Pliego, M. U., Marín, R. M., Shi, N., Shibayama, T., Leth, U., & Sacristán, M. M. (2025). Transport-related surface detection with machine learning: Analyzing temporal trends in Madrid and Vienna. Remote Sensing Applications, 101503-. https://doi.org/10.1016/j.rsase.2025.101503

Transport-Related Surface Detection with Machine Learning: Analyzing Temporal Trends in Madrid and Vienna

Miguel Ureña Pliego[1], Rubén Martínez Marín[1], Nianfang Shi[2], Takeru Shibayama[2], Ulrich Leth[2], Miguel Marchamalo Sacristán[1]

[1]Department of Land Morphology and Engineering, Civil Engineering School, Universidad Politécnica de Madrid, C. del Prof. Aranguren 3, 28040 Madrid, Spain

[2]Institut für Verkehrswissenschaften, TU Wien, Karlsplatz 13/230, A-1040 Wien, Austria

Contact: miguel.urena@upm.es

08/2024

Abstract

This study explores the integration of machine learning into urban aerial image analysis, with a focus on identifying infrastructure surfaces for cars and pedestrians and analyzing historical trends. It emphasizes the transition from convolutional architectures to transformer-based pre-trained models, underscoring their potential in global geospatial analysis. A workflow is presented for automatically generating geospatial datasets, enabling the creation of semantic segmentation datasets from various sources, including WMS/WMTS links, vectorial cartography, and OpenStreetMap (OSM) overpass-turbo requests. The developed code allows a fast dataset generation process for training machine learning models using openly available data without manual labelling. Using aerial imagery and vectorial data from the respective geographical offices of Madrid and Vienna, two datasets were generated for car and pedestrian surface detection. A transformer-based model was trained and evaluated for each city, demonstrating good accuracy values. The historical trend analysis involved applying the trained model to earlier images predating the availability of vectorial data 10 to 20 years, successfully identifying temporal trends in infrastructure for pedestrians and cars across different city areas. This technique is applicable for municipal governments to gather valuable data at a minimal cost.

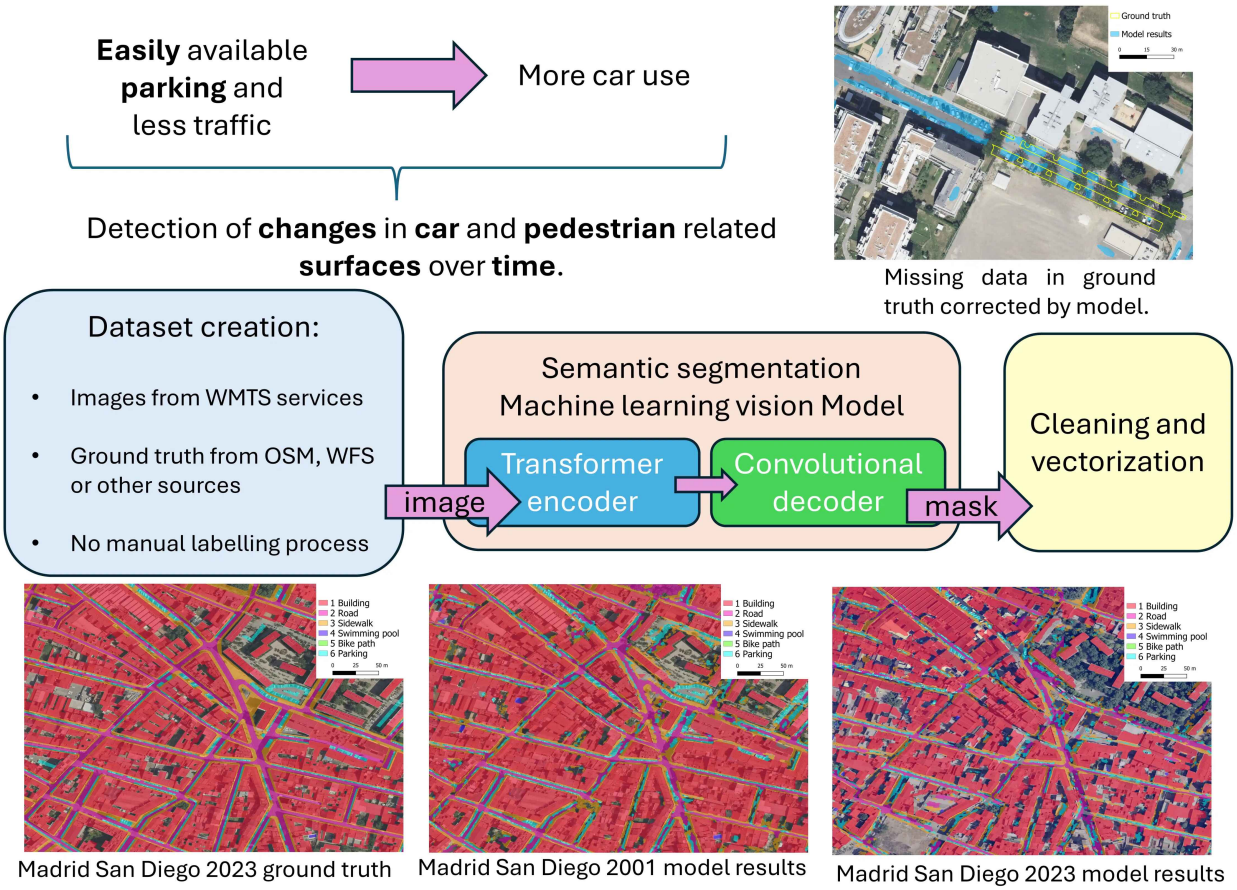
Keywords: image vision, OSM, parking, geospatial dataset, urban land use, sustainable urban transport

This document is a preprint. Peer reviewed version available under `https://doi.org/10.1016/j.rsase.2025.101503`

Highlights:

- General purpose semantic segmentation models do not provide accurate results when working with aerial imagery.
- Large training datasets for machine learning vision models were created using openly available aerial imagery and vectorial data from OpenStreetMap and official sources.
- The dataset creation did not require manual human labelling and the code created for this study is easy to use and fast.
- The vision models detected street, sidewalk, parking, and other classes with medium to high accuracy.
- The accuracy and usability of the vision models were tested in the cities of Madrid and Vienna, detecting temporal tendencies in the evolution of transportation-related surfaces.
- It was possible to detect a different tendency in the evolution of the amount of parking and road surface, with a decrease in the centre and an increase in the suburbs of the city of Madrid.

Graphical Abstract:



Graphical abstract illustrating the workflow and results.

1 Introduction and Objectives

Satellite imagery has been globally accessible for the past two decades, while high-resolution aerial imagery from airplanes has been available since the mid-20th century. Both current and historical images are freely accessible for research via services like Google Earth Engine [1] and national or local cartography providers.

Vectorial cartography is created by tracing aerial imagery. OpenStreetMap (OSM), the largest crowd-sourced project, relies on this method for global vectorial cartography [2]. However, this manual process is labor-intensive, resulting in incomplete data, especially in non-common categories like on-street parking spaces. Historical urban analysis is challenging due to scarce, less detailed historical cartography.

Aerial imagery is widely available, but vectorial cartography data is more accessible in developed regions with Spatial Data Infrastructures (SDIs) like the European Union's INSPIRE scheme, which standardizes geospatial datasets [3, 4]. However, categories may lack clear definitions. For instance, parking falls under RoadService without further subdivisions. In major European cities, openly accessible vector cartography following the INSPIRE scheme usually undergo annual updates [5–9].

OSM, as it is maintained by volunteers, does not follow the INSPIRE scheme but offers diverse geospatial data based on local needs [4]. OSM enables collaboration for rapid cartography, especially useful for humanitarian mapping post-disasters [10]. OSM has 88931 different keys, but the dataset is not comprehensive, and the absence of numerous objects can be anticipated. The data's quality varies, posing challenges for scientific studies [11, 12].

Advancements in computer vision and object detection [13] have enabled semantic segmentation and object detection in aerial imagery, useful across many fields [14]. Given the challenges of training general models and the availability of high-resolution imagery and cartography, it is important to establish a workflow for creating specific training datasets and training vision models for object detection and semantic segmentation in aerial imagery.

Analyzing surfaces dedicated to cars and pedestrians is crucial for evaluating car use in neighborhoods [15]. Although government and OSM data exist, they often lack quality and historical data are often not available. Machine learning can address these gaps.

Our research aims to create a workflow for developing datasets from global aerial imagery and vectorial cartography, applied to Madrid and Vienna for car and pedestrian surfaces. This workflow accelerates dataset creation, the most time-consuming part of a machine learning project. We want to demostrate the possibility of adapting and fine tuning existing semantic segmentation models for gathering data on urban infrastructure, especially for classes like parking spaces and demonstrate the viability of extracting historical trends using this method.

2 Background

Machine learning has seen extensive application in remote sensing and object detection over the last decade [13]. In the earlier stages, traditional statistical methods were prevalent for tasks like automatic road extraction [16]. However, machine learning methods have surpassed traditional approaches in recent years [13]. Various datasets and benchmarks for semantic segmentation tasks related to building detection, land use and cover extraction, and road geometry are available [17–20].

Machine learning models working with aerial imagery in the optical spectrum have been developed for diverse applications, including land use detection [21, 22], building detection [23–26], and road geometry extraction [27, 28].

In most cases, these machine learning models are based on image convolutions. They play a vital role in annotating maps [29,30] and exhibit high accuracy when tested in environments similar to the training data, but a low generability to unknown environments. Additionally, machine learning models can be utilized to correct OpenStreetMap (OSM) data [31].

Researchers have explored the labeling and creation of more diverse datasets to enhance the generability of aerial imagery segmentation models across diverse global environments [32]. However, addressing this challenge remains an ongoing effort. Recent advancements include the development of general purpose models that leverage transformer-based machine learning architectures for geospatial analysis [33]. A commercially available plugin for QGIS is now accessible [34] to run such models directly.

Datasets like GBSS [35] by Google and ML Building Footprints by Microsoft [36] provide building footprints worldwide. Both datasets where created with machine learning models trained with OSM data. The datasets may not have a very high quality [37], especially when precise annotations are required or when models are tested in regions outside the Western world.

While there are methods for detecting vehicles and monitoring parking space occupancy, they do not specifically extract the parking and road surface from individual images as it is intended in this study. These methods focus on detecting the cars rather than the parking surface itself. There are techniques involving surveillance cameras [38] and UAV high-resolution images, with resolutions ranging from 5 to 1.5 cm per pixel [39, 40] and multispectral high resolution aerial imagery [41]. Our objective is to create an inventory using openly available data with much lower resolution and to segment surfaces independently of the presence of cars.

Additionally, there is a method using satellite imagery that statistically evaluates image histograms to extract parking space occupancy without detecting individual cars [42]. However, this method is not suitable for our purpose.

Another intriguing area of research involves utilizing OpenStreetMap (OSM) data to generate datasets for tasks similar to ours. Some studies have created datasets to analyze transportation infrastructure, particularly parking and road surfaces [43, 44], aiming to develop a comprehensive parking space inventory. These studies achieved good results with Intersection over Union (IoU) values exceeding 60. In these cases, OSM data were refined and manually annotated, which is both time-consuming and expensive and the generability of the model has not been evaluated, so applications in areas different from the training set

are not suitable.

2.1 The relevance of road and parking surface in encouraging car usage

When studying the reasons people commute by car, research has long shown [45, 46] that the main factors are related to self-interest. Specifically, people choose to drive because it is the most convenient way for them to commute. Additionally, individuals who are more morally aware tend to avoid using cars. Those who do choose to drive often feel more guilt or responsibility regarding the environmental and social problems associated with car-centric behavior [45].

The amount of car-related infrastructure is therefore one of the main factors encouraging car use and diverting financing away from more sustainable alternatives [47]. Parking space availability, in particular, has long been recognized as one of the most critical factors influencing the choice of travel by car [48–51]. This is acknowledged by the European Commission in its technical guide for parking policy [52]. Parking space availability in urban environments can be indirectly inferred from the amount of paved surface [53] or directly evaluated by segmenting the area dedicated to that purpose from aerial images.

For another important factor in the modal split, the surface dedicated to active modes in already developed urban areas competes directly with surface dedicated to cars [54]. The relationship between surfaces dedicated to cars and pedestrians has a major and direct influence on walkability and the choice of active modes instead of driving [54]. In new urban developments, dedicating too much surface to cars makes distances and destinations farther apart, encouraging urban sprawl and car use [55, 56]. New approaches to street design consider reallocating urban space from cars to pedestrians and cyclists, redesigning streets to consider the needs of all users and encourage sustainable behavior [57–59]. This new paradigm has been adopted by the European Commission and is encouraged in its technical topic guides for developing Sustainable Urban Mobility Plans (SUMP) [60, 61].

Therefore, the main interests for the purpose of this study are pedestrian spaces, road surfaces, and car parking surfaces.

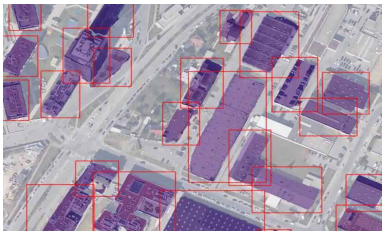
2.2 Semantic segmentation

The most widely used deep learning models for image segmentation, fully convolutional neural networks [62], employ image convolutions where the parameters of the kernel are training parameters. Semantic segmentation models usually follow a encoder-decoder structure [63].

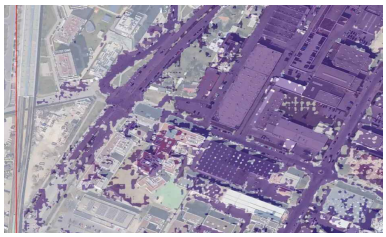
Until recently fully convolutional neural networks where the standard for semantic segmentation tasks. Recently language related tasks with deep learning saw a large improvement with transformer based models [64]. For image related tasks a vision transformer was developed [65]

During the year 2023 two models revolutionized the research regarding image vision with deep learning. The DaTaSeg model [66] and the segment anything model SAM [67] both are one of the first general purpose semantic segmentation models. Those models are trained with very large datasets using computing resources only available for the biggest technology

companies. The SAM model includes a GPT module allowing any user to input an image and a text prompt. The model segments the image according to the text given by the user (one shot segmentation). Another option provided by the model is to segment unknown objects in the image without providing any prompt or additional training (zero shot segmentation). A geospatial version SAM is available [33] through the SamGeo python package but employing the same trained model. The training dataset, the SA-1B dataset, comprises general images sourced from the internet [67] using information such as alt texts to train text promts, while the application involves aerial images, which possess markedly distinct characteristics and segmentation classes compared to the training data. General purpose semantic segmentation models such as SAM are effective in the context of geospatial analysis for general and easy tasks as segmenting buildings or detecting forests [68], but for a specific task like detecting parking spaces it fails completely [fig 1].



(a) SAM model results when segmenting building roofs.



(b) SAM model results when segmenting parking spaces.

Figure 1: The SAM model is able to fulfil only more general and common tasks.

Transfer learning and fine tuning in the context of computer vision involves taking a pre-trained model, which has been trained on a general and diverse dataset and freezing the weights of the pre-trained model's encoder and train a new decoder. The underlying assumption is that a general understanding of image patterns and features is transferable, independent of the specific dataset chosen, no matter how different the datasets might be [69]. However, it's important to note that this assumption has been empirically proven for years but the mathematical background is still under development [70].

3 Methodology

3.1 Machine learning implementation details

3.1.1 The model

For a deep learning approach an encoder-decoder structure is chosen. To achieve high quality results with limited resources, general purpose models such as SAM can be utilized using transfer learning for further training. The pre-trained encoder of the SAM model [67] was employed, with only a new convolutional decoder being trained in accordance with a U-Net like structure [fig 2]. The pre-trained encoder from the SAM model is a vision transformer

trained with masked autoencoders on the SA-1B dataset [67, 71] for images with 1024x1024 pixel resultion. The implementation utilizes the pytorch-lightning [72] and segmentation models [73] Python libraries. For the initial results, the basic vision transformer encoder with vit-b weights was utilized.

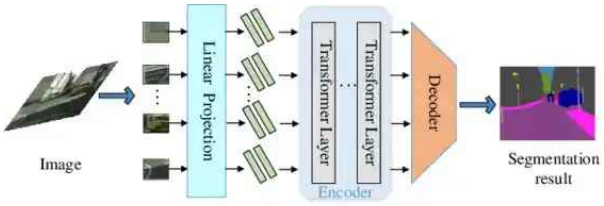


Figure 2: Transformer-based model [65]

Loss functions for semantic segmentation tasks can be categorized into pixel, boundary, and region-level loss functions [74]. In the case of training datasets featuring unbalanced classes, such as in this study, where smaller surfaces are detected over a main background class, the asymmetrical unified focal loss function [75] is an adecuate loss function for model training. This combined loss integrates the Tversky loss (region loss) with the focal loss (pixel loss) function.

The diversity of the images was measured calculating the dot product of the class count vector (number of pixels of each class in the image) of each image respect to the median value of the dataset. Images with lower diversity values are repeated more often during the training process.

3.1.2 Segmentation mask cleaning

Segmentation results are cleaned using morphological operations from the `scikit-image` library [76]. Holes in masks are closed, and noise is removed. Erosion thins objects by deleting border pixels, while dilation thickens objects by adding border pixels. Opening (erosion followed by dilation) removes noise, and closing (dilation followed by erosion) fills holes.

Erosion and dilation use image convolutions with specific kernels. Different kernels enhance various features [77, 78]. A round kernel rounds mask edges, a rectangular kernel creates rectangular shapes, and an octagon kernel sharpens edges optimally without limiting to rectangles [79].

Segmentation classes have known area ranges, used to further clean results. For example, parking spaces should be at least 3 m^2 in area and 1.5 m in width. Similar logic is applied to other classes.

The cleaning process is:

1. Erosion with a kernel of $\frac{1}{4}$ minimum width.

2. Deleting detections smaller than half the minimum area using binary opening.

3. Dilation.

Then, the inverse:

1. Dilation.
2. Deleting dark spots smaller than $\frac{1}{4}$ minimum area using binary opening.
3. Erosion.

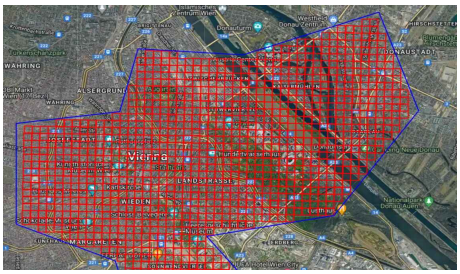
This process is repeated for each class, treating other pixels as background.

3.2 Aerial imagery datasets

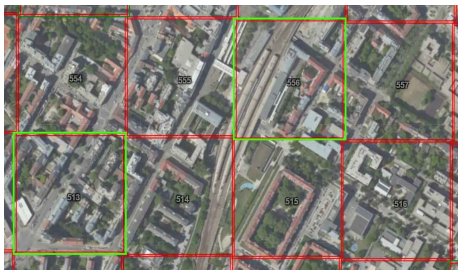
To illustrate the functionality of the developed code, two datasets were generated for the purpose of parking space detection in the cities of Vienna and Madrid. Creating the dataset in both scenarios can be accomplished with just a few lines of code, utilizing the library developed alongside this study. To generate the dataset, it is necessary to define a link to a WMTS service, specify the dataset area, and provide either a vectorial geometry file containing the ground truth or an overpass-turbo request.

3.2.1 Dataset grid

Neural networks based on Transformers need input images with a particular pixel size, whereas convolutional networks can handle images of different sizes. It is recommended for all images in the dataset to have similar sizes and resolutions in both scenarios, as these aspects can influence the visual characteristics of specific features in the images. To form a dataset, one or more polygons are assigned as the dataset region, and a grid structure is created [fig 3].



(a) Dataset region (blue) and generated dataset grid (red).



(b) Detail of the tiles in the grid. In green tiles randomly selected for the dataset. Numbers are the tile ids. Overlap is set to 0 m.

Figure 3: Dataset grid example.

The procedure involves defining a grid in UTM coordinates, shaped as a rectangle, based on the overall bounds of the dataset region. This grid is established with a specified tile size and overlap between tiles. As minimum and maximum values for x and y coordinates are needed to georeference a matrix and save it as a image, tiles are defined as minimum

and maximum x and y values. After forming the grid, all coordinates are converted to the coordinate system of the image provider. This conversion might adjust the grid bounds, as they could appear rotated. The reason for this is that the x and y axis can rotate during the coordinate conversion and, to cover the minimum and maximum x and y values from the previous coordinate system a larger area has to be selected and tiles can overlap slightly. In figure 3 tiles exhibit this situation due to the aerial image being in geographic coordinates while the grid is in UTM resulting in a small misalignment and overlap of the tile bounds. The grid is then converted to the coordinates of the image to accurately preserve the image bounds. Similarly, the mask in vector format undergoes conversion to the image's coordinate system. Incorporating overlap between tiles could be important for precise results at the image borders, helping prevent errors caused by partially visible objects. The overlap can also be configured accordingly.

3.2.2 Aerial imagery

Aerial optical imagery is globally accessible. Satellites provide lower-resolution images, with NASA's Landsat 8 and 9 offering 15-meter resolution and an 8-day revisit time [80], and ESA's Sentinel 2 providing 10-meter resolution with a 5-day revisit time [81]. Both agencies offer WMTS services for true-color images covering the entire Earth. Private companies like DigitalGlobe provide higher-resolution imagery (up to 30 cm), but access is limited and requires payment [82].

Drone or airplane images offer higher resolutions, typically 20 to 5 cm per pixel. For semantic segmentation and urban feature detection, resolutions finer than 1 meter are often needed [fig 4]. Our parking space detection project required images better than 20 cm per pixel, which can be challenging in regions with only satellite imagery.

(a) Sentinel 2 10 m resolution image in Vienna [81].

(b) 20 cm resolution orthophoto in the same area [6].

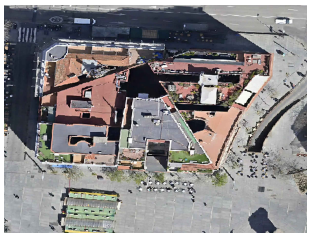
Figure 4: The importance of image resolution.

Google Earth [83] offers global aerial photography with resolutions from 15 meters to 10 cm. However, controlling the date and quality of images is challenging due to stitching from multiple sources.

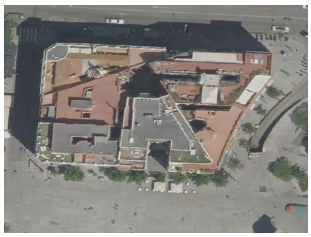
European cities provide public vector cartography and orthophotography with resolutions of 20 to 5 cm, updated annually [5–9]. These datasets are available via cloud services following Open Geospatial Consortium standards [84]. Historic aerial images since 1950 are often available. Overall, sufficient resolution imagery is available in most regions.

Aerial images have distortions from camera tilt or terrain topography. Orthorectification

removes these distortions, allowing precise measurements. This is crucial in urban environments, as buildings can occlude streets and distort building footprints [fig 5].



(a) 1.5 cm/pix UAV image.



(b) 10 cm/pix Aerial image.

(c) 30 cm/pix Satellite (WorldView3) image.

Figure 5: Remaining distortion in orthoimages of a building in Madrid [9] taken by UAV (least distortion), plane, and satellite (most distortion).

Image resolution and orthorectification are vital for classes like road surface, parking spaces, or sidewalks. Objects near buildings may be occluded, and trees pose challenges. Winter images have larger shadows, while summer images have less shadows but may obscure streets with trees [fig 6].

(a) Vienna winter image 2021 [83].

(b) Vienna summer image 2022 [6].

Figure 6: Issues with different image providers.

3.2.3 Ground truth

Ground truth data must be in the form of vector data, represented as polygons that delineate the objects targeted for segmentation. The use of vector data allows for seamless compatibility with various coordinate systems and resolutions. It is crucial to maintain the coordinate system used by the image provider consistently, necessitating the conversion of both the grid and ground truth to the coordinate system of the images. If the ground truth is initially presented as raster data, it can be transformed into vector data using tools such as rasterio [85].

A challenge arises when cartography is available as line data instead of polygons. Consequently, polygons need to be generated by connecting the lines that enclose the desired objects. This adds inaccuracies in the data.

The main inaccuracies found on the data used for this project are:

- Different or unclear criteria from the responsible organization when classifying the data.
- Even though the organization states the data where updated, new infrastructure or changes shown on the images are missing.
- Mistakes or inaccuracies in the boundaries of object.
- Inaccuracies or missing data when converting from line to polygon geometry.

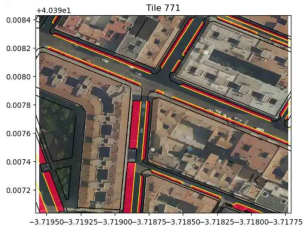
3.2.4 Madrid dataset

Two version of the dataset were created [figure 10a]. The main version has 6 classes: background, building, road, sidewalk, swimming pool, bike path and parking. The swimming pool class was chosen as it is considered an important factor for suburbanization and car-centric developments in Spain. The second version only has parking spaces. The selection of the training and testing areas aimed for diversity, encompassing poorer neighborhoods in the southern periphery characterized by dense and unorganized urban structures, newer neighborhoods from the 2000s in the eastern region featuring grid street patterns and multi-family housing with shared swimming pools, richer areas in the north with high rise buildings from the 80s and 90s, as well as neighborhoods from the historic medieval center and 19th-century developments.

The train dataset (4778 images and masks with a size of 1024x1024 pixels) encompasses all tiles contained inside the official boundaries of the following neighbourhoods of Madrid: Berruguete, Costillares, El Viso,Castellana, Quintana, Embajadores, Puerta del Angel, Los Rosales, Acacias, Goya, Numancia, Palomeras Bajas, Valderrivas, Orcasitas, Almendrales, Universidad, Almagro, Bellas Vistas and Hispanoamerica.

The test dataset (1238 images and masks with a size of 1024x1024 pixels) encompasses all tiles contained inside the official boundaries of the following neighbourhoods of Madrid. It is used exclusively for evaluation: Gaztambide, Cuatro Caminos, Pilar, Arcos and San Diego.

(a) Madrid's cartography provides parking spaces as lines (yellow) [9].



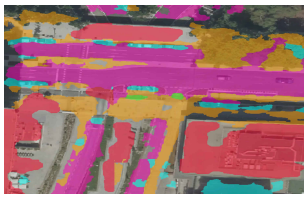
(b) Tile 771 of the dataset. In red the parking polygons in raster format created using the available line geometry.

(c) High resolution historic image from 1941 [9].

Figure 7: Madrid dataset details.

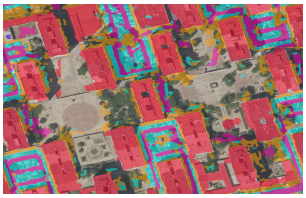
For the conversion from line to polygon geometry for the parking class the accuracy of the data was measured calculating the length of the parking lines contained in the parking

(a) Ground truth data in the test set is missing the new bike lane. Parking and sidewalk were displaced too.



(b) Model's results show inaccurate results. Some bike lanes are detected, but most are missing.

(c) Ground truth shows big areas classified as road that are wrong.



(d) The model correctly does not classify the wrong areas as road but this will be negatively evaluated.

Figure 8: Some examples about mistakes in the ground truth from the Madrid test set.

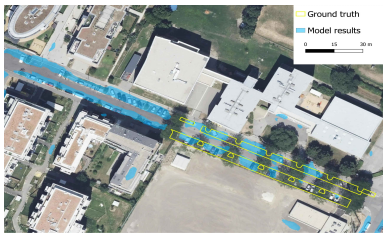
polygon boundaries. The ratio over the total length of the parking lines show that the worst neighbourhood passes 85% and that usual values where over 90%.

The inaccuracies found in the Madrid dataset are especially related to unclear criteria to classify data and data that is not up to date even though the 2023 cartography was used. The figure 8 shows some examples of the problems found in the dataset and the models output.

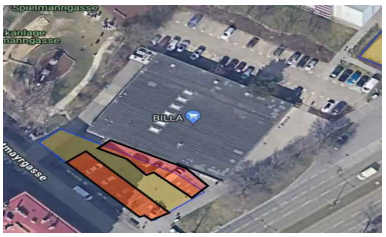
3.2.5 Vienna Dataset

The Vienna dataset encompasses a rectangular region situated to the east and west of the Danube, combining old and dense urbanisation with newer developments and even single family homes [fig 12a].

Two versions of the Vienna Dataset were created. The main version has 7 categories: background, public road, tram or train tracks, crosswalk, on street parking, private road surface, sidewalk and separated pedestrian or bicycle path. The dataset has 1826 images for training and 156 for testing all with a size of 1024x1024 pixels. Another version with 2 classes (background and on street parking) was created only to detect parking spaces.



(a) Ground truth data in the test set is missing some parking spaces. The model's results correctly classify all parking spaces as parking, but it will be negatively evaluated as most of the parking is missing in the ground truth.



(b) Overpass-turbo request results (in yellow) are not accurate enough compared to actual parking spaces (in red), and the data is incomplete. (In the top right corner, there is an unlabeled parking area)

Figure 9: Examples of mistakes in the ground truth from the Vienna test set.

Especially the OSM requests have many mistakes, as shown in figure 9b. In the case

of the Vienna dataset, most issues are related to missing data, particularly for the parking class, both in OSM and in the government's data [fig 9].

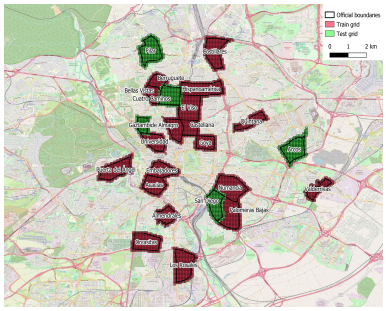
4 Results

The metrics chosen for model evaluation are:

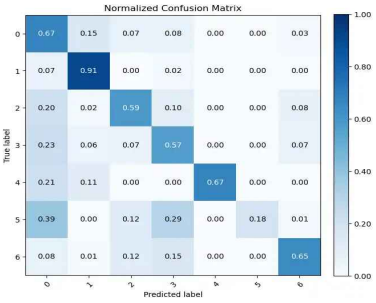
- **IoU**: Jaccard Index or Intersection over Union (IoU).
- **F1 score**: F1 score or Dice score.
- **IoU_200**: IoU using a 200-centimeter buffer for the ground truth before calculating the intersection.
- **Street ratio**: Amount of the model's output falling inside one of the classes related to streets in ground truth data (classes 2 and 6).
- **Pedestrian ratio**: Amount of the model's output falling inside one of the classes related to pedestrians in ground truth data (classes 3 and 5).
- $\frac{\textbf{model area}}{\textbf{GT area}}$: Area of the model output with class i divided by the area of the ground truth with class i. This ratio measures the amount of over or under detection in the model.

4.1 Madrid

4.1.1 Model evaluation



(a) Training and testing neighbourhoods.



(b) Normalized confusion matrix for the Madrid test set (Class 0 is background).

Figure 10: Madrid dataset and model results.

The building class shows the highest accuracy among all classes. For the rest of the classes, the IoU and F1 score is in a medium to high range [fig 10 and tab 1]. However IoU_200 indicates that the accuracy improves significantly if a deviation of a few centimeters or meters around the ground truth is allowed. The street and pedestrian ratios reveal minor

Table 1: Evaluation results for the Madrid test set.

Class id	Class name	model area	GT area	IoU	IoU_200	F1	street ratio	pedestrian ratio	$\frac{modelarea}{GTarea}$
1	building	2,769,163	2,601,307	0.73	0.82	0.86	0.01	0.02	1.07
2	road	1,343,829	1,776,690	0.50	0.55	0.67	0.81	0.06	0.77
3	sidewalk	1,103,500	1,127,623	0.40	0.52	0.56	0.22	0.57	0.99
4	pool	21,403	12,624	0.25	0.33	0.47	0.07	0.02	2.61
5	bike path	11,250	18,060	0.06	0.08	0.19	0.18	0.32	0.75
6	parking	541,439	438,223	0.36	0.52	0.55	0.75	0.14	1.26

errors in distinguishing between infrastructure intended for vehicles and pedestrians. Most inaccuracies occur between classes within similar categories.

The bike lane and swimming pool classes have the lowest accuracies [fig 10 and tab 1]. However, the area covered by these classes in the ground truth is minimal or none in many training and testing areas, leading to inconsistent results for these classes. Nonetheless, as explained in the following section, correct trends over time were observed for all classes, including pool and bike lane. Overall, the model's results demonstrate that it is effective for predicting exact building boundaries and providing general statistical data for the other classes.

Even though there can be significant differences (over 15%) in the area picked up by the model in comparison with ground truth for some classes, the assumption is made that the mistakes or differences between ground truth and model will not occur between the results of the same model evaluated with different images. For the comparison between 2001 and 2023, the model will apply the same criteria, overrepresenting or underrepresenting the classes in the images from both times by the same amount, allowing for a valid comparison. This is the reason why the model's output from 2001 is not compared directly to the ground truth from 2023.

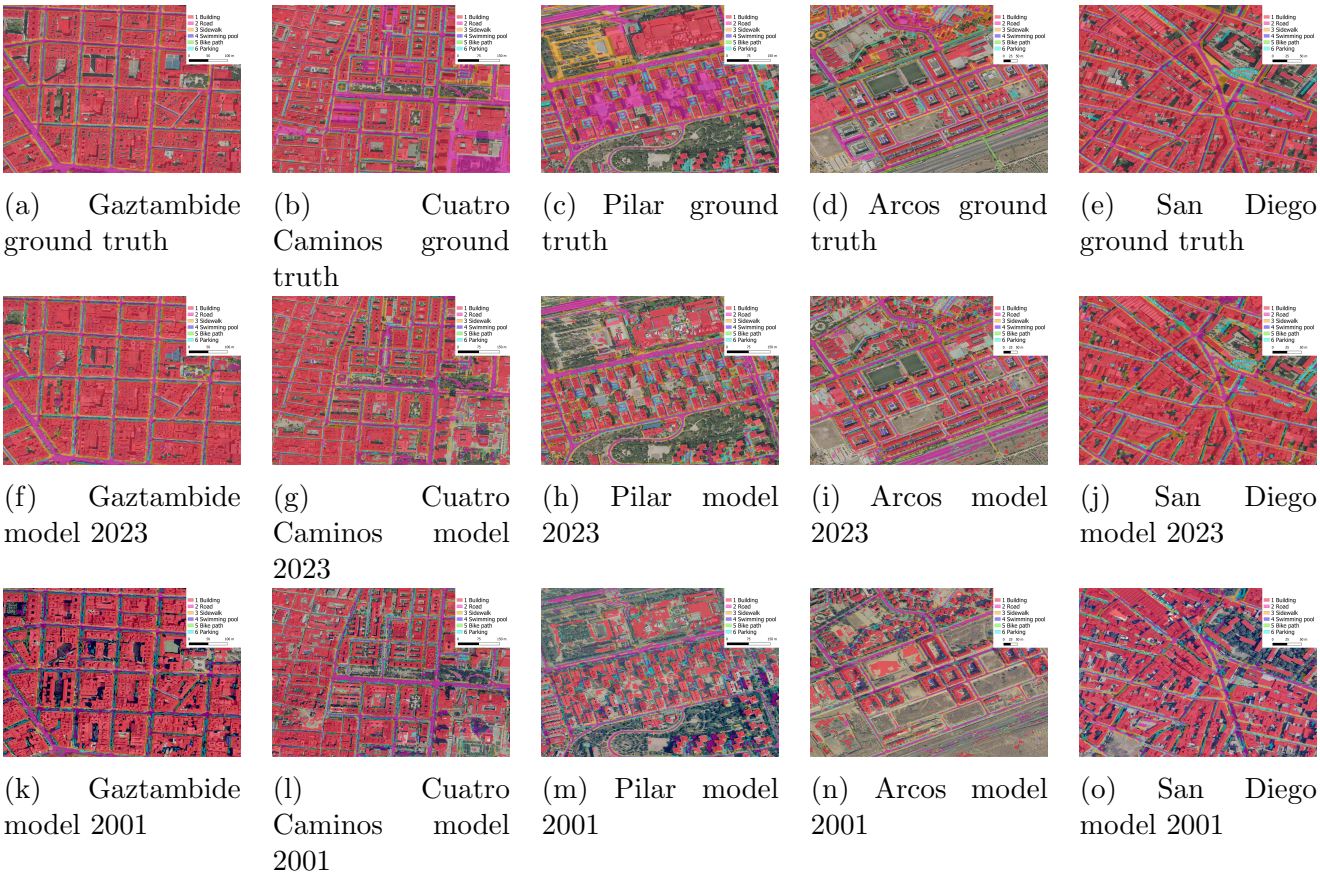
4.1.2 Temporal trends (2001-2023)

Images from the same testing neighborhoods but from the year 2001 (RGB images with 10 cm per pixel resolution) were inputted into the model that was trained with images from 2023. The model's output from 2001 was compared to the model's output from 2023, evaluated previously. The IoU values indicate if the classes overlap and show the amount of variation that occurred during the timeframe. Very low IoU values (below 0.1), especially if the areas have not changed much, can indicate that the model is not providing accurate predictions for that class. The change in area shows the trend over time.

Table 2: Temporal trend (area in 2023 / area in 2001) for the Madrid test dataset and iou_200 (in red) between the 2001 and 2023 geometries to show the validity of the results.

Dataset	1 building	2 road	3 sidewalk	4 pool	5 bike Path	6 parking
Gaztambide	1.11 (0.80)	1.45 (0.51)	1.27 (0.46)	0.87 (0.08)	0.11 (0.00)	0.81 (0.48)
C. Caminos	1.13 (0.71)	1.36 (0.57)	1.22 (0.45)	1.07 (0.10)	1.00 (0.00)	0.79 (0.38)
Pilar	1.15 (0.52)	1.67 (0.41)	1.65 (0.29)	1.31 (0.23)	2.17 (0.02)	1.40 (0.31)
Arcos	1.22 (0.53)	1.64 (0.46)	1.54 (0.32)	2.04 (0.20)	2.88 (0.01)	1.41 (0.31)
San Diego	1.10 (0.69)	1.85 (0.41)	1.41 (0.41)	1.02 (0.08)	2.01 (0.05)	1.34 (0.35)
ALL	1.13 (0.65)	1.58 (0.47)	1.42 (0.39)	1.34 (0.14)	2.35 (0.02)	1.15 (0.37)

In the center of the city [tab 2], specifically in the Gaztambide and Cuatro Caminos areas, the trend for parking spaces is decreasing, by around 20%. However, there is an increase in



(a) Gaztambide ground truth (b) Cuatro Caminos ground truth (c) Pilar ground truth (d) Arcos ground truth (e) San Diego ground truth

(f) Gaztambide model 2023 (g) Cuatro Caminos model 2023 (h) Pilar model 2023 (i) Arcos model 2023 (j) San Diego model 2023

(k) Gaztambide model 2001 (l) Cuatro Caminos model 2001 (m) Pilar model 2001 (n) Arcos model 2001 (o) San Diego model 2001

Figure 11: Madrid test dataset. An example region is shown for each testing neighbourhood. Ground truth from 2023 and model output for images from 2023 and 2001 are shown.

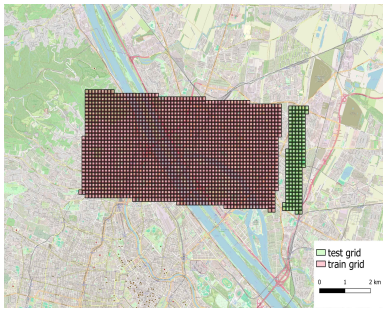
the surface of roads and sidewalks, with a growth of 30 to 40%. The number of buildings remains mostly constant, with a slight increase of 10%.

In the periphery of the city [tab 2], including Pilar, Arcos, and San Diego, the trends for parking, roads, and sidewalks are increasing, with about a 50% increase in surface area. Building growth varies, with a slight increase of 10 to 15% in Pilar and San Diego, and a moderate increase of 22% in Arcos. Arcos, the newest neighborhood developed in the early 2000s, shows a significant increase in the number of swimming pools (104%), a trend not seen in other neighborhoods, where the increase is less than 30%. This is due to the tendency at that time to build multifamily housing with shared swimming pools in new developments.

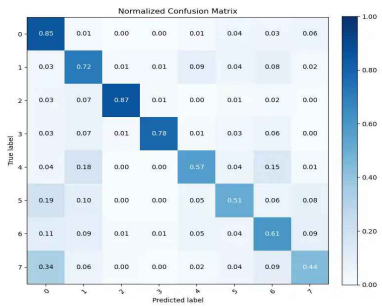
There are no significant differences in the trends in road and parking surface between richer neighborhoods, such as Pilar, and poorer neighborhoods, such as San Diego. The most notable differences are observed between the city's periphery and its center.

4.2 Vienna

Public road and tram tracks show the best accuracies. For the rest of the classes the IoU is lower, but the rest of the indices can validate the results [tab 3]. IoU_200 (23-GT calculated with ground truth and model results for 2023 images) show that the accuracy increases vastly



(a) Training and testing areas.



(b) Normalized confusion matrix for the Vienna test set (Class 0 is background).

Figure 12: Vienna dataset and model results.

if a deviation of a few centimeters or meters around the ground truth is tolerated. The street and pedestrian ratios show small amounts of errors between detecting infrastructure for cars or for pedestrians. Most of the inaccuracies arise between similar looking classes, like parking and road.

Table 3: Evaluation metrics and temporal trend (2014-2023) for the Vienna model.

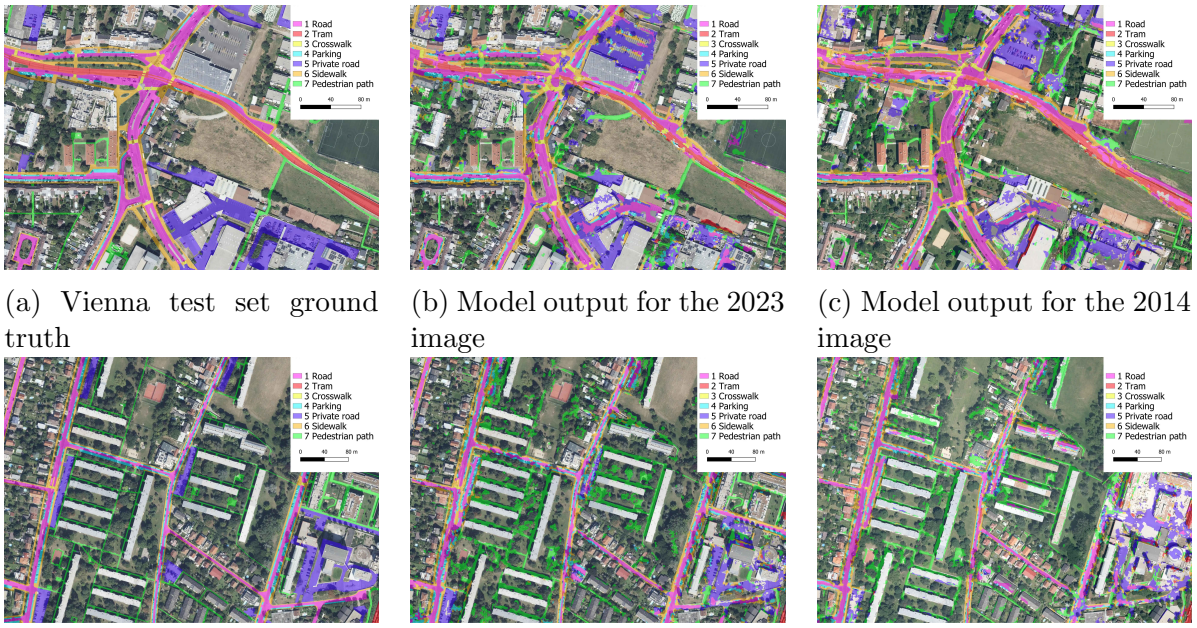
Class id	Class name	model area	GT area	iou	iou_200 (23-GT)	F1	street ratio	pedestrian ratio	$\frac{modelarea}{GTarea}$	iou_200 (14-23)	$\frac{area23}{area14}$
1	public road	163,676	159,214	0.47	0.56	0.69	0.79	0.11	1.03	0.59	1.09
2	rail tracks	31,696	28,819	0.55	0.66	0.76	0.09	0.04	1.10	0.46	0.69
3	crosswalk	4,781	2,999	0.33	0.57	0.51	0.37	0.52	1.59	0.40	1.37
4	parking	71,956	50,593	0.23	0.37	0.45	0.61	0.13	1.42	0.35	1.61
5	private road	157,414	166,984	0.32	0.36	0.51	0.55	0.06	0.94	0.37	0.89
6	sidewalk	153,034	97,993	0.25	0.44	0.45	0.20	0.40	1.56	0.42	1.33
7	pedestrian path	195,103	145,929	0.20	0.33	0.37	0.08	0.34	1.34	0.39	0.99

The model trained with 2023 imagery was tested with images from 2014, the first available RGB images. The outputs for both time periods are compared in table 3 (columns iou_200 (14-23 calculated with model results from 2014 and 2023) and $\frac{area23}{area14}$) and figure 13. Results show a decrease in tram tracks and private roads. The amount of road surfaces, bike paths, and pedestrian paths remains constant. Sidewalks increase, but the most significant growth is observed in public on-street parking, confirming the trend towards car-centric developments in the newer areas of Vienna.

To validate the transfer learning approach, a model was trained on this dataset using the SAM encoder, while another UNet model was trained without utilizing SAM's encoder weights. The transfer learning SAM model outperformed the UNet model in the mean IoU metric by 7%, even though it was trained with only 10 epochs compared to the UNet's 30 epochs.

4.3 Parking models

Two models were trained using only the parking datasets, each with two classes. One model was trained with Madrid data and evaluated in Vienna, while the other model was trained



(a) Vienna test set ground truth

(b) Model output for the 2023 image

(c) Model output for the 2014 image

(d) Vienna test set ground truth

(e) Model output for the 2023 image

(f) Model output for the 2014 image

Figure 13: Vienna test dataset. The two regions cover most of the test set. Ground truth from 2023, model output for images from 2023 and 2014.

in Vienna and evaluated in Madrid. The Madrid-trained model achieved an IoU of 0.18 and an IoU_200 of 0.31 in the Vienna test set, which is a decrease of only 0.06 compared to the Vienna-trained model. Similarly, the Vienna-trained model achieved an IoU of 0.2 and an IoU_200 of 0.3 in the Madrid test set. In that case the decrease was higher, 0.2, but it has to be considered that the IoU_200 value for that model in Vienna was 0.37. This demonstrates the generalizability of the models and the validity of the method when working with images from cities that look very different. Images from different times but from the same city as the training set are more similar, so the results should be more accurate.

5 Discussion

Results indicate that the models are accurate enough to gather statistical data, such as the overall area in a neighborhood used for various purposes like housing, cars, parking, or pedestrians. The models exhibit sufficient generalizability to work in environments different from the training dataset as it was shown running tests on the Madrid-trained model in Vienna and on the Vienna-trained model in Madrid. It was possible to compare results from 2023 and 2014 for Vienna and from 2001 for Madrid, dates before government official data is available. This allowed us to compare trends over time and distinguish increases, decreases, and stable tendencies, though exact values exhibit errors if more than one decimal of precision is desired. The difference in the area between the models and ground truth is significant for some classes, with values over 15%. However, as the ground truth can have missing data, overestimations in area might actually be correct. Unfortunately, it was not

possible to provide exact estimates for the amount of over- or under-detection for the classes with the worst performance. To achieve this, a small dataset with manual labels could be created, but this is a very time-consuming task and is not desirable for the objectives of this study.

Overall, evaluating the precision of the models remains a challenge, as ground truth data is inaccurate. For certain classes, the models' results are accurate, but the ground truth is wrong, as shown in figures 8 and 9. This issue makes testing results appear worse than those of other research [43] (which reported IoU values around 0.6), where the test and train datasets were refined with extensive manual work. Image resolution appears to be an influential factor, as the Madrid-trained models with images at 10 cm per pixel achieved much better results than the Vienna-trained models with images at 15 cm per pixel.

OSM data proved to be valuable in creating datasets covering topics not included in official data, such as private parking surfaces in Vienna. However, evaluating the model remains a challenge. Without a high-quality test set, it is impossible to determine the validity of the model. Therefore, if OSM data is to be used, manual labeling for a small test set would be necessary to establish the model's validity.

The transfer learning approach provided better results with less training epochs than a non-transfer learning approach.

6 Conclusions

The study has demonstrated the feasibility of creating training dataset for machine learning vision models employing openly available data as ground truth without the need for human manual labeling. The process is fast and easily adaptable to other regions, tasks or data. It is feasible to generate datasets for less common tasks, such as segmenting parking surfaces, given the availability of data. While OpenStreetMap (OSM) serves as a valuable source for ground truth in areas lacking comprehensive cartography, its incompleteness and errors can impede precise model outcomes in some instances. Despite inherent challenges related to the accuracy and completeness of ground truth data producing discrepancies between model predictions and ground truth, our models have shown sufficient accuracy and adaptability across diverse urban landscapes.

Validating models trained with incomplete data remains an open challenge. However, one potential solution is to test these models in a different city from where the training data was collected. This approach was applied in the case of parking models for Vienna and Madrid.

General purpose semantic segmentation models like SAM may lack precision for non-common tasks or images beyond the Western world. Therefore, the methodology developed in this study, using transfer learning with models like SAM, could contribute to the development of a larger amount of models suitable for a wider range of tasks and diverse environments. Transformer-based models, when employing pretrained encoders, outperform convolutional-based models with less training.

It was possible to detect temporal trends differentiating the tendencies in different areas in the city, contributing to the advancement of urban analytics. This research could help city administrations in assessing whether their policies to reduce car usage have effectively altered the city's land use and road and parking surface.

6.1 Further research

To enhance the dataset, the utilization of semi-supervised methods [86] could be employed, as these methods can help correct errors within the training dataset. In the training workflow examples where the models prediction differs vastly from the ground truth can be automatically excluded if the models prediction probability is high. Self and semi-supervised methods specifically developed for aerial image semantic segmentation can be explores to implement such workflows [87]. Accounting for class uncertainties during the labeling process can be factored into establishing weights in the loss function [88]. In order to give exact estimates for the models accuracy a small test set could be manually labelled if resources are available.

The models predictions could be used by official agencies to correct wrong data or manually supervise data where model and ground truth differ. Advancements in machine learning technology could motivate official agencies to establish more clear and accurate criteria when elaborating their cartography.

Acknowledgments

Universidad Politécnica de Madrid (www.upm.es) provided computing resources on the Magerit Supercomputer.

Miguel Ureña is supported by a contract funded by the Industrial Doctorates of the Community of Madrid (IND2020/TIC-17528 and IND2023/TIC-28743).

We want to thank Javier Sempere for proofreading this article.

Code

The code developed for this paper is accessible on github:

Dataset creator and downloader: `https://github.com/GeomaticsCaminosUPM/GeoVisionDataset`

Machine learning model: `https://github.com/GeomaticsCaminosUPM/GeoVisionModels`

An example notebook to create a dataset: `https://github.com/GeomaticsCaminosUPM/GeoVisionDataset/blob/main/examples/vienna_transport_dataset.ipynb`

A notebook to try the SAM model: `https://github.com/GeomaticsCaminosUPM/GeoVisionModels/blob/main/examples/sam_text_based_segmentation_example.ipynb`

References

[1] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717302900

[2] M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, Oct. 2008. [Online]. Available: https://ieeexplore.ieee.org/document/4653466

[3] "Directive 2007/2/EC," Mar. 2007. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2007/2/oj

[4] M. Minghini, A. Kotsev, and M. Lutz, "COMPARING INSPIRE AND OPENSTREETMAP DATA: HOW TO MAKE THE MOST OUT OF THE TWO WORLDS," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4/W14, pp. 167–174, Aug. 2019. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLII-4-W14/167/2019/

[5] "City of London Web Mapping." [Online]. Available: https://www.mapping.cityoflondon.gov.uk/geocortex/mapping/

[6] "Stadtvermessung Wien (MA 41)." [Online]. Available: https://www.wien.gv.at/stadtentwicklung/stadtvermessung/index.html

[7] P. C. Council, "Paris map - Maps Paris (Île-de-France - France)." [Online]. Available: http://maps-paris.com/

[8] "Luftbilder Berlin." [Online]. Available: https://daten.berlin.de/anwendungen/luftbilderberlin

[9] Ayuntamiento de Madrid, "Geoportal." [Online]. Available: https://geoportal.madrid.es/

[10] B. Herfort, S. Lautenbach, J. Porto De Albuquerque, J. Anderson, and A. Zipf, "The evolution of humanitarian mapping within the OpenStreetMap community," *Scientific Reports*, vol. 11, no. 1, p. 3037, Feb. 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-82404-z

[11] M. Minghini and F. Frassinelli, "OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date?" *Open Geospatial Data, Software and Standards*, vol. 4, no. 1, p. 9, Sep. 2019. [Online]. Available: https://doi.org/10.1186/s40965-019-0067-x

[12] M. Haklay, "How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets," *Environment and Planning B: Planning and Design*, vol. 37, no. 4, pp. 682–703, Aug. 2010. [Online]. Available: http://journals.sagepub.com/doi/10.1068/b35097

[13] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10028728/

[14] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications*, vol. 169, p. 114417, May 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417420310836

[15] K. Nurul Habib, C. Morency, and M. Trépanier, "Integrating Parking Behaviour in Activity-Based Travel Demand Modelling: Investigation of the Relationship between Parking Type Choice and Activity Scheduling Process," *Transportation Research Part A*, vol. 46, pp. 154–166, Jan. 2012.

[16] J. B. Mena, "State of the art on automatic road extraction for GIS update: a novel classification," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3037–3058, Dec. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865503001648

[17] S. W. Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. S. Khan, F. Zhu, L. Shao, G.-S. Xia, and X. Bai, "iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images," Aug. 2019. [Online]. Available: http://arxiv.org/abs/1905.12886

[18] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. Valencia: IEEE, Jul. 2018, pp. 204–207. [Online]. Available: https://ieeexplore.ieee.org/document/8519248/

[19] V. Mnih, "Machine Learning for Aerial Image Labeling," phd, University of Toronto, Toronto, Canada, 2013. [Online]. Available: https://www.cs.toronto.edu/~vmnih/docs/Mnih_Volodymyr_PhD_Thesis.pdf

[20] D. Marmanis, J. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla, "SEMANTIC SEGMENTATION OF AERIAL IMAGES WITH AN ENSEMBLE OF CNNS," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 473–480, Jun. 2016.

[21] S. Talukdar, P. Singha, S. Mahato, Shahfahad, S. Pal, Y.-A. Liou, and A. Rahman, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, p. 1135, Jan. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/7/1135

[22] V. Nasiri, A. Deljouei, F. Moradi, S. M. M. Sadeghi, and S. A. Borz, "Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods," *Remote Sensing*, vol. 14, no. 9, p. 1977, Apr. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/9/1977

[23] Q. Zhu, C. Liao, H. Hu, X. Mei, and H. Li, "MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction From Remote Sensed Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6169–6181, Jul. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9212557/

[24] X. Dong, J. Cao, and W. Zhao, "A review of research on remote sensing images shadow detection and application to building extraction," *European Journal of Remote Sensing*, vol. 57, no. 1, p. 2293163, Dec. 2024. [Online]. Available: https://doi.org/10.1080/22797254.2023.2293163

[25] Z. Shao, P. Tang, Z. Wang, N. Saleem, S. Yam, and C. Sommai, "BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images," *Remote Sensing*, vol. 12, no. 6, p. 1050, Mar. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/6/1050

[26] P. Liu, X. Liu, M. Liu, Q. Shi, J. Yang, X. Xu, and Y. Zhang, "Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network," *Remote Sensing*, vol. 11, no. 7, p. 830, Apr. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/7/830

[27] G. Máttyus, W. Luo, and R. Urtasun, "DeepRoadMapper: Extracting Road Topology from Aerial Images," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3458–3466. [Online]. Available: https://ieeexplore.ieee.org/document/8237634

[28] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, May 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8309343

[29] J. Jiao, "Machine Learning Assisted High-Definition Map Creation," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. Tokyo, Japan: IEEE, Jul. 2018, pp. 367–373. [Online]. Available: https://ieeexplore.ieee.org/document/8377682/

[30] S. Abdurakhmonov, K. Bekanov, S. Ochilov, S. Tukhtamishev, and Y. Karimov, "Advances in cartography: a review on employed methods," *E3S Web of Conferences*, vol. 389, p. 03057, 2023. [Online]. Available: https://www.e3s-conferences.org/10.1051/e3sconf/202338903057

[31] J. E. Vargas-Munoz, S. Srivastava, D. Tuia, and A. X. Falcao, "OpenStreetMap: Challenges and Opportunities in Machine Learning and Remote Sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 1, pp. 184–199, Mar. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9119753/

[32] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2017,

pp. 3226–3229. [Online]. Available: https://ieeexplore.ieee.org/document/8127684/footnotes#footnotes-id-fn2

[33] Q. Wu and L. P. Osco, "samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM)," *Journal of Open Source Software*, vol. 8, no. 89, p. 5663, Sep. 2023. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.05663

[34] P. Aszkowski, B. Ptak, M. Kraft, D. Pieczyński, and P. Drapikowski, "Deepness: Deep neural remote sensing plugin for QGIS," *SoftwareX*, vol. 23, p. 101495, Jul. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352711023001917

[35] Y. Hu, X. Huang, J. Li, and Z. Zhang, "GBSS:a global building semantic segmentation dataset for large-scale remote sensing building extraction," Jan. 2024. [Online]. Available: http://arxiv.org/abs/2401.01178

[36] Microsoft, "GlobalMLBuildingFootprints," Jan. 2024. [Online]. Available: https://github.com/microsoft/GlobalMLBuildingFootprints

[37] J. J. Gonzales, "Building-Level Comparison of Microsoft and Google Open Building Footprints Datasets (Short Paper)," *LIPIcs, Volume 277, GIScience 2023*, vol. 277, pp. 35:1–35:6, 2023, tex.copyright: Creative Commons Attribution 4.0 International license, info:eu-repo/semantics/openAccess. [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.GIScience.2023.35

[38] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep Learning for Decentralized Parking Lot Occupancy Detection," *Expert Systems with Applications*, vol. 72, Oct. 2016.

[39] N. Audebert, B. L. Saux, and S. Lefevre, "Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images," *Remote Sensing*, vol. 9, no. 4, 2017, tex.copyright: Copyright MDPI AG 2017. [Online]. Available: https://www.proquest.com/docview/1905789870/abstract/39EE60F6089043F1PQ/1

[40] S. Kumar, A. Jain, S. Rani, H. Alshazly, S. Idris, and S. Bourouis, "Deep Neural Network Based Vehicle Detection and Classification of Aerial Images," *Intelligent Automation & Soft Computing*, vol. 34, no. 1, pp. 119–131, 2022. [Online]. Available: https://www.techscience.com/iasc/v34n1/47359

[41] N. Merkle, S. M. Azimi, S. Pless, and F. Kurz, "Semantic Vehicle Segmentation in Very High Resolution Multispectral Aerial Images Using Deep Neural Networks," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5045–5048, Jul. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8898513/

[42] S. Drouyer, "Parking Occupancy Estimation on Planetscope Satellite Images," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa, HI, USA: IEEE, Sep. 2020, pp. 1098–1101. [Online]. Available: https://ieeexplore.ieee.org/document/9323104/

[43] C. Henry, J. Hellekes, N. Merkle, S. M. Azimi, and F. Kurz, "CITYWIDE ESTIMATION OF PARKING SPACE USING AERIAL IMAGERY AND OSM DATA FUSION WITH DEEP LEARNING AND FINE-GRAINED ANNOTATION," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B2-2021, pp. 479–485, Jun. 2021. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLIII-B2-2021/479/2021/

[44] J. Hellekes, A. Kehlbacher, M. L. Díaz, N. Merkle, C. Henry, F. Kurz, and M. Heinrichs, "Parking space inventory from above: Detection on aerial images and estimation for unobserved regions," *IET Intelligent Transport Systems*, vol. 17, no. 5, pp. 1009–1021, 2023, tex.copyright: © 2022 The Authors. IET Intelligent Transport Systems published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1049/itr2.12322

[45] W. Abrahamse, L. Steg, R. Gifford, and C. Vlek, "Factors influencing car use for commuting and the intention to reduce it: A question of self-interest or morality?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 4, pp. 317–324, Jul. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847809000230

[46] A. Carse, A. Goodman, R. L. Mackett, J. Panter, and D. Ogilvie, "The factors influencing car use in a cycle-friendly city: the case of Cambridge," *Journal of Transport Geography*, vol. 28, no. 100, pp. 67–74, Apr. 2013, tex.pmcid: PMC4060748.

[47] G. Mattioli, C. Roberts, J. K. Steinberger, and A. Brown, "The political economy of car dependence: A systems of provision approach," *Energy Research & Social Science*, vol. 66, p. 101486, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214629620300633

[48] B. P. Feeney, "A review of the impact of parking policy measures on travel demand," *Transportation Planning and Technology*, vol. 13, no. 4, pp. 229–244, Apr. 1989. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03081068908717403

[49] P. Christiansen, O. Engebretsen, N. Fearnley, and J. Usterud Hanssen, "Parking facilities and the built environment: Impacts on travel behaviour," *Transportation Research Part A: Policy and Practice*, vol. 95, pp. 198–206, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0965856416301525

[50] C. T. McCahill, N. Garrick, C. Atkinson-Palombo, and A. Polinski, "Effects of Parking Provision on Automobile Use in Cities: Inferring Causality," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2543, no. 1, pp. 159–165, Jan. 2016. [Online]. Available: http://journals.sagepub.com/doi/10.3141/2543-19

[51] A. Pandhe and A. March, "Parking availability influences on travel mode: Melbourne CBD offices," *Australian Planner*, vol. 49, no. 2, pp. 161–171, Jun. 2012. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07293682.2011.616177

[52] Tom Rye, Susan Tully, Glenn Godin, Niklas Schmalholz, and Martina Hertel, "Parking and SUMP. Using parking management to achieve your SUMP objectives effectively and sustainably," Aug. 2022. [Online]. Available: https://urban-mobility-observatory.transport.ec.europa.eu/document/download/4ae9e061-dfc9-4f92-afe2-30d70de76580_en?filename=parking_and_sump.pdf&prefLang=it

[53] T. Litman, "Why and how to reduce the amount of land paved for roads and parking facilities," *Environmental Practice*, vol. 13, pp. 38–46, Apr. 2011.

[54] R. Gerike, C. Koszowski, B. Schröter, R. Buehler, P. Schepers, J. Weber, R. Wittwer, and P. Jones, "Built Environment Determinants of Pedestrian Activities and Their Consideration in Urban Street Design," *Sustainability*, vol. 13, no. 16, p. 9362, Jan. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/16/9362

[55] T. O. Alshammari, A. M. Hassan, Y. Arab, H. Hussein, F. Khozaei, M. Saeed, B. Ahmed, M. Zghaibeh, W. Beitelmal, and H. Lee, "The Compactness of Non-Compacted Urban Developments: A Critical Review on Sustainable Approaches to Automobility and Urban Sprawl," *Sustainability*, vol. 14, no. 18, p. 11121, Jan. 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/18/11121

[56] E. L. Glaeser and M. E. Kahn, "Chapter 56 - Sprawl and Urban Growth," in *Handbook of Regional and Urban Economics*, ser. Cities and Geography, J. V. Henderson and J.-F. Thisse, Eds. Elsevier, Jan. 2004, vol. 4, pp. 2481–2527. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574008004800130

[57] C. Halpern, "Road Space Reallocation for Sustainable Urban Mobility in the EU," *LIEPP Policy Brief*, vol. 60, Jun. 2022. [Online]. Available: https://sciencespo.hal.science/hal-03701457/document

[58] A. Rodriguez-Valencia and H. A. Ortiz-Ramirez, "Understanding Green Street Design: Evidence from Three Cases in the U.S," *Sustainability*, vol. 13, p. 1916, Feb. 2021.

[59] B. Schröter, S. Hantschel, C. Koszowski, R. Buehler, P. Schepers, J. Weber, R. Wittwer, and R. Gerike, "Guidance and Practice in Planning Cycling Facilities in Europe—An Overview," *Sustainability*, vol. 13, no. 17, p. 9560, Jan. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/17/9560

[60] Fabian Küster, "Practitioner Briefings: Cycling. Supporting and encouraging cycling in Sustainable Urban Mobility Planning," Sep. 2019. [Online]. Available: https://urban-mobility-observatory.transport.ec.europa.eu/document/download/ea316d2f-7155-4297-b673-11a514726d53_en?filename=supporting_and_encouraging_cycling_in_sumps.pdf&prefLang=it

[61] Jim Walker, Bronwen Thornton, and Lina Marcela Quiñones, "Supporting and Encouraging Walking in Sustainable Urban Mobility Planning," Oct. 2019. [Online]. Available: https://urban-mobility-observatory.transport.ec.

europa.eu/document/download/6c00c382-42a9-4cd8-9327-33c0cfbbc345_en?filename=supporting_and_encouraging_walking_in_sumps.pdf&prefLang=it

[62] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Mar. 2015. [Online]. Available: http://arxiv.org/abs/1411.4038

[63] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," Oct. 2016. [Online]. Available: http://arxiv.org/abs/1511.00561

[64] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Long Beach, CA, USA, Aug. 2023. [Online]. Available: http://arxiv.org/abs/1706.03762

[65] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[66] X. Gu, Y. Cui, J. Huang, A. Rashwan, X. Yang, X. Zhou, G. Ghiasi, W. Kuo, H. Chen, L.-C. Chen, and D. A. Ross, "DaTaSeg: Taming a Universal Multi-Dataset Multi-Task Segmentation Model," Jun. 2023. [Online]. Available: http://arxiv.org/abs/2306.01736

[67] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," Apr. 2023. [Online]. Available: http://arxiv.org/abs/2304.02643

[68] L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Li, and J. M. Junior, "The Segment Anything Model (SAM) for Remote Sensing Applications: From Zero to One Shot," Oct. 2023. [Online]. Available: http://arxiv.org/abs/2306.16623

[69] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9134370/

[70] H. Cao, H. Gu, and X. Guo, "Feasibility of Transfer Learning: A Mathematical Framework," May 2023. [Online]. Available: http://arxiv.org/abs/2305.12985

[71] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, Jun. 2022, pp. 15 979–15 988. [Online]. Available: https://ieeexplore.ieee.org/document/9879206/

[72] W. Falcon and T. P. L. team, "PyTorch Lightning," Dec. 2023. [Online]. Available: https://zenodo.org/records/10419201

[73] P. Iakubovskii, "segmentation_models_pytorch," 2019, tex.copyright: MIT. [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch

[74] R. Azad, M. Heidary, K. Yilmaz, M. Hüttemann, S. Karimijafarbigloo, Y. Wu, A. Schmeink, and D. Merhof, "Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook," Dec. 2023. [Online]. Available: http://arxiv.org/abs/2312.05391

[75] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611121001750

[76] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and t. s.-i. contributors, "scikit-image: Image processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014. [Online]. Available: http://arxiv.org/abs/1407.6245

[77] N. Jamil, T. M. T. Sembok, and Z. A. Bakar, "Noise removal and enhancement of binary images using morphological operations," in *2008 International Symposium on Information Technology*. Kuala Lumpur, Malaysia: IEEE, 2008, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/4631954/

[78] J. Serra, "Introduction to mathematical morphology," *Computer Vision, Graphics, and Image Processing*, vol. 35, no. 3, pp. 283–305, Sep. 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0734189X86900022

[79] R. Srisha and A. Khan, "Morphological Operations for Image Processing : Understanding and its Applications," Dec. 2013. [Online]. Available: https://www.researchgate.net/publication/272484795_Morphological_Operations_for_Image_Processing_Understanding_and_its_Applications

[80] "USGS (United States Geological Survey)." [Online]. Available: https://www.usgs.gov/

[81] "Copernicus." [Online]. Available: https://copernicus.eu

[82] DigitalGlobe, "WorldView-3." [Online]. Available: https://worldview3.digitalglobe.com/

[83] "Google Earth." [Online]. Available: https://earth.google.com/

[84] P. D. Maidment, B. Domenico, A. Gemmell, K. Lehnert, D. Tarboton, and I. Zaslavsky, "The Open Geospatial Consortium and EarthCube," Oct. 2011. [Online]. Available: https://www.ogc.org/standards/technical-papers/

[85] S. Gillies and others, "Rasterio: geospatial raster I/O for Python programmers," 2013. [Online]. Available: https://github.com/mapbox/rasterio

[86] A. Peláez-Vegas, P. Mesejo, and J. Luengo, "A Survey on Semi-Supervised Semantic Segmentation," Feb. 2023. [Online]. Available: http://arxiv.org/abs/2302.09899

[87] M. Tang, K. Georgiou, H. Qi, C. Champion, and M. Bosch, "Semantic Segmentation in Aerial Imagery Using Multi-level Contrastive Learning with Local Consistency," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE, Jan. 2023, pp. 3787–3796. [Online]. Available: https://ieeexplore.ieee.org/document/10030272/

[88] P. O. Bressan, J. M. Junior, J. A. Correa Martins, M. J. De Melo, D. N. Gonçalves, D. M. Freitas, A. P. Marques Ramos, M. T. Garcia Furuya, L. P. Osco, J. De Andrade Silva, Z. Luo, R. C. Garcia, L. Ma, J. Li, and W. N. Gonçalves, "Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 108, p. 102690, Apr. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243422000162