GENCP: GAN-Based Ground Control Point Generation for Satellite Image Georeferencing

Elodie Guasch , Ilyas Yalcin , Sébastien Saunier , Leonardo de Laurentiis , Philippe Goryl , Sultan Kocaman *

Posted Date: 17 April 2026

doi: 10.20944/preprints202604.1240.v1

Keywords: satellite optical imagery; geometric calibration; validation; generative adversarial networks; vector reference data

 doi:10.20944/preprints202604.1240.v1

Article

GENCP: GAN-Based Ground Control Point Generation for Satellite Image Georeferencing

Elodie Guasch [1], Ilyas Yalcin [2], Sebastien Saunier [1], Leonardo de Laurentiis [3], Philippe Goryl [3] and Sultan Kocaman [4,*]

[1] Telespazio France, 26 Avenue Jean François Champollion, 31100 Toulouse, France
[2] Baskent OSB Technical Sciences Vocational School, Hacettepe University, 06909 Sincan, Ankara, Türkiye
[3] European Space Agency, ESRIN, Via Galileo Galilei, 1, 00044 Frascati, RM, Italy
[4] Nexa Geo GmbH, Am Kesselhaus 3, 79576 Weil am Rhein, Germany
* Correspondence: sultankocaman@nexageo.de

Highlights

This study introduces a GAN-based approach to generate reference image chips from vector data for satellite georeferencing. The results show that vector-derived imagery can provide geometrically consistent and scalable reference data, with performance influenced by data quality and resolution. Developed within the ESA-supported GenCP initiative, the approach supports sensor-independent reference generation. It offers a pathway toward automated and flexible Cal/Val workflows across multi-mission systems.

What are the main findings?

- GAN-based generation of vector-derived image chips enables geometrically consistent reference data suitable for automated image matching and georeferencing.
- Model performance is strongly governed by spatial resolution and vector–image alignment.

What are the implications of the main findings?

- The approach provides a scalable, sensor-independent alternative to traditional GCP acquisition, supporting multi-mission Cal/Val workflows.
- It enables both global reference data generation from open vector sources and sensor-specific model training for optimized, operational use.

Abstract

Remote sensing has become a core technology for environmental and climate monitoring, supported by expanding sensor constellations, advanced processing capabilities, and coordination frameworks established by the European Space Agency (ESA), Global Earth Observation System of Systems (GEOSS), and Committee on Earth Observation Satellites (CEOS). Ensuring consistency across missions requires robust geometric and radiometric calibration and validation. However, traditional reliance on ground control points (GCPs) is limited by sparse global coverage, temporal instability, and dependence on surveyed accuracy. While alternative geospatial datasets, including satellite and aerial imagery, Light Detection and Ranging (LiDAR) point clouds, and vector databases, can serve as references, challenges remain in data access, automation, and cross-sensor applicability. This study proposes a generative adversarial network (GAN)–based approach to generate geometrically consistent image chips from vector maps. Two models were trained at 50 cm and 10 m resolution within the ESA-supported Generative Ground Control Point (GenCP) study, using Sentinel-2 and very-high-resolution RGB imagery. The generated GenCP image chips are evaluated using image-based similarity (radiometric consistency), geometric, and model-performance metrics. Results demonstrate their suitability for automated Cal/Val workflows and their potential as scalable, fit-for-purpose reference datasets.

 doi:10.20944/preprints202604.1240.v1

Keywords: satellite optical imagery; geometric calibration; validation; generative adversarial networks; vector reference data

1. Introduction

Remote sensing (RS) has grown into a central technology for governments and the private sector, driven by rapid expansion in sensor types, orbital platforms, and advanced data-processing methods. Owing to its synoptic, repeatable, and multiscale observation capabilities, RS now supports monitoring of numerous Sustainable Development Goal (SDG) indicators, with analyses showing that Earth Observation (EO) contributes directly or indirectly to 72 targets and 30 indicators, particularly for SDGs 6, 11, 14, and 15 [1]. In Europe, among other institutions and programs, the European Space Agency [2], the Global Earth Observation System of Systems [3], and the Copernicus Programme [4] collectively advance SDG monitoring by providing coordinated satellite missions, interoperable data infrastructures, and free, high-quality EO datasets that enable consistent environmental assessment and reporting through integrated data policies and long-term continuity of observations. To track the climate change and a broader set of planetary challenges, including biodiversity loss, land degradation, hydrological stress, and oceanic change, high-quality, consistent observations are essential. Thus, geometric calibration and validation (Cal/Val) of satellite EO sensors are essential for maintaining coherence among multi-mission data, thereby enabling reliable long-term monitoring [5].

To maintain coherence across EO missions and align them with the broader EO strategy, the ESA established a coordination framework for Third-Party Missions (TPM) under the Earthnet Programme, which has supported the GEOSS for more than 40 years [6]. For over 20 years, ESA has actively supported the Cal/Val activities for both ESA and non-ESA missions as well as TPMs (e.g., see, [2,7,8]). The Committee on Earth Observation Satellites (CEOS), founded in 1984, also provides an international coordination framework to align satellite-based Earth observation efforts that no single agency or nation can meet alone. Its work focuses on harmonizing mission planning, data standards, and policies to ensure complementarity, cost-effectiveness, and effective global use of spaceborne observations across environmental and climate applications [9].

Optical sensor geometric Cal/Val have been a fundamental subject in photogrammetry for decades. Traditionally, ground control points (GCPs) serve as the primary reference for sensor orientation and calibration, while independent check points provide further ground truth for accuracy assessment. This paradigm carried over to satellite sensor Cal/Val, where GCPs remained the dominant reference source, even though their availability, quality, and task-specific suitability have long been subjects of debate. Supported by ESA Earthnet Data Assessment Project (EDAP) [2], CEOS has developed the Cal/Val Portal, notably providing radiometric and geometric calibration datasets across multiple sites and sensor types [10]. The portal also includes several GCP datasets tailored for very high-resolution (VHR) optical missions, which have been utilized in recent data quality assessment and calibration studies (e.g., [8,11]. However, achieving truly global GCP coverage for the full range of sensors and spatial resolutions remains distant, and the temporal stability of existing points poses an additional challenge, requiring periodic resurveying to maintain their validity.

On the other hand, a re-conceptualization of GCPs is also needed, expanding the notion of control from ground surveyed control points to a broader suite of geospatial reference datasets, including satellite and aerial imagery, Light Detection and Ranging (LiDAR) and stereo-derived point clouds, and vector databases, enabling extended uses such as Digital Elevation Model (DEM) validation, cross-sensor consistency checking, and multi-source georeferencing. In addition, the automatic use of such reference datasets for geometric Cal/Val relies strongly on the employed methodology, with most approaches depending on image-matching techniques. Data access constraints, particularly licensing and usage policies, limit the development of globally accessible GCP datasets. In addition, temporal variability of reference features introduces further challenges, as

 doi:10.20944/preprints202604.1240.v1

changes in land cover and infrastructure can degrade the long-term reliability of GCPs. Suitability of a GCP or reference dataset across different spatial resolutions and sensor types represents another significant limitation for this purpose.

To address these challenges, we propose a generative adversarial network (GAN)–based approach for producing dynamic image chips that can serve as geometrically reliable reference data for satellite Cal/Val applications. The concept builds on generating imagery from vector maps, which are widely available in national geodatabases with high positional accuracy. Using the established pix2pix framework [12], we implemented a prototype study to assess the feasibility and performance of this approach using an openly available vector map source, the Open Street Map [13]. As part of the Generative Ground Control Point (GenCP) Study [14,15], we trained two separate GAN models corresponding to spatial resolutions (50 cm and 10 m). For the 10 m resolution model, we used Sentinel-2 imagery, whereas the 50 cm model was trained primarily with EDAP datasets, including data from Maxar (now rebranded as VANTOR) WorldView-3, SkySat, and Pleiades sensors. The methodology, datasets, and results are presented in this article, followed by a discussion of the limitations and future prospects. The implementation of the proposed methodology and the associated datasets are publicly available [16,17].

2. Background on GAN-Based Studies in Remote Sensing

GANs introduced by Goodfellow et al. [18], have revolutionized generative modeling by pitting a generator against a discriminator in a minimax game, enabling the synthesis of realistic data distributions. In RS, GANs have rapidly gained traction since around 2017, addressing key challenges like data scarcity, resolution limitations, and environment-related degradation in satellite and aerial imagery [19]. Their adversarial training excels at producing high-fidelity synthetic RS images, outperforming traditional methods such as interpolation or dictionary learning in perceptual quality and detail recovery.

Early applications focused on super-resolution (SR), where GANs improve the appearance of low-resolution RS images while preserving textures and structures. Ledig et al. [20] pioneered SRGAN for natural images, inspiring RS adaptations like enhanced SRGAN (ESRGAN) variants [21]. A meta-analysis of 231 papers showed image classification as the top use (especially semantic segmentation), with pix2pix [12] and CycleGAN [22] popular for unpaired data translation. Multi-sensor SR advanced with conditional GANs (cGANs). Zhang et al. [23] proposed a cGAN for Landsat-8 OLI to Sentinel-2 MSI upscaling (×3 factor), achieving superior Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) on real-world datasets via solution-space reduction.

GANs have also been applied to image enhancement. Jia et al. [24] introduced a physically constrained GAN for remote-sensing dehazing that integrates texture-recovery mechanisms to eliminate atmospheric scattering, achieving superior performance on Sentinel imagery compared with conventional baselines. Cloud removal has also benefited from GAN-based approaches. Enomoto et al. [25] demonstrated that pix2pix-style cGANs can effectively replace thin clouds in multispectral WorldView-2 scenes, reconstructing obscured regions with contextually coherent spectral and spatial structures from the near-infrared band data. More recent diffusion–GAN hybrid models such as DEGAN [26] combine denoising diffusion processes with adversarial refinement to enhance very-high-resolution super-resolution products, yielding sharper detail while controlling spectral noise.

Pix2pix [12] is an image-to-image translation method [27] and has also been used in various studies, including day-night image conversion, colorization of grayscale images, and generating satellite images from map images [22]. Zuo and Li [28] and Amitrano [29] translated synthetic aperture radar (SAR) images into optical data with pix2pix. Xu et al. [30] adapted the method for solid waste detection. Hu et al. [31] implemented a wildfire monitoring framework based on SAR-optical translation using pix2pix. Šidlauskas et al. [32] also used OSM data to produce Sentinel-2-like images with pix2pix with the aim of generating continuous images. They emphasized the importance of addressing class imbalance to achieve higher accuracy.

 doi:10.20944/preprints202604.1240.v1

The main contribution of this study is the assessment of vector maps as a source of ground reference data for satellite image Cal/Val through automatic image matching. The research design explores two spatial resolutions (50 cm and 10 m), incorporates data cleaning to evaluate feasibility, and optimizes the pix2pix architecture for training both models. Validation combines multiple accuracy measures, including model-performance metrics, radiometric analyses (Peak Signal-to-Noise Ratio – PSNR – and histogram evaluation), geometric assessments via image matching, and qualitative visual inspection. The GenCP study represents an initial step toward generating reference data with high geometric fidelity. Unlike high-resolution satellite imagery, which are often affected by geometric and radiometric artefacts such as relief displacement, saturation, and shadows; vector databases typically offer stable and high-quality geometric data. The proposed approach also allows targeted model training for specific sensors, seasons, geographies, or acquisition geometries, thereby increasing the likelihood of successful image matching.

3. Methodology

Within the framework of the GenCP study, synthetic image generation was performed in the form of GAN-based GCP image chips. In this context, satellite images with two different spatial resolutions were used for training, namely high resolution (HR) with 10 m spatial resolution and VHR with 50 cm spatial resolution. Vector datasets derived from the OSM [13] were employed as reference inputs to guide the synthetic image generation process. However, extensive manual refinement of the VHR vector data was performed to thoroughly evaluate the potential and the performance of the proposed methodology. Although the experiments with the HR and VHR data were carried out with different data and resolutions, a single processing pipeline was implemented in order to ensure a consistent performance assessment of the model. The pipeline is depicted in Figure 1. Under the following sub-headings, the data preparation, model training and validation methodology are explained in detail.

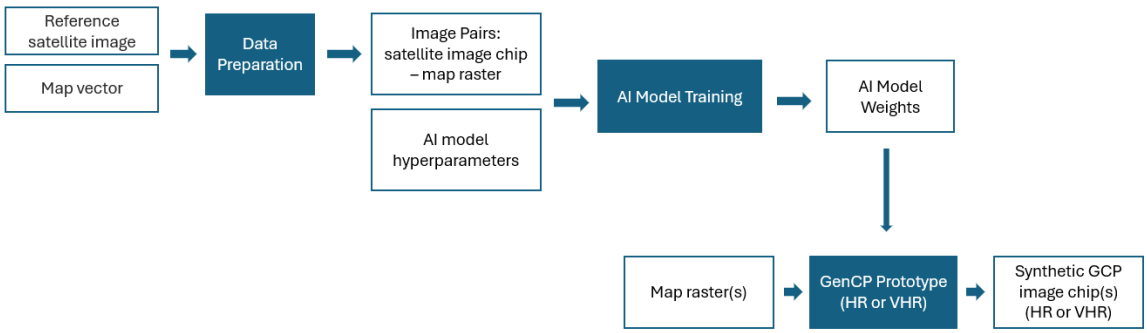


Figure 1. The overall workflow of the study.

3.1. The HR Image Data Pre-Processing

The preparation of vector data and the generation of paired datasets followed a systematic workflow. First, relevant vector datasets covering the geographical extent of the study area were downloaded from the OSM website. These vector features were then adapted by adjusting their properties, such as width and color, to ensure consistent representation across all features. Subsequently, the vector layers were rasterized using a hierarchical layer structure (background followed by semantic features such as buildings and vegetation), with scale-dependent generalization applied to ensure consistency at the target resolution. Sentinel-2 image tiles were then paired with the corresponding rasterized OSM tiles to construct the training data for the model. Finally, data cleaning was performed, including cloud filtering and harmonization of the image pairs, to ensure consistency and quality across the dataset. Figure 2 illustrates the HR pre-processing steps conducted in this study. The sensor and data characteristics and the pre-processing approach are given below.

 doi:10.20944/preprints202604.1240.v1

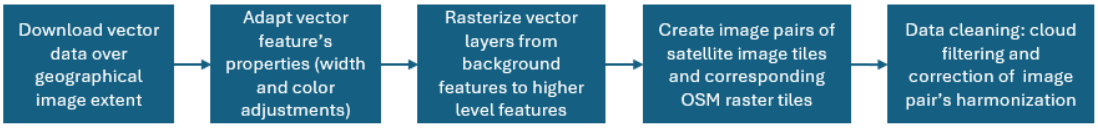


Figure 2. The HR image data preparation workflow.

3.1.1. The HR Sensor and Data

The HR dataset is primarily based on Sentinel-2 Level-2A (L2A) orthorectified images with 12 spectral bands. With a temporal resolution 5 days and spatial resolution of 10 m for the visible spectrum [33], Sentinel-2 has been an essential data source for many environmental applications. The HR training data were selected and compiled in two steps. In the first step, images from five mid-latitude European sites were used. In the second step, 18 additional European sites were incorporated into the HR dataset to obtain more training patches and introduce greater variability. Each site contains several GCPs from the Sentinel-2 Global Reference Image (GRI) database [34] that can be used to extract relevant training patches for the HR dataset. All training sites included in the dataset are shown in Figure 3 together with their Military Grid Reference System (MGRS) codes.

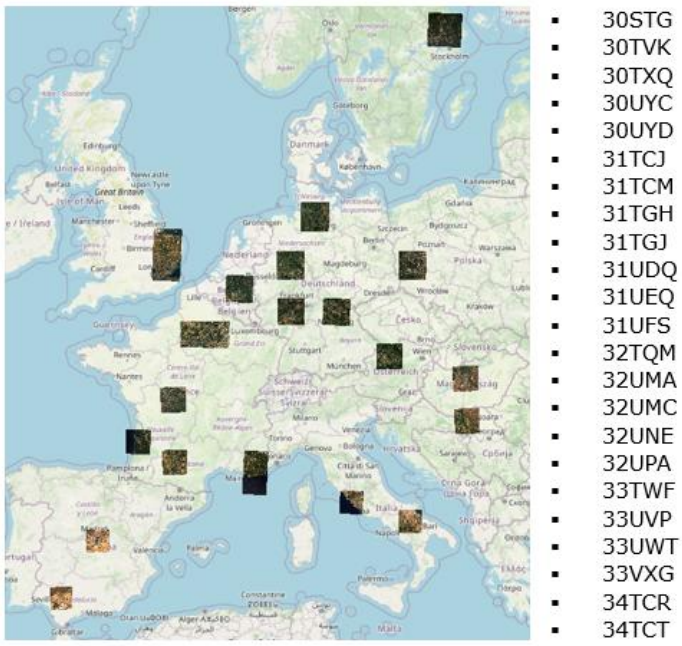


Figure 3. The HR training sites used in this study and the list of their MGRS codes.

Two HR datasets were created for the training phase. The first dataset consists of RGB images generated from the true color composite (TCI) band combination. In the second dataset, band B04 was additionally included, as it carries more radiometric information than bands B03 and B02 and is also used in the Sentinel-2 database.

Although vector datasets can be obtained from various geodatabases, the OSM was preferred here due to its global coverage and ease of access. The data were extracted and classified into three geometry types commonly used to represent physical features: points, lines, and polygons. Point features include discrete objects such as landmarks and addresses, line features correspond to linear transportation networks such as roads, tracks, and highways, while polygon features represent areal units such as buildings, surfaces, and administrative boundaries (see OSM wiki for further information). The OSM vectors were retrieved in GeoJSON format using the Python library OSMnx [35] by specifying a geographical bounding box and selecting the relevant OSM tags. The selection of these tags was performed with consideration of the target resolution, based on a visual inspection of

 doi:10.20944/preprints202604.1240.v1

which OSM features were visible in the HR satellite imagery. It is worth noting that the positional accuracy of the OSM is comparable to a scale of approximately 1:5000 according to several studies conducted over European sites [36–38], which is a sufficient match for the 10 m spatial resolution of Sentinel-2.

3.1.2. The HR Training Data Preparation

HR Vector Data Preparation

An HR vector data preparation step was performed to comply with the data input format required by the implemented DL architecture. After selecting and downloading the OSM features, several color maps were defined to assign a unique color to each feature for investigations. This step was carried out iteratively by testing different color maps and assessing their impact on model training. In addition to the OSM data, Corine Land Cover (CLC) data were used for the HR dataset to complement missing features in some areas. The CLC dataset is available as a 10 m resolution raster over Europe [39]. The color map shown in Figure 4 was defined to classify the CLC pixel values and to match the color map used for the OSM features. After defining the color map, rasterization was performed feature by feature. For the HR data, direct rasterization at 10 m provided a coarse approximation of vector features. To reduce aliasing, rasterization was temporarily carried out at 5 m, and the resulting map was subsequently downsampled to 10 m using a resampling algorithm. Several methods were tested, as shown in Figure 5, and the Lanczos filter [40] was preferred for its ability to preserve sharpness and minimize blurring and aliasing artifacts.

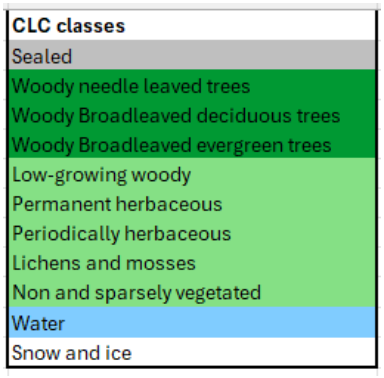


Figure 4. Color encoding of the CLC classes used for rasterization for the HR dataset.

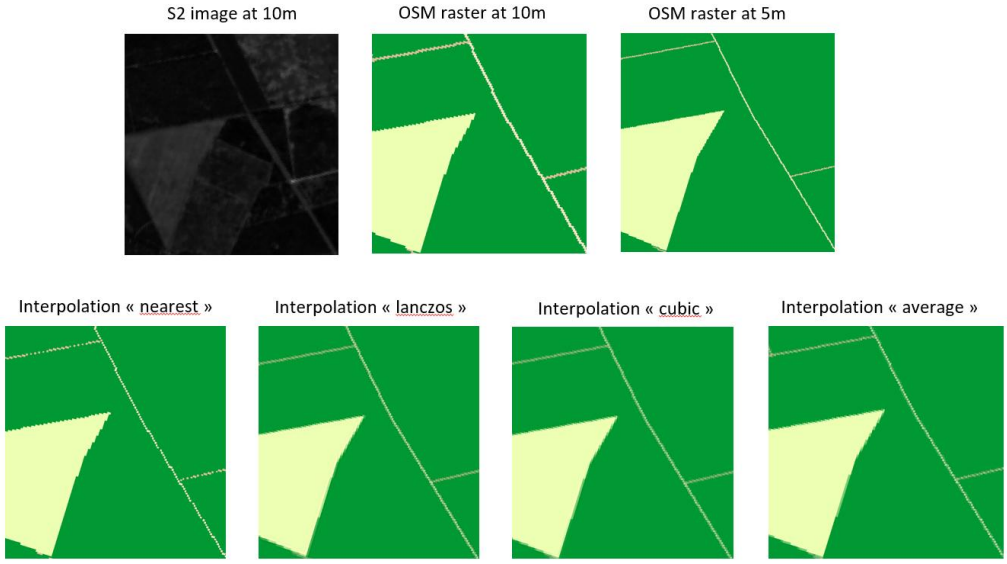


Figure 5. Results from different downsampling methods used for HR training data preparation.

 doi:10.20944/preprints202604.1240.v1

Since farmland, greenfield, and scrub area classes often share similar radiometric characteristics, which may introduce confusion during model training, statistical analyses were performed to obtain average pixel values for each OSM feature type. The objective was to verify whether the pre-defined classes exhibit sufficient separability. Figure 6 illustrates an example derived from two Sentinel-2 images. Similar analyses were conducted for OSM "highway" and "natural" classes. The results indicated that certain feature groups could be merged in the color mapping to reduce model uncertainty and dimensionality. The classes used for rasterization and the colors assigned to them are shown in Figure 7, with examples in Figure 8. Given sensor-specific radiometric responses, this analysis should be repeated or adapted for each sensor type to ensure consistent class separability.

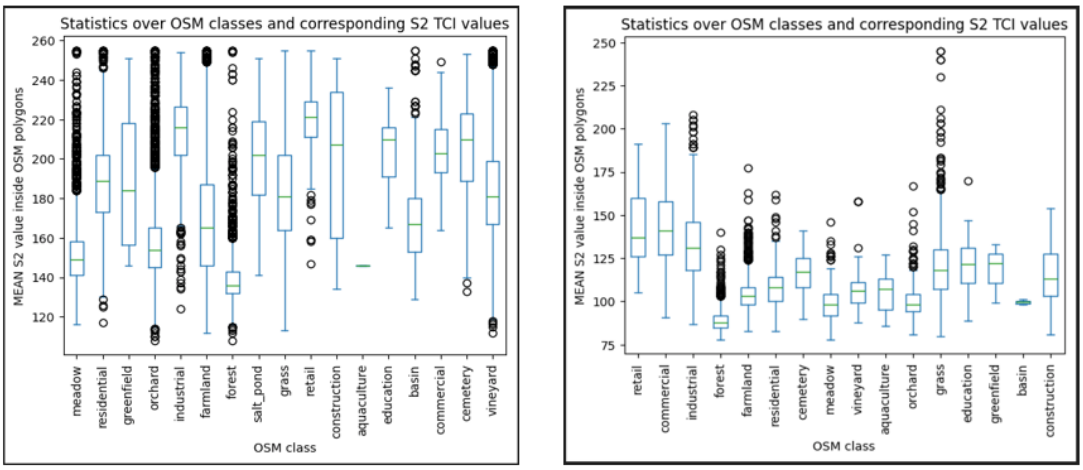


Figure 6. Mean pixel values analyzed in two Sentinel-2 images for OSM "landuse" features.

value
commercial
education
retail
industrial
construction
residential
forest
farmland
greehouse_horticulture
vineyard
greenfield
paddy
grass
port
aquaculture
basin
salt_pond

key	value
natural	fell
natural	grassland
natural	scrub
natural	shrubbery
natural	tundra
natural	wood
natural	valley
natural	hill
natural	heath
natural	bay
natural	wetland
natural	strait
natural	water
natural	shoal
natural	beach
natural	reef
natural	sand
natural	bare_rock
natural	blockfield
natural	scree

key	value
highway	motorway
highway	trunk
highway	primary
highway	secondary
highway	tertiary
highway	unclassified
highway	residential
highway	track
highway	footway

Figure 7. Color map used for rasterizing the HR classes.

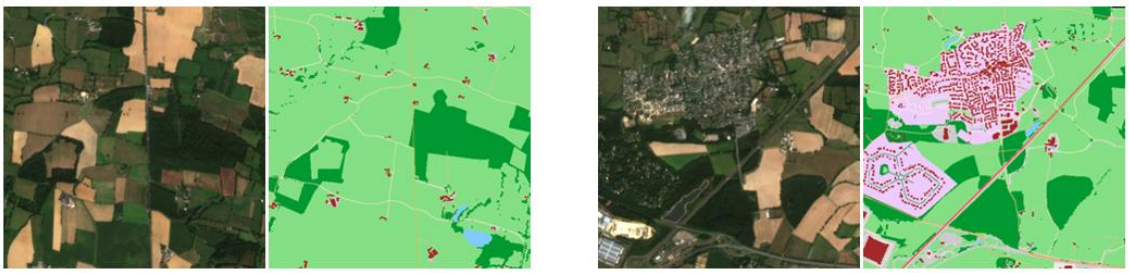


Figure 8. Two Sentinel-2 L2A and rasterized OSM data pair examples.

 doi:10.20944/preprints202604.1240.v1

A threshold-based approach was introduced to eliminate the OSM features that are not visible in the corresponding satellite imagery. This filter mainly targets highway-related features. Band B04 was used for threshold definition, as paths and tracks are visually more distinguishable in B04 than B03 or B02 (Figure 9). The pixel value diagrams previously computed for OSM feature analysis were used on Band B04 to determine the optimal threshold.

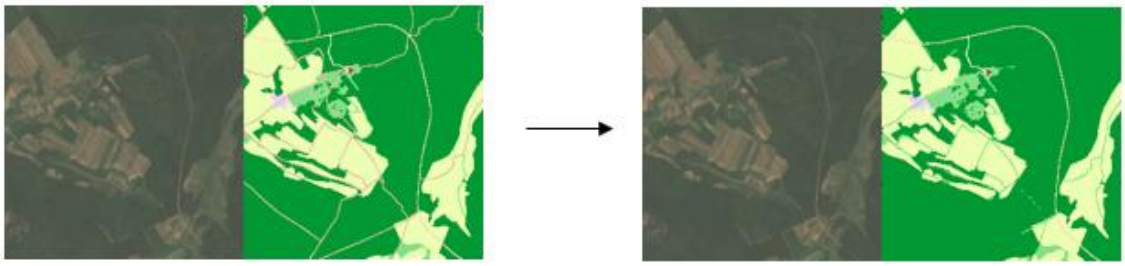


Figure 9. Example of highway feature filtering using the B04-based threshold.

HR Raster Data Preparation

A min–max normalization was applied for 16-bit to 8-bit conversion. A total of 5500 image patches with a size of 256 × 256 pixels were extracted from the 23 sites described above. Of these, 80% were used for training while the remaining 20% constituted a first test set composed of unseen image patches from the same regions as those used for training. An additional independent test set was created using image patches from an unseen European site (MGRS tile 30TXT). For both test sets, grayscale (Band B04) and RGB (TCI) versions were used to validate the outputs of the mono-band and multi-band HR prototypes.

3.2. The VHR Image Data Pre-Processing

The VHR dataset is derived from images taken from three different satellites over three different sites with 30 cm–60 cm ground sampling distances (GSDs). The corresponding vector maps were compiled from the OSM and extensive manual editing was carried out for obtaining consistent images and vector maps. The sensors, data and the training data preparation steps are explained next.

3.2.1. VHR Sensors and Datasets

In the VHR section of the study, three study areas were identified: Salon-de-Provence in France, Munich in Germany, and Ankara in Türkiye (Figure 10). The approximate geographical coordinates of these study areas are Salon-de-Provence (43.627735°, 5.084599°), Munich (48.32077°, 11.582156°), and Ankara (39.932015°, 32.857808°). The Salon-de-Provence site is located in the southern part of France and comprises various LULC features, including agricultural fields, buildings, and roads. The Munich site, located in the Bavarian state of Germany in south, characterized by a notably high density of urban settlements and tree-covered regions. The Ankara site, on the other hand, is located in the Central Anatolia region of Türkiye, and has continental climate. Alongside man-made structures such as buildings, roads, and dams, agricultural fields and steppe vegetation enhances the LULC diversity in model training. Figure 10 shows the site locations and the overviews of the satellite images.

The Ankara study site consists of 14 VHR images: nine Maxar (60 cm GSD), four SkySat (50 cm GSD), and one Pleiades (50 cm GSD) image. The Salon-de-Provence site consists of 21 Maxar images at 60 cm GSD, and the Munich site includes 4 Maxar images at 30 cm GSD. All images were resampled to a spatial resolution of 50 cm. Images were taken at varying daylight conditions acquisition angles. Such differences led to variations in both brightness levels and image detail. The detailed descriptions of the image features used in the study with acquisition date, product level as defined by the vendor,

 doi:10.20944/preprints202604.1240.v1

Mean Off Nadir Angle (MON), Mean Satellite Azimuth Angle (MSA), Mean Sun Elevation Angle (SEA), Mean Sun Azimuth Angle (SAA) and GSD are given in Table 1.

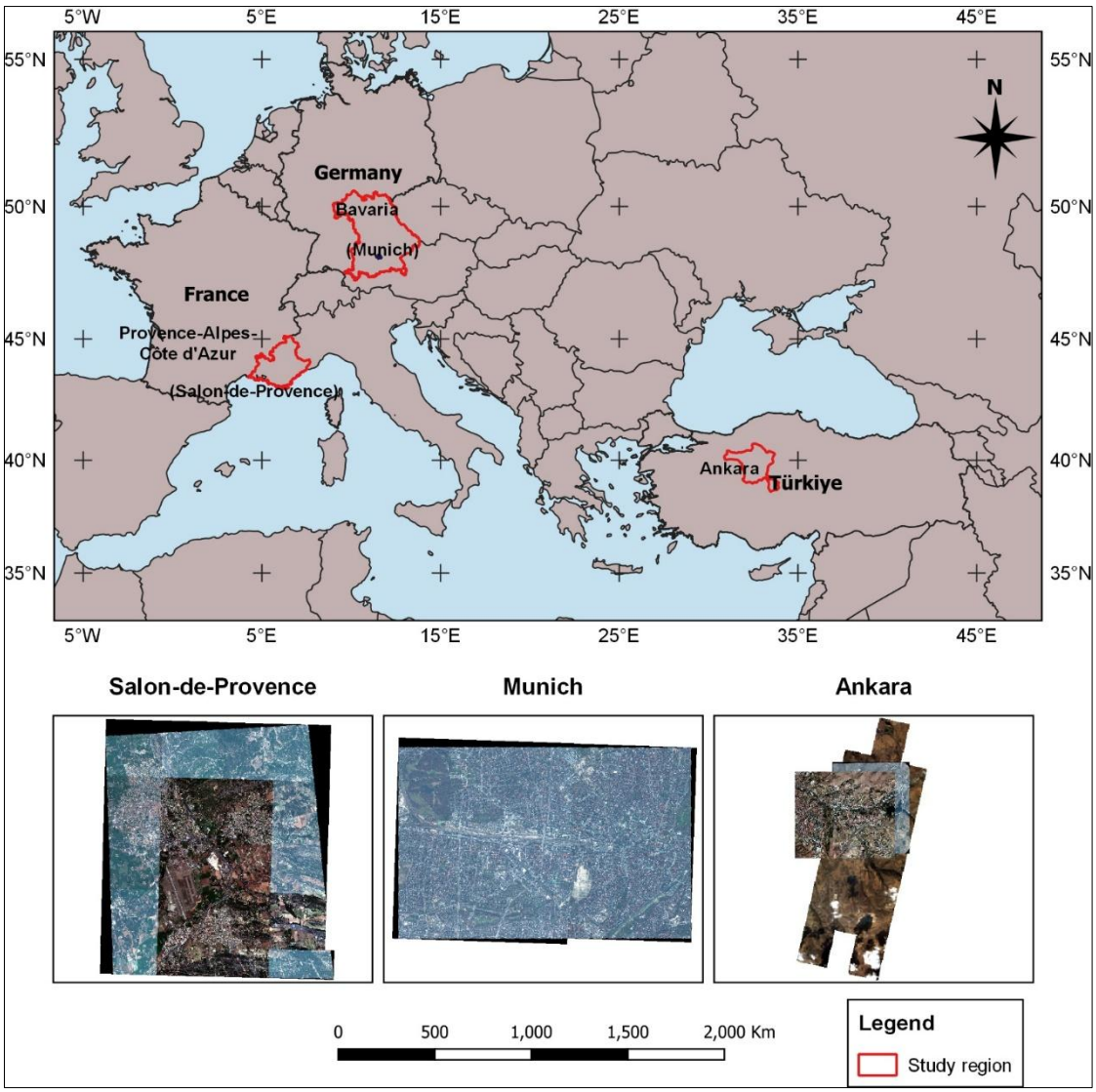


Figure 10. Locations of the VHR study sites.

Table 1. Characteristics of the VHR satellite image datasets for Ankara, Salon-de-Provence and Munich study sites.

Sensor & Site	Acquisition Date	Product Level	MON	MSA	SEA	SAA	GSD
Maxar Ankara	13 Oct 2020, 8:37	Orthorectified (LV3D)	21.9°	57.8°	40.5°	161.4°	1.41 m
SkySat-4 Ankara for image 1	18 Jul 2020, 8:28	Ortho Collect	20.2°	348.4°	63.5°	129.7°	0.72 m
SkySat-4 Ankara for image 2	18 Jul 2020, 8:28	Ortho Collect	16.6°	348.4°	63.6°	129.5°	0.70 m
SkySat-7 Ankara for image 3	23 Jul 2020, 11:23	Ortho Collect	3.8°	310.8°	43.2°	216.9°	0.66 m
SkySat-7 Ankara for image 4	23 Jul 2020, 11:23	Ortho Collect	1.8°	310.8°	43.3°	216.9°	0.66 m
Pleiades Ankara	12 May 2021, 8:41	Orthorectified	9.1°	179.8°	64.5°	142.9°	0.71 m

 doi:10.20944/preprints202604.1240.v1

Maxar Salon-de-Provence	3 Jul 2017, 10:57	Orthorectified (LV3D)	27.6°	179.2°	67.4°	151.4°	1.53 m
Maxar Munich	19 Mar 2020, 10:28	ORStandard2A (LV2A)	18.2°	14.8°	40.4°	162.4°	0.33 m

3.2.2. VHR Training Data Preparation

VHR Vector Data Preparation

In this phase, the OSM data was downloaded, projected and geometrically corrected to ensure positional accuracy. Furthermore, manual content adaptation (e.g., correction of mislabeled or incomplete vector features and alignment with image content) and radiometric image processing were applied. Finally, image tiling and mask generation were performed (Figure 11). All processing steps on the OSM data were performed in Python programming environment, using the same libraries and tools as in the HR data preparation workflow.

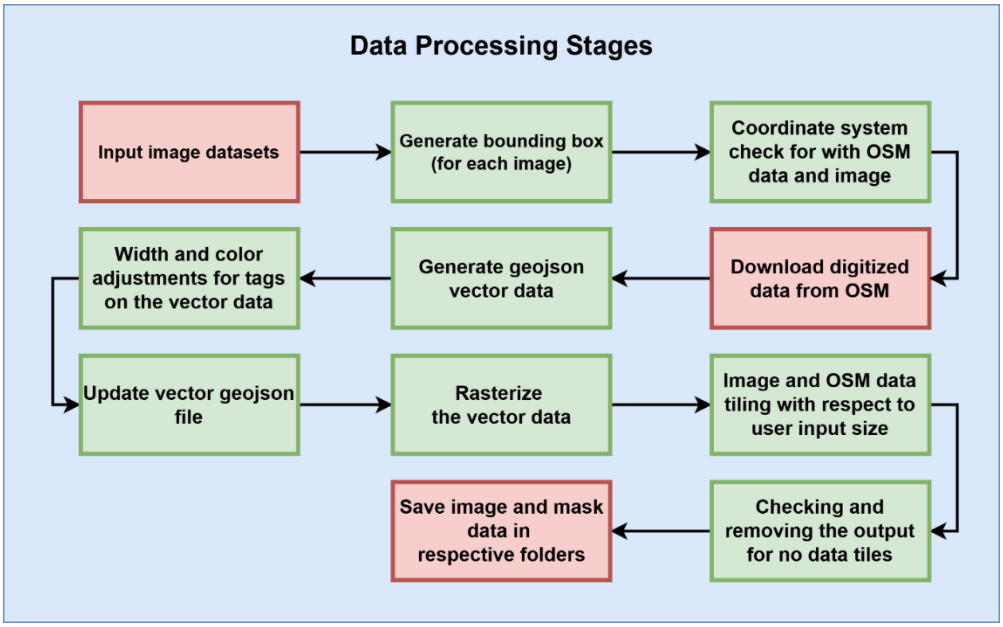


Figure 11. The data pre-processing workflow for the VHR training data preparation.

The OSM dataset contains 29 primary features (OSM, 2026) with polygon and line vector data. The polygon features include building, crop, forest, grass, residential, and water classes; while motorway, path, and service–residential classes were line vectors. The features were selected based on visual inspection of images and the OSM data to ensure consistency. The colors assigned to the restructured classes are provided in Tables 2 and 3 for polygon and line features, respectively. Line thickness parameter is also given in Table 3. Examples of the images and corresponding masks generated as a result of this phase are presented in Figure 12.

Table 2. Standard features and color values for polygon vector data.

Feature Key	R	G	B	Color Definition
Building	255	255	255	White
Crop	255	165	0	Orange
Forest	0	128	0	Green
Grass	107	142	35	Olivedrab
Water	30	144	255	Dodgerblue

 doi:10.20944/preprints202604.1240.v1

Table 3. Standard features and color values for line vector data.

Feature Key	R	G	B	Color Definition	Line Thickness
Motorway	0	0	255	Blue	3
Path	255	0	255	Magenta	1
Service-Residential	255	0	0	Red	2

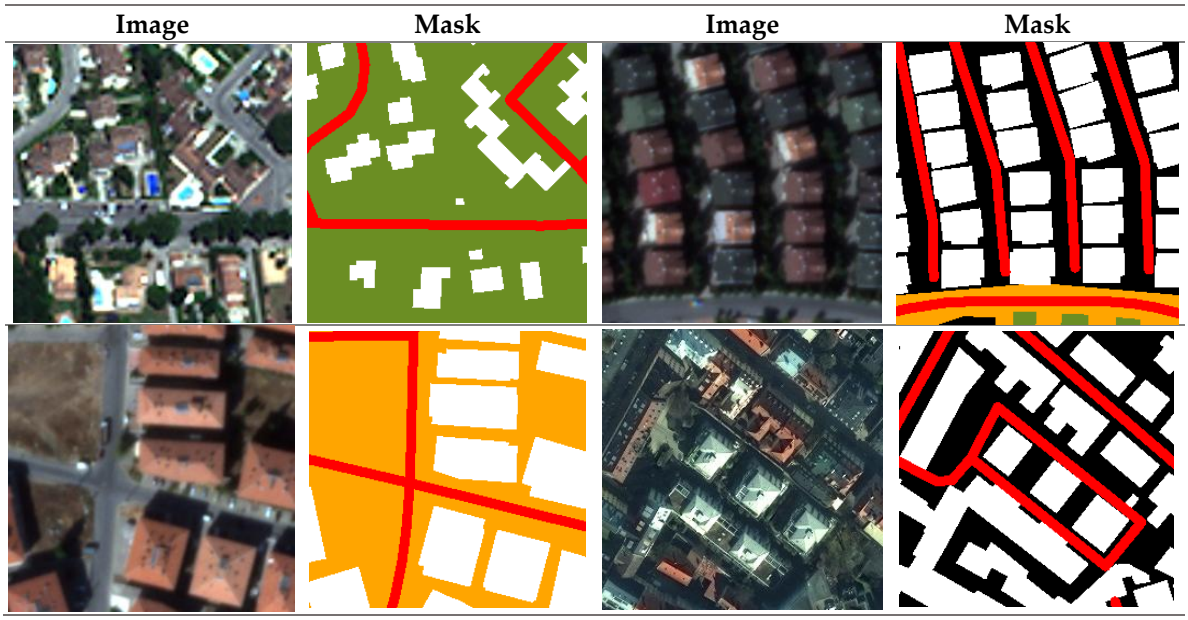


Figure 12. Images and corresponding masks from the Salon-de-Provence dataset (top left), Ankara dataset (top right and bottom left), and Munich dataset (bottom right).

A major challenge encountered at this step was the geometric discrepancies between the OSM data and the images (shift and rotation). Consequently, affine transformation was performed for image geometric correction and manual editing was performed to ensure image-object compatibility in all study areas. Figure 13 depicts an example from the Maxar imagery over the Salon-de-Provence site, illustrating this issue.

Figure 13. An example to the inconsistencies between the OSM data (red polygons) and the images. The corrected data is shown with green polygons.

 doi:10.20944/preprints202604.1240.v1

VHR Raster Data Preparation

Since the VHR satellite images used in this study have dynamic ranges of 11-bit and 12-bit, with actual bin occupation (histogram width) of approximately 2000 pixels, a downsampling to 8-bit required histogram analysis of each image and manual parameter identification prior to histogram normalization. In addition, all images were visually inspected. As a result, histogram thresholds were defined. Being part of EDAP datasets, cloud cover was minimal in the satellite imagery and no cloud-mask was applied.

For training data preparation, both the images and the vector maps (masks) were divided into tiles with 256 x 256 pixels. Additionally, image patches with no data or all zeros were removed. Furthermore, any image patch with more than 256 pixels with zero was also removed. No data augmentation was performed. Table 4 summarizes the training datasets used to train four DL models in this study. Dataset 1 (DS1) contains single-class masks, Dataset 2 (DS2) excludes them, Dataset 3 (DS3) includes non-compliant masks due to the lack of manual editing in the Munich data and Dataset 4 (DS4) consists of fully compliant manually edited masks with the highest image–mask compliance.

Table 4. Overview of the VHR training samples used in this study.

Dataset	Source / Site	Satellite(s)	Train Samples	Test Samples
DS1	Salon-de-Provence	Maxar	3,955	439
DS2	Salon-de-Provence	Maxar	2,908	323
DS3	Salon-de-Provence, Munich, Ankara	Maxar, SkySat, Pleiades	14,049	333
DS4	Salon-de-Provence, Ankara	Maxar, SkySat	7,325	288

3.3. GAN Model Training

A GAN is a form of deep learning model consisting of two neural networks: a generator and a discriminator. The generator produces sample images in accordance with the training data, while the discriminator tries to distinguish between the training data and the samples produced by the generator [41]. Proposed by Goodfellow et al. [18], this model is realized by training the generator to maximize the prediction error of the discriminator in the adversarial process. Figure 14 presents the workflow of the GAN architecture employed in this study. Pix2pix, an image-to-image translation process, is one of the widely used GAN models. Unlike the others, pix2pix employs the U-Net architecture as a generator, based on cGAN [12].

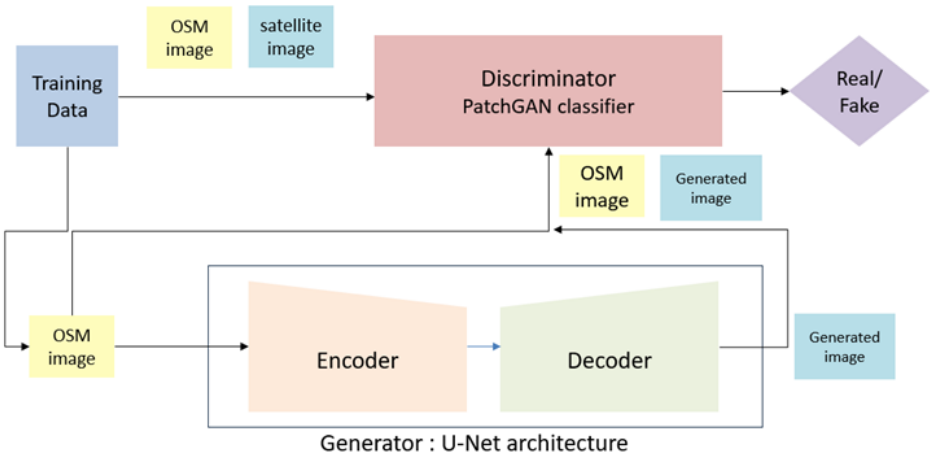


Figure 14. Overview of the GAN based synthetic image generation process in the GenCP study.

 doi:10.20944/preprints202604.1240.v1

As explained before, vector layers (e.g., roads and buildings) were rasterized, and HR/VHR satellite images were used to build the pix2pix training dataset. During the training process, the generator and discriminator networks were iteratively updated according to the model outputs and the data processing pipeline was also refined. The hyperparameters were determined based on visual inspection and the model's loss functions. The Adaptive Moment Estimation (Adam) optimizer, which is a learning algorithm that utilizes the gradient descent approach [42], was used throughout the study. It simultaneously reduces memory consumption while ensuring fast convergence [43]. Additionally, Rectified Linear Unit (ReLU) and sigmoid activations were investigated in the study. According to Nair and Hinton, [44], the ReLU has demonstrated superior performance and efficiency compared to other activation functions on large-scale datasets. Binary classification was performed by incorporating the sigmoid function in the final layer of the discriminator. Since the input dataset was normalized between [-1,1], the tanh function was used in the last layer of the generator. In this study, the HR model utilized Binary Cross Entropy (BCE). In Table 5, the parameters used for the HR and VHR model training are listed. In the VHR configuration, the DS1 and DS2 were trained for 150 epochs with a batch size of 64, whereas the others were trained for 100 epochs with a batch size of 32. Furthermore, to improve the performance of the HR model, the L1 reconstruction loss used in the classical Pix2Pix formulation was replaced by a Learned Perceptual Image Patch Similarity (LPIPS) loss [45].

Table 5. The hyperparameters configuration of the HR and VHR model.

Parameters	HR model parameters	VHR model parameters
Learning rate	0.0002	0.0002
Epochs	120	100 - 150
Discriminator Loss	BCE	BCE
Lambda (λ)	100	100
Batch Size	[1,4,16]	32 - 64
Backbone (generator)	U-Net	U-Net
GAN Loss	Adversarial Loss + λ LPIPS	Adversarial Loss + λL1
Optimizer	Adam	Adam
Dropout	True	True
Normalization	Batch Normalization	Batch Normalization
Activation Layer	ReLU, Sigmoid, Tanh	ReLU, Sigmoid, Tanh
Image size	256 x 256	256 x 256

3.4. Validation

Model performance for both HR and VHR configurations was assessed by comparing the generated outputs with reference test images that were not used during training. The validation framework included three components: visual similarity, radiometric quality, and geometric accuracy.

3.4.1. Image Visual Similarity

In evaluating the visual similarity of images, contrast and brightness effects may be disadvantage. Therefore, the Multi-Scale Structural Similarity Index (MS-SSIM), which reduces the influence of these effects and enables a structural assessment of images, was preferred in this study. The MS-SSIM allows for a multi-scale evaluation of structural components within the image. The MS-SSIM metric takes values in the range [−1,1], where values closer to 1 indicate higher structural similarity between images.

$$MS-SSIM(x,y) = [lm(x,y)]^{\alpha M} \cdot \prod_{j=1}^{M} [cj(x,y)]^{\beta j} [sj(x,y)]^{\gamma j} \quad (1)$$

where l, c and s denote luminance, contrast and structure, respectively. M is the number of scales and α, β, γ are the weighting factors [46].

 doi:10.20944/preprints202604.1240.v1

In addition, the LPIPS measure, developed by Zhang et al. [45] from datasets of human judgments trained with deep learning architectures, was also used. Pre-trained models such as AlexNet and VGG were utilized in the development of this measure. Unlike the SSIM, the LPIPS also includes human perception [47] and measures perceptual difference as the distance between image features. In this respect, small values represent high similarity, while high values represent low similarity. Here, the LPIPS was computed between the synthetic and reference images using the PyTorch TorchMetrics library [48].

Another measure used in this study is the Zero-Mean Normalized Cross-Correlation (ZNCC), which is commonly employed in template matching applications [49]. The ZNCC measures similarity based on the correlation of pixel intensity values and takes values in the range [−1,1], where negative values indicate negative correlation and positive values indicate positive correlation.

3.4.2. Radiometric Quality

In this study, the noise level and consequently the radiometric quality between the generated synthetic images and the reference images were evaluated using the PSNR, which utilizes the maximum pixel value of the image together with the Mean Squared Error (MSE). A small MSE value leads the PSNR to approach infinity [50]. The PSNR was computed here with 8-bit images using Equation (2) [51]:

$$PSNR = 10\log_{10}\left(\frac{Max^2}{MSE}\right) \quad (2)$$

where the Max value was defined as 255 and the MSE was evaluated on a per-image basis.

Furthermore, histograms of the test and the generated images were compared as an additional radiometric assessment.

3.4.3. Geometric Quality

The Kanade–Lucas–Tomasi (KLT)-based Algorithm for Registration of Images from Observing Systems (KARIOS), developed within the scope of the ESA/EDAP project, was used to evaluate the geometric quality of synthetic images. KARIOS generates geometric deformations and reports in optical and radar images using an image matching method based on the Kanade–Lucas–Tomasi (KLT) algorithm. Developed with the OpenCV library, KARIOS is available as open source [52]. The geometric accuracy report includes root MSE (RMSE), minimum and maximum error, mean error, standard deviation error, circular error at 90%. KARIOS reports provide these values for both the east–west and north–south directions. The configurations required for the KARIOS tool to function are given in Table 6.

Table 6. The parameter configuration of the KARIOS for HR and VHR.

Parameters	HR	VHR
Confidence value	0.8	0.40
MinDistance	1	1
Blocksize	5	25
MaxCorners	0	2000000
Matching_winsize	15	55
Quality_level	0.1	0.1
XStart	0	0
Tile_size	4000	5000
Laplacian_kernel_size	7	7
Outliers_filtering	True	True

 doi:10.20944/preprints202604.1240.v1

4. Results

In this section, the results obtained from the two GAN-models are presented and evaluated through visual inspection, and visual similarity, radiometric, and geometric quality measures under the HR and VHR sections.

4.1. The HR Model Results

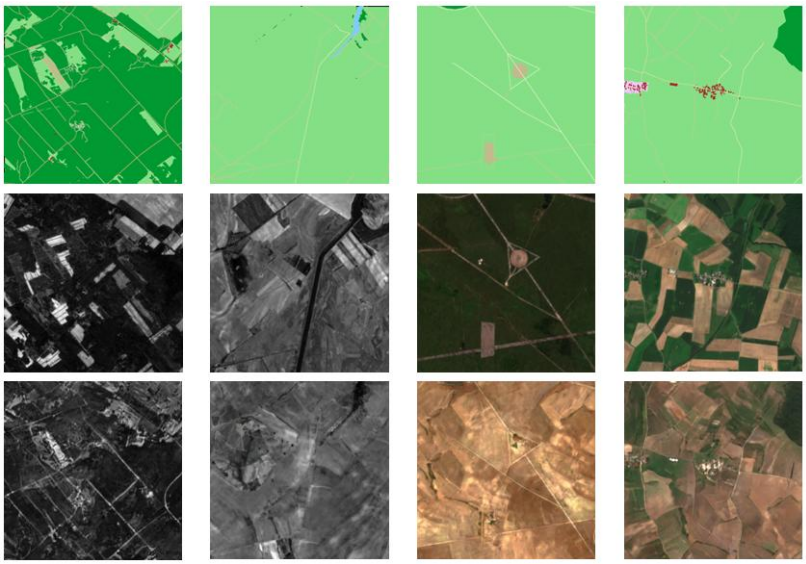
4.1.1. Visual Inspection of the HR Model Predictions

Figure 15 presents samples of synthetic images generated from Sentinel-2 imagery at 10 m resolution along with the corresponding OSM-derived masks. The figure also illustrates the final mono-band (B04) and multi-band (TCI) configurations used to produce the synthetic images. As shown in the figure, the test and generated images exhibit similar colors. However, the objects in the generated image align with the mask rather than the original test image. It must be noted that, unlike the VHR training dataset, no extensive manual editing was performed to ensure object–HR image alignment in the study.

Figure 15. Examples from the results of Sentinel-2 **(a)** TCI band, and **(b)** B04 band.

In Figure 16, further examples of the generated images are given. Here, the generated images largely deviate from the original ones. Notably, the objects in the synthetic images are rendered largely according to their representation in the masks, suggesting that the masks (i.e., vector maps) do not accurately capture the actual scene content. As a result, substantial discrepancies arise between the generated and real images. Nevertheless, the synthetic and the original images exhibit comparable color tones, indicating that the model successfully reproduces the global color distribution despite local structural inaccuracies. However, as can be seen in the third column of Figure 16, the radiometry of the generated image may differ from that of the original one. This discrepancy arises from radiometric variations within the training dataset (e.g., different sites and acquisition conditions) that are not captured by the OSM-derived rasters. In addition, seasonality of the images are not taken into account.

 doi:10.20944/preprints202604.1240.v1



(**a**) (**b**)

Figure 16. Samples from poor results in **(a)** B04 and **(b)** TCI bands (top row: mask, middle row: original images, bottom row: generated images).

4.1.2. Visual Similarity and Radiometry Measures for the HR Data

The visual similarity between the synthetic and reference images was evaluated using MS-SSIM, LPIPS, and ZNCC, while radiometric accuracy was assessed using PSNR. Two test sets were used for image quality assessment, such as;

i. unseen patches extracted from the training sites (within-domain) in both B04 and TCI bands,
ii. and test patches extracted from an independent MGRS tile (30TXT), also in B04 and TCI, from one Sentinel-2 L2A product.

Some examples are shown in Figure 17 and Figure 18, where the synthetic images are visually close to the real ones. Both figures correspond to samples from the first test site; Figure 17 shows results obtained using the TCI band, while Figure 18 shows results obtained using the B04 band.

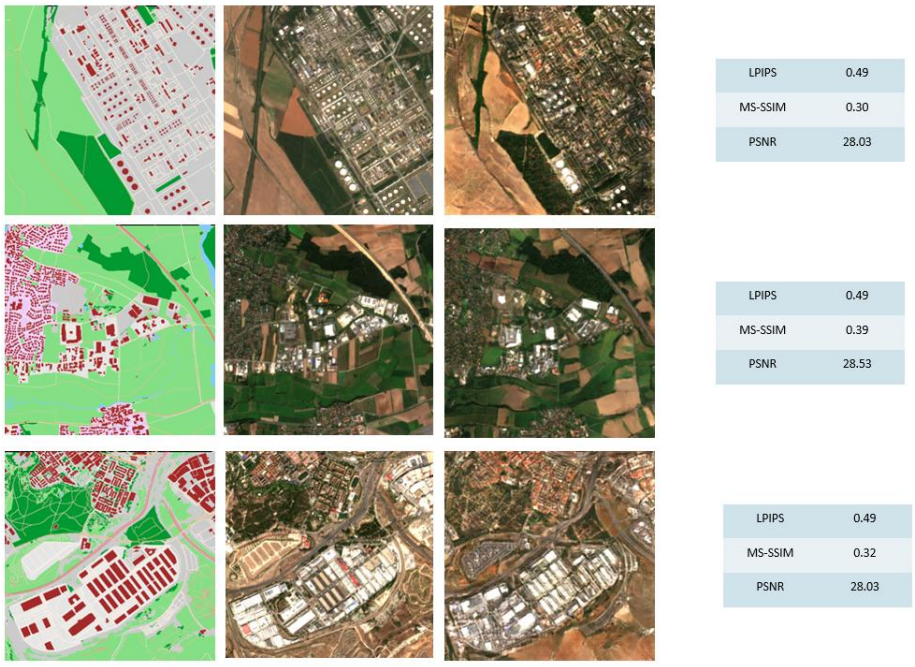


Figure 17. Samples of generated patches in TCI band from 1st test set (masks, real images, synthetic images) and associated quality measure values.

 doi:10.20944/preprints202604.1240.v1

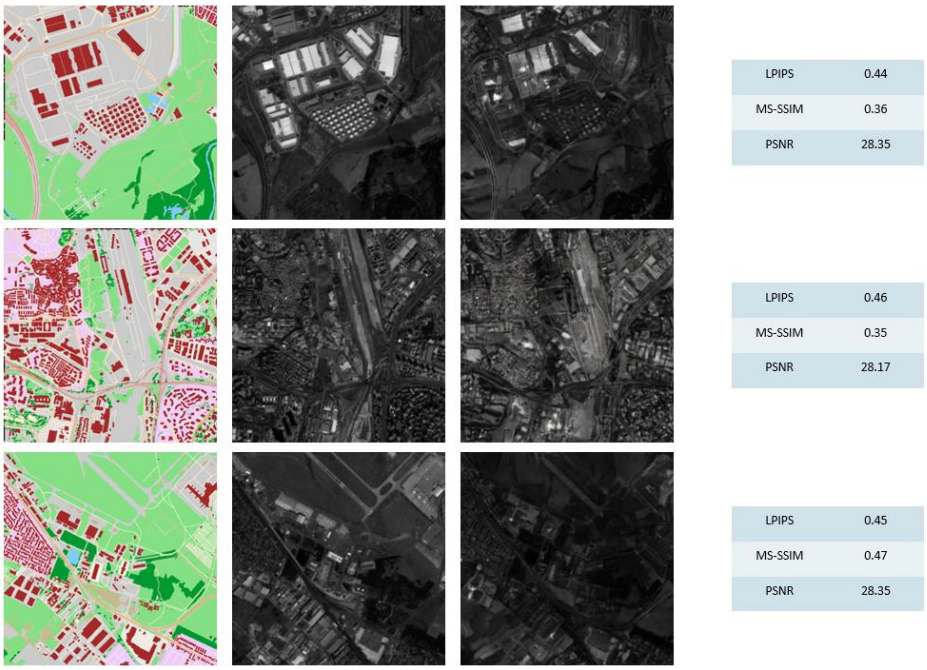


Figure 18. Samples of generated patches in B04 band from 1st test set (mask, real image, synthetic image) and associated quality measure values.

When comparing the B04 and TCI results, the B04 configuration shows slightly better performance. However, the low LPIPS values for both configurations indicate high perceptual similarity, while the PSNR and MS-SSIM values remain within acceptable ranges. The histograms of the reference and generated images (from the second test set) in Figure 19 show a good agreement in terms of distribution. However, greater discrepancies may arise if the acquisition conditions of the reference imagery or the geographical characteristics of the site differ from those represented in the training dataset.

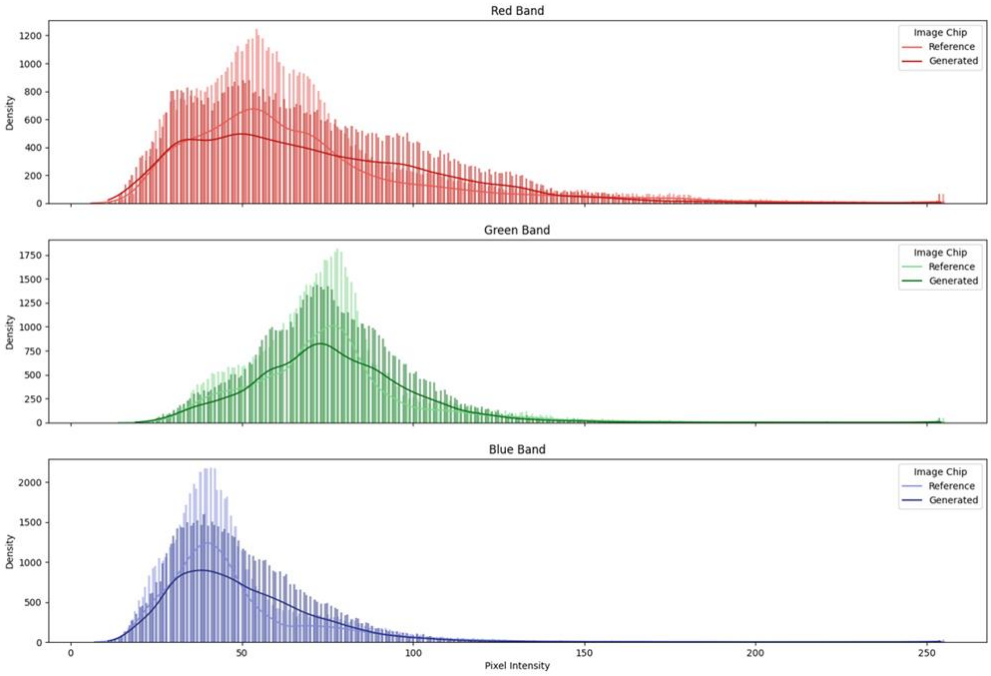


Figure 19. Band histograms for reference and generated images from the HR model.

 doi:10.20944/preprints202604.1240.v1

Figure 20 presents ZNCC-based cross-correlation results computed on test patches from both test sets in rural and urban environments. High correlation levels are observed for several features, particularly linear structures such as roads. The center of the image chips corresponds salient points identified as GCP in the Sentinel-2 GRI database. The ZNCC scores at the central pixel for the three examples are 0.46, 0.89, and 0.90, respectively, indicating strong correspondence at the GCP locations for the second and third cases.

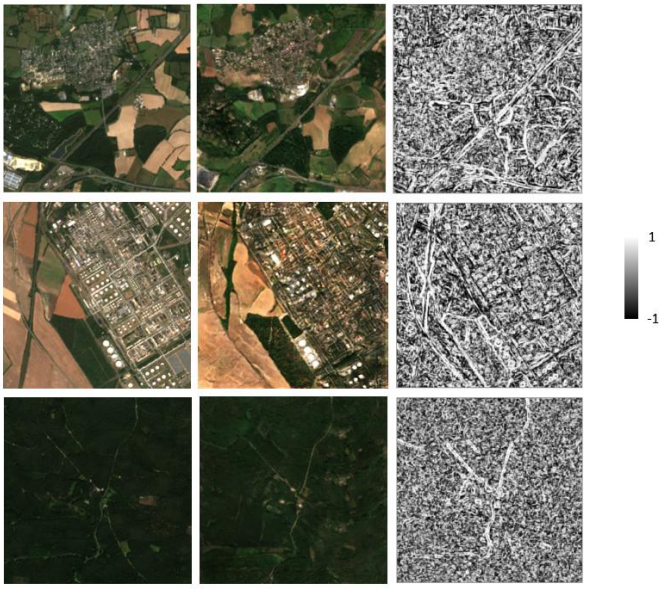


Figure 20. Cross-correlation samples (real image, synthetic image, ZNCC results) from 2nd test set (top row) and 1st test set (middle and bottom rows).

The mean values given in Table 7 show that the PSNR remains consistent across both test sets, while MS-SSIM and LPIPS show greater variability and yield slightly better values on the mixed-sites test set than on the independent test set (tile 30TXT).

Table 7. Visual similarity and radiometric assessment averaged over all patches in the HR test sets.

Image Pairs	MS-SSIM	PSNR	LPIPS
Mix Sites - TCI	0.414	28.586	0.520
Mix Sites - B04	0453	29.16	0.468
30TXT - TCI	0.312	28.431	0.542
30TXT - B04	0.336	28.625	0.494

4.1.3. Geometric Assessment Results of the HR Model

Two S2 L2A products from two different test sites were used for the geometric assessment. The first corresponds to MGRS tile 31TFJ (training site) in B04 and TCI, from the S2B L2A product (S2B_MSIL2A_20240915T102559_N0511_R108_T31TFJ_20240915T131207). The second corresponds to MGRS tile 30TXT (independent site), also in B04 and TCI, from the S2A L2A product (S2A_MSIL2A_20240919T105731_N0511_R094_T30TXT_20240919T171547). Figures 21 and 22 show the KARIOS matching results for MGRS tile 31TFJ for both multi-band and mono-band prototypes. The geometric results indicate good centering accuracy. Most of the East and North residual components fall within approximately 20 m, resulting in a global positional RMSE of about 22 m (considering a spatial resolution of 10 m per pixel). Despite the presence of outliers, the overall centering precision remains acceptable.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2026 doi:10.20944/preprints202604.1240.v1

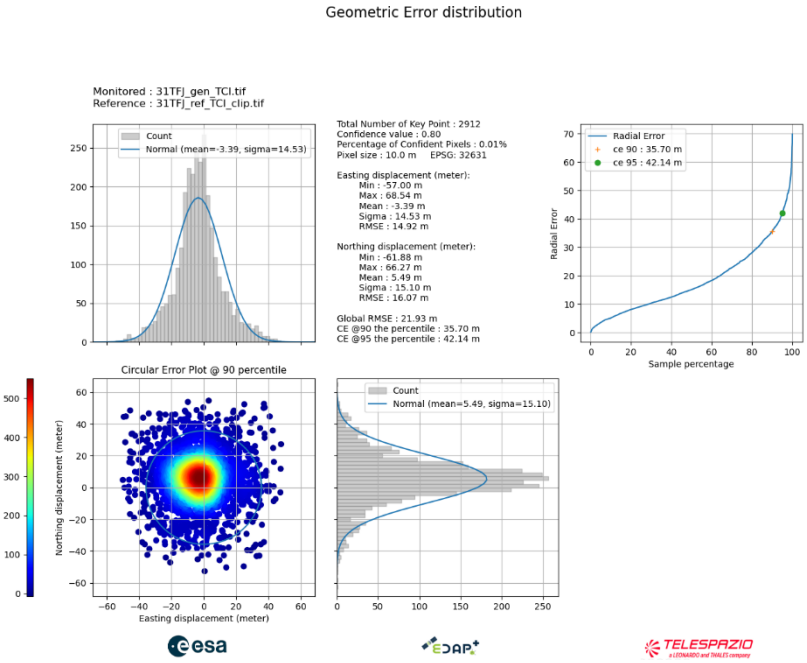


Figure 21. KARIOS results of the HR Model for tile 31TFJ in TCI band.

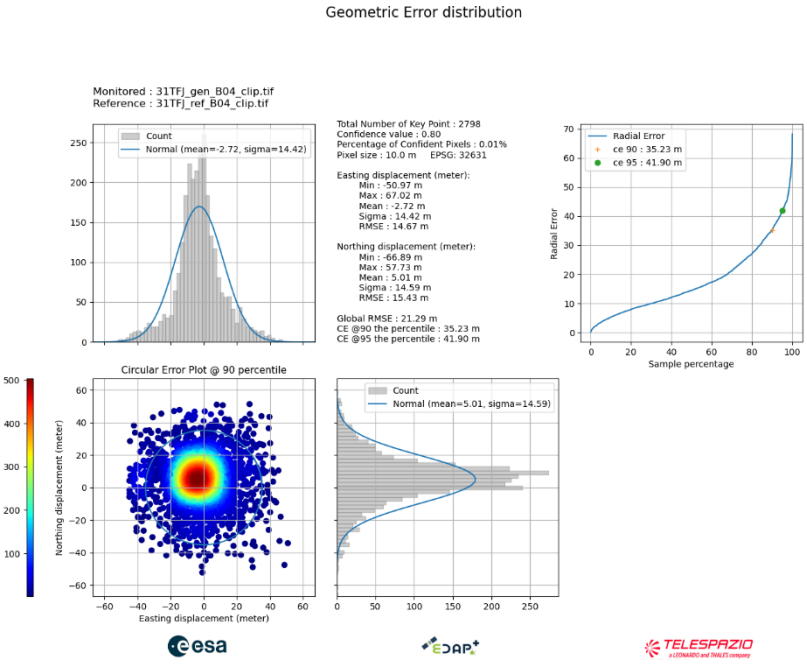


Figure 22. KARIOS results of the HR Model for tile 31TFJ in B04 band.

The results from an independent site are shown in Figures 23 and 24. Compared to the training site, the independent site generalizes less effectively. These differences can be attributed to the varying content of the patches; i.e., patches located near the center cover a major urban area, whereas patches located elsewhere cover rural areas where the OSM raster contains less information. Despite this fact, the global RMSE remains on the order of approximately 25 m.

Table 8 summarizes the geolocation performance under two configurations: Training/Test and an independent Sentinel-2 image excluded from model training. The results show a mean registration accuracy within 0.5 pixels (5.0 m). The RMSE is mainly driven by precision loss, as indicated by the standard deviation of the residuals.

 doi:10.20944/preprints202604.1240.v1

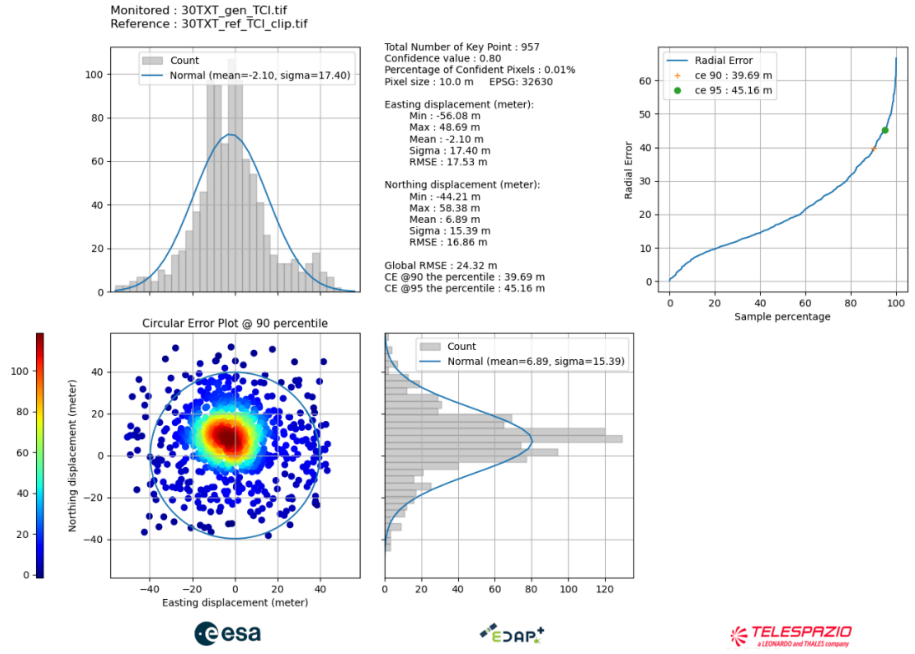


Figure 23. KARIOS results of the HR Model for tile 30TXT in TCI band.

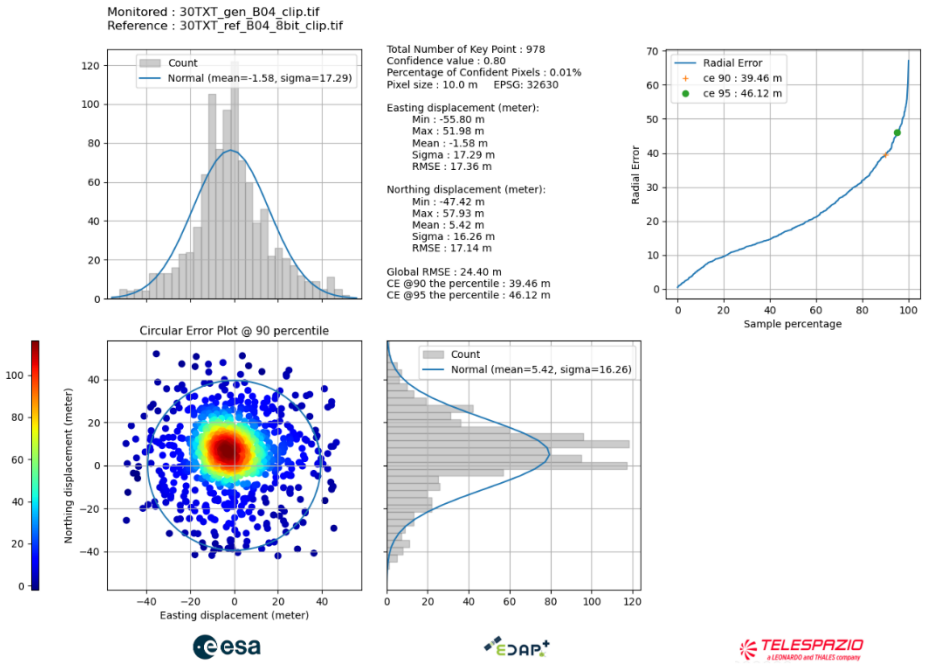


Figure 24. KARIOS results of the HR Model for tile 30TXT in B04 band.

Table 8. Summary of HR geolocation error performance.

	Training Site B04	Training Site TCI	Independent Site B04	Independent Site TCI
Mean Error	5.7 m	6.4 m	5.6 m	7.2 m
STD Error	20.5 m	21 m	23.7 m	23.2 m
RMSE	21.3 m	21.9 m	4.4 m	4.3 m

 doi:10.20944/preprints202604.1240.v1

4.2. The VHR Model Results

4.2.1. Visual Inspection

This section presents the results obtained from the four training datasets. Due to the 50 cm resolution and complex urban structures, masks required manual refinement. DS1 contains Maxar imagery from Salon-de-Provence, with some masks composed entirely of green surfaces. Test results are shown in Figure 25.

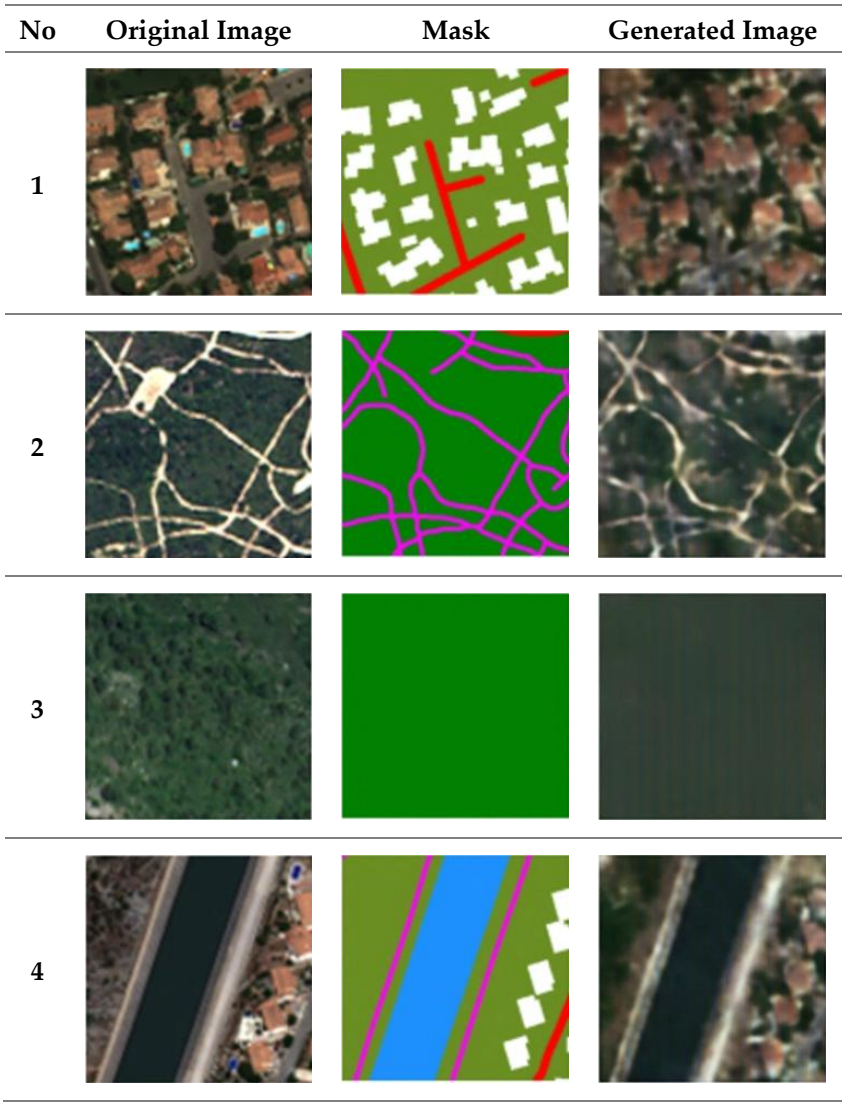


Figure 25. Results for DS1 (Maxar, Salon-de-Provence site), showing the original test image on the left, the corresponding mask in the middle, and the generated image on the right.

As shown in Figure 25, the model trained with DS1 dataset exhibit noticeable blurring artifacts. To address this issue, uniformly colored patches (as shown in the third row in Figure 25) were excluded before re-training. The revised dataset, DS2, comprises 3,231 images, with representative outputs presented in Figure 26.

Using the DS3, the model was trained with images from different satellite sensors and geographic regions, and tested on images from various locations. The results are shown in Figure 27. It was observed that the model performance was poor in homogeneous areas. Thus, the DS4 was compiled using images from the Salon-de-Provence and Ankara sites after extensive manual editing. Similarly, the test dataset was selected from these regions, comprising a total of 288 images. The DS4 contains a total of 7,325 training images. The model results obtained using this dataset are presented in Figure 28.

 doi:10.20944/preprints202604.1240.v1

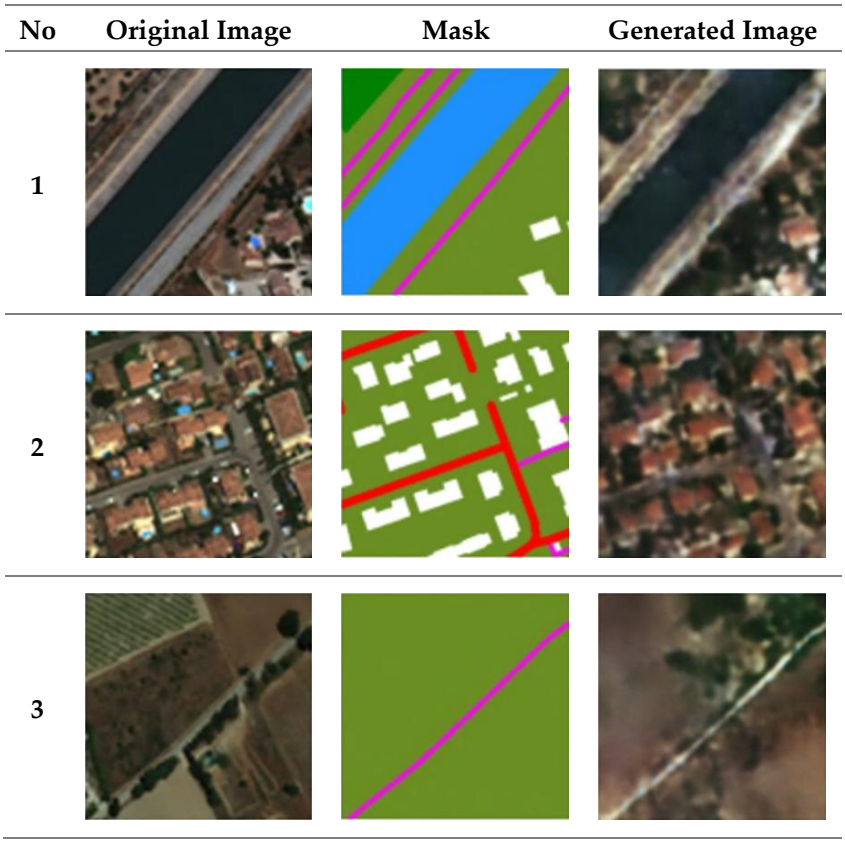


Figure 26. Results for DS2 (Maxar, Salon-de-Provence, after removing the single class masks), showing the test image on the left, the corresponding mask in the middle, and the generated image on the right.

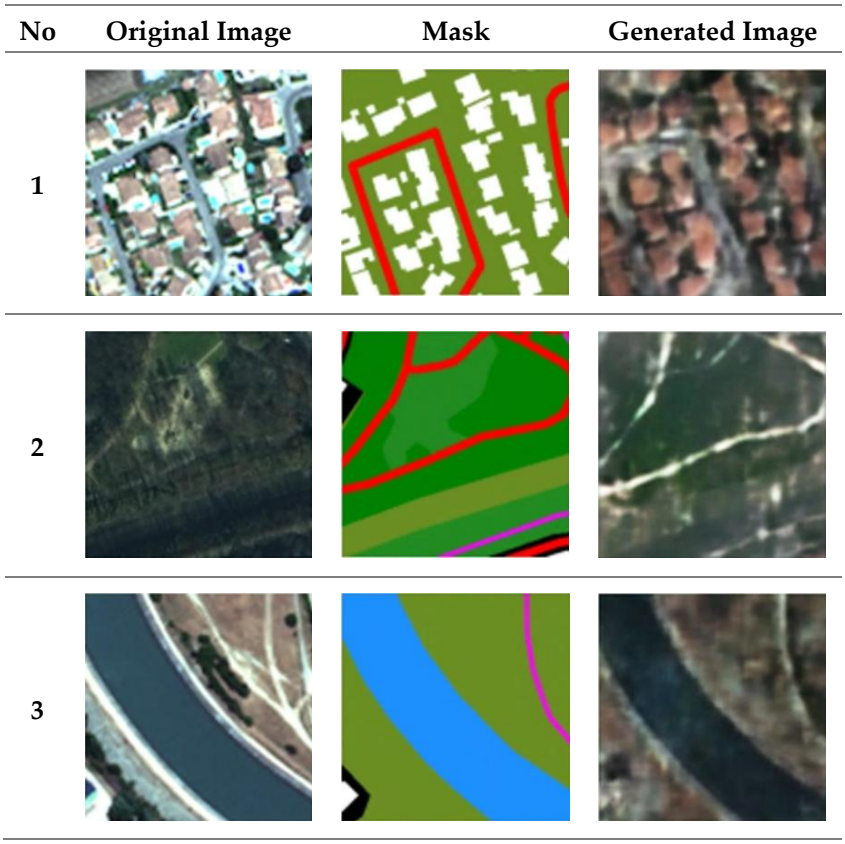


Figure 27. Examples from the DS3 model.

 doi:10.20944/preprints202604.1240.v1

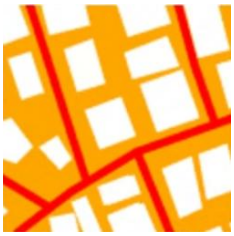
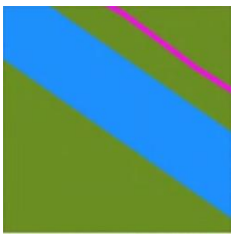
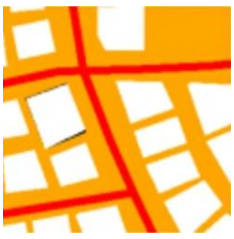
No	Image	Mask	Generate Image
1			
2			
3			

Figure 28. Examples from the DS4 model.

4.2.2. Visual Similarity and Radiometric Assessment

Based on visual inspection, DS4 produced the best results. Therefore, the synthetic image dataset used for further quality assessments was generated using the weights obtained from the model trained with the DS4.

A test area within the study site was selected, where Maxar, SkySat, and Pleiades imagery overlap. This training dataset does not contain Maxar and Pleiades images over Ankara test site. Thus, the model weights allow to investigate model performance across different satellite sensors and in an unseen area. The OSM mask and generated image for the test area are depicted in Figure 29. The rasterized vector layer, produced based on the SkySat satellite images, includes the OSM data with manual editing.

Figure 29. The map raster produced with the SkySat base (top) and the corresponding synthetic image generated (below).

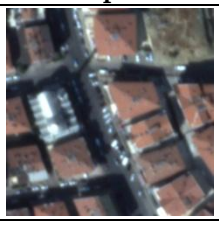
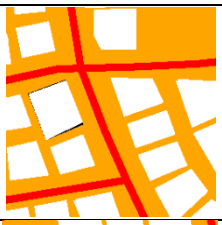
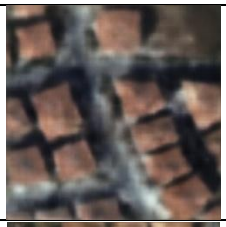
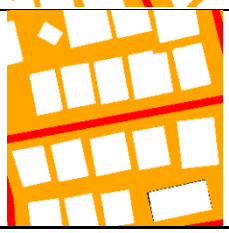
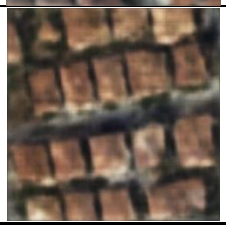
 doi:10.20944/preprints202604.1240.v1

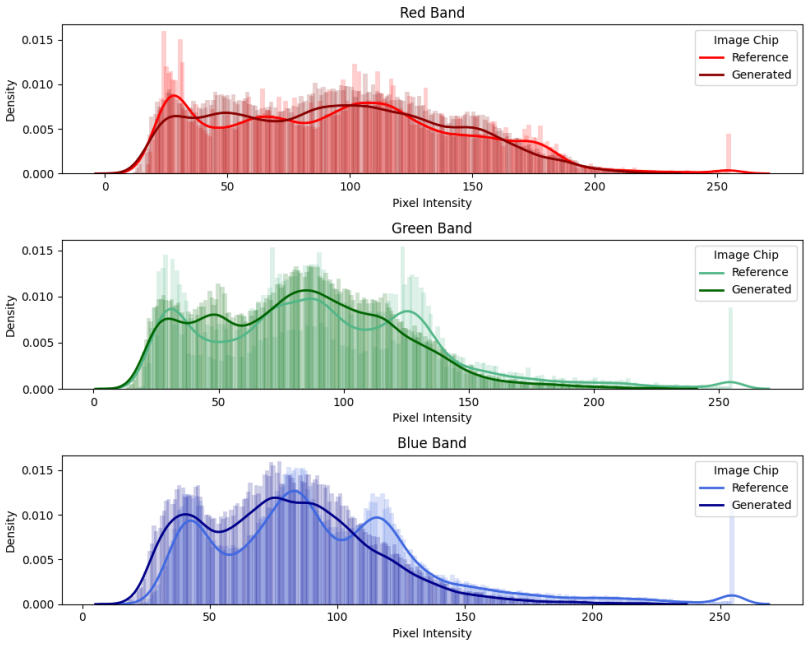
The quality assessment was conducted between the synthetic image produced for the test area and the original input data of Maxar, SkySat, and Pleiades satellite images, and using performance tests based on MS-SSIM, PSNR, LPIPS, and ZNCC metrics. As shown in Table 9, the best performance was observed for the SkySat satellite image. The Maxar and Pleiades satellite images follow SkySat in performance, respectively. The generated image demonstrates radiometric differences. These differences can be attributed to the lack of diversity in features within the rasterized vector layers. Despite the removal of uniform rasterized vector layer image from the training / test dataset (one class), some with limited feature diversity are still existing. Table 10 summarizes the SkySat image results, showing improved performance in urban areas at the chip scale. Histogram values for the samples labeled as 1 are shown in Figure 30, where comparison with Kernel Density Estimation curves reveals similarities between the reference and generated bands.

Table 9. Visual similarity and radiometric assessment averaged over all patches in the VHR test sets.

	MS-SSIM	PSNR	LPIPS
Maxar	0.256	27.961	0.691
SkySat	0.321	28.082	0.645
Pleiades	0.188	27.914	0.702

Table 10. Image quality assessment of SkySat image chip samples.

No	Input	Mask	Generated	MS-SSIM	PSNR	LPIPS
1				0.45	28.23	0.55
3				0.56	28.29	0.56

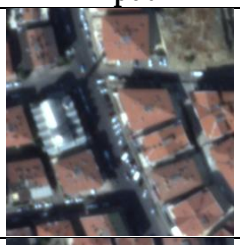
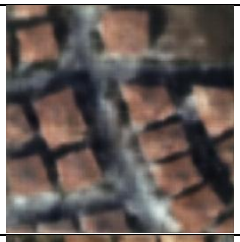
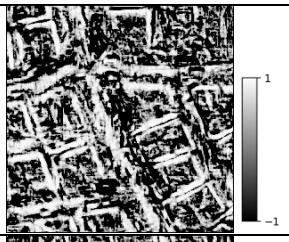
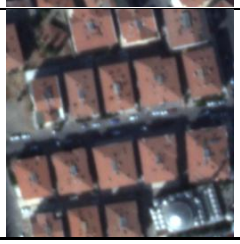
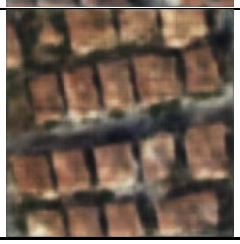
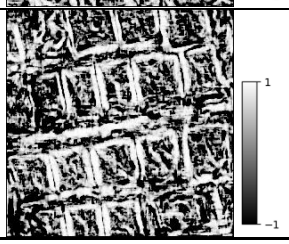


 doi:10.20944/preprints202604.1240.v1

Figure 30. Histogram graphics of the image bands labelled as 1 and the corresponding generated image bands.

Table 11 presents the ZNCC results for SkySat image chip samples. This table, which contains detailed images of areas with dense urban settlements, shows that the ZNCC values are high, indicating a high degree of similarity between the images.

Table 11. ZNCC results for SkySat image chip samples.

Input	Generated	ZNCC
		
		

4.2.3. Geometric Assessment Results of the VHR Model

The geometric quality assessment was conducted on the test area shown in Figure 29 (Ankara). The KARIOS matching results for Maxar, SkySat, and Pleiades are shown in Figures 31, 32, and 33, respectively.

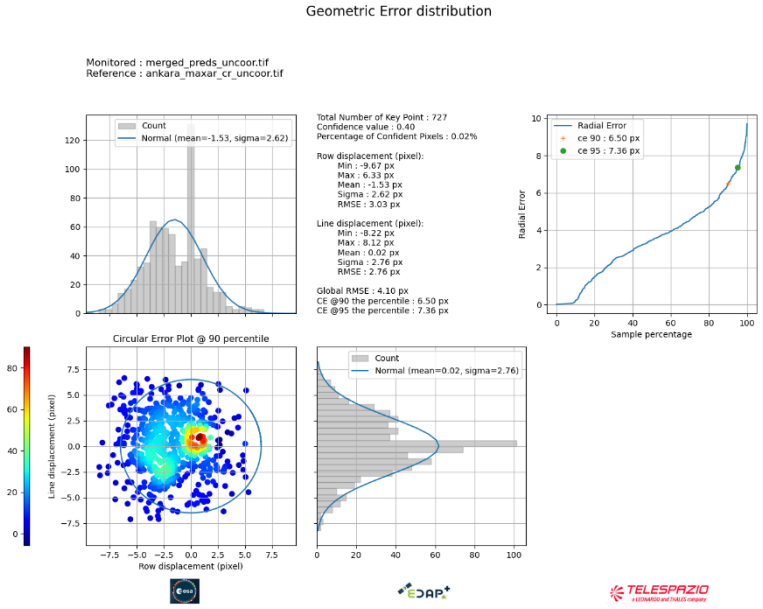


Figure 1. KARIOS results of the VHR Model w.r.t. the Maxar data.

Table 12 shows the geolocation error results for SkySat, Maxar and Pleiades data over Ankara. Although the results of SkySat and Pleiades are similar, the results of Maxar is worse due to the relief displacements at the building roofs caused by the off-nadir imaging angle and sparse digital surface model used for standard orthorectification. The building footprints in raster maps do not overlay with rooftops, which yields to noise in matching.

 doi:10.20944/preprints202604.1240.v1

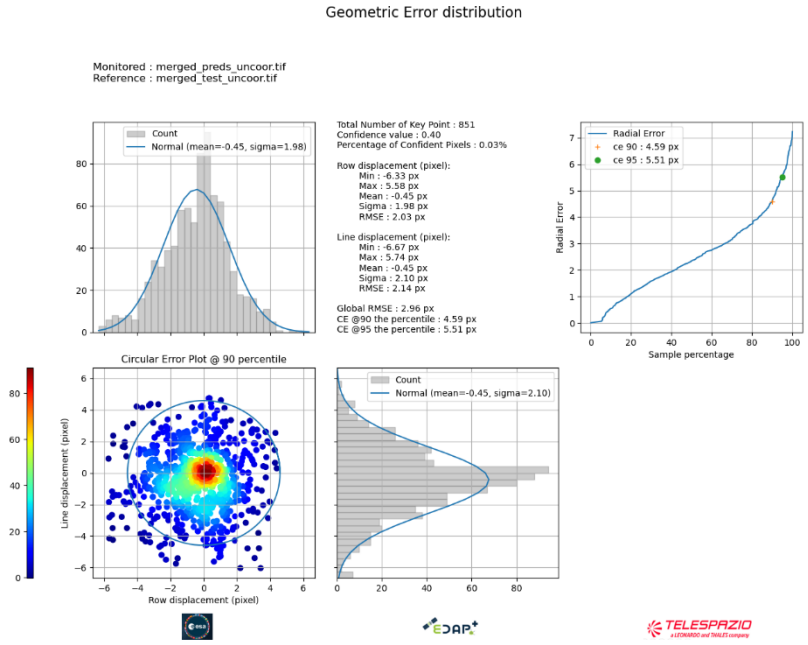


Figure 32. KARIOS results of the VHR Model w.r.t. the SkySat data.

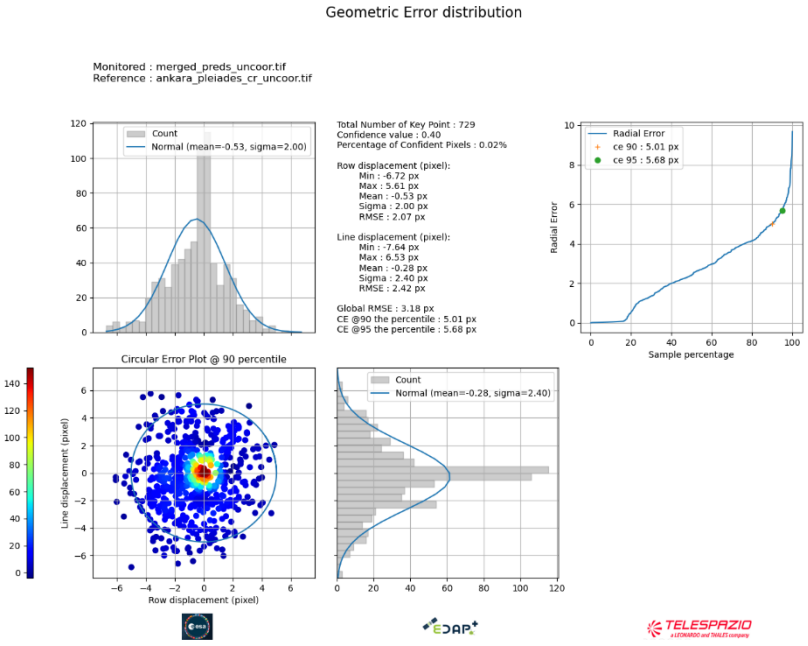


Figure 33. KARIOS results of the VHR Model w.r.t. the Pleiades data.

Table 12. Summary of VHR geolocation error performance based on DS4.

	SkySat Ankara test site	Pleiades Ankara test site	Maxar Ankara test site
Mean Error	0.2 m	0.2 m	0.5 m
STD Error	1.0 m	1.1 m	1.3 m
RMSE	1.5 m	1.6 m	2.1 m

5. Discussion

This study evaluated the feasibility of generating GCP–like reference image chips using a GAN-based framework trained on vector-derived inputs across two spatial resolutions (10 m and 50 cm).

 doi:10.20944/preprints202604.1240.v1

The results demonstrate that the proposed approach can produce geometrically consistent and visually coherent image representations that are suitable for automated image matching and georeferencing tasks.

A key aspect influencing the performance of the proposed approach is spatial resolution. The results indicate a clear contrast between the HR (10 m) and VHR (50 cm) configurations. At VHR, the model achieves significantly higher geometric accuracy and improved matching performance, particularly in urban environments. This can be attributed to the richer structural detail and higher positional fidelity of vector data at finer resolutions, where features such as buildings and roads are explicitly represented and directly constrain the generated imagery. In contrast, the HR configuration is more affected by abstraction and generalization effects, as multiple land cover types are aggregated within a single pixel. This reduces the distinctiveness of features and limits the ability of the model to reproduce precise spatial structures, leading to increased variability in matching performance and higher residual errors.

Moreover, the alignment between vector data and image content becomes more critical at higher resolutions. While small geometric inconsistencies may be negligible at 10 m, they introduce significant errors at 50 cm, necessitating manual refinement of vector datasets. These findings suggest that the proposed approach is inherently resolution-sensitive, with performance depending on both the level of detail in the input vector data and the degree of semantic and geometric alignment with the target imagery. Consequently, resolution-specific preprocessing strategies, including feature selection, generalization, and alignment procedures, are essential for achieving robust and transferable results.

The geometric assessment using KARIOS confirms that the generated products preserve spatial consistency at a level compatible with Cal/Val applications. For the HR configuration, sub-pixel mean accuracy (~0.5 pixel) was achieved, although the overall RMSE is influenced by residual variability linked to input data inconsistencies and scene heterogeneity. At VHR, significantly higher geometric accuracy (RMSE ~1.5–2.1 m) was observed, particularly in well-structured urban environments, confirming that the approach benefits from richer and more precise vector information.

Radiometric and perceptual evaluations indicate that the generated images reproduce global spectral characteristics and structural patterns of the reference imagery. However, discrepancies arise in cases where the vector representation does not fully capture scene complexity, particularly in rural or homogeneous regions. This highlights a fundamental limitation, namely that the generated outputs are constrained by the completeness and semantic richness of the input vector data. The absence of temporal, seasonal, and illumination information in the vector domain further contributes to radiometric inconsistencies.

A key finding of this study is that image–vector alignment quality is a dominant factor controlling model performance. In the VHR experiments, datasets with extensive manual correction (DS4) significantly outperformed those with limited alignment, demonstrating that geometric consistency between rasterized vectors and imagery is critical. This introduces a trade-off between automation and data preparation effort, which remains a central challenge for scaling the approach globally.

In this study, the OSM data were used as the primary vector source due to their global availability and open access, enabling large-scale dataset preparation. However, the positional accuracy and completeness of the OSM data were not systematically validated, as such an assessment falls outside the scope of this work. The objective of this study is not to evaluate a specific vector dataset, but to assess the feasibility of generating reference imagery from vector data. Consequently, the proposed methodology is not inherently tied to OSM and can be applied to any vector data source, including authoritative geospatial databases, provided that sufficient geometric accuracy and semantic consistency are ensured. This highlights the broader potential of the approach for generating high-quality, geometry-driven reference data across different data ecosystems. In practice, the use of higher-quality or authoritative vector datasets is expected to further improve the geometric fidelity of the generated products.

 doi:10.20944/preprints202604.1240.v1

Another important observation is the sensor dependency of the learned representations. Variations in radiometric response, acquisition geometry, and spatial resolution affect both training and inference performance. As a result, preprocessing steps such as class generalization, color mapping, and normalization require adaptation when transferring the methodology across sensors. This limits immediate generalization but also suggests that sensor-specific or domain-adaptive training strategies may be necessary for operational deployment.

Despite these limitations, the results support a broader conceptual shift from static, survey-based GCPs toward dynamic, data-driven reference representations. By leveraging widely available vector datasets, the proposed approach offers a scalable alternative to traditional GCP acquisition, particularly in regions where ground truth is sparse or inaccessible. Because vector data can be updated from diverse sources and is not tied to a specific sensor, it enables sensor-independent ground reference generation and facilitates consistent use across multiple missions. The demonstrated capability for successful image matching indicates that such generated products are fit for purpose for automated Cal/Val workflows, especially in multi-mission and multi-sensor contexts.

From an operational perspective, the current results demonstrate that the GenCP approach can already support automated Cal/Val workflows, particularly for image matching and georeferencing tasks where dense and globally available reference data are required. While the present implementation requires careful data preparation and is influenced by vector–image alignment quality, it provides a viable basis for semi-automated integration into existing processing chains. With the planned improvements in model generalization, multi-sensor adaptation, and automated quality control, the approach is expected to evolve toward fully operational use, enabling large-scale, automated generation of reference datasets for multi-mission environments.

6. Conclusions

In conclusion, this study, conducted within the framework of the ESA-supported GenCP initiative, demonstrates that GAN-based generation of vector-derived image chips constitutes a viable pathway toward scalable and adaptable reference datasets for satellite image georeferencing. While not a replacement for high-accuracy surveyed GCPs, the proposed method provides a complementary solution that enhances automation, coverage, and flexibility in Cal/Val workflows, with demonstrated suitability for image matching and georeferencing tasks. The performance of the approach is primarily governed by the quality and alignment of the input vector data, as well as the target spatial resolution, highlighting the importance of consistent data preparation.

In this context, the proposed approach contributes to operational Cal/Val frameworks supported by ESA and coordinated within CEOS, while also extending to commercial and institutional satellite data providers, where scalable, harmonized, and sensor-independent reference data are critical for ensuring consistency and interoperability across multi-mission and multi-vendor sensor systems. In addition, the methodology supports sensor-specific model training, enabling each provider to generate optimized reference datasets tailored to their own sensor characteristics. The approach can also be extended to other sensing modalities, including SAR, provided that modality-specific characteristics such as speckle and imaging geometry are properly accounted for during training.

Future work will be further advanced within the framework of the ongoing ESA-supported follow-up GenCP activities. These efforts will focus on improving both the robustness and operational readiness of the approach through several key developments: (i) enhancement of VHR image quality via dedicated deblurring and model optimization strategies, (ii) integration of Digital Elevation Models (DEMs) in the HR workflow to better account for topographic effects, (iii) extension of the methodology to additional geographic regions and sensor types, (iv) automation of vector-to-raster data generation workflows, particularly from OSM and similar sources, and (v) systematic refinement of the training process, including dataset preparation, class definition, and model configuration. These developments are expected to improve generalization, reduce manual intervention, and enable scalable, sensor-adaptive deployment in operational Cal/Val environments.

 doi:10.20944/preprints202604.1240.v1

Supplementary Materials: The code and the associated datasets are publicly available at https://github.com/telespazio-tim/GenCP and https://doi.org/10.5281/zenodo.15044428.

Author Contributions: Conceptualization, S.K., S.S., P.G. and L.L.; methodology, S.K., E.G., I.Y., S.S. and L.L.; software, E.G. and I.Y.; validation, S.K. and S.S.; formal analysis, L.L.; investigation, S.K., E.G., I.Y. and S.S. resources, P.G. and L.L.; data curation, visualization, E.G. and I.Y.; writing—original draft preparation, S.K., I.Y. and E.G.; writing—review and editing, S.K., S.S. and L.L.; supervision, S.K. and S.S.; project administration, S.S. and L.L.; funding acquisition, L.L. and P.G. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research was funded by European Space Agency under the Generative Control Point Study Contract.

Data Availability Statement: The code and the associated datasets are publicly available at https://github.com/telespazio-tim/GenCP and https://doi.org/10.5281/zenodo.15044428.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ekmen, O.; Kocaman, S. Remote sensing for UN SDGs: A global analysis of research and collaborations. *Egypt. J. Remote Sens. Space Sci.* **2024**, *27*, 329–341.
2. EDAP. The Earthnet Data Assessment Project (EDAP+). Available online: https://earth.esa.int/eogateway/activities/edap (accessed on 25 January 2026).
3. GEOSS. Global Earth Observation System of Systems (GEOSS). Available online: https://old.earthobservations.org/geoss.php (accessed on 25 January 2026).
4. Copernicus. Copernicus Programme. Available online: https://www.copernicus.eu/en/about-copernicus (accessed on 25 January 2026).
5. Kocaman, S.; Seiz, G. Contribution of Photogrammetry for Geometric Quality Assessment of Satellite Data for Global Climate Monitoring. *Remote Sens.* **2023**, *15*, 4575. https://doi.org/10.3390/rs15184575
6. Mannan, R.; Halsall, K.; Albinet, C.; Ottavianelli, G.; Goryl, P.; Boccia, V.; Melchiorre, A.; Piro, A.; Giudici, D.; Fox, N.; et al. ESA's Earthnet data assessment pilot: Paving the way for new space players. In Proceedings of the International Society for Optics and Photonics, Sensors, Systems, and Next-Generation Satellites XXIII, Strasbourg, France, 10 October 2019; p. 11151.
7. Saunier, S.; Goryl, P.; Chander, G.; Santer, R.; Bouvet, M.; Collet, B.; Mambimba, A.; Aksakal, S.K. Radiometric, geometric, and image quality assessment of ALOS AVNIR-2 and PRISM sensors. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3855–3866.
8. Saunier, S.; Karakas, G.; Yalcin, I.; Done, F.; Mannan, R.; Albinet, C.; Goryl, P.; Kocaman, S. SkySat Data Quality Assessment within the EDAP Framework. *Remote Sens.* **2022**, *14*, 1646. https://doi.org/10.3390/rs14071646
9. CEOS. CEOS EO Handbook. Available online: https://eohandbook.com/eohb2011/ceos.html (accessed on 25 January 2026).
10. CEOS Cal/Val Portal. The CEOS Cal/Val Portal: Calibration and Validation activities around space-borne sensors. Available online: https://calvalportal.ceos.org/ (accessed on 25 January 2026).
11. Yalcin, I.; Kocaman, S.; Saunier, S.; Albinet, C. Radiometric Quality Assessment for Maxar HD Imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2021, XLIII-B3-2021, 797–804. https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-797-2021
12. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
13. OpenStreetMap (OSM). Available online: https://wiki.openstreetmap.org/ (accessed on 18 January 2026).
14. Guasch, E.; Yalcin, I.; Kocaman, S.; Saunier, S.; de Laurentiis, L. GenCP – Development of AI-generated Ground Control Point (GCP) for EO Satellites Cal/Val. In Proceedings of the ESA VH-RODA 2024 Workshop, Frascati, Italy, 2–6 December 2024.

 doi:10.20944/preprints202604.1240.v1

15. Guasch, E.; Yalcin, I.; Kocaman, S.; Saunier, S. Generative Ground Control Point Study Algorithm Theoretical Basis Document. Technical report submitted to ESA, Nr. GENCP-TPZ-001-ATBD, Issue 1.0, 15 January 2025.
16. GenCP GitHub Repository. Available online: https://github.com/telespazio-tim/GenCP (accessed on 13 April 2026).
17. GenCP Dataset and Code. Zenodo. 2026. https://doi.org/10.5281/zenodo.15044428 (accessed on 13 April 2026).
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
19. Jozdani, S.; Chen, D.; Pouliot, D.; Johnson, B.A. A review and meta-analysis of generative adversarial networks and their applications in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102734.
20. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
21. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
22. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
23. Zhang, H.; Song, Y.; Han, C.; Zhang, L. Remote sensing image spatiotemporal fusion using a generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4273–4286.
24. Jia, Y.; Yu, W.; Zhao, L. Generative adversarial networks with texture recovery and physical constraints for remote sensing image dehazing. *Sci. Rep.* **2024**, *14*, 31426.
25. Enomoto, K.; Sakurada, K.; Wang, W.; Fukui, H.; Matsuoka, M.; Nakamura, R.; Kawaguchi, N. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 48–56.
26. Li, R.; Wen, L.; Shao, S.; Yu, M.; Mohaisen, L. A novel generative adversarial network framework for super-resolution reconstruction of remote sensing. *Front. Earth Sci.* **2025**, *13*, 1578321.
27. Pang, Y.; Lin, J.; Qin, T.; Chen, Z. Image-to-image translation: Methods and applications. *IEEE Trans. Multimed.* **2021**, *24*, 3859–3881.
28. Zuo, Z.; Li, Y. A SAR-to-optical image translation method based on PIX2PIX. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3026–3029.
29. Amitrano, D. Multitemporal SAR-to-Optical Image Translation Using Pix2Pix with Application to Vegetation Monitoring. *IEEE Access* **2024**, *12*, 14521–14535.
30. Xu, X.; Zhao, B.; Tong, X.; Xie, H.; Feng, Y.; Wang, C.; Xiao, C.; Ke, X.; Du, J. A data augmentation strategy combining a modified pix2pix model and the copy-paster operator for solid waste detection with remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8484–8491.
31. Hu, X.; Zhang, P.; Ban, Y.; Rahnemoonfar, M. GAN-based SAR and optical image translation for wildfire impact assessment using multi-source remote sensing data. *Remote Sens. Environ.* **2023**, *289*, 113522.
32. Šidlauskas, A.; Kriščiūnas, A.; Čalnerytė, D. Continuous Satellite Image Generation from Standard Layer Maps Using Conditional Generative Adversarial Networks. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 448.
33. Sentinel Online. Available online: https://sentinels.copernicus.eu/web/sentinel/sentinel-data-access/sentinel-products/sentinel-2-data-products/collection-1-level-2a (accessed on 18 January 2026).
34. Copernicus GCPs Database. Available online: https://s2gri.csgroup.space/#/GCP_L1C/collection.json?.language=en (accessed on 18 January 2026).
35. OSMnx. Available online: https://osmnx.readthedocs.io/en/stable/ (accessed on 18 January 2026).

 doi:10.20944/preprints202604.1240.v1

36. Brovelli, M.A.; Minghini, M.; Molinari, M.E.; Zamboni, G. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: The method and a case study. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2016, XLI-B2, 615–620.
37. Brovelli, M.A.; Minghini, M.; Molinari, M.E. An Automated Grass-Based Procedure To Assess the Geometrical Accuracy of the Openstreetmap Paris Road Network. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 2016, XLI-B7, 919–925.
38. Brovelli, M.A.; Zamboni, G. A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 289.
39. CORINE. Available online: https://land.copernicus.eu/en/products/corine-land-cover (accessed on 18 January 2026).
40. Duchon, C.E. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol.* **1979**, *18*, 1016–1022.
41. De Souza, V.L.T.; Marques, B.A.D.; Batagelo, H.C.; Gois, J.P. A review on generative adversarial networks for image generation. *Comput. Graph.* **2023**, *114*, 25–41.
42. Kingma, D.P.; Ba, J.L. ADAM: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
43. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Ye, X.; Mutlag, A.H.; Mohammed, M.A. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53.
44. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
45. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.
46. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.
47. Fan, L.; Khairuddin, A.S.M.; Liu, H.; Hasikin, K.B. Perceptual Carlini-Wagner Attack: A Robust and Imperceptible Adversarial Attack Using LPIPS. *IEEE Access* **2025**, *13*, 1234–1245.
48. TorchMetrics. Available online: https://pypi.org/project/torchmetrics/ (accessed on 18 January 2026).
49. Wang, X.; Wang, X.; Han, L. A novel parallel architecture for template matching based on zero-mean normalized cross-correlation. *IEEE Access* **2019**, *7*, 186626–186636.
50. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
51. Setiadi, D.R.I.M. PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444.
52. KARIOS. Available online: https://github.com/telespazio-tim/karios?tab=readme-ov-file (accessed on 18 January 2026).