Chapter 1

Introduction

REMOTE sensing involves collecting information from a distance without touching the object physically. It collects electromagnetic radiation from the Earth's surroundings using geospatial technology, allowing remote feature detection and monitoring. This technique generally employs active or passive sensors found on satellites or aircraft.

Remote sensing proves valuable across diverse fields like crop mapping, building recognition, vegetation index calculation, and change detection. However, optical remote sensing faces a perennial challenge in the form of persistent cloud cover. Dense clouds, appearing impenetrable in optical frequency bands, distort the spectral reflectance signal by obstructing the view of the underlying land. Studies conducted by the International Satellite Cloud Climatology Project (ISCCP) indicate that over 66% of Earth's surface is cloud-covered when observed from satellites. Consequently, distorted images reveal significant data gaps in both temporal and spatial space. This cloud cover poses a major obstacle for applications requiring continuous time series data, such as agricultural management and disaster monitoring. Hence, removing clouds from satellite images is essential to enhance the efficiency of cloud removal algorithms.

Researchers have explored a range of traditional and state-of-the-art algorithms in the domain of cloud removal from satellite imagery. However, several datasets used for this purpose have

been either based on simulated cloud cover or have lacked adequate size despite utilizing real cloud data. The introduction of machine learning techniques in image processing has facilitated the development of models for cloud removal. Nonetheless, these models demand substantial datasets for effective training. Creating well-organized and carefully handled datasets is crucial when developing machine learning models for extracting information from remote sensing data. The research community has shared many publicly available cloud removal datasets, but these datasets come with certain limitations. These limitations include restricted spatial coverage, a lack of temporal data, or a shortage of available samples. Traditional shallow learning techniques can be applied to smaller datasets, such as the well-known Indian Pines scene [1]. However, with the rise of modern deep learning, substantial data volumes are essential to achieving the desired level of generalization [2]. Yet, computer vision typically works with regular photos of everyday objects, while interpreting remote sensing data is more versatile and challenging. Therefore, compared to conventional image databases like ImageNet [3], comprehensively labelled repositories for remote sensing need to be improved. Although initial steps have been taken, including an overview of some existing datasets presented in Table 1.1, some datasets exist, and an ongoing effort persists to bridge this gap, with additional datasets provided mainly through machine learning competitions, often accessible through private repositories [4].

For the task of cloud removal and making algorithms/models that remove clouds from a given dataset (satellite image), we require cloud masks of where the cloud is present in any image. There are traditional approaches that can help you find a cloud mask. We have Physics-based methods such as Fmask [4], LaSRC [5], Sen2Cor [6], MAJA [7], and others [8]. Physics-based techniques function well when the satellite product has the necessary signals (such as thermal band) and the physical presumptions (such as the presence of observable parallax) are satisfied [5] [9]. However, when we do not have a thermal band in our images, in the case of high-resolution images, e.g., Planetscope, then Cloud detection/masking becomes difficult using the above-mentioned algorithms. Therefore, we must derive other methods to detect cloud, shadow and haze. Other methods may include machine learning-based methods. Previous attempts [10] [11] [12] focused mostly on conventional machine learning techniques, including decision trees, Bayesian approaches, SVM, random forests, and more. Recent advancements have focused on deep learning models, such as CNN-based classification frameworks [13] and semantic segmentation using residual learning [12]. Cloud-free scenes were also produced using Generative

TABLE 1.1: List of the organized dataset with details

Dataset	**No. of Images**	**Size of Images**	**Source**	**Dataset details**
Brazilian Coffee [15] Scenes Dataset	51,004	64 × 64	SPOT	Two classes with imbalance distributions, representing instances of non-coffee and coffee
SAT-4 [16]	5,00,000	28 × 28	Color aerial imagery (R, G, B, NIR)	Four agricultural categories across the continental United States
SAT-6 [16]	4,05,000	28 × 28	Color aerial imagery (R, G, B, NIR)	Across the continental United States, there are six distinct land cover classifications
USGS SIRI-WHU [17]	1	10,000×9,000	Color aerial image	Within Montgomery Country, Ohio, USA, there are four distinct classification categories
SIRI-WHU [18]	200	200 × 200	Google Earth	China's urban areas are characterized by twelve distinct classes.
Inria Aerial Image Labeling Dataset [19]	360	1,500 × 1,500	Color aerial imagery	For 10 cities in USA and Austria, there are two categories i.e. buildings and non-building areas.
UC Merced Land Use Dataset [20]	2,100	256 × 256	Color aerial images	21 categories of land use
DOTA [21]	2,806	800 × 800 to 4,000 × 4,000	Color aerial images	There are 188,282 occurrences of 15 object categories, each marked by a quadrilateral label.
EuroSAT [22]	27,000	64 × 64	Sentinel-2	10 categories
SEN1-2 [23]	5,64,768	256 × 256	Sentinel-1 and Sentinel-2	Paired sets of SAR (single-polarization intensity) and optical (RGB) images, lacking any annotations
DeepGlobe: Land Cover Classification [24]	1,146	2,448 × 2,448	Worldview-2/-3, GeoEye-1 (R, G, B)	7 categories of land cover
DeepGlobe: Road Extraction [24]	8,570	1,024 × 1,024	Worldview-2/-3, GeoEye-1 (R, G, B)	roads over India and Thailand (1 class)
2017 IEEE GRSS Data Fusion Contest [25]	57	from 447 × 377 to 1,461 × 1,222	Sentinel-2, Landsat, OpenStreetMap	17 classes representing local climate zones
BigEarthNet [26]	5,90,326	up to 120 × 120	Sentinel-2	43 Land Cover classifications across the European region
38-Cloud [27]	17,601	384 × 384	Landsat 8 (R, G, B, NIR)	clouds as target class

Adversarial Networks (GANs) [14] However, they rely on large datasets with accurate information for training.

Chapter 2

Dataset Essentials

IN our research, we utilized publicly accessible data from the PlanetScope satellite provided by Planet Labs. Focusing our investigation on the region of Ropar in Punjab, India, we meticulously collected and curated a dataset.

2.1 Planetscope data details

PlanetScope, a satellite constellation managed by Planet Labs Inc. for Earth observation, provides frequent and detailed imagery of the Earth's surface. Consisting of small satellites in low-altitude orbits, the constellation captures high-resolution multispectral data on a daily basis. These satellites excel in optical imagery acquisition, delivering high-resolution snapshots with a spatial resolution of 3 meters per pixel, contingent on specific configurations.

PlanetScope records data across multiple spectral bands, encompassing Coastal Blue (431 - 452 nm), Green I (513 - 549 nm), Green (547 – 583 nm), Yellow (600 - 620 nm), Red (650 – 680 nm), RedEdge (697 – 713 nm), Blue (465 – 515 nm), and NIR (845 – 885 nm). It's important to note that not all scenes downloaded from PlanetScope comprise 8 bands; some are in a 4-band TOAR (Top of Atmosphere Reflectance) format. Consequently, extracting data at the 8-band level in

bulk is not feasible. As a result, we have curated our dataset with 4-band images, namely Red, Blue, Green, and NIR.

These distinct channels facilitate a comprehensive analysis of the Earth's surface. They support various applications, such as monitoring coastal features using Coastal Blue and assessing vegetation health and composition with RedEdge and NIR. The Blue and Green channels provide detailed visual interpretation, while the Red and NIR bands play a crucial role in land cover classification and vegetation monitoring.

The multispectral capabilities of PlanetScope enable Earth observation across different spectral bands, contributing to applications like agriculture monitoring, environmental assessment, urban planning, and disaster response.

A notable feature of PlanetScope is its ability to deliver near-real-time data and frequent revisits to specific locations, facilitating dynamic monitoring of surface changes. This high revisit frequency proves beneficial for applications requiring timely information, such as tracking crop development, assessing environmental shifts, and responding to natural disasters. Furthermore, the subscription-based model for accessing PlanetScope's data enhances its accessibility, catering to a diverse user base, including researchers, government agencies, and commercial entities. With its combination of high-resolution imagery, frequent revisits, and global coverage, PlanetScope emerges as a valuable resource for monitoring and analyzing diverse aspects of the Earth's surface, offering a scale of observation previously unattainable with traditional satellite imaging systems.

Chapter 3

Dataset Curation

THE collection of data products that are made available for download is created by applying several processing stages [28] to PlanetScope pictures.

3.1 Calibration of Sensors and Radiometry

In the domain of sensor and radiometric calibration, several crucial procedures are involved. The Darkfield/Offset Correction plays a key role in addressing sensor bias and reducing dark noise. This means that by aggregating on-orbit darkfield imaging data across temperature categories, consolidated (master) offset tables must be created. These tables are then utilized during scene processing, considering the charge-coupled device (CCD) temperature at the time of image acquisition. Another essential step is the Flat Field Correction, which requires the collection of flat fields for each optical instrument before satellite deployment. To match the sensor's ideal response area, these fields are employed to manage CCD element effects and adjust image illumination. Importantly, flat fields undergo regular updates over the satellite's operational lifespan.

Moving on, the Camera Acquisition parameters Correction makes sure that every picture has a constant radiometric response. This adjustment is made globally, taking into account differences in exposure duration, camera temperature, TDI stage count, gain, and other pertinent camera factors. Absolute Calibration, which transforms geographically and temporally adjusted datasets, is the last step in the calibration process. This conversion translates digital number values into physically based radiance values, meticulously scaled to $\mathrm{W}/(\mathrm{m}^2\mathrm{str}\mu\mathrm{m}) * 100$. One can consult the comprehensive documentation on On-Orbit Radiometric Calibration in Planet Satellite Fleet for more technical details.

3.2 Terrain Rectification (Orthorectification)

The subsequent pivotal step in assembling datasets is Terrain Rectification, a procedure crafted to eradicate distortions in the landscape. This process transpires through a pair of consecutive stages. Initially, the rectification anchor process pinpoints tie points throughout the original images. Concurrently, an assemblage of benchmark images encompassing NAIP, ALOS, and Landsat is employed to formulate RPCs, also called Rational Polynomial Coefficients. These RPCs, in sequence, facilitate the effective rectification of scenes, eliminating terrain distortions. Essentially, the landscape model utilized in the terrain rectification procedure is derived from diverse origins, such as Intermap, SRTM, NED, and various regional elevation datasets. This landscape model undergoes periodic revisions to ensure precision. Preserved snapshots of the elevation datasets utilized play a pivotal role in recognizing the specific Digital Elevation Model (DEM) applied for any given scene at a particular juncture.

3.2.1 Surface Reflectance Product Processing

The final step in the dataset collection journey is Surface Reflectance Product Processing, primarily focused on mitigating atmospheric effects. This intricate process unfolds through three distinct steps. At first, the at-sensor radiance product's coefficients (RPCs) are used to compute the Top of Atmosphere reflectance (TOA).

After that, the 6SV2.1 radiative transfer algorithm is used to create a Lookup Table (LUT), which is then enhanced by MODIS(satellite) close to real-time inputs. For the TOA reflectance to be

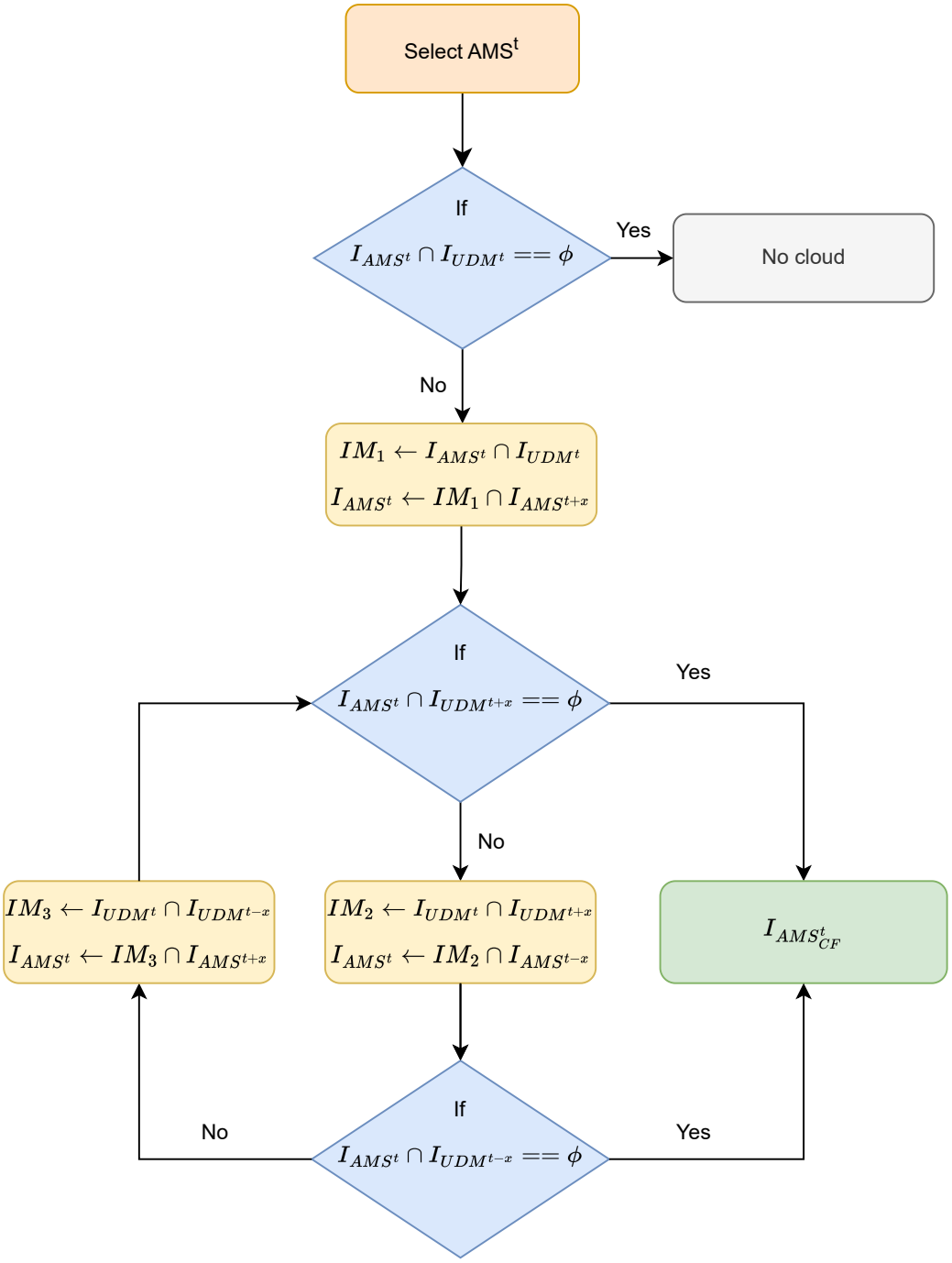

FIGURE 3.1: Dataset preparation process: A comprehensive flowchart illustrating the process of dataset creation, from initial data collection through preprocessing.

converted to surface reflectance later on, this LUT development is essential. The conversion process considers all conceivable combinations of selected physical conditions and accounts for the unique spectral response of each satellite sensor type, alongside estimates of the prevailing atmospheric state.

3.3 Dataset Preparation

In the below subsections, we delved into three key aspects of satellite imagery processing: Image Mosaicking, Cloud Removal, and Image Tiling. Image mosaicking involves merging multiple satellite images into a seamless composite. Challenges arise when dealing with different scenes for the same date, necessitating georeferencing, resampling, and blending for a visually consistent result. After image mosaicking, we obtained a single image of a given date.

Cloud removal is essential for obtaining clear and usable remote sensing data. Leveraging temporal data and a UDM2 cloud mask, this process identifies and eliminates cloud-affected pixels. Temporal analysis, including techniques like temporal filtering and image composition, is crucial for generating cloud-free composites.

Image tiling is a strategy to efficiently manage and process large satellite datasets. By dividing the imagery into smaller tiles, each section can be processed independently, facilitating parallel processing and enhancing computational efficiency. Tiling is particularly useful for handling extensive datasets and optimizing resource utilization in image processing workflows.

These subsections collectively address key challenges in satellite image processing, providing insights into creating seamless mosaics, ensuring clear data through cloud removal, and optimizing processing through image tiling strategies.

3.3.1 Mosaicking of Planetscope Images dataset

In the initial phase of our dataset acquisition, we meticulously selected a Region of Interest (ROI) to narrow our study focus to a specific geographic area. The data sourced from Planet.com provided a diverse collection of distinct scenes or patches for the same date, spanning a duration of 4-5 years within the designated ROI. Notably, certain dates were missing from our dataset,

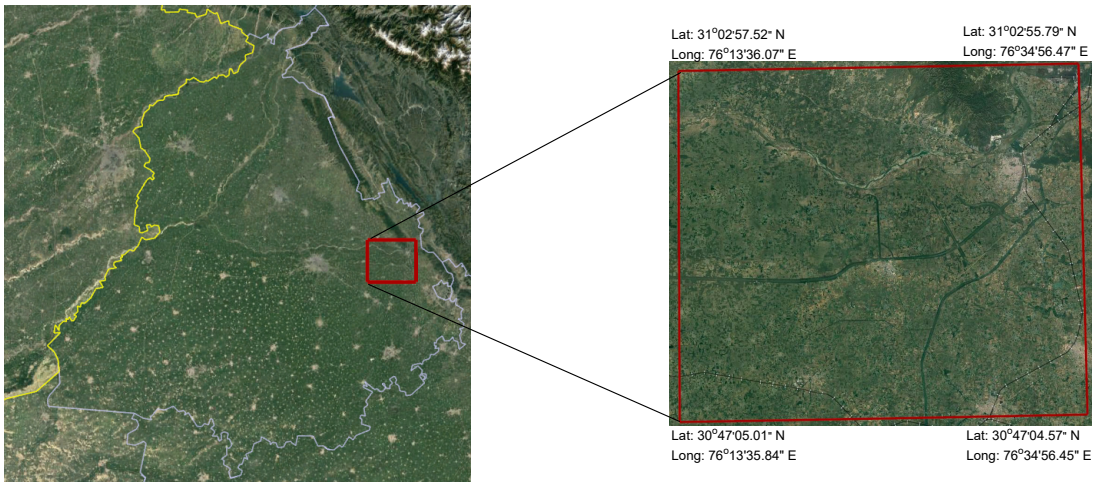

FIGURE 3.2: Representation of the ROI for Ropar region of Punjab, India. The red rectangle represents the large tile, which was further divided into sub-tiles.

either due to the absence of satellite coverage on those specific dates or complete coverage with limited visibility. To ensure the temporal coherence of our dataset and facilitate a comprehensive analysis, we undertook the task of mosaicking these diverse scenes of the same date. This involved combining multiple scenes for a given date to create a single image viz. AnalyticMS (Multispectral) data and its corresponding Usable Data Mask (UDM2) cloud mask. To integrate this information efficiently, we leveraged the rasterio.merge library, playing a crucial role in the seamless creation of mosaics for corresponding UDM2 and Analytic Multi-spectral images captured on the same date and within the designated ROI. By merging these scenes of each date, our objective was to curate a seamless and unified dataset, overcoming variations inherent in individual patches and enabling a more accurate representation of the temporal evolution of the designated area.

3.3.2 Image tiling

In our study, we utilized a high-resolution image of dimensions $11,461 \times 9,942$ pixels, which was subsequently partitioned into tiles of size 256×256. This partitioning resulted in 460 individual images for a single date, thereby facilitating the generation of an extensive dataset. Subsequently, we meticulously curated pairs from these tiles, specifically focusing on instances featuring cloud formations.

3.3.3 Cloud Removal

Introducing an innovative cloud removal algorithm for satellite imagery processing, this method addresses the persistent challenge of cloud cover interference in optical remote sensing. The algorithm, consisting of four key steps, begins by extracting a cloud mask from the Usable Data Mask (UDM2) image, providing detailed information about cloud presence for each pixel shown in Figure 3.1. Subsequently, it strategically selects a specific date for the AnalyticMS image, serving as the foundation for cloud replacement. Through a meticulous process, the algorithm then replaces cloudy AnalyticMS images with those from the following day, and, in cases of persistent clouds, it further refines the replacement by checking the cloud mask from the previous day. This robust methodology offers a systematic approach to enhance the quality and utility of satellite observations by effectively mitigating the impact of cloud cover.

3.3.4 Curation of the Dataset

The following steps illustrate the process of dataset creation from the satellite images:

Step 1: Obtain Cloud Mask from Usable Data Mask (UDM2) Image ($\mathrm{t\_udm}$):

Obtain the cloud mask from the UDM2 image ($\mathrm{t\_udm}$) downloaded from Planetscope for each date in the dataset. The UDM2 image provides information about the cloud presence for each pixel.

Step 2: Select AnalyticMS Image of a specific date ($\mathrm{t\_ams}$):

Choose a specific date ($\mathrm{t\_ams}$) for the AnalyticMS image. This image serves as the base image for cloud replacement.

Step 3: Replace Cloudy AnalyticMS Images with next day ($\mathrm{t} + 1\mathrm{\_ams}$):

For each date where clouds are present in the UDM2 cloud mask ($\mathrm{t\_udm}$), replace the AnalyticMS image ($\mathrm{t\_ams}$) with the AnalyticMS image from the next day ($\mathrm{t} + 1\mathrm{\_ams}$) where clouds are also present. This replacement is performed by checking the intersection of the cloud masks ($\mathrm{t\_udm}$ and $\mathrm{t} + 1\mathrm{\_udm}$).

Step 4: Check for Persistent Clouds and Replace with previous day ($\mathrm{t} - 1\mathrm{\_ams}$):

If clouds are still present in both the cloud masks $\mathrm{t\_udm}$ and $\mathrm{t} + 1\mathrm{\_udm}$ after the

replacement, check the cloud mask from the previous day $\mathrm{t}-1\_\mathrm{udm}$. Replace the AnalyticMS image $\mathrm{t}+1\_\mathrm{ams}$ with the AnalyticMS image from the previous day $\mathrm{t}-1\_\mathrm{ams}$ where clouds are persistently present in both t_udm and $\mathrm{t}+1\_\mathrm{udm}$.

Following this process, we iteratively replace cloudy AnalyticMS images with images from adjacent days until a cloud-free image is obtained. This approach leverages the UDM2 cloud masks to make informed decisions about cloud presence and ensures the selection of the most suitable images for a cloud-free dataset.

Algorithm 1 Cloud Removal Process

Input: $\{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t}}, \mathrm{UDM}^{\mathrm{t}}, \mathrm{UDM}^{\mathrm{t+x}}, \mathrm{AMS}_{\mathrm{C}}^{\mathrm{t+x}}, \mathrm{AMS}_{\mathrm{CF}}^{\mathrm{t-x}}\}; \mathrm{x} \in \{1,2,3\}$

Output : $\mathrm{I}(\mathrm{AMS}_{\mathrm{CF}}^{\mathrm{t}})$

Could removal using UDM2(Planetscope)

Objective:

$\mathrm{I}_{\mathrm{C}}^{\mathrm{t}} \rightarrow \mathrm{I}_{\mathrm{CF}}^{\mathrm{t}}$

Procedure:

while $\mathrm{I}_{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t}}} \neq \mathrm{I}_{\mathrm{CF}}^{\mathrm{t}}$ **do** ▷

$\mathrm{IM}_1 \leftarrow \mathrm{I}_{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t}}} \cap \mathrm{I}_{\mathrm{UDM}^{\mathrm{t}}}$

$\mathrm{I}_{\mathrm{AMS}^{\mathrm{t}}} \leftarrow \mathrm{IM}_1 \cap \mathrm{I}_{\mathrm{AMS}^{\mathrm{t+x}}}$

$\mathrm{IM}_2 \leftarrow \mathrm{I}_{\mathrm{UDM}^{\mathrm{t}}} \cap \mathrm{I}_{\mathrm{UDM}^{\mathrm{t+x}}}$

$\mathrm{I}_{\mathrm{AMS}_{\mathrm{CF}}^{\mathrm{t}}} \leftarrow \mathrm{I}_{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t-x}}} \cap \mathrm{IM}_2$

$\mathrm{IM}_3 \leftarrow \mathrm{I}_{\mathrm{UDM}^{\mathrm{t}}} \cap \mathrm{I}_{\mathrm{UDM}^{\mathrm{t-x}}}$

$\mathrm{I}_{\mathrm{AMS}_{\mathrm{CF}}^{\mathrm{t}}} \leftarrow \mathrm{I}_{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t+x}}} \cap \mathrm{IM}_3$

The notation $\mathrm{I}_{\mathrm{AMS}_{\mathrm{C}}^{\mathrm{t}}}$ is decoded as follows: 'I' signifies any satellite image, 'AMS' denotes the AnalyticMS image of date 't'. Similarly, 'UDM' represents the Combined Mask of date 't'. The superscript '$\mathrm{t} \pm \mathrm{x}$' signifies the image before or after date 't', with 'x' indicating the number of days to move back or forth. The subscript 'C' designates a Cloudy image, and 'CF' stands for a cloud-free image. The variables IM_1 and IM_2 represent the masked portion of the image after superimposing the UDM mask on the AMS image. The intersection operation $(\cap)$ refers to the superimposition of one mask or image on the other. In Algorithm 1, the process begins by finding the cloud mask (Usable Data Mask (UDM)) for the cloud image as Analytic Multispectral (AMS) at time t. Then, the cloudy area of AMS at time t is determined using the

corresponding UDM. To address cloud coverage, a search is conducted for a cloud-free image for the previous and next day, i.e., $\mathrm{t}-1$ or $\mathrm{t}+1$. The next step involves replacing the cloudy region in AMS at time t with the cloud-free region from the previous or next day $\mathrm{AMS}_{\mathrm{t}-1}$ or $\mathrm{AMS}_{\mathrm{t}+1}$, respectively. If the selected day provides a cloud-free image, the process stops. However, if there is an intersection in the cloud mask for the previous or next day, the algorithm iteratively repeats the process for $\mathrm{t}+2$ or $\mathrm{t}-2$, replacing the cloudy areas of AMS at time t with those from $\mathrm{t}-2$ or $\mathrm{t}+2$. This iteration continues until a cloud-free region is successfully identified, providing an effective method for cloud removal in the AMS image at time t.

3.4 The PLA4MS Dataset

The PLA4MS dataset comprises 100,000 pairs of cloud and cloud-free .GEOTIFF images captured by the Planetscope satellite over the Ropar region in Punjab, India. These image pairs provide valuable data for research and development in the field of satellite imagery processing, particularly in addressing challenges associated with cloud cover. The dataset is a valuable resource for training and testing algorithms designed to enhance the quality of satellite observations by effectively handling and removing clouds from the imagery. Figure 3.3 provides an overview of Regions of Interest (ROI) and offers comprehensive insights into the dataset.

The dataset size is nearly 95GB, emphasizing the substantial volume of data available within the PLA4MS dataset. This ample size suggests a comprehensive and diverse collection of cloud and cloud-free images, providing a rich source for various applications, such as machine learning-based cloud removal algorithms, land cover classification, and environmental monitoring. Researchers and practitioners can leverage the extensive content of the dataset to develop and evaluate methodologies that contribute to improved satellite imagery analysis.

The accompanying Figure 3.2 visually conveys the selected Region of Interest (ROI) within the PLA4MS dataset. Researchers can use this visual representation to better interpret and contextualize the dataset's content, facilitating more targeted analysis and application in diverse remote sensing projects.

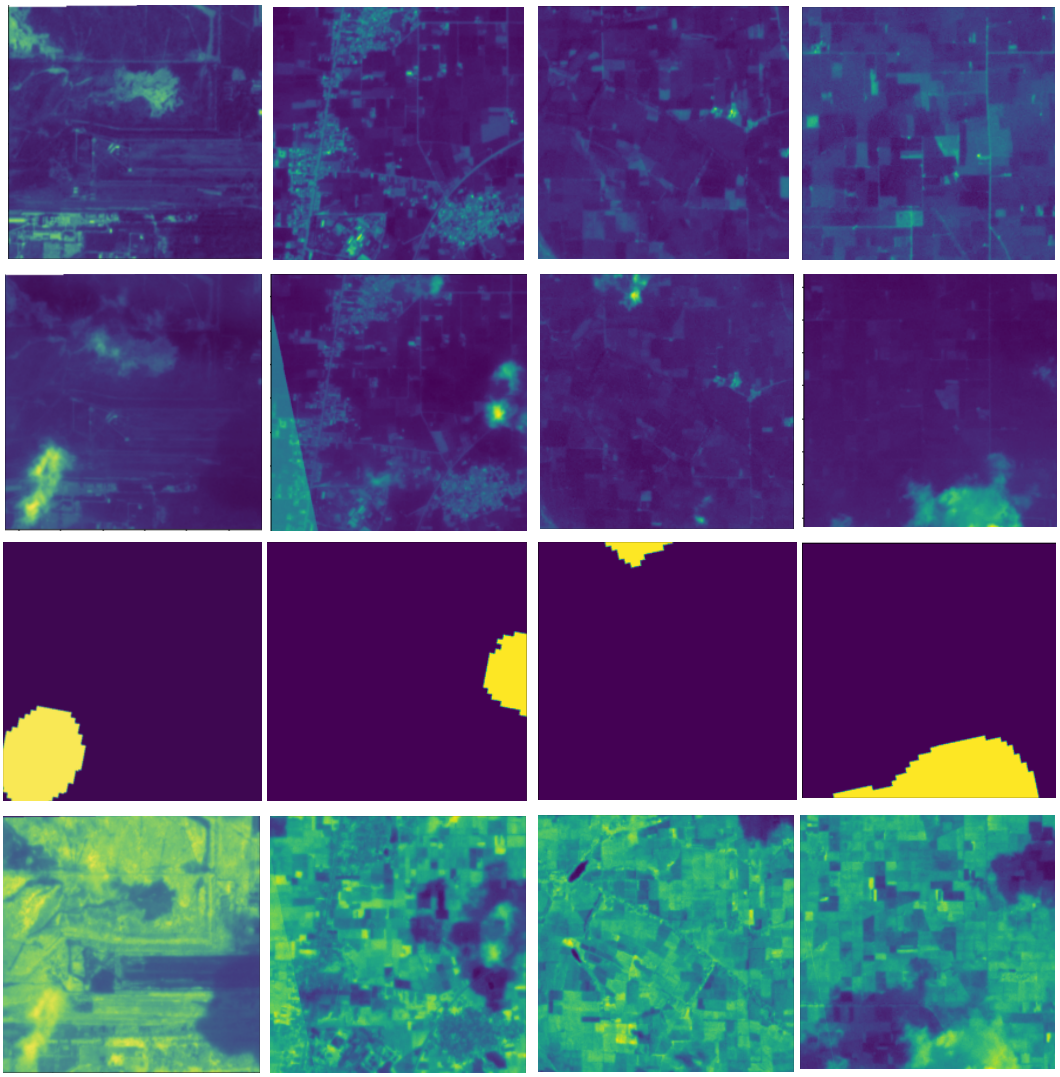

FIGURE 3.3: Example Dataset featuring four Landmarks. Each column denotes four Regions of Interest (ROIs). The initial row illustrates cloud-free images, the second row depicts cloudy images from the same locations, the third row showcases the corresponding cloud masks for the second row, and the fourth row displays the Near-Infrared (NIR) band of the corresponding cloud-free image.

3.4.1 Final Dataset Structure

This research leverages satellite imagery acquired from the Planetscope constellation, accessible through the platform planet.com. The dataset employed herein is composed of paired images, each capturing identical geographical locations on the same date, with one image depicting cloud-covered conditions and its counterpart showcasing a cloud-free scene. The focal area of investigation is the Ropar region, situated in Punjab, India. The images were collected across multiple years, spanning from 2018 to 2023, allowing for a comprehensive analysis of temporal changes in the study area.

Each image in the dataset initially had a resolution of $11,461 \times 9,942$ pixels. This implies that, due to a spatial resolution of 3 mts, which translates to a length of $34,383 \times 29,826$ mts for the Region of Interest (ROI). To enhance processing efficiency and facilitate detailed analysis, these images were subsequently divided into smaller tiles, each measuring 256×256 pixels. This division translates to a spatial coverage of approximately 1.5×1.5 sq km per tile. This dataset structure not only enables a more manageable and standardized unit for analysis but also provides a valuable resource for studying land cover dynamics, monitoring environmental changes, crop growth monitoring, and assessing the impact of cloud cover on satellite observations over the specified region.

The file naming convention is structured as PLA4MS.XXX.tif, where PLA signifies Planetscope, indicating that the images originate from the Planetscope satellite. The numeral 4 denotes the presence of four channels, while MS designates the images as Multispectral. The variable XXX functions as the unique image identifier within this naming scheme. This standardized format ensures clarity and provides key metadata information related to the Planetscope satellite images, facilitating effective organization and identification of data in research endeavours.

3.5 Utility of the Dataset

The dataset primarily facilitates the enhancement and restoration of cloudy images by utilizing their clear counterparts as references, employing techniques such as image fusion and dehazing. Furthermore, this dataset proves instrumental in training machine learning models, enabling the development of robust algorithms for tasks like object detection and segmentation under varied atmospheric scenarios. Additionally, the comparative analysis of these image pairs supports change detection studies, allowing for the identification of alterations in land use, urban development, and environmental dynamics. Beyond this, the dataset contributes to cloud cover pattern analysis, aiding meteorological investigations, and serves as a valuable resource for optical remote sensing calibration, enhancing the accuracy of quantitative analyses.

Chapter 4

Baseline Model Implementation

THE used model 4.1, called DSen2-CR[29], is based on the super-resolution Deep Sentinel-2 (DSen-2) ResNet presented in Lanaras et al[30]. (2018), which is itself derived from the state-of-the-art single-image super-resolution EDSR network (Lim et al., 2017)[31]. Similarly to superresolution, cloud removal can be seen as an image reconstruction task, where missing spatial and spectral information has to be integrated into the image to restore the complete information content.

4.1 Baseline - DSen2CR

A SAR picture is used by DSen2-CR as a kind of prior. Sentinel-1 data from the same scene is added as input to the network for this purpose. The SAR channels of the picture are merely appended to the other optical image channels. The cloud identification and treatment, together with the extremely non-linear SAR-to-optical translation, are implicitly learned and executed within the network. A cloud-free picture of the identical scene is sent via the network as a goal for the loss of computation during the end-to-end training process. A representation of the DSen2-CR model and the applied residual block design is presented in Figure 3. Additional characteristics and features of our model is detailed below:

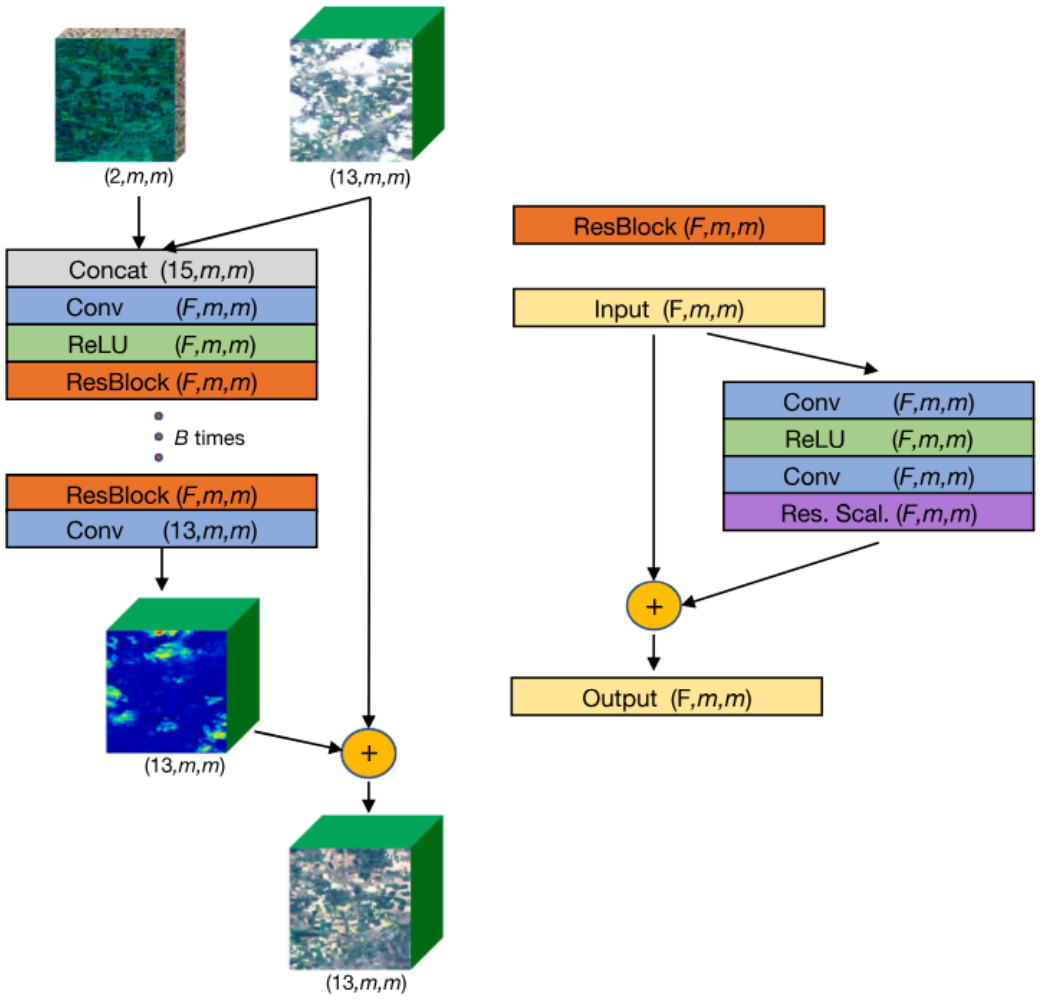

FIGURE 4.1: DSen2-CR model diagram on left and design of residual blocks on right. Both the spatial dimensions and the total quantity of layers for each component of the network are shown inside parentheses. The network may receive input images of any spatial dimension m during training and prediction times because it is fully convolutional. The chosen feature dimension is denoted by F, while the number of residual blocks that comprise the network is indicated by B.

• Long-Skip connection: As first suggested by Lanaras et al. (2018)[30], an additive shortcut transfers the input fuzzy image to a separate addition layer just prior to the final output. This essentially means that every pixel of an input hazy image is being corrected, and the network as a whole is learning to forecast a residual map. In the event of clear sky inputs or thin clouds, there won't be any or very few expected corrections. Conversely, there will be more corrections for dense clouds that appear bright. • Residual blocks: A series of sequentially stacked residual units make up the majority of the network. One hyperparameter that determines the network's depth is the precise number of units B. Four layers make up each residual unit, plus an additional layer to form the residual connection. A 2D convolution layer, with subsequent ReLU

activation, another 2D convolution layer, and an additional residual scaling layer are the four layers that are skipped (see the next point). Since the network is meant to predict adjustments that might be both positive and negative, only one ReLU activation is employed after the initial layer of convolution and not after the second. Three-by-three kernels are employed for both convolutional layers, according to the wider community trend of using smaller kernels in deeper models (Lanaras et al., 2018[30]). The number of filters, or the output characteristic dimension F, is a hyperparameter that remains constant for all units. The spatial dimensions associated with the data are always preserved across the network by using a stride of one pixels and zero padding. Residual units with F = 256 features were chosen as a baseline for the DSen2-CR design, taking into account both personal trials and the documented experience in Lanaras et al. (2018)[30] and Lim et al. (2017)[31]. This represented a compromise between symbolic capacity and computing complexity.

Residual scaling: A custom layer that multiplies its inputs by a constant scalar is known as the residual scaling layer. Originally suggested according to Szegedy et al. (2017), the training is stabilized by this activations scaling without the need for extra parameters, as those in batches of normalization layers. In this work, 0.1 is used as the scaling constant value.

In our case, we are taking only one input of 4 band image instead of 2 input images having 13 and 2 band data each. This change has been done in the architecture of the model too for training and testing purposes.

4.1.1 Cloud-adaptive regularized loss

The original cloudy image and the goal cloud-free optical images were taken on separate days throughout the same meteorological season, as detailed in the dataset section. Even with a small time difference, variations in the surface parameters between the photos are frequently discernible, particularly in agricultural settings. The goal of a cloud removal method is to recover ground information beneath clouds while leaving unaltered areas; hence, it is critical that the output retains as much of the original image as possible. A customized training loss was created in this work to reduce the impact of ground modifications on the target picture.

Because of the Sentinel-2 data's strong dynamic range and resilience to huge deviations, the L1 measure (mean absolute error) was chosen as the fundamental error function after the advice

of Lanaras et al. (2018)[30]. The traditional target loss T based on the straightforward $\mathrm{L1}$ distances between predicted & target can be expressed as follows, where the predicted output image is defined as P and the expected cloud-free objective image as T.

$$\mathrm{T} = \frac{1}{\mathrm{N_{tot}}} \sum_{\mathrm{i}=1}^{\mathrm{N_{tot}}} |\mathrm{P_i} - \mathrm{T_i}|$$

where $\mathrm{N_{tot}}$ represents the total quantity of pixels across all optical image channels. The network's training on multi-temporal input with shifting ground conditions is to learn, forecast, and apply undesired surface changes. This makes the optimization of this basic $\mathrm{L1}$ loss easy to understand, but it has a disadvantage. We proposed a novel loss theory to minimize these artefacts. The objective is to maximize the retention of input data by guiding the learning process with the use of a binary cloud plus cloud-shadow mask (CSM) integrated into the loss computation. We refer to this customized loss as Cloud-Adaptive Regularized Loss (CARL), and it is expressed as follows:

$$\mathrm{CARL} = \frac{1}{\mathrm{N_{tot}}} \left(\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{P_i} - \mathrm{T_i}| + \sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{P_i} - \mathrm{I_i}| \cdot \mathrm{CSM_i} \right)$$

where P, T, and I stand for the anticipated, target, and input optical images, respectively. The CSM mask has pixels with values of 1 for clouds and shadows and 0 for uncorrupted pixels, with the same spatial dimensions as the pictures. A matrix of ones having identical spatial dimensions like the pictures and the CSM is indicated by the symbol $\mathbf{1}$. The element-wise multiplications applied over all channels, denoted by $\cdot$, are made by combining the CSM and the image differences. In the cloud-adaptive section, the mean absolute error loss is calculated with respect to the input image itself for clear-sky pixels and with respect to the target image for overcast or shadowed pixels.

4.2 Experiments

Schmitt et al.'s 2019 SEN12MS dataset is expanded upon by the SEN12MS-CR dataset utilized in this study. The original SEN12MS dataset was produced for remote sensing applications like

scene categorization and land cover mapping. It is publically accessible and consists of a collection of Sentinel-2 optical and SAR images and MODIS land cover maps. While SEN12MS-CR uses the same methodology as its predecessor, it is tailored to train deep-learning models for the purpose of removing clouds from photos.

SEN12MS-CR has 169 unique areas of interest (ROIs) sampled in all seasons and spread throughout all inhabited continents. Due to the fact that these regions are chosen from two uniform distributions—one that covers the entire land area and the other that concentrates on urban areas—urban landscapes, which are highly interesting to remote sensing because of their complexity, are favoured. An average of 52 km by 40 km is covered by each ROI, which corresponds to photos with a ground sampling distance of 10 meters. To reduce surface changes, each ROI set comprises a matching Sentinel-1 image, three orthorectified with geo-referenced triplets for cloudy and cloud-free Sentinel-2 photos, and all of these images were taken within a single meteorological season. A cloud detector that was published by Schmitt et al. in 2019 is used to measure the amount of cloud cover in optical images. Less than 10% of Sentinel-2 cloud-free photographs exhibit cloud cover, whereas 20% to 70% of cloud-covered images do. The ROIs were split into smaller 256 X 256 pixel tiles using a stride of 128 px, leading of a 50% overlap between adjacent patches, in order to prepare the pictures for input into a CNN. In order to maximize the number of patches retrieved from each image while preserving a respectable degree of independence between them, this method was selected. To get rid of any mosaicking artifacts and other corrupted areas, both automated and manual inspections were performed on the resulting patches. There are 157,521 patch triplets in the finalized quality-controlled SEN12MS-CR dataset, with a total of 28 layers in each.

4.2.1 Evaluation Metric

In evaluating the performance of image reconstruction and enhancement models, several key metrics are commonly employed: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Correlation Coefficient. Each metric offers unique insights into different aspects of image quality and fidelity.

Peak Signal-to-Noise Ratio (PSNR)

PSNR is a widely used metric for assessing the quality of reconstructed images, especially in the context of image compression and denoising. It is defined as the logarithmic ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Mathematically, PSNR is expressed as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

where MAX is the maximum possible pixel value of the image (e.g., 255 for an 8-bit image), and MSE is the Mean Squared Error between the original and reconstructed images. Higher PSNR values indicate better image quality, as they suggest that the reconstructed image is closer to the original image.

PSNR is particularly useful for quantifying the overall fidelity of image reconstruction, but it may not always align with human visual perception, especially when the types of artifacts vary.

Structural Similarity Index Measure (SSIM)

SSIM is designed to address some of the limitations of PSNR by considering changes in structural information, luminance, and contrast between the original and reconstructed images. The SSIM index ranges from -1 to 1, where 1 indicates perfect similarity. SSIM is calculated as follows:

$$\text{SSIM}(\mathrm{x}, \mathrm{y}) = \frac{(2\mu_\mathrm{x}\mu_\mathrm{y} + \mathrm{C}_1)(2\sigma_\mathrm{xy} + \mathrm{C}_2)}{(\mu_\mathrm{x}^2 + \mu_\mathrm{y}^2 + \mathrm{C}_1)(\sigma_\mathrm{x}^2 + \sigma_\mathrm{y}^2 + \mathrm{C}_2)}$$

where x and y are the original and reconstructed images, μ_x and μ_y are the mean intensities, σ_x^2 and σ_y^2 are the variances, σ_xy is the covariance, and C_1 and C_2 are constants to stabilize the division.

SSIM is particularly effective in measuring perceptual quality, as it models human visual perception more closely by considering structural information and image attributes.

Correlation Coefficient

The Correlation Coefficient is another important metric for evaluating the similarity between the original and reconstructed images. It measures the linear correlation between the corresponding pixel values of the two images, indicating how well the pixel intensities of the reconstructed image match those of the original. The Correlation Coefficient, ρ, is defined as:

$$\rho(\mathrm{x}, \mathrm{y}) = \frac{\mathrm{Cov}(\mathrm{x}, \mathrm{y})}{\sigma_{\mathrm{x}} \sigma_{\mathrm{y}}}$$

where $\mathrm{Cov}(\mathrm{x}, \mathrm{y})$ is the covariance between the original and reconstructed images, and σ_{x} and σ_{y} are the standard deviations of the original and reconstructed images, respectively. The Correlation Coefficient ranges from -1 to 1, with 1 indicating perfect positive correlation, -1 indicating perfect negative correlation, and 0 indicating no correlation.

A high Correlation Coefficient signifies that the reconstructed image maintains the statistical relationship of pixel intensities with the original image, which is crucial for applications where preserving the spatial relationship of pixel values is important.

In summary, PSNR, SSIM, and Correlation Coefficient comprehensively assess image reconstruction quality. PSNR quantifies the overall fidelity, SSIM captures perceptual quality by considering structural information, and the Correlation Coefficient evaluates the statistical relationship between pixel intensities. Together, these metrics offer a robust framework for evaluating and comparing the performance of image reconstruction models.

4.2.2 Experimentation Results

Taking 1000 pair of images and the modified DSen2CR model we were able to reproduce the below result shown in table 4.1 which is better than the baseline models figures.

TABLE 4.1: Comparison of datasets

Dataset	**SEN12MS-CR**	**PLA4MS**
SSIM	0.878	0.860
PSNR	29	36

4.3 Discussion

The analysis of Table 1 reveals that when solely considering the quantity of image patches, PLA4MS ranks among the five most extensive datasets. Nevertheless, PLA4MS surpasses its competitors in overall size, primarily attributable to its larger patch dimensions of 256×256 pixels, in contrast to the 28×28 (SAT-4/6) or 120×120 (BigEarthNet) pixel dimensions of its counterparts. Additionally, PLA4MS exhibits significantly higher spatial information content, featuring multi-spectral Planetscope images with a spatial resolution of 3 meters.

4.4 Conclusion

In this research paper, we introduce the PLA4MS dataset, comprising 100,000 pairs of multispectral satellite images from Planetscope with a spatial resolution of 3 meters. Noteworthy for its substantial patch dimensions, diverse temporal scene distribution, and comprehensive crop information, it can be regarded as the most extensive agricultural remote-sensing dataset presently accessible. Our aim is that this dataset will catalyze the advancement of machine learning models capable of robust and comprehensive automatic analysis of Planetscope satellite data.