Remote Sensing Applications: Society and Environment 37 (2025) 101503

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Transport-related surface detection with machine learning: Analyzing temporal trends in Madrid and Vienna

Miguel Ureña Pliego [a,*], Rubén Martínez Marín [a], Nianfang Shi [b], Takeru Shibayama [b], Ulrich Leth [b], Miguel Marchamalo Sacristán [a]

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

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

ARTICLE INFO

Keywords:
Image vision
OSM
Parking
Geospatial dataset
Foundational models
Sustainable urban transport

ABSTRACT

This study investigates the integration of machine learning into urban aerial image analysis, focusing on identifying mobility-related surfaces for cars and pedestrians, such as parking spaces, road surfaces, and sidewalks, while also analyzing historical trends. It highlights the potential of foundational models and the advantages of fine-tuning them for global geospatial analysis. A workflow is proposed for automatically generating geospatial datasets, facilitating the creation of semantic segmentation datasets from diverse sources, including WMS/WMTS links, vector-based cartography, and OpenStreetMap (OSM) overpass-turbo requests. The developed code streamlines dataset generation for training machine learning models using publicly available data, eliminating the need for manual labeling.

Two datasets for car- and pedestrian-related surface detection were generated using aerial imagery and vector data from the geographical offices of Madrid and Vienna, with a particular focus on parking surfaces. A transformer-based model was trained and evaluated for each city, achieving strong accuracy with F1 scores exceeding 0.5 in most classes in just less than 10 epochs. Historical trend analysis was conducted by applying the trained model to images from 10 to 20 years ago, prior to the availability of vector data, successfully uncovering temporal infrastructure trends for cars and pedestrians in different city areas. The trends discovered in Madrid and Vienna showed distinct differences: Madrid's periphery is experiencing a slower growth in parking surface, while Vienna is seeing a greater increase in parking compared to newly built road surfaces.

This method offers a cost-effective solution for municipal governments to gather valuable urban data.

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 (Gorelick et al., 2017) 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 (Haklay and Weber, 2008). However, this manual process is labor-intensive, resulting

* Corresponding author.
E-mail address: miguel.urena@upm.es (M.U. Pliego).

https://doi.org/10.1016/j.rsase.2025.101503
Received 5 August 2024; Received in revised form 22 December 2024; Accepted 24 February 2025
Available online 6 March 2025

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 (EC, 2007; Minghini et al., 2019). 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 (City, 2025; Stadtvermessung, 2025; Paris City Council, 2025; Luftbilder, 2025; Ayuntamiento de Madrid, 2025).

OSM, as it is maintained by volunteers, does not follow the INSPIRE scheme but offers diverse geospatial data based on local needs (Minghini et al., 2019). OSM enables collaboration for rapid cartography, especially useful for humanitarian mapping post-disasters (Herfort et al., 2021). OSM has 88 931 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 (Minghini and Frassinelli, 2019; Haklay, 2010).

Classical and machine learning-enhanced remote sensing methods utilizing multispectral imagery (Sharifi and Hosseingholizadeh, 2020; Sharifi and Felegari, 2022; Nejad et al., 2023; Esmaeili et al., 2023; Jalayer et al., 2023) are insufficient for identifying parking surfaces. These techniques primarily classify surface materials, textures, and patterns, which often make parking areas indistinguishable from roads. Since parking spaces are typically inferred through additional contextual information, a specialized machine learning approach is necessary. This approach should focus on features such as road markings, texture variations, vehicle positions, and other contextual indicators. Another advantage of utilizing optical imagery and machine learning is the availability of high-quality imagery within the optical spectrum.

Advancements in computer vision and object detection (Zou et al., 2023) have enabled significant progress in semantic segmentation and object detection for aerial imagery, benefiting a wide range of fields (Yuan et al., 2021). Until recently, vision models were predominantly based on convolutional neural networks and were trained for specific tasks using topic-specific datasets with a fixed number of classes (Thisanke et al., 2023; Maurício et al., 2023). The limitations of these models included their difficulty in handling images significantly different from the training context, an issue that domain adaptation techniques sought to address (Singhal et al., 2023), as well as their inflexibility in labeling schemes, where introducing new labels was not straightforward.

In contrast, recent transformer-based vision models, known as foundational models (Awais et al., 2023), are trained on extensive and diverse datasets sourced from the internet. These models can perform a broad range of tasks, such as general object segmentation, across various types of images. Models such as SAM (Kirillov et al., 2023) incorporate text-processing capabilities to interpret image labels and are designed to handle an unfixed number of classes. Moreover, they can potentially identify new labels without additional training, provided these labels are sufficiently similar to those in the training dataset.

However, these models often lack precision when applied to tasks that differ significantly from their training data, such as analyzing aerial imagery or segmenting uncommon elements like parking surfaces. With the increasing availability of high-resolution aerial imagery and detailed cartography, it is essential to establish automated workflows for creating specialized training datasets using open data and developing fine-tuning methodologies to adapt foundational models for specific tasks and diverse environments. Fine-tuning transformer-based vision models has already been explored in other contexts (Sandler et al., 2022).

Analyzing surfaces dedicated to cars and pedestrians is crucial for evaluating car use in neighborhoods (Nurul Habib et al., 2012). Although government and OSM data exist, they often lack quality and historical data are often not available. Image vision models can address these gaps. Other approaches rely on openly available data combined with an intensive manual labeling post-process to create datasets for training convolutional vision models specifically targeted at this task (Henry et al., 2021; Hellekes et al., 2023). In this study, we demonstrate the flexibility of newly available foundational models to achieve similar results without the need for any manual labeling in the creation of training data.

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 demonstrate 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 (Zou et al., 2023). In the earlier stages, traditional statistical methods were prevalent for tasks like automatic road extraction (Mena, 2003). However, machine learning methods have surpassed traditional approaches in recent years (Zou et al., 2023). Various datasets and benchmarks for semantic segmentation tasks related to building detection, land use and cover extraction, and road geometry are available (Zamir et al., 2019; Helber et al., 2018; Mnih, 2013; Marmanis et al., 2016).

Machine learning models working with aerial imagery in the optical spectrum have been developed for diverse applications, including land use detection (Talukdar et al., 2020; Nasiri et al., 2022), building detection (Zhu et al., 2021; Dong et al., 2024; Shao et al., 2020; Liu et al., 2019), urban mapping (Mohammadi and Sharifi, 2021), road geometry extraction (Máttyus et al., 2017; Zhang et al., 2018).

In most cases, these machine learning models are based on image convolutions. They play a vital role in annotating maps (Jiao, 2018; Abdurakhmonov et al., 2023) 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 (Vargas-Munoz et al., 2021).

Researchers have explored the labeling and creation of more diverse datasets to enhance the generability of aerial imagery segmentation models across diverse global environments (Maggiori et al., 2017). 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 (Wu and Osco, 2023). A commercially available plugin for QGIS is now accessible (Aszkowski et al., 2023) to run such models directly.

Datasets like GBSS (Hu et al., 2024) by Google and ML Building Footprints by Microsoft (Microsoft, 2024) 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 (Gonzales, 2023), 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 (Amato et al., 2016) and UAV high-resolution images, with resolutions ranging from 5 to 1.5 cm per pixel (Audebert et al., 2017; Kumar et al., 2022) and multispectral high resolution aerial imagery (Merkle et al., 2019). 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 (Drouyer, 2020). 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 (Henry et al., 2021; Hellekes et al., 2023), 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. The generability of the model was not evaluated, so applications in areas different from the training set could not be 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 (Abrahamse et al., 2009; Carse et al., 2013) 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 (Abrahamse et al., 2009).

The amount of car-related infrastructure is therefore one of the main factors encouraging car use and diverting financing away from more sustainable alternatives (Mattioli et al., 2020). Parking space availability, in particular, has long been recognized as one of the most critical factors influencing the choice of travel by car (Feeney, 1989; Christiansen et al., 2017; McCahill et al., 2016; Pandhe and March, 2012). This is acknowledged by the European Commission in its technical guide for parking policy (Rye et al., 2022). Parking space availability in urban environments can be indirectly inferred from the amount of paved surface (Litman, 2011) 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 (Gerike et al., 2021). 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 (Gerike et al., 2021). In new urban developments, dedicating too much surface to cars makes distances and destinations farther apart, encouraging urban sprawl and car use (Alshammari et al., 2022; Glaeser and Kahn, 2004). 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 (Halpern, 2022; Rodriguez-Valencia and Ortiz-Ramirez, 2021; Schröter et al., 2021). 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) (Küster, 2019; Walker et al., 2019).

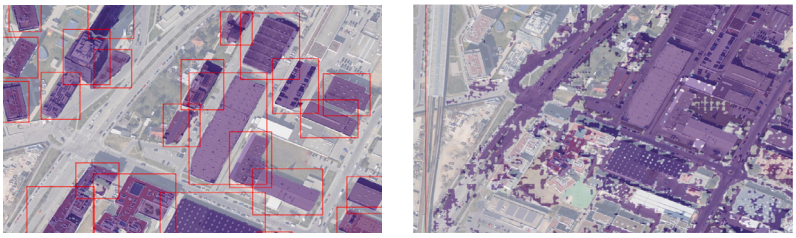
Therefore, the main interests for the purpose of this study are pedestrian spaces, road surfaces, and car parking surfaces.

2.2. Semantic segmentation

Semantic segmentation consist in assigning a label to each pixel of an image. The output are raster values and the technique is not suitable for producing vectorial geometries. The most widely used deep learning models for image segmentation, fully convolutional neural networks (Long et al., 2015), employ image convolutions where the parameters of the kernel are training parameters. Semantic segmentation models usually follow a encoder–decoder structure (Badrinarayanan et al., 2016).

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 (Vaswani et al., 2023). For image related tasks a vision transformer was developed (Dosovitskiy et al., 2020)

The availability of large datasets combined with the computational power of transformer models has led to a new trend: the development of foundational models (Awais et al., 2023). These are models trained on massive datasets using computational



(a) SAM model results when segmenting building roofs.

(b) SAM model results when segmenting parking spaces.

Fig. 1. The SAM model is able to fulfill only more general and common tasks.

resources typically accessible only to the largest technology companies, designed for general or multi-purpose tasks. Two notable examples of foundational models in the field of image vision are the DaTaSeg model (Gu et al., 2023) and the Segment Anything Model (SAM) (Kirillov et al., 2023). These models enable automatic segmentation of objects and instances, without being tied to a specific topic or dataset. They also facilitate the creation of semi-supervised and coarsely annotated datasets. 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 (Wu and Osco, 2023) 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 (Kirillov et al., 2023) using information such as alt texts to train text prompts, 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 (Osco et al., 2023), but for a specific task like detecting parking spaces it fails completely [Fig. 1].

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 (Zhuang et al., 2021). However, it is important to note that this assumption has been empirically proven for years but the mathematical background is still under development (Cao et al., 2023).

3. Methodology

3.1. Machine learning implementation details

3.1.1. The model

This paper aims to develop a practical machine learning vision workflow that delivers high-quality results without requiring training on a large compute cluster. The approach developed in this paper is to fine-tune a pre-trained foundational model using a task-specific dataset. Among the top-performing models currently available is the SAM model (Kirillov et al., 2023). The pre-trained image encoder of the SAM model was employed, with only a new convolutional decoder being trained in accordance with a U-Net like structure [Fig. 2]. The convolutional decoder consists of four layers with channel depths of 256, 128, 64, and 32, respectively. It is integrated with a pre-trained image encoder from the SAM base model, which also has a depth of four. To enhance feature integration, four skip connections link corresponding layers of the image encoder and the decoder. The image encoder is a vision transformer, with its weights initialized using the publicly available checkpoint vit-b from a model trained with masked autoencoders on the SA-1B dataset (He et al., 2022; Kirillov et al., 2023), specifically for images with a resolution of 1024×1024 pixels. The implementation utilizes the pytorch-lightning (Falcon and team, 2023) and segmentation models (Iakubovskii, 2019) Python libraries.

Loss functions for semantic segmentation tasks can be categorized into pixel, boundary, and region-level loss functions (Azad et al., 2023). 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 (Yeung et al., 2022) is an adequate 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.

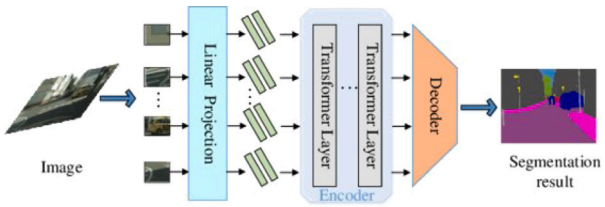


Fig. 2. Transformer-based model (Dosovitskiy et al., 2020).

3.1.2. Segmentation mask cleaning

Segmentation results are cleaned using morphological operations from the `scikit-image` library (van der Walt et al., 2014). 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 (Jamil et al., 2008; Serra, 1986). A round kernel rounds mask edges, a rectangular kernel creates rectangular shapes, and an octagon kernel sharpens edges optimally without limiting to rectangles (Srisha and Khan, 2013).

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].

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 Fig. 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.

(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.

Fig. 3. Dataset grid example.

(a) Sentinel 2 10 m resolution image in Vienna (92).

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

Fig. 4. The importance of image resolution.

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-m resolution and an 8-day revisit time (USGS, 2025), and ESA's Sentinel 2 providing 10-m resolution with a 5-day revisit time (Copernicus, 2025). 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 (DigitalGlobe, 2025).

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 m 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.

Google Earth (Google, 2025) offers global aerial photography with resolutions from 15 m 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 (City, 2025; Stadtvermessung, 2025; Paris City Council, 2025; Luftbilder, 2025; Ayuntamiento de Madrid, 2025). These datasets are available via cloud services following Open Geospatial Consortium standards (Maidment et al., 2011). 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].

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].

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 (Gillies et al., 2013).

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.

(a) 1.5 cm/pix UAV image. (b) 10 cm/pix Aerial image. (c) 30 cm/pix Satellite (World-View3) image.

Fig. 5. Remaining distortion in orthoimages of a building in Madrid (Ayuntamiento de Madrid, 2025) taken by UAV (least distortion), plane, and satellite (most distortion).

(a) Vienna winter image 2021 (94). (b) Vienna summer image 2022 (6).

Fig. 6. Issues with different image providers.

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 [Fig. 10(a)]. 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 80 s 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 1024 × 1024 pixels) encompasses all tiles contained inside the official boundaries of the following neighborhoods 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.

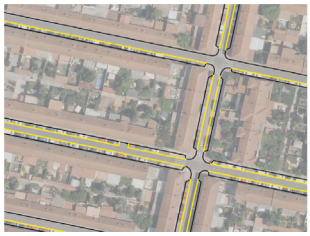
The test dataset (1238 images and masks with a size of 1024 × 1024 pixels) encompasses all tiles contained inside the official boundaries of the following neighborhoods of Madrid. It is used exclusively for evaluation: Gaztambide, Cuatro Caminos, Pilar, Arcos and San Diego (see Fig. 7).

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

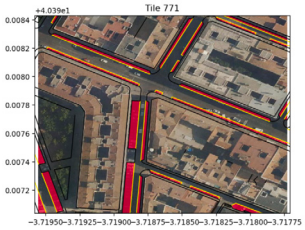
The inaccuracies found in the Madrid dataset are especially related to unclear criteria to classify data and data that is not up to date even though the 2023 cartography was used. The Fig. 8 shows some examples of the problems found in the dataset and the models output.

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 urbanization with newer developments and even single family homes [Fig. 12(a)].



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

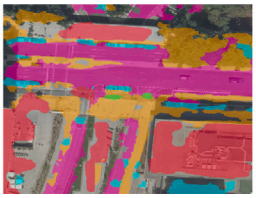


(b) Tile 771 of the dataset. In red the parking polygons in raster format created using the available line geometry.

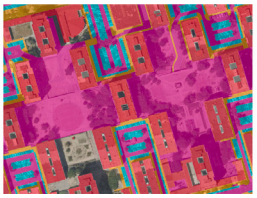
(c) High resolution historic image from 1941 (9).

Fig. 7. Madrid dataset details.

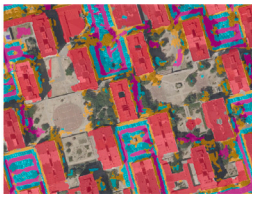
(a) Ground truth data in the test set is missing the new bike lane. Parking and sidewalk were displaced too.



(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.

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

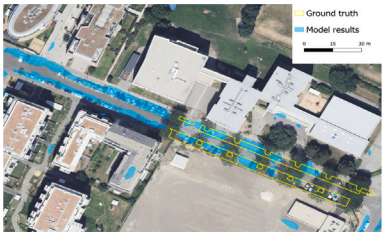
Two versions of the Vienna Dataset were created. The main version has 7 categories: background, public road, tram or train tracks, crosswalk, on street parking, private road surface, sidewalk and separated pedestrian or bicycle path. The dataset has 1826 images for training and 156 for testing all with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

Especially the OSM requests have many mistakes, as shown in Fig. 9(b). 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].

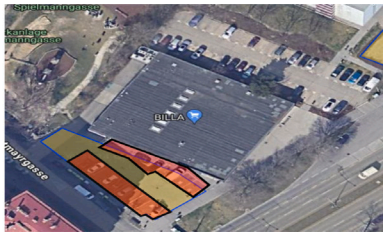
4. Results

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-cm 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).
- $\frac{\mathbf{area(model)}}{\mathbf{area(GT)}}$: 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.

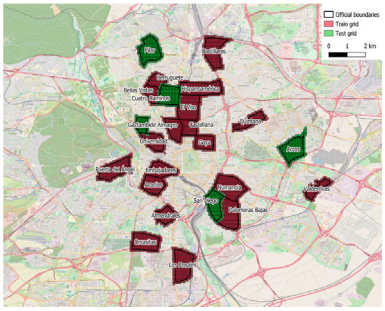


(a) Ground truth data in the test set is missing some parking spaces. The model's results correctly classify all parking spaces as parking, but it will be negatively evaluated as most of the parking is missing in the ground truth.

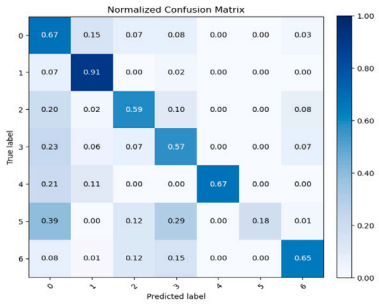


(b) Overpass-turbo request results (in yellow) are not accurate enough compared to actual parking spaces (in red), and the data is incomplete. (In the top right corner, there is an unlabeled parking area)

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



(a) Training and testing neighbourhoods.



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

Fig. 10. Madrid dataset and model results.

Table 1
Evaluation results for the Madrid test set.

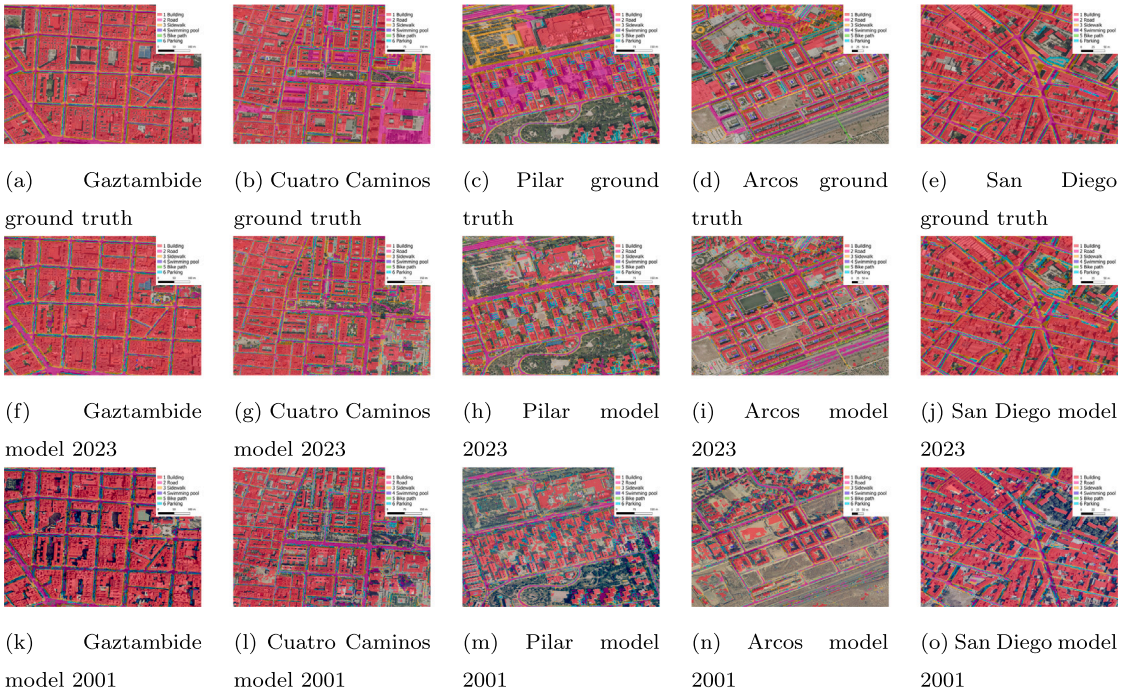
Class id	Class name	Model area	GT area	IoU	IoU_200	F1	Street ratio	Pedestrian ratio	$\frac{\text{area (model)}}{\text{area (GT)}}$
1	Building	2,769,163	2,601,307	0.73	0.82	0.86	0.01	0.02	1.07
2	Road	1,343,829	1,776,690	0.50	0.55	0.67	0.81	0.06	0.77
3	Sidewalk	1,103,500	1,127,623	0.40	0.52	0.56	0.22	0.57	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

4.1. Madrid

4.1.1. Model evaluation

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 Table 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 errors in distinguishing between infrastructure intended for vehicles and pedestrians. Most inaccuracies arise between classes within closely related categories, such as confusions between parking areas and roads or between bike paths and sidewalks.

The bike lane and swimming pool classes have the lowest accuracies [Fig. 10 and Table 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.



(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

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

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)

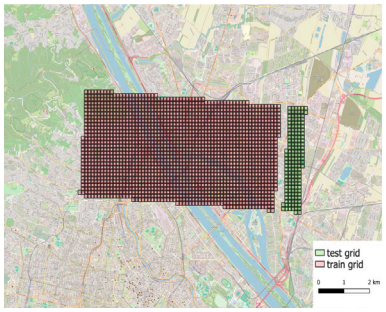
Even though there can be significant differences (over 15%) in the area picked up by the model in comparison with ground truth for some classes, the assumption is made that the mistakes or differences between ground truth and model will not occur between the results of the same model evaluated with different 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 (see Fig. 11).

4.1.2. Temporal trends (2001–2023)

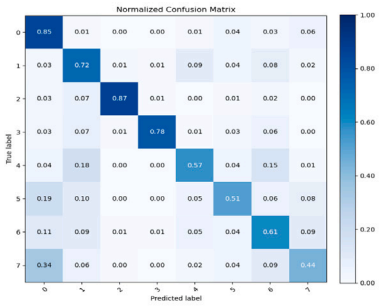
Images from the same testing neighborhoods but from the year 2001 (RGB images with 10 cm per pixel resolution) were inputted into the model that was trained with images from 2023. The model's output from 2001 was compared to the model's output from 2023, evaluated previously. The IoU values indicate 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.

In the city center [Table 2], particularly in the Gaztambide and Cuatro Caminos areas, parking spaces show a declining trend, with a reduction of approximately 20%. In contrast, the surface areas dedicated to roads and sidewalks have increased by 30%–40%. The number of buildings has remained largely stable, with a modest growth of 10%, consistent with urban development patterns in these areas over the past two decades. This slight building increase is attributed to urban regeneration projects initiated by the city council in the early 2000s.

The reduction in parking spaces reflects transformations of roads into mixed-use spaces for pedestrians and vehicles, with many parking areas removed to expand sidewalk widths. The presence of bike paths and swimming pools in these neighborhoods is negligible (less than 0.01% of the total area), making temporal trends in these features insignificant.



(a) Training and testing areas.



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

Fig. 12. Vienna dataset and model results.

Table 3
Evaluation metrics and temporal trend (2014–2023) for the Vienna model (GT: ground truth).

Class id	Class name	Model area	GT area	IoU	IoU_200 (23-GT)	F1	Street ratio	Pedestrian ratio	$\frac{\text{area (model)}}{\text{area (GT)}}$	iou_200 (14–23)	$\frac{\text{area}-23}{\text{area}-14}$
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	4781	2999	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

In the periphery of the city [Table 2], including Pilar, Arcos, and San Diego, trends indicate a notable increase in surface areas dedicated to parking, roads, and sidewalks, with approximately a 50% growth. Building expansion varies across neighborhoods, with Pilar and San Diego showing modest growth of 10%–15%, while Arcos experiences a more pronounced 22% increase. This growth aligns with the developmental timeline of each neighborhood: most buildings in Pilar and San Diego date back to before 1990, whereas Arcos, a newly developed area, has seen many empty lots from the early 2000s now built upon.

Arcos stands out with a significant 104% increase in swimming pools, reflecting the trend during its development toward multifamily housing complexes with shared pools. This contrasts with Pilar and San Diego, where swimming pool growth remains below 30%, highlighting disparities between wealthier areas like Arcos and Pilar and less affluent ones like San Diego, where pools are less common.

The construction of a bike path, the Anillo Verde Ciclista in 2006, is evident in the temporal trends for Pilar, Arcos, and San Diego. Road and sidewalk surface areas show comparable growth, generally surpassing the increase in parking spaces. This suggests that on-street parking provision is lower in newly developed areas.

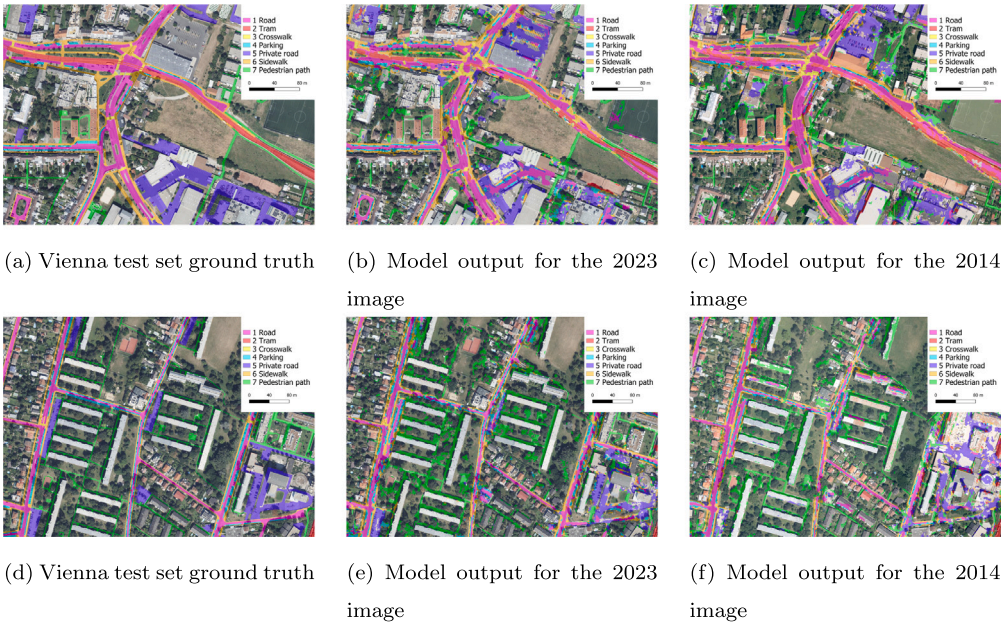
There are no significant differences in the trends in road and parking surface between richer neighborhoods, such as Pilar, and poorer neighborhoods, such as San Diego. The most notable differences are observed between the city's periphery and its center.

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 [Table 3]. IoU_200 (23-GT calculated with ground truth and model results for 2023 images) 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 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.

The model trained with 2023 imagery was tested using images from 2014, the earliest available RGB images. The outputs for both time periods are compared in Table 3 and Fig. 13. The results reveal a decline in tram tracks and private roads, while the areas dedicated to road surfaces, bike paths, and pedestrian paths remain unchanged. Sidewalks and crosswalks have seen a more significant increase than road surfaces, indicating a growing effort to enhance walkability and pedestrian-friendly environments.

The most notable growth is observed in public on-street parking, which has increased at a rate higher than that of road surfaces, highlighting the trend toward car-centric development in newer areas of Vienna. This trend is more pronounced in Vienna than in Madrid, where, unlike Vienna, on-street parking provision is not increasing.



(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

Fig. 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.

4.3. Additional performance evaluation

To further validate the transfer learning approach, a model was trained on the Vienna 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%, despite being trained with only 10 epochs compared to the UNet's 30 epochs.

Two models were trained using only the parking datasets, each with two classes. One model was trained on Madrid data and evaluated on Vienna, while the other was trained on Vienna data and evaluated on Madrid. The Madrid-trained model achieved an IoU of 0.18 and an IoU_200 of 0.31 in the Vienna test set, showing 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 this case, the decrease was more substantial (0.2), but it is important to note that the IoU_200 value for the Vienna-trained model in Vienna was 0.37. These results demonstrate the generalizability of the models and the effectiveness of the method when applied to images from cities with significantly different characteristics. Images from the same city but different time periods are generally more similar, suggesting that results would be more accurate in such cases.

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 Figs. 8 and 9. This issue makes testing results appear worse than those of other research (Henry et al., 2021) (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.

The trends observed in Madrid show an increase in walkability and the replacement of parking spaces with sidewalks in the urban center, along with a slight decrease in parking provision in the newly developed peripheral areas. The only significant difference between rich and poor areas during the period from 2001 to 2023 is in the number of swimming pools built. When it comes to walkability and parking provision, the primary distinction is between neighborhoods in the city center and those on the periphery.

In Vienna, the neighborhood analyzed is located on the periphery of the city. Results for the 2014–2023 period show an increase in pedestrian-related infrastructure. However, the expansion of parking surfaces exceeds the growth of road infrastructure, indicating a rise in parking provision in newer developments and a trend toward car-centric urban planning.

6. Conclusions

This study has demonstrated the feasibility of creating training datasets for machine learning vision models using openly available data as ground truth, eliminating the need for manual labeling. The proposed process is both efficient and adaptable to diverse regions, tasks, and datasets. This methodology was successfully employed to fine-tune a foundational model for segmenting parking-related surfaces in high-quality aerial imagery. While OpenStreetMap (OSM) and official open data provide valuable sources of ground truth in areas with limited cartographic resources, their incompleteness and inaccuracies present challenges that may impact model outcomes. Despite these limitations, the models developed in this study showed strong performance and adaptability across varied urban landscapes.

Nevertheless, the experimental approach has certain limitations. The reliance on incomplete or inconsistent ground truth data can result in discrepancies between model predictions and actual conditions, making the validation of trained models a significant challenge. A promising direction for addressing this issue is cross-validation using data from cities different from those where the training datasets were generated, as demonstrated in this study with parking models for Vienna and Madrid.

Foundational image vision models like SAM may struggle with non-standard tasks or data from non-Western contexts. This highlights the importance of domain-specific fine tuning methodologies, as proposed in this study, for extending applicability to diverse environments.

It was possible to detect temporal trends differentiating the tendencies in different areas of the city, contributing to the advancement of urban analytics. Additionally, the findings provide valuable insights into how urban development is shaping the city's infrastructure, with a particular focus on walkability and the increasing car-centric trends in newer developments. Differences between two cities in different countries were also established, offering a comparative perspective on urban planning. This research could help city administrations assess whether their policies to reduce car usage have effectively altered the city's land use, as well as road and parking surface distribution. The ability to track such trends provides a foundation for informed policy adjustments aimed at fostering more sustainable urban environments and improving mobility options.

6.1. Future research

To improve the dataset, future research could explore semi-supervised and self-supervised methods (Peláez-Vegas et al., 2023), particularly those tailored for aerial image semantic segmentation. These approaches could help correct errors in the training dataset and implement workflows where high-probability model predictions automatically flag discrepancies with ground truth for review (Tang et al., 2023). Additionally, incorporating class uncertainties into the loss function could improve model robustness and performance (Bressan et al., 2022).

For precise accuracy assessments, manually labeling a small test set remains an option, albeit resource-intensive. Collaboration with official agencies represents another promising avenue, enabling the use of model predictions to identify and correct errors in existing data or supervise areas where discrepancies arise. Advancements in machine learning could encourage agencies to establish clearer criteria and standards for cartography, fostering improved integration of automated methods with traditional practices.

CRediT authorship contribution statement

Miguel Ureña Pliego: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rubén Martínez Marín:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition. **Nianfang Shi:** Software, Methodology, Formal analysis, Data curation, Conceptualization. **Takeru Shibayama:** Validation, Supervision, Resources, Project administration, Methodology, Conceptualization. **Ulrich Leth:** Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Miguel Marchamalo Sacristán:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition.

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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.

Data availability

Link to code shared in the latex source file. All datasets are publically available.

References

Abdurakhmonov, S., Bekanov, K., Ochilov, S., Tukhtamishev, S., Karimov, Y., 2023. Advances in cartography: a review on employed methods. In: Pletnev, D., Nguyen Khanh, B., Kankhva, V. (Eds.), E3S Web Conf. 389, 03057. http://dx.doi.org/10.1051/e3sconf/202338903057, URL https://www.e3s-conferences.org/10.1051/e3sconf/202338903057.

Abrahamse, W., Steg, L., Gifford, R., Vlek, C., 2009. Factors influencing car use for commuting and the intention to reduce it: A question of self-interest or morality? Transp. Res. Part F: Traffic Psychol. Behav. 12 (4), 317–324. http://dx.doi.org/10.1016/j.trf.2009.04.004, URL https://www.sciencedirect.com/science/article/pii/S1369847809000230.

Alshammari, T.O., Hassan, A.M., Arab, Y., Hussein, H., Khozaei, F., Saeed, M., Ahmed, B., Zghaibeh, M., Beitelmal, W., Lee, H., 2022. The compactness of non-compacted urban developments: A critical review on sustainable approaches to automobility and urban sprawl. Sustain. 14 (18), 11121. http://dx.doi.org/10.3390/su141811121, URL https://www.mdpi.com/2071-1050/14/18/11121, Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., Vairo, C., 2016. Deep learning for decentralized parking lot occupancy detection. Expert Syst. Appl. 72, http://dx.doi.org/10.1016/j.eswa.2016.10.055.

Aszkowski, P., Ptak, B., Kraft, M., Pieczyński, D., Drapikowski, P., 2023. Deepness: Deep neural remote sensing plugin for QGIS. SoftwareX 23, 101495. http://dx.doi.org/10.1016/j.softx.2023.101495, URL https://www.sciencedirect.com/science/article/pii/S2352711023001917.

Audebert, N., Saux, B.L., Lefevre, S., 2017. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. Remote. Sens. 9 (4), http://dx.doi.org/10.3390/rs9040368, URL https://www.proquest.com/docview/1905789870/abstract/39EE60F6089043F1PQ/1, Place: Basel, Switzerland Publisher: MDPI AG.

Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.-H., Khan, F.S., 2023. Foundational models defining a new era in vision: A survey and outlook. http://dx.doi.org/10.48550/arXiv.2307.13721, URL http://arxiv.org/abs/2307.13721, arXiv:2307.13721 [cs].

Ayuntamiento de Madrid, 2025. Geoportal. URL https://geoportal.madrid.es/.

Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., Merhof, D., 2023. Loss functions in the era of semantic segmentation: A survey and outlook. URL http://arxiv.org/abs/2312.05391, arXiv:2312.05391 [cs].

Badrinarayanan, V., Kendall, A., Cipolla, R., 2016. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. http://dx.doi.org/10.48550/arXiv.1511.00561, URL http://arxiv.org/abs/1511.00561, arXiv:1511.00561 [cs] version: 3.

Bressan, P.O., Junior, J.M., Correa Martins, J.A., De Melo, M.J., Gonçalves, D.N., Freitas, D.M., Marques Ramos, A.P., Garcia Furuya, M.T., Osco, L.P., De Andrade Silva, J., Luo, Z., Garcia, R.C., Ma, L., Li, J., Gonçalves, W.N., 2022. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. Int. J. Appl. Earth Obs. Geoinf. 108, 102690. http://dx.doi.org/10.1016/j.jag.2022.102690, URL https://linkinghub.elsevier.com/retrieve/pii/S0303243422000162.

Cao, H., Gu, H., Guo, X., 2023. Feasibility of transfer learning: A mathematical framework. URL http://arxiv.org/abs/2305.12985, arXiv:2305.12985 [cs].

Carse, A., Goodman, A., Mackett, R.L., Panter, J., Ogilvie, D., 2013. The factors influencing car use in a cycle-friendly city: the case of Cambridge. J. Transp. Geogr. 28 (100), 67–74. http://dx.doi.org/10.1016/j.jtrangeo.2012.10.013.

Christiansen, P., Engebretsen, O., Fearnley, N., Usterud Hanssen, J., 2017. Parking facilities and the built environment: Impacts on travel behaviour. Transp. Res. Part A: Policy Pr. 95, 198–206. http://dx.doi.org/10.1016/j.tra.2016.10.025, URL https://www.sciencedirect.com/science/article/pii/S0965856416301525.

City of London web mapping, 2025. URL https://www.mapping.cityoflondon.gov.uk/geocortex/mapping/.

Copernicus, 2025. URL https://copernicus.eu.

DigitalGlobe, 2025.WorldView-3. URL https://worldview3.digitalglobe.com/.

Dong, X., Cao, J., Zhao, W., 2024. A review of research on remote sensing images shadow detection and application to building extraction. Eur. J. Remote. Sens. 57 (1), 2293163. http://dx.doi.org/10.1080/22797254.2023.2293163, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/22797254.2023.2293163.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. URL https://openreview.net/forum?id=YicbFdNTTy.

Drouyer, S., 2020. Parking occupancy estimation on planetscope satellite images. In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Waikoloa, HI, USA, pp. 1098–1101. http://dx.doi.org/10.1109/IGARSS39084.2020.9323104, URL https://ieeexplore.ieee.org/document/9323104/.

2007. Directive 2007/2/EC. URL https://eur-lex.europa.eu/eli/dir/2007/2/oj.

Esmaeili, M., Abbasi-Moghadam, D., Sharifi, A., Tariq, A., Li, Q., 2023. Hyperspectral image band selection based on CNN embedded GA (CNNeGA). IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16, 1927–1950. http://dx.doi.org/10.1109/JSTARS.2023.3242310, URL https://ieeexplore.ieee.org/document/10037182, Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Falcon, W., team, T.P.L., 2023. PyTorch lightning. http://dx.doi.org/10.5281/zenodo.10419201, URL https://zenodo.org/records/10419201.

Feeney, B.P., 1989. A review of the impact of parking policy measures on travel demand. Transp. Plan. Technol. 13 (4), 229–244. http://dx.doi.org/10.1080/03081068908717403, URL http://www.tandfonline.com/doi/abs/10.1080/03081068908717403.

Gerike, R., Koszowski, C., Schröter, B., Buehler, R., Schepers, P., Weber, J., Wittwer, R., Jones, P., 2021. Built environment determinants of pedestrian activities and their consideration in urban street design. Sustain. 13 (16), 9362. http://dx.doi.org/10.3390/su13169362, URL https://www.mdpi.com/2071-1050/13/16/9362, Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

Gillies, S., et al., 2013. Rasterio: geospatial raster I/O for Python programmers. URL https://github.com/mapbox/rasterio.

Glaeser, E.L., Kahn, M.E., 2004. Chapter 56 - Sprawl and urban growth. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. In: Cities and Geography, vol. 4, Elsevier, pp. 2481–2527. http://dx.doi.org/10.1016/S1574-0080(04)80013-0, URL https://www.sciencedirect.com/science/article/pii/S1574008004800130.

Gonzales, J.J., 2023. Building-level comparison of microsoft and google open building footprints datasets (short paper). LIPIcs, Vol. 277, GIScience 2023 277, 35:1–35:6. http://dx.doi.org/10.4230/LIPICS.GISCIENCE.2023.35, URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.GIScience.2023.35, Artwork Size: 6 pages, 914745 bytes ISBN: 9783959772884 Medium: application/pdf Publisher: Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Google earth, 2025. URL https://earth.google.com/.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. http://dx.doi.org/10.1016/j.rse.2017.06.031, URL https://www.sciencedirect.com/science/article/pii/S0034425717302900.

Gu, X., Cui, Y., Huang, J., Rashwan, A., Yang, X., Zhou, X., Ghiasi, G., Kuo, W., Chen, H., Chen, L.-C., Ross, D.A., 2023. DaTaSeg: Taming a universal multi-dataset multi-task segmentation model. URL http://arxiv.org/abs/2306.01736, arXiv:2306.01736 [cs].

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ. Plan. B: Plan. Des. 37 (4), 682–703. http://dx.doi.org/10.1068/b35097, URL http://journals.sagepub.com/doi/10.1068/b35097.

Haklay, M., Weber, P., 2008. OpenStreetMap: User-generated street maps. IEEE Pervasive Comput. 7 (4), 12–18. http://dx.doi.org/10.1109/MPRV.2008.80, URL https://ieeexplore.ieee.org/document/4653466, Conference Name: IEEE Pervasive Computing.

Halpern, C., 2022. Road space reallocation for sustainable urban mobility in the EU. LIEPP Policy Brief 60, URL https://sciencespo.hal.science/hal-03701457/document.

He, K., Chen, X., Xie, S., Li, Y., Dollar, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, New Orleans, LA, USA, pp. 15979–15988. http://dx.doi.org/10.1109/CVPR52688.2022.01553, URL https://ieeexplore.ieee.org/document/9879206/.

Helber, P., Bischke, B., Dengel, A., Borth, D., 2018. 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. IEEE, Valencia, pp. 204–207. http://dx.doi.org/10.1109/IGARSS.2018.8519248, URL https://ieeexplore.ieee.org/document/8519248/.

Hellekes, J., Kehlbacher, A., Díaz, M.L., Merkle, N., Henry, C., Kurz, F., Heinrichs, M., 2023. Parking space inventory from above: Detection on aerial images and estimation for unobserved regions. IET Intell. Transp. Syst. 17 (5), 1009–1021. http://dx.doi.org/10.1049/itr2.12322, URL https://onlinelibrary.wiley.com/doi/abs/10.1049/itr2.12322, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1049/itr2.12322.

Henry, C., Hellekes, J., Merkle, N., Azimi, S.M., Kurz, F., 2021. Citywide estimation of parking space using aerial imagery and osm data fusion with deep learning and fine-grained annotation. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLIII-B2-2021, 479–485. http://dx.doi.org/10.5194/isprs-archives-XLIII-B2-2021-479-2021, URL https://isprs-archives.copernicus.org/articles/XLIII-B2-2021/479/2021/.

Herfort, B., Lautenbach, S., Porto De Albuquerque, J., Anderson, J., Zipf, A., 2021. The evolution of humanitarian mapping within the OpenStreetMap community. Sci. Rep. 11 (1), 3037. http://dx.doi.org/10.1038/s41598-021-82404-z, URL https://www.nature.com/articles/s41598-021-82404-z.

Hu, Y., Huang, X., Li, J., Zhang, Z., 2024. GBSS:a global building semantic segmentation dataset for large-scale remote sensing building extraction. http://dx.doi.org/10.48550/arXiv.2401.01178, URL http://arxiv.org/abs/2401.01178, arXiv:2401.01178 [cs].

Iakubovskii, P., 2019. Segmentation_models_pytorch. URL https://github.com/qubvel/segmentation_models.pytorch, original-date: 2019-03-01T16:21:21Z.

Jalayer, S., Sharifi, A., Abbasi-Moghadam, D., Tariq, A., Qin, S., 2023. Assessment of spatiotemporal characteristic of droughts using in situ and remote sensing-based drought indices. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16, 1483–1502. http://dx.doi.org/10.1109/JSTARS.2023.3237380, URL https://ieeexplore.ieee.org/document/10018194, Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Jamil, N., Sembok, T.M.T., Bakar, Z.A., 2008. Noise removal and enhancement of binary images using morphological operations. In: 2008 International Symposium on Information Technology. IEEE, Kuala Lumpur, Malaysia, pp. 1–6. http://dx.doi.org/10.1109/ITSIM.2008.4631954, URL http://ieeexplore.ieee.org/document/4631954/.

Jiao, J., 2018. Machine learning assisted high-definition map creation. In: 2018 IEEE 42nd Annual Computer Software and Applications Conference. COMPSAC, IEEE, Tokyo, Japan, pp. 367–373. http://dx.doi.org/10.1109/COMPSAC.2018.00058, URL https://ieeexplore.ieee.org/document/8377682/.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. URL http://arxiv.org/abs/2304.02643, arXiv:2304.02643 [cs].

Kumar, S., Jain, A., Rani, S., Alshazly, H., Idris, S., Bourouis, S., 2022. Deep neural network based vehicle detection and classification of aerial images. Intell. Autom. Soft Comput. 34 (1), 119–131. http://dx.doi.org/10.32604/iasc.2022.024812, URL https://www.techscience.com/iasc/v34n1/47359, Publisher: Tech Science Press.

Küster, F., 2019. Practitioner briefings: Cycling. Supporting and encouraging cycling in sustainable urban mobility planning. URL 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.

Litman, T., 2011. Why and how to reduce the amount of land paved for roads and parking facilities. Environ. Pr. 13, 38–46. http://dx.doi.org/10.1017/S1466046610000530.

Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., Zhang, Y., 2019. Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. Remote. Sens. 11 (7), 830. http://dx.doi.org/10.3390/rs11070830, URL https://www.mdpi.com/2072-4292/11/7/830.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. http://dx.doi.org/10.48550/arXiv.1411.4038, URL http://arxiv.org/abs/1411.4038, arXiv:1411.4038 [cs].

Luftbilder Berlin, 2025. URL https://daten.berlin.de/anwendungen/luftbilderberlin.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In: 2017 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, (ISSN: 2153-7003) pp. 3226–3229. http://dx.doi.org/10.1109/IGARSS.2017.8127684, URL https://ieeexplore.ieee.org/document/8127684/footnotes#footnotes-id-fn2.

Maidment, P.D., Domenico, B., Gemmell, A., Lehnert, K., Tarboton, D., Zaslavsky, I., 2011. The open geospatial consortium and EarthCube. URL https://www.ogc.org/standards/technical-papers/.

Marmanis, D., Wegner, J., Galliani, S., Schindler, K., Datcu, M., Stilla, U., 2016. Semantic segmentation of aerial images with an ensemble of cnns. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. III-3, 473–480. http://dx.doi.org/10.5194/isprs-annals-III-3-473-2016.

Mattioli, G., Roberts, C., Steinberger, J.K., Brown, A., 2020. The political economy of car dependence: A systems of provision approach. Energy Res. Soc. Sci. 66, 101486. http://dx.doi.org/10.1016/j.erss.2020.101486, URL https://www.sciencedirect.com/science/article/pii/S2214629620300633.

Máttyus, G., Luo, W., Urtasun, R., 2017. DeepRoadMapper: Extracting road topology from aerial images. In: 2017 IEEE International Conference on Computer Vision. ICCV, (ISSN: 2380-7504) pp. 3458–3466. http://dx.doi.org/10.1109/ICCV.2017.372, URL https://ieeexplore.ieee.org/document/8237634.

Maurício, J., Domingues, I., Bernardino, J., 2023. Comparing vision transformers and convolutional neural networks for image classification: A literature review. Appl. Sci. 13 (9), 5521. http://dx.doi.org/10.3390/app13095521, URL https://www.mdpi.com/2076-3417/13/9/5521, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

McCahill, C.T., Garrick, N., Atkinson-Palombo, C., Polinski, A., 2016. Effects of parking provision on automobile use in cities: Inferring causality. Transp. Res. Rec. 2543 (1), 159–165. http://dx.doi.org/10.3141/2543-19, Publisher: SAGE Publications Inc.

Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. Pattern Recognit. Lett. 24 (16), 3037–3058. http://dx.doi.org/10.1016/S0167-8655(03)00164-8, URL https://www.sciencedirect.com/science/article/pii/S0167865503001648.

Merkle, N., Azimi, S.M., Pless, S., Kurz, F., 2019. Semantic vehicle segmentation in very high resolution multispectral aerial images using deep neural networks. In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 5045–5048. http://dx.doi.org/10.1109/IGARSS.2019.8898513, URL https://ieeexplore.ieee.org/document/8898513/, Conference Name: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium ISBN: 9781538691540 Place: Yokohama, Japan Publisher: IEEE.

Microsoft, 2024. GlobalMLBuildingFootprints. URL https://github.com/microsoft/GlobalMLBuildingFootprints, original-date: 2022-04-22T22:09:24Z.

Minghini, M., Frassinelli, F., 2019. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? Open Geospatial Data Softw. Stand. 4 (1), 9. http://dx.doi.org/10.1186/s40965-019-0067-x.

Minghini, M., Kotsev, A., Lutz, M., 2019. Comparing inspire and openstreetmap data : How to make the most out of the two worlds. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLII-4/W14, 167–174. http://dx.doi.org/10.5194/isprs-archives-XLII-4-W14-167-2019, URL https://isprs-archives.copernicus.org/articles/XLII-4-W14/167/2019/.

Mnih, V., 2013. Machine Learning for Aerial Image Labeling (Ph.D. thesis). University of Toronto, Toronto, Canada, URL https://www.cs.toronto.edu/~vmnih/docs/Mnih_Volodymyr_PhD_Thesis.pdf.

Mohammadi, M., Sharifi, A., 2021. Evaluation of convolutional neural networks for urban mapping using satellite images. J. Indian Soc. Remote. Sens. 49 (9), 2125–2131. http://dx.doi.org/10.1007/s12524-021-01382-x.

Nasiri, V., Deljouei, A., Moradi, F., Sadeghi, S.M.M., Borz, S.A., 2022. Land use and land cover mapping using sentinel-2, landsat-8 satellite images, and google earth engine: A comparison of two composition methods. Remote. Sens. 14 (9), 1977. http://dx.doi.org/10.3390/rs14091977, URL https://www.mdpi.com/2072-4292/14/9/1977.

Nejad, S.M.M., Abbasi-Moghadam, D., Sharifi, A., Farmonov, N., Amankulova, K., László, M., 2023. Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16, 254–266. http://dx.doi.org/10.1109/JSTARS.2022.3223423, URL https://ieeexplore.ieee.org/document/9956008, Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Nurul Habib, K., Morency, C., Trépanier, M., 2012. Integrating parking behaviour in activity-based travel demand modelling: Investigation of the relationship between parking type choice and activity scheduling process. Transp. Res. Part A 46, 154–166. http://dx.doi.org/10.1016/j.tra.2011.09.014.

Osco, L.P., Wu, Q., de Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Junior, J.M., 2023. The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. URL http://arxiv.org/abs/2306.16623, arXiv:2306.16623 [cs].

Pandhe, A., March, A., 2012. Parking availability influences on travel mode: Melbourne CBD offices. Aust. Plan. 49 (2), 161–171. http://dx.doi.org/10.1080/07293682.2011.616177, Publisher: Routledge _eprint: https://doi.org/10.1080/07293682.2011.616177.

Paris City Council, 2025. Paris open data. URL https://opendata.paris.fr/map/.

Peláez-Vegas, A., Mesejo, P., Luengo, J., 2023. A survey on semi-supervised semantic segmentation. http://dx.doi.org/10.48550/arXiv.2302.09899, URL http://arxiv.org/abs/2302.09899, arXiv:2302.09899 [cs].

Rodriguez-Valencia, A., Ortiz-Ramirez, H.A., 2021. Understanding green street design: Evidence from three cases in the U.S.. Sustain. 13, 1916. http://dx.doi.org/10.3390/su13041916.

Rye, T., Tully, S., Godin, G., Schmalholz, N., Hertel, M., 2022. Parking and SUMP. Using parking management to achieve your SUMP objectives effectively and sustainably. URL https://urban-mobility-observatory.transport.ec.europa.eu/document/download/4ae9e061-dfc9-4f92-afe2-30d70de76580_en?filename=parking_and_sump.pdf&prefLang=it.

Sandler, M., Zhmoginov, A., Vladymyrov, M., Jackson, A., 2022. Fine-tuning image transformers using learnable memory. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, New Orleans, LA, USA, pp. 12145–12154. http://dx.doi.org/10.1109/CVPR52688.2022.01184, URL https://ieeexplore.ieee.org/document/9879120/.

Schröter, B., Hantschel, S., Koszowski, C., Buehler, R., Schepers, P., Weber, J., Wittwer, R., Gerike, R., 2021. Guidance and practice in planning cycling facilities in Europe—An overview. Sustain. 13 (17), 9560. http://dx.doi.org/10.3390/su13179560, URL https://www.mdpi.com/2071-1050/13/17/9560, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

Serra, J., 1986. Introduction to mathematical morphology. Comput. Vis. Graph. Image Process. 35 (3), 283–305. http://dx.doi.org/10.1016/0734-189X(86)90002-2, URL https://www.sciencedirect.com/science/article/pii/0734189X86900022.

Shao, Z., Tang, P., Wang, Z., Saleem, N., Yam, S., Sommai, C., 2020. BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images. Remote. Sens. 12 (6), 1050. http://dx.doi.org/10.3390/rs12061050, URL https://www.mdpi.com/2072-4292/12/6/1050.

Sharifi, A., Felegari, S., 2022. Nitrogen dioxide (NO2) pollution monitoring with sentinel-5P satellite imagery over during the coronavirus pandemic (case study: Tehran). Remote. Sens. Lett. 13 (10), 1029–1039. http://dx.doi.org/10.1080/2150704X.2022.2120780, URL https://www.tandfonline.com/doi/full/10.1080/2150704X.2022.2120780.

Sharifi, A., Hosseingholizadeh, M., 2020. Application of sentinel-1 data to estimate height and biomass of rice crop in astaneh-ye Ashrafiyeh, Iran. J. Indian Soc. Remote. Sens. 48 (1), 11–19. http://dx.doi.org/10.1007/s12524-019-01057-8.

Singhal, P., Walambe, R., Ramanna, S., Kotecha, K., 2023. Domain adaptation: Challenges, methods, datasets, and applications. IEEE Access 11, 6973–7020. http://dx.doi.org/10.1109/ACCESS.2023.3237025, URL https://ieeexplore.ieee.org/document/10017290/?arnumber=10017290, Conference Name: IEEE Access.

Srisha, R., Khan, A., 2013. Morphological operations for image processing : Understanding and its applications. URL https://www.researchgate.net/publication/272484795_Morphological_Operations_for_Image_Processing_Understanding_and_its_Applications.

Stadtvermessung Wien (MA 41), 2025. URL https://www.wien.gv.at/stadtentwicklung/stadtvermessung/index.html.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. Remote. Sens. 12 (7), 1135. http://dx.doi.org/10.3390/rs12071135, URL https://www.mdpi.com/2072-4292/12/7/1135, Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

Tang, M., Georgiou, K., Qi, H., Champion, C., Bosch, M., 2023. 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, IEEE, Waikoloa, HI, USA, pp. 3787–3796. http://dx.doi.org/10.1109/WACV56688.2023.00379, URL https://ieeexplore.ieee.org/document/10030272/.

Thisanke, H., Deshan, C., Chamith, K., Seneviratne, S., Vidanaarachchi, R., Herath, D., 2023. Semantic segmentation using vision transformers: A survey. Eng. Appl. Artif. Intell. 126, 106669. http://dx.doi.org/10.1016/j.engappai.2023.106669, URL https://www.sciencedirect.com/science/article/pii/S0952197623008539.

USGS (United States Geological Survey), 2025. URL https://www.usgs.gov/.

Vargas-Munoz, J.E., Srivastava, S., Tuia, D., Falcao, A.X., 2021. OpenStreetMap: Challenges and opportunities in machine learning and remote sensing. IEEE Geosci. Remote. Sens. Mag. 9 (1), 184–199. http://dx.doi.org/10.1109/MGRS.2020.2994107, URL https://ieeexplore.ieee.org/document/9119753/.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need. Long Beach, CA, USA, http://dx.doi.org/10.48550/arXiv.1706.03762, URL http://arxiv.org/abs/1706.03762, arXiv:1706.03762 [cs].

Walker, J., Thornton, B., Quiñones, L.M., 2019. Supporting and encouraging walking in sustainable urban mobility planning. URL 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.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., contributors, t.s.-i., 2014. Scikit-image: Image processing in Python. PeerJ 2, e453. http://dx.doi.org/10.7717/peerj.453, URL http://arxiv.org/abs/1407.6245, arXiv:1407.6245 [cs].

Wu, Q., Osco, L.P., 2023. Samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). J. Open Source Softw. 8 (89), 5663. http://dx.doi.org/10.21105/joss.05663, URL https://joss.theoj.org/papers/10.21105/joss.05663.

Yeung, M., Sala, E., Schönlieb, C.-B., Rundo, L., 2022. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput. Med. Imaging Graph. 95, 102026. http://dx.doi.org/10.1016/j.compmedimag.2021.102026, URL https://www.sciencedirect.com/science/article/pii/S0895611121001750.

Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. 169, 114417. http://dx.doi.org/10.1016/j.eswa.2020.114417, URL https://www.sciencedirect.com/science/article/pii/S0957417420310836.

Zamir, S.W., Arora, A., Gupta, A., Khan, S., Sun, G., Khan, F.S., Zhu, F., Shao, L., Xia, G.-S., Bai, X., 2019. iSAID: A large-scale dataset for instance segmentation in aerial images. URL http://arxiv.org/abs/1905.12886, arXiv:1905.12886 [cs].

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-Net. IEEE Geosci. Remote. Sens. Lett. 15 (5), 749–753. http://dx.doi.org/10.1109/LGRS.2018.2802944, URL https://ieeexplore.ieee.org/document/8309343, Conference Name: IEEE Geoscience and Remote Sensing Letters.

Zhu, Q., Liao, C., Hu, H., Mei, X., Li, H., 2021. MAP-Net: Multiple attending path neural network for building footprint extraction from remote sensed imagery. IEEE Trans. Geosci. Remote Sens. 59 (7), 6169–6181. http://dx.doi.org/10.1109/TGRS.2020.3026051, URL https://ieeexplore.ieee.org/document/9212557/, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2021. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76. http://dx.doi.org/10.1109/JPROC.2020.3004555, URL https://ieeexplore.ieee.org/document/9134370/.

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: A survey. Proc. IEEE 111 (3), 257–276. http://dx.doi.org/10.1109/JPROC.2023.3238524, URL https://ieeexplore.ieee.org/document/10028728/.