.971	36.825
solar energy	5.098	0.971	37.047
NDVI, DEM	5.138	**0.975**	36.830
NDVI, LC	5.150	0.974	36.835
DEM, LC	5.148	0.971	36.751
NDVI, DEM, LC	5.001	0.974	36.932
NDVI, DEM, LC, doy	**4.947**	0.974	**37.316**
NDVI, DEM, LC, dew	5.222	0.973	36.667
NDVI, DEM, LC, solar energy	5.098	0.972	36.744

Table 5.15: Results of EDSR 2x with bicubic interpolation applied before, trained with Sentinel-3 LST after BT2LST LR images

SRGAN

The SRGAN experimental results can be distinguished from the previous neural networks by the significant variability in metrics outcomes. For instance, MAE ranges from 1.471 K to 13.200 in Table 5.16; this is mostly due to the inherent instability of GAN-based models that rely on the adversarial training process where generator and discriminator balance is crucial.

Despite being more recent than SRCNN and EDSR, SRGAN does not appear to bring any significant improvement compared to the older models. Tables 5.16, 5.17 and 5.18 mark results that are aligned compared to the other neural network, with the combinations of NDVI, DEM, and/or LC serving the best scores.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.804	0.976	44.079
	NDVI	2.132	0.969	42.173
	DEM	1.532	0.981	45.603
	LC	**1.471**	**0.983**	**46.008**
	doy	1.718	0.977	44.635
	dew	13.200	0.709	27.724
	solar energy	2.176	0.975	43.263
	NDVI, DEM	2.176	0.974	42.027
	NDVI, LC	1.934	0.975	43.317
	DEM, LC	1.519	0.982	45.674
	NDVI, DEM, LC	1.658	0.979	44.859
	NDVI, DEM, LC, doy	1.723	0.978	44.480
	NDVI, DEM, LC, dew	1.746	0.978	44.363
	NDVI, DEM, LC, solar energy	5.286	0.908	35.621
Sentinel-3 S8 BT	-	5.693	0.970	34.941
	LC	**5.610**	**0.977**	**46.400**
Sentinel-3 LST after BT2LST	-	8.667	0.968	32.084
	LC	**8.290**	**0.975**	**32.622**

Table 5.16: Results of SRGAN trained with Landsat 8 LR images tested on different input types

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.548	0.967	36.143
NDVI	7.285	0.925	33.521
DEM	5.754	0.954	35.454
LC	5.354	0.967	**36.964**
doy	5.734	0.969	36.468
dew	6.689	0.908	33.994
solar energy	18.066	0.543	25.168
NDVI, DEM	**5.257**	**0.973**	37.347
NDVI, LC	5.273	**0.973**	36.906
DEM, LC	8.561	0.782	30.733
NDVI, DEM, LC	6.139	0.970	36.197
NDVI, DEM, LC, doy	5.353	0.972	36.952
NDVI, DEM, LC, dew	5.507	0.971	36.869
NDVI, DEM, LC, solar energy	6.300	0.963	35.710

Table 5.17: Results of SRGAN trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.208	0.971	36.517
NDVI	5.163	0.968	**36.649**
DEM	5.687	0.971	36.229
LC	5.402	0.969	36.211
doy	6.629	0.963	35.030
dew	18.875	0.543	24.179
solar energy	8.311	0.923	32.364
NDVI, DEM	5.268	0.962	35.968
NDVI, LC	5.816	0.967	35.587
DEM, LC	**5.119**	0.970	36.600
NDVI, DEM, LC	5.233	0.971	36.446
NDVI, DEM, LC, doy	5.417	**0.974**	36.433
NDVI, DEM, LC, dew	6.471	0.971	35.056
NDVI, DEM, LC, solar energy	6.463	0.958	34.825

Table 5.18: Results of SRGAN trained with Sentinel-3 LST after BT2LST LR images

SE-EDSR

The addition of the SE module to the EDSR architecture results in an improvement compared to the original model. These SE block dynamically recalibrate channel-wise feature importance, enabling the model to focus on the most relevant features.

As highlighted in Table 5.20, the MAE decreases from 4.915K to 4.740K, while SSIM and PSNR remain nearly unchanged, with only minor oscillations of lower magnitude. SE-EDSR performs also well in the standard SR task, analyzed in Table 5.19, that indicates generally lower MAE and higher SSIM and PSNR, for different combinations of data. Increases in performances are also visible in Table 5.21, whose results are for the most part better than the EDSR counterparts, and records a MAE of 4.847 K against 4.947 K.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.710	0.976	44.714
	NDVI	1.623	0.980	45.149
	DEM	1.763	0.977	44.265
	LC	**1.422**	**0.984**	46.273
	doy	1.674	0.977	44.854
	dew	1.751	0.975	44.504
	solar energy	1.766	0.975	44.438
	NDVI, DEM	1.579	0.981	45.339
	NDVI, LC	1.520	0.981	45.709
	DEM, LC	1.453	**0.984**	46.236
	NDVI, DEM, LC	1.440	**0.984**	46.188
	NDVI, DEM, LC, doy	1.501	0.982	45.818
	NDVI, DEM, LC, dew	1.428	**0.984**	**46.383**
	NDVI, DEM, LC, solar energy	1.448	**0.984**	46.191
Sentinel-3 S8 BT	-	5.729	0.971	36.117
	LC	**5.617**	**0.979**	**36.441**
Sentinel-3 LST after BT2LST	-	**8.142**	0.969	**32.890**
	LC	8.249	**0.977**	32.653

Table 5.19: Results of SE-EDSR 2x with bicubic applied before, trained with Landsat 8 LR images tested on different input types

Auxiliary Data	**MAE** ↓	**SSIM** ↑	**PSNR** ↑
-	5.198	0.971	37.249
NDVI	5.146	**0.975**	37.305
DEM	5.246	0.971	37.213
LC	5.328	0.971	36.990
doy	**4.740**	0.971	37.456
dew	5.048	0.970	37.202
solar energy	4.949	0.971	37.679
NDVI, DEM	5.786	0.971	36.480
NDVI, LC	5.082	0.972	37.59
DEM, LC	5.328	0.969	37.238
NDVI, DEM, LC	5.798	**0.975**	37.099
NDVI, DEM, LC, doy	4.881	0.974	37.588
NDVI, DEM, LC, dew	5.074	0.974	37.635
NDVI, DEM, LC, solar energy	4.932	0.974	**37.902**

Table 5.20: Results of SE-EDSR 2x with bicubic applied before, trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE** ↓	**SSIM** ↑	**PSNR** ↑
-	5.112	0.971	36.826
NDVI	4.965	0.974	37.128
DEM	5.016	0.971	36.895
LC	4.938	0.972	**37.191**
doy	5.337	0.971	36.496
dew	4.931	0.971	37.034
solar energy	4.867	0.970	37.072
NDVI, DEM	5.087	**0.975**	37.011
NDVI, LC	5.100	0.971	36.771
DEM, LC	5.245	0.971	36.609
NDVI, DEM, LC	5.063	0.974	37.146
NDVI, DEM, LC, doy	**4.847**	0.973	37.382
NDVI, DEM, LC, dew	5.220	0.973	36.969
NDVI, DEM, LC, solar energy	5.102	0.965	36.589

Table 5.21: Results of SE-EDSR 2x with bicubic applied before, trained with Sentinel-3 LST after BT2LST LR images

DR-SE-SRCNN

The DR-SE-SRCNN model enhances the SE-SRCNN by incorporating a deeper recursive layer into the feature extractor. This modification allows the model to better capture hierarchical and contextual features from the input data, leading to overall improved performance across different auxiliary data combinations.

Tables 5.22, 5.23, and 5.24 summarize the results of DR-SE-SRCNN when trained and tested on different input types and auxiliary data combinations. The model trained with Landsat 8 LR images demonstrates the consistent superiority of DR-SE-SRCNN in terms of MAE, SSIM, and PSNR when compared with SE-SRCNN. Conversely the model trained on Sentinel-3 S8 band and BT2LST product do not seem to drastically change what the simpler architecture already achieved, resulting in similar results that can oscillate easily on several nuances.

LR Input	**Auxiliary Data**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Landsat LST resampled at 1 km	-	1.738	0.976	44.575
	NDVI	1.604	0.980	45.172
	DEM	1.607	0.980	45.275
	LC	1.604	0.979	45.300
	doy	1.760	0.975	44.452
	dew	1.756	0.976	44.506
	solar energy	1.739	0.976	44.578
	NDVI, DEM	1.579	**0.981**	45.353
	NDVI, LC	1.590	0.980	45.276
	DEM, LC	**1.561**	0.980	45.551
	NDVI, DEM, LC	**1.561**	**0.981**	**45.490**
	NDVI, DEM, LC, doy	1.578	**0.981**	45.403
	NDVI, DEM, LC, dew	1.582	**0.981**	45.315
	NDVI, DEM, LC, solar energy	1.575	**0.981**	45.408
Sentinel-3 S8 BT	-	5.745	0.971	36.088
	NDVI, DEM, LC	**5.564**	**0.976**	**36.491**
Sentinel-3 LST after BT2LST	-	**8.398**	0.969	**32.528**
	NDVI, DEM, LC	8.445	**0.975**	32.440

Table 5.22: Results of DR-SE-SRCNN 2x with bicubic applied before, trained with Landsat 8 LR images tested on different input types

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	5.209	0.971	36.843
NDVI	5.001	**0.972**	37.878
DEM	5.061	**0.972**	37.630
LC	4.981	0.971	37.670
doy	4.846	0.971	37.850
dew	5.249	0.971	37.296
solar energy	**4.733**	0.971	**38.032**
NDVI, DEM	5.110	0.974	37.673
NDVI, LC	4.895	**0.972**	37.824
DEM, LC	5.189	0.971	37.257
NDVI, DEM, LC	4.944	**0.972**	37.900
NDVI, DEM, LC, doy	5.089	0.971	37.554
NDVI, DEM, LC, dew	4.966	0.971	37.775
NDVI, DEM, LC, solar energy	5.243	**0.972**	37.536

Table 5.23: Results of DR-SE-SRCNN 2x with bicubic applied before, trained with Sentinel-3 S8 BT band LR images

Auxiliary Data	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
-	**4.870**	0.971	**37.376**
NDVI	5.116	0.972	36.606
DEM	5.142	0.971	36.840
LC	5.228	0.970	36.702
doy	5.088	0.971	36.923
dew	5.020	0.971	36.840
solar energy	4.946	0.971	37.119
NDVI, DEM	5.193	0.973	36.850
NDVI, LC	4.888	0.972	37.156
DEM, LC	5.171	0.971	36.822
NDVI, DEM, LC	5.108	0.972	37.031
NDVI, DEM, LC, doy	4.974	0.973	37.166
NDVI, DEM, LC, dew	4.906	0.973	37.286
NDVI, DEM, LC, solar energy	4.944	**0.974**	37.282

Table 5.24: Results of DR-SE-SRCNN 2x with bicubic applied before, trained with Sentinel-3 LST after BT2LST LR images

5.5 Test on the unseen city of Istanbul

In this section we test all the best configurations for each model for the city of Istanbul with input Sentinel-3 LST obtained with BT2LST. These experiments aim to measure performance for an unseen city and verify that the neural networks can be applied for out of distribution data.

The first observation is the drastic improvement brought by neural networks in contrast to bicubic interpolation, for cities whose samples were used during training; bicubic interpolation reaches a MAE of 6.290 K, meanwhile SE-SRCNN yields the best results with only 4.606 K.

For the unseen city of Istanbul instead, the comparison between bicubic interpolation and the proposed neural networks is closer. Surprisingly DR-SE-SRCNN obtains best scores reaching a 5.130 K of MAE, reducing errors of almost 1 K.

Model	**Auxiliary Data**	**Cities**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
Bicubic Interpolation	-	I	6.290	0.970	34.721
		Istanbul	5.975	0.953	34.635
SRCNN	NDVI, DEM, LC, doy	I	4.941	0.973	37.259
		Istanbul	5.993	0.950	34.006
SE-SRCNN	doy	I	4.606	0.971	37.760
		Istanbul	5.908	0.948	34.371
EDSR 2x Bic. Inter. before	NDVI, DEM, LC, doy	I	4.947	0.974	37.316
		Istanbul	5.628	0.952	34.846
SRGAN	DEM, LC	I	5.199	0.970	36.600
		Istanbul	5.723	0.947	34.304
SE-EDSR 2x Bic. Inter. before	NDVI, DEM, LC, doy	I	4.847	0.973	37.382
		Istanbul	6.861	0.933	33.296
DR-SE-SRCNN	-	I	4.870	0.971	37.376
		Istanbul	5.139	0.948	35.336

Table 5.25: Performance of best models for the unseen city of Istanbul, compared to the original performance on the cities of Milan, Turin and Budapest

5.6 Gaussian Distributions for Cloud Masks

This section analyzes the performances of the previous models using cloud masks generated with Method II described in Subsection 4.1.1. The cloud masks in this scenario are obtained by masking pixels that do not resemble expected patterns that are shown on a cloud-free reference raster. A detailed description of the method, however, is digressed in the relative subsection.

The best configurations, reported in Table 5.25, are trained from scratch using Method II, that excludes cloud pixels for loss and metrics calculation. Their performances are then compared to the ones of the same networks, but instead trained using Method I. More precisely the exact pre-trained networks of Table 5.25 are tested using Method II as metrics evaluation, in this way the tests' comparison is made on the very same test set and with same evaluation techniques.

Table 5.26, which gathers the results, suggests that there are both instances of better and worse performances when training with Method II, and that a deeper and more accurate validation technique is necessary to demonstrate the effectiveness of the proposed method. Regardless of the validity of the method, SE-EDSR 2x with bicubic interpolation applied beforehand yields the best MAE results for this cloud masks technique with 6.455 °C of error.

Additionally, it can be seen that the MAE values are slightly higher when using this criterion to calculate metrics. This is because, by excluding more clouds, the evaluation becomes stricter as the MAE calculation does not account for the easier reconstructions of clouds, but rather applies to more complex elements such as the city's morphology.

Model	**Auxiliary Data**	**Cloud Masks Method**	**MAE ↓**	**SSIM ↑**	**PSNR ↑**
SRCNN	NDVI, DEM, LC, doy	I	7.215	0.973	35.307
		II	6.610	0.973	36.181
SE-SRCNN	doy	I	6.475	0.972	36.570
		II	6.733	0.948	35.961
EDSR 2x Bic. Inter. before	NDVI, DEM, LC, doy	I	6.685	0.974	36.276
		II	7.401	0.973	34.972
SRGAN	DEM, LC	I	6.869	0.971	35.464
		II	6.721	0.970	34.691
SE-EDSR 2x Bic. Inter. before	NDVI, DEM, LC, doy	I	6.665	0.973	36.189
		II	6.455	0.973	36.135
DR-SE-SRCNN	-	I	6.582	0.971	36.092
		II	6.582	0.971	35.529

Table 5.26: Comparison between models trained using Method I and II for cloud masks, where Method II was used to generate data for the test set and to calculate metrics

6 Conclusions

The thesis analyzed multiple SR deep learning networks for enhancing spatial resolution of Sentinel-3 BT using Landsat 8 LST as target. The study focuses also on the ablation of different auxiliary data sources that can help drive the task. The key idea was to leverage high temporal resolution of Sentinel-3 and high spatial resolution of Landsat 8 so that future stakeholder can produce daily Sentinel-3 HR images that are of extreme importance in microclimate analysis, that leads to better urban planning, people's health and environment monitoring.

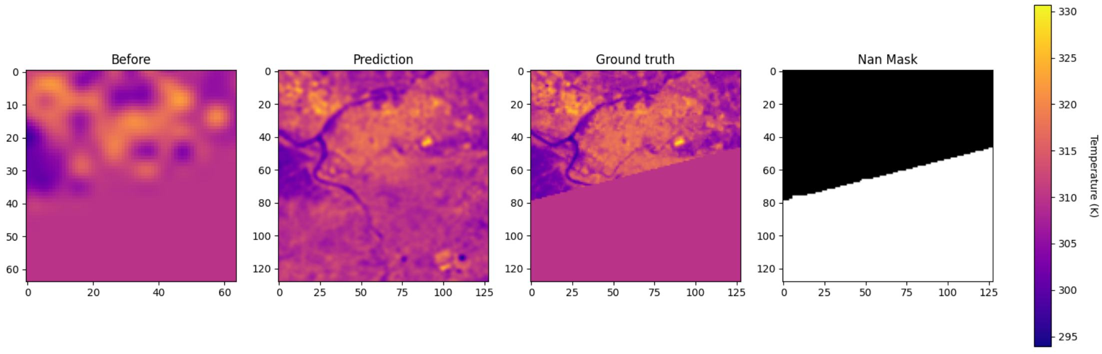
The analysis conducted with LR Landsat 8 images as training inputs demonstrates that the proposed models, i.e. SE-EDSR and DR-SE-SRCNN, tend to achieve better results than their predecessors. Nevertheless the best configuration for this task is EDSR 2x with bicubic interpolation applied before the network, that obtains MAE, SSIM and PSNR of 1.410 K, 0.983 and 46.380 respectively. Examples of Landsat 8 LR images enhanced using this configuration are shown in Figure 6.1.

On the other hand, for the task of the thesis — enhancing Sentinel-3 images with Landsat 8 as target — it is recorded that best metrics are yield by SE-SRCNN using BT2LST LR procuct and day of the year as additional data, with a MAE of 4.606 K, which is more than 1.5 K less than standard bicubic interpolation.

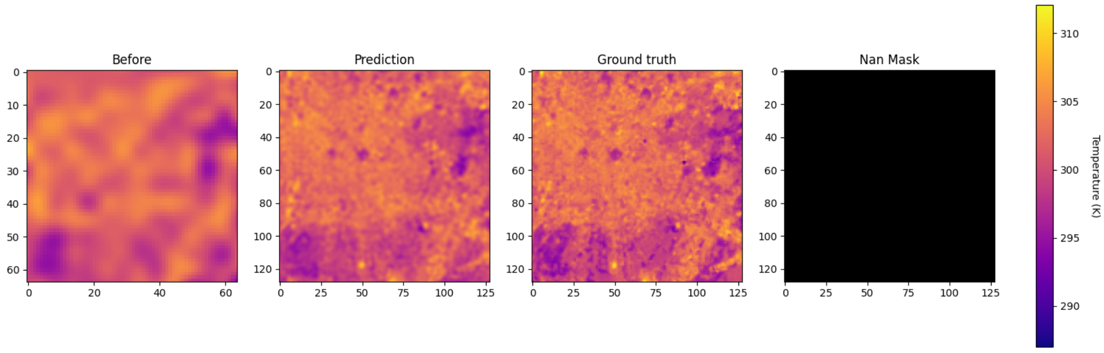
Furthermore, for unseen data, and more specifically for unseen cities such Istanbul, the best results are achieved by DR-SE-SRCNN with a MAE of 5.139 K, almost 0.8 K less than bicubic interpolation.

Lastly, the large amount of tests reported in this thesis confirm that overall additional data is beneficial and most importantly CNN-based neural networks outperform the baseline utilized for comparison, bicubic interpolation.

On a final note, the proposed method involving Gaussian distributions for the task of cloud masks is promising but requires further validation to fully establish its reliability in lack of cloud masks datasets.



(a) Turin



(b) Milan

Figure 6.1: Two examples of Landsat 8 LR image of Turin and Milan enhanced using EDSR 2x with bicubic interpolation before the network

6.1 Contribution

The key contributions of this thesis include:

1. **Review of established SR networks for remote sensing.** The study revisits well-established networks, such as SRCNN, SE-SRCNN, EDSR and SRGAN, for optical images SR task and apply them to the domain of remote sensing and in particular to BT and LST measurements.

2. **EDSR 10x.** The thesis explore a new upsampler module that performs 10x scaling for the EDSR architecture. Results show that the proposed module yield similar results to techniques that comprehend a combination of smaller upsamplers (2x, 3x and 4x) and bicubic interpolation, demonstrating its potential for future applications.

3. **SE-EDSR and DR-SE-SRCNN.** SE-EDSR architecture enhances EDSR by including the SE attention mechanism, that was proposed in the SE-

SRCNN work. Similarly DR-SE-SRCNN explores the performances of a deeper version of SE-SRCNN.

4. **SR with input and target differences.** This work analyzes for on of the first times the usage of SR networks where the LR input is not produced as a function of the HR counterpart, but rather come from a different dataset.
5. **Clouds Masks with Gaussian Distributions.** The thesis proposes a new method to mask clouds utilizing Gaussian distributions of the pixel-wise difference with respect to a cloud-free reference raster. Although it was impossible to validate numerically the methodology, visual analysis proved its potential.
6. **BT2LST methods.** The study implements a new method for image pixel-wise translation, that focuses on translating BT into LST. The method provide a valid alternative to the direct mapping between Sentinel-3 BT to Lansat 8 LST.

6.2 Future Works

This section expresses some suggestions on additional future works, that could extend the research of this thesis and further advance the field. Firstly, it would beneficial to limit, if not eliminate, all the dataset limitations described in the thesis, specifically those related to cloud cover and discrepancies in datasets measurements (i.e. BT and LST). Cloud-free images or cloud masks datasets directly downloaded from the satellite missions' databases could improve the performances and their reliability as valid results; however a deeper analysis on the proposed Gaussian Distribution Cloud Mask method could reveal the latter to be a valid alternative. For what concern the discrepancies in measurements it is suggested to either have Sentinel-3 LST product or calculate the values using more advanced methods that, though require additional physic quantities, can yield more accurate transformation from BT to LST.

As far as the super resolution task, new neural networks for optical images may emerge in the future that could be applied to remote sensing data. These networks could potentially decrease training and inference time, while also enhancing spatial resolution by factors greater than 10x, or achieving better results than the architectures explored in this research.

Finally it could be beneficial to increase dataset size, including more cities and rasters from different years.

Acknowledgments

I would like to express my deepest gratitude to my co-supervisors, Giacomo, Lorenzo, and Luca, for their guidance, support, and encouragement throughout this thesis. Their advices have been crucial for shaping this work. Additionally, this research would not have been possible without the LINKS Foundation, which supported me with the necessary resources.

I am profoundly grateful to have had Professor Giovanni Squillero as my supervisor, for his mentorship over the past year and his guidance not only on the academic path and beyond. His feedback provided a strong foundation for both my research and personal development.

I could not have undertaken this journey without the support of my family, especially my parents. Their support in my decisions, both past and recent, has been of extreme importance. I am also deeply grateful for their financial support, which has been essential throughout my studies. Above all, they have always believed in me, never for once doubting my abilities, and their faith has been a constant source of strength.

Many thanks to all of my friends, who have been a great source of fun throughout this journey. Whether it was sharing a laugh or simply taking a break from studying, they made sure I knew the importance of balance between work and leisure.

Last but not least, I'd like to acknowledge my cats, Altea, Oscar, and Bice, for all the entertainment and emotional support.

Bibliography

[1] Ying Sun. "Retrieval and application of land surface temperature". In: *Geo. Utexas. Edu* 1.1 (2008), pp. 1–27.

[2] Jiajia Yang et al. "Investigation and validation of algorithms for estimating land surface temperature from Sentinel-3 SLSTR data". In: *International Journal of Applied Earth Observation and Geoinformation* 91 (2020), p. 102136.

[3] *Landsat Missions | U.S. Geological Survey*. Sept. 2024. URL: `https://www.usgs.gov/landsat-missions`.

[4] *MODIS*. Nov. 2024. URL: `http://modis.gsfc.nasa.gov`.

[5] *Sentinel-3*. May 2024. URL: `https://sentiwiki.copernicus.eu/`.

[6] Nguyen Thanh Son et al. "Monitoring agricultural drought in the Lower Mekong Basin using MODIS NDVI and land surface temperature data". In: *International Journal of Applied Earth Observation and Geoinformation* 18 (2012), pp. 417–427.

[7] Stenka Vulova et al. "Summer nights in Berlin, Germany: modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 5074–5087.

[8] Huanfeng Shen et al. "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data". In: *Remote Sensing of Environment* 240 (2020), p. 111692.

[9] Hung Chak Ho et al. "Mapping maximum urban air temperature on hot summer days". In: *Remote Sensing of Environment* 154 (2014), pp. 38–45.

[10] Yuki Yasuda et al. "Super-resolution of near-surface temperature utilizing physical quantities for real-time prediction of urban micrometeorology". In: *Building and Environment* 209 (2022), p. 108597.

[11] Jingchao Yang et al. "A high spatiotemporal resolution framework for urban temperature prediction using IoT data". In: *Computers & Geosciences* 159 (2022), p. 104991.

[12] Jintong Han et al. "Microclimate spatio-temporal prediction using deep learning and land use data". In: *Building and Environment* (2024), p. 111358.

[13] Jianjun Zhang and Jing Li. *Spatial Cognitive Engine Technology.* Elsevier, 2023.

[14] Linwei Yue et al. “Image super-resolution: The techniques, applications, and future”. In: *Signal processing* 128 (2016), pp. 389–408.

[15] Bobby R Hunt. “Super-resolution of images: Algorithms, principles, performance”. In: *International Journal of Imaging Systems and Technology* 6.4 (1995), pp. 297–304.

[16] Jianchao Yang et al. “Image super-resolution via sparse representation”. In: *IEEE transactions on image processing* 19.11 (2010), pp. 2861–2873.

[17] Marco Bevilacqua et al. “Low-complexity single-image super-resolution based on nonnegative neighbor embedding”. In: (2012).

[18] Chao Dong et al. “Learning a deep convolutional network for image super-resolution”. In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13.* Springer. 2014, pp. 184–199.

[19] Chao Dong, Chen Change Loy, and Xiaoou Tang. “Accelerating the super-resolution convolutional neural network”. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14.* Springer. 2016, pp. 391–407.

[20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. “Accurate image super-resolution using very deep convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 1646–1654.

[21] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. “Deeply-recursive convolutional network for image super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 1637–1645.

[22] Bee Lim et al. “Enhanced deep residual networks for single image super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops.* 2017, pp. 136–144.

[23] Christian Ledig et al. “Photo-realistic single image super-resolution using a generative adversarial network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017, pp. 4681–4690.

[24] Xintao Wang et al. “Esrgan: Enhanced super-resolution generative adversarial networks”. In: *Proceedings of the European conference on computer vision (ECCV) workshops.* 2018.

[25] Huang G Liu Z Weinberger KQ. *Densely connected convolutional networks.* 2018.

[26] Lukas Liebel and Marco Körner. “Single-image super resolution for multispectral remote sensing data using convolutional neural networks”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41 (2016), pp. 883–890.

[27] R Nomura and K Oki. *Downscaling of MODIS NDVI by Using a Convolutional Neural Network-Based Model with Higher Resolution SAR Data. Remote Sens 13: 732.* 2021.

[28] Manmeet Singh et al. “DownScaleBench for developing and applying a deep learning based urban climate downscaling-first results for high-resolution urban precipitation climatology over Austin, Texas”. In: *Computational Urban Science* 3.1 (2023), p. 22.

[29] Darren Pouliot et al. “Landsat super-resolution enhancement using convolution neural networks and Sentinel-2 for training”. In: *Remote Sensing* 10.3 (2018), p. 394.

[30] Tomoki Izumi et al. “Super-resolution of sea surface temperature with convolutional neural network-and generative adversarial network-based methods”. In: *Journal of Water and Climate Change* 13.4 (2022), pp. 1673–1683.

[31] Binh Minh Nguyen et al. “Convolutional neural network modelling for modis land surface temperature super-resolution”. In: *2022 30th European Signal Processing Conference (EUSIPCO)*. IEEE. 2022, pp. 1806–1810.

[32] *European Space Agency*. Aug. 2024. URL: `https://www.esa.int`.

[33] *Landsat 8*. Jan. 2023. URL: `https://landsat.gsfc.nasa.gov/satellites/landsat-8`.

[34] Nathalie Pettorelli et al. “Using the satellite-derived NDVI to assess ecological responses to environmental change”. In: *Trends in ecology & evolution* 20.9 (2005), pp. 503–510.

[35] Copernicus Data Space Ecosystem. *Copernicus DEM - Global and European Digital Elevation Model*. Sept. 2024. URL: `https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM`.

[36] *Home - Copernicus Contributing Missions Online*. Oct. 2024. URL: `https://spacedata.copernicus.eu`.

[37] *Urban Atlas Land Cover/Land Use 2018 (vector), Europe, 6-yearly, Jul. 2021*. Jan. 2017. URL: `https://sdi.eea.europa.eu/catalogue/copernicus/api/records/fb4dffa1-6ceb-4cc0-8372-1ed354c285e6?language=all`.

[38] *Copernicus Land Monitoring Service*. Oct. 2024. URL: `https://land.copernicus.eu/en`.

[39] Visual Crossing Corporation. *Weather Data & Weather API | Visual Crossing*. June 2024. URL: `https://www.visualcrossing.com`.