ZenSVI: An Open-Source Software for the Integrated Acquisition, Processing and Analysis of Street View Imagery Towards Scalable Urban Science

Koichi Ito[a], Yihan Zhu[a], Mahmoud Abdelrahman[a], Xiucheng Liang[a], Zicheng Fan[a], Yujun Hou[a], Tianhong Zhao[b], Rui Ma[a,c], Kunihiko Fujiwara[a,d], Jiani Ouyang[a,e], Matias Quintana[f], Filip Biljecki[a,g,*]

[a]*Department of Architecture, National University of Singapore, 4 Architecture Drive, Singapore, 117356, Singapore*
[b]*College of Big Data and Internet, Shenzhen Technology University, 3002 Lantian Road, Shenzhen, 518118, China*
[c]*Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, 999077, China*
[d]*Research & Development Institute, Takenaka Corporation, 1-5-1 Otsuka, Inzai, Chiba, 270-1352, Japan*
[e]*State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China*
[f]*Singapore-ETH Centre, Future Cities Lab Global Programme, CREATE campus, 1 Create Way, #06-01 CREATE Tower, 138602, Singapore*
[g]*Department of Real Estate, National University of Singapore, 15 Kent Ridge Drive, Singapore, 119245, Singapore*

Abstract

Street view imagery (SVI) has been instrumental in many studies in the past decade to understand and characterize street features and the built environment. Researchers across a variety of domains, such as transportation, health, architecture, human perception, and infrastructure have employed different methods to analyze SVI. However, these applications and image-processing procedures have not been standardized, and solutions have been implemented in isolation, often making it difficult for others to reproduce existing work and carry out new research. Using SVI for research requires multiple technical steps: accessing APIs for scalable data collection, preprocessing images to standardize formats, implementing computer vision models for feature extraction, and conducting spatial analysis. These technical requirements create barriers for researchers in urban studies, particularly those without extensive programming experience. We developed ZenSVI, a free and open-source Python package that integrates and implements the entire process of SVI analysis, supporting a wide range of use cases. Its end-to-end pipeline

*Corresponding author. Email: filip@nus.edu.sg

includes downloading SVI from multiple platforms (e.g., Mapillary and KartaView) efficiently, analyzing metadata of SVI, applying computer vision models to extract target features, transforming SVI into different projections (e.g., fish-eye and perspective) and different formats (e.g., depth map and point cloud), visualizing analyses with maps and plots, and exporting outputs to other software tools. We demonstrated its use in Singapore through a case study of data quality assessment and clustering analysis in a streamlined manner. Our software improves the transparency, reproducibility, and scalability of research relying on SVI and supports researchers in conducting urban analyses efficiently. Its modular design facilitates extensions of the package for new use cases. This package is openly available at `https://github.com/koito19960406/ZenSVI`, and it is supported by documentation including tutorials (`https://zensvi.readthedocs.io/en/latest/examples/index.html`).

Keywords: Street-level Imagery, Python Package, Computer Vision, FAIR, Reproducibility

---

1. Introduction

Street view imagery (SVI) is a popular geospatial dataset in various fields pertaining to urban science, such as transportation, health, morphology, and urban visual perception, thanks to its scalability, coverage, and spatiotemporal granularity (Biljecki and Ito, 2021; Rzotkiewicz et al., 2018; Kang et al., 2020; Ito et al., 2024b; Wang et al., 2024b). Machine learning techniques to process image data have enabled automated assessments of SVI on large scales, and such scalability has been leveraged with the growing coverage of SVI data platforms. Also importantly, SVI's ability to capture street-level information and its spatiotemporal granularity has supported many micro-scale assessments of streets and neighborhoods over time that were difficult or impossible with conventional spatial data, such as remote sensing imagery and field surveys (Fan et al., 2023; Kang et al., 2020; Wang et al., 2024c; Liang et al., 2023).

The strengths of SVI have enabled numerous applications in urban studies and data-driven planning. Researchers have developed various indices to quantify urban phenomena, from the green view index (Li et al., 2015; Yang et al., 2009) to sky view and building view factors (Liang et al., 2017; Wang et al., 2022b; Biljecki et al., 2023), and comprehensive walkability and bikeability indices (Koo et al., 2022; Ito and Biljecki, 2021). These SVI-derived indicators have been used to analyze diverse urban phenomena including land use patterns, housing prices (Law et al., 2019; Zhao et al., 2023a), transportation flows (Jiao and Wang, 2023), and socioeconomic status (Fan et al., 2023; Ning et al., 2022; Cao et al., 2018; Hu

et al., 2024; Fujiwara et al., 2024). SVI has also been utilized to predict or estimate objective and subjective characteristics of streets, such as road condition assessment (Ai et al., 2018; Shu et al., 2021; de Mesquita et al., 2022; Arya et al., 2021), building classification (Kang et al., 2018; Li et al., 2018b; Sun et al., 2022), accessibility evaluation (Weld et al., 2019), and assessment of subjective perceptions including safety (Kang et al., 2023), sound intensity (Zhao et al., 2023b), architectural design (Liang et al., 2024), and attractiveness (Dubey et al., 2016; Zhang et al., 2018).

Besides the achievements in maximizing usability, research relying on SVI owes its developments and popularity also to the availability of data. For example, major commercial providers (e.g., Google Street View) now have extensive coverage around the world, in particular, in North America, East, and Southeast Asia, and Europe. Many researchers have taken advantage of such coverage, standardized data, and scalability to conduct large-scale studies comparing multiple cities, regions, and countries (Biljecki and Ito, 2021; Ito et al., 2024a). The increasing popularity of crowdsourced SVI has also been a notable development — anyone, anywhere can collect street-level imagery with consumer devices while moving and share it openly on platforms such as Mapillary, KartaView, and Panoramax. This data stream has advantages such as free license, availability of additional points of view (e.g., imagery from cyclists), and spatiotemporal coverage that commercial services may not offer in some areas (Biljecki et al., 2023; Helbich et al., 2024; Sánchez and Labib, 2024; Hou and Biljecki, 2022; Bendixen et al., 2023; Hou et al., 2024).

As much as SVI has been a popular and powerful instrument in sensing the built environment, there remain technical difficulties in conducting studies relying on it, especially for researchers without advanced analytical and technical skills. In fact, such difficulties have even been on the rise, as large-scale big data analyses with deep learning models are becoming the norm (Ito et al., 2024b). As a consequence, domain experts who are less technical find it challenging to conduct such studies, which hampers equity and participation in research (Ito and Biljecki, 2021; Bromm et al., 2020; Zhu et al., 2025), and may also impede the development of new use cases and the representation of certain geographical areas. Further, most publications are not accompanied by openly released code to allow reproducibility, compounding the issue (Biljecki and Ito, 2021).

This challenge has been common across many domains in urban analytics, including urban morphology. However, the growing adoption of open science practices in urban studies has begun to address this issue, with free and open-source software packages being developed to support various analytical tasks. These tools have democratized urban analytics by enabling researchers to conduct sophisticated analyses without requiring advanced technical expertise (Yap et al., 2022).

3

OSMnx by Boeing (2017) is a quintessential example — it is a Python package that has facilitated many applications that require network analysis or OpenStreetMap data by integrating them as a *one-stop* software. Trackintel by Martin et al. (2023a) is also a Python package and provides many functions of GPS-trajectory data analysis with standardized models, thereby enhancing the transparency, transferability, and reproducibility of research. Another notable Python package is Urbanity developed by Yap et al. (2023), which allows users to download ready-to-use feature-rich network data. There have also been domain-specific tools, e.g., an R package by Mahajan (2024) computes the green index based on multiple sources of data, such as OpenStreetMap and SVI.

As Yap et al. (2022) highlighted in their comprehensive systematic review, there is still a need for integrative open-source tools that can comparatively evaluate urban environments on a large scale and do so with relative ease of use and accessibility. This need is particularly acute in SVI research, where even though researchers occasionally share their code, these contributions often address specific niche problems, leading to a fragmented landscape of tools that are difficult to adapt or integrate into broader workflows. Despite the increasing number of applications, trends, and challenges mentioned above, researchers have only made a limited effort to standardize, facilitate, or integrate the workflow of SVI analysis, i.e., an integrated open-source solution to support urban studies relying on SVI.

Therefore, to bridge such a clear gap and facilitate reproducible SVI research, we introduce ZenSVI — a framework and implementation that integrates downloading, processing, transforming, and visualizing SVI, which will benefit and encourage studies by improving the transparency, methodology, and ease of conducting research that requires SVI and associated elements such as computer vision models. This study is the first to create such an integrated software not only in SVI research but also in urban imagery. ZenSVI simplifies many layers of SVI analysis, such as data acquisition and processing, and will therefore accelerate innovative urban science research by mitigating barriers for less experienced researchers and making their research reproducible. In a broader scope, its framework based on open data and open models presents a novel contribution to the open science current of urban analytics, and its modular framework may contribute to its extensions or integrations in other workflows.

The paper is organized as follows: Section 2 discusses existing studies, their research gaps, and our study's value, Section 3 provides an overview of the package's framework, Section 4 shows how the package can be used in typical research settings, Section 5 offers insights into how the package can further be extended or

used by exporting data to other software applications, and lastly Section 6 concludes this paper by highlighting the contribution of our work.

2. Related work

2.1. Open-source software for urban science

Table 1: Comparison of urban analysis tools and libraries.

Title	Primary Application	Download	ML Applications	Visualization
OSMnx (Boeing, 2017)[a]	Network Analysis	•		•
Urbanity (Yap et al., 2023) [b]	Network Analysis	•	•	•
global-indicators (Higgs et al., 2023) [c]	Mobility	•		
PCT (Lovelace et al., 2017) [d]	Mobility	•		•
MovingPandas (Graser, 2019) [e]	Mobility			•
scikit-mobility (Pappalardo et al., 2022) [f]	Mobility			•
rpi-urban-mobility-tracker (Wojke et al., 2017; Wojke and Bewley, 2018) [g]	Mobility	•	•	
Trackintel (Martin et al., 2023b) [h]	Mobility	•		•
GWmodelS (Lu et al., 2023a) [i]	Spatial analysis	•		•
raster vision[j]	Satellite imagery		•	

[a] https://github.com/gboeing/osmnx
[b] https://github.com/winstonyym/urbanity
[c] https://github.com/healthysustainablecities/global-indicators
[d] https://github.com/ITSLeeds/pct
[e] https://github.com/movingpandas/movingpandas
[f] https://github.com/scikit-mobility/scikit-mobility
[g] https://github.com/nathanrooy/rpi-urban-mobility-tracker
[h] https://github.com/mie-lab/trackintel
[i] https://github.com/GWmodel-Lab/GWmodelS
[j] https://github.com/azavea/raster-vision

Open-source software tools are increasingly used in urban studies, enabling researchers to analyze complex urban phenomena efficiently and collaboratively (Yap et al., 2022; Pönkänen et al., 2025). These tools have contributed to the advancement of urban science by providing freely and easily accessible platforms for acquiring, processing, analyzing, and visualizing urban data. For example, OSMnx has facilitated the study of urban networks by streamlining the extraction, construction, and analysis of street networks from OpenStreetMap. Such tools allow researchers to tackle large-scale spatial analyses with minimal technical barriers.

The strength of these open-source tools lies not in introducing new scientific methods, but rather in their ability to democratize existing methodologies and foster collaborative improvement through community engagement. Their success stems from implementing established scientific approaches while remaining open to enhancement by a community of like-minded developers and researchers. These tools excel at facilitating essential tasks, such as downloading large datasets and visualizing complex spatial information, often serving as comprehensive one-stop solutions for specific analytical needs. Additionally, some have integrated machine learning (ML) approaches for tasks such as feature extraction without the need to build models from scratch (Table 1). By embedding cutting-edge algorithms

5

and workflows in user-friendly interfaces, these tools empower researchers across disciplines, including those without extensive programming experience, to conduct innovative studies. This inclusivity expands the reach and impact of urban science.

2.2. Open-source software for street view imagery analysis

SVI has a wide range of applications in urban studies (Biljecki and Ito, 2021), including urban planning and management (Wang et al., 2023; Li et al., 2022c; Xia et al., 2021a; Lu, 2019; Maniat et al., 2021; Wang et al., 2022b), environmental monitoring and evaluation (Cao et al., 2018; He and Li, 2021; Xia et al., 2021a; Li et al., 2022d; Yin and Wang, 2016; Chen and Biljecki, 2023; Li et al., 2022b; Liang et al., 2023; Lu et al., 2023b; Sanusi et al., 2016; Li et al., 2018a), transportation research (Campbell et al., 2019; Cai et al., 2022), and socioeconomic studies (Al-hasoun and González, 2019; Zhang and Dong, 2018). Although these applications solve specific problems, they share many of the same approaches and data sources, but no effort has been made to integrate these applications into one platform so that other researchers can reproduce their work easily.

SVI data have diverse sources and acquisition methods. The most widely used SVI comes from commercial GIS service platforms such as Google[1], Amap[2] and Baidu[3], which capture high-resolution geotagged panoramic images via vehicles (Mahabir et al., 2020; Chen et al., 2019). These platforms offer a wide coverage suitable for urban research but are subject to terms of use and privacy policies (Hou and Biljecki, 2022; Helbich et al., 2024). Open projects like KartaView and Mapillary consolidate and provide user-uploaded SVI, offering free, scalable alternatives, although often with lower consistency and quality (Hara et al., 2013; Zheng and Amemiya, 2023; Juhász and Hochmair, 2016; Hou et al., 2024). Additionally, some city governments and public institutions collect and release SVI, and social media platforms (e.g., Flickr and Instagram) are also significant sources of imagery (Bahrehdar et al., 2020; Chen et al., 2020b; Callau et al., 2019).

Processing SVI typically involves complex computer vision techniques to extract and analyze valuable information, broadly categorized into traditional methods and deep learning-based approaches. The former includes edge detection, texture analysis, color feature extraction, and shape feature extraction (Bansal et al., 2011; de Souza-Daw et al., 2015). With advancements in deep learning, object detection, semantic segmentation, and image classification have become predominant. Object detection identifies and locates objects (e.g., vehicles, pedestrians, traffic signs)

---

[1]https://www.google.com/maps

[2]https://www.amap.com/

[3]https://map.baidu.com/

using models such as YOLO, Faster R-CNN, and R-CNN (Krylov et al., 2018; Campbell et al., 2019; Kang et al., 2020; Ren et al., 2023; Zhang et al., 2020). Semantic segmentation assigns pixel-level labels (e.g., roads, buildings, vegetation) via models such as U-Net, SegNet, and DeepLab (Xia et al., 2021b; Aikoh et al., 2023; Wang et al., 2022a; Gou et al., 2022; Zhao et al., 2023b). Image classification categorizes entire images (e.g., commercial, residential, industrial) using models such as ResNet and VGG (Law et al., 2020; Fang et al., 2022; Huang et al., 2024).

SVI data processing is a multistage process, with each stage requiring specific libraries. The initial stage of SVI data processing involves the downloading of data, which typically entails extracting and downloading high-quality SVI from online services such as the Google Street View API. Traditional computer vision processing libraries such as OpenCV[4] and Pillow[5] are widely used for processing these images. OpenCV offers a rich set of image processing tools for tasks such as cropping, scaling, and rotation, and excels in computer vision tasks such as edge detection and feature extraction. Pillow, on the other hand, is an easy-to-use image processing library suitable for image format conversion, basic image processing, and batch image operations. As processing demands increase, deep learning processing libraries become essential. These libraries can perform complex tasks such as semantic segmentation and scene recognition. TensorFlow[6] and PyTorch[7] are currently the most commonly used deep learning frameworks. They support the construction and training of deep neural networks, effectively handling large amounts of SVI data. Through these libraries, advanced tasks such as object detection and semantic segmentation can be achieved, allowing for a deeper understanding of the content within SVI.

In addition to conventional image processing and deep learning libraries, there are specialized libraries designed for SVI data processing and customized functionalities. These libraries often incorporate deep learning applications adapted from existing studies. Table 2 summarizes common libraries that target SVI data. In terms of programming languages, the primary languages are Python, C++, and R, with some libraries having developed QGIS plugins. Some plugins support a pipeline of functions from data downloading to data processing. For example, Li et al. (2015) developed Treepedia, which uses computer vision techniques applied to Google Street View images to calculate the GVI for measuring and mapping vegetation coverage along urban streets. It establishes a workflow encompassing sampling

---

[4]https://opencv.org/
[5]https://pypi.org/project/pillow/
[6]https://www.tensorflow.org/
[7]https://pytorch.org/

Table 2: Comparison of SVI data processing libraries.

Package	Platform & language	Download	Quality assessment	Metadata extraction	Image transformation	Deep learning	Application	Data source
streetlearn (Mirowski et al., 2018)[a]	C++, Python	✓	✗	✓	✗	✓	Navigation	Google
street-view-pipeline[b]	Python	✓	✗	✗	✗	✗	Calculating GVI	Google, OSM
streetview[c]	Python	✓	✗	✓	✗	✗	Retrieving Google Street View	Google
CODING (Zhao et al., 2021)[d]	Python	✗	✗	✓	✗	✓	Building detection	Google
urban-perceptions (Muller et al., 2022)[e]	Python, QGIS	✓	✗	✗	✓	✓	Street quality assessment	Google, Mapillary, others
Treepedia[f]	Python	✓	✗	✓	✗	✗	Calculating GVI	Google
Green View Index[g]	Python, QGIS	✓	✗	✓	✓	✗	Calculating GVI	Google
go2mapillary[h]	Python, QGIS	✓	✗	✓	✗	✗	Metadata extraction	Mapillary
googleway[i]	R	✓	✗	✓	✗	✗	Accessing and plotting	Google
StreetView-NatureVisibility[j]	Python	✓	✓	✗	✓	✓	Calculating GVI	Google
percept (Danish et al., 2025)[k]	Python	✓	✓	✓	✓	✗	Survey for human perception	Mapillary
streetscape (Yang et al., 2025)[l]	R	✓	✓	✓	✗	✓	GVI and human perception	Mapillary
ZenSVI[m] (*Our work*)	Python	✓	✓	✓	✓	✓	Acquisition, processing and analysis of SVI	Mapillary, KartaView, Amsterdam SVI

[a] https://github.com/google-deepmind/streetlearn
[b] https://github.com/datasciencecampus/street-view-pipeline
[c] https://github.com/robolyst/streetview
[d] https://github.com/kyle-one/Context-Encoding-of-Detected-Buildings
[e] https://github.com/emilymuller1991/urban-perceptions
[f] https://github.com/mittrees/Treepedia_Public
[g] https://github.com/kowalski93/Green-View-Index-for-QGIS
[h] https://github.com/enricofer/go2mapillary
[i] https://github.com/SymbolixAU/googleway
[j] https://github.com/Spatial-Data-Science-and-GEO-AI-Lab/StreetView-NatureVisibility
[k] https://github.com/Spatial-Data-Science-and-GEO-AI-Lab/percept
[l] https://github.com/land-info-lab/streetscape
[m] https://github.com/koito19960406/ZenSVI

point construction, SVI acquisition from Google Street View, metric calculation, and result visualization.

Different libraries support various applications such as data exploration or metadata acquisition, green view index calculation, street quality assessment, and building feature exploration. The methods for computing street view features include traditional computer vision techniques (e.g., edge detection, grayscale calculation) and deep learning-based methods (e.g., scene classification, semantic segmentation). Most libraries use Google Street View data as their primary data source, but some libraries support the import of other data sources. For instance, the 'urban-perceptions' and 'percept' libraries support not only Google Street View data but also other sources such as Mapillary (Muller et al., 2022; Danish et al., 2025). Prior to analysis, data cleaning is essential to ensure the usability of SVI. This involves the provision of metadata, quality assessment to determine whether images are suitable for analysis, and image transformation to convert a panoramic image into a regular image.

However, the existing specialized libraries are typically designed for single-purpose tasks, there is still a lack of a one-stop solution that provides unified support for downloading, cleaning, and multi-purpose SVI analytics, supporting a wide range of use cases. Our integrated Python package aspires to contribute to many domains by filling this gap.

3. ZenSVI framework and implementation

ZenSVI is conceived as a one-stop open-source solution for SVI analytics and is built with a modular architecture to ensure maintainability, extensibility, and code quality. The package is developed with comprehensive unit tests, automated continuous integration/continuous deployment (CI/CD) pipelines through GitHub Actions and detailed API documentation with usage examples. Further, as an open-source project, ZenSVI welcomes contributions from the community, enabling researchers and developers to extend its functionality, improve existing features, or address specific use cases. This continuous and collaborative development approach ensures that the package can evolve alongside the changing needs of the urban research community while maintaining high standards of code quality and documentation.

The framework is designed with compatibility in mind, leveraging widely used Python packages in the geospatial ecosystem. Core dependencies include GeoPandas[8] for spatial operations (Jordahl et al., 2020), OSMnx for street network analysis, and PyTorch for deep learning implementations. This foundation ensures seamless integration with existing geospatial workflows while maintaining forward compatibility through semantic versioning. The framework's reliance on these well-maintained, community-supported packages helps ensure long-term continuity and stability.

For computer vision capabilities, we integrated established models from other research efforts, maintaining their original architectures and parameters to ensure reproducibility and validated performance. Also, each implemented model is documented with citations to its source publication.

The package architecture consists of five main components (Figure 1), which are described separately in the subsequent sections:

- Download — Acquisition of imagery from multiple sources (Section 3.1).

- Metadata Analysis — Computation of metadata based on spatiotemporal information at different spatial levels (Section 3.2).

- Transformation — Transformation of images' projections and formats (Section 3.3).

- Computer Vision — Multi-faceted analysis through traditional and deep learning techniques (Section 3.4).

- Visualization — Presentation of results through various maps and plot types (Section 3.5).

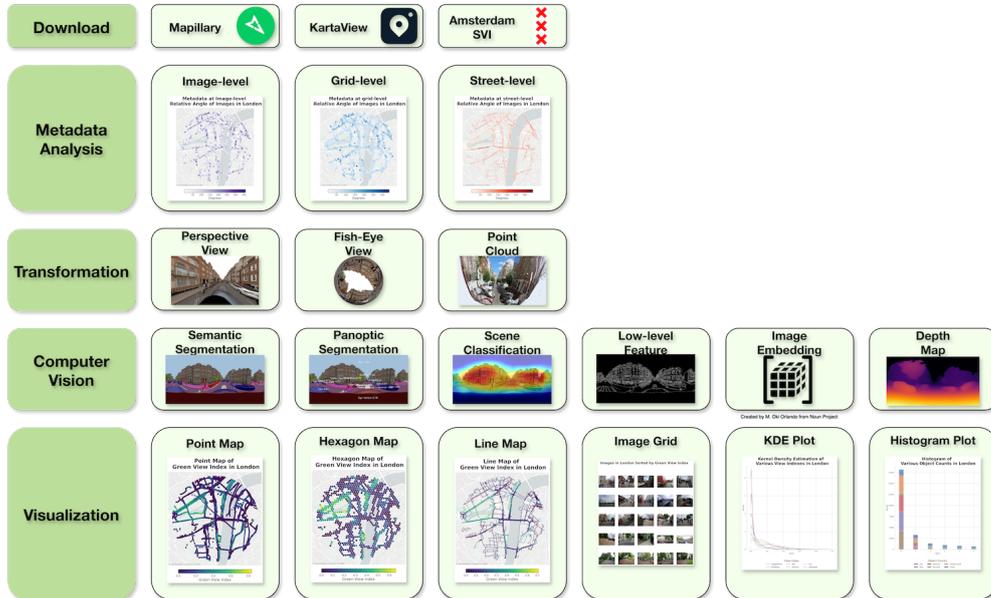
[8]https://github.com/geopandas/geopandas

Figure 1: Overview of the framework of ZenSVI package, consisting of five sub-packages: 1) download, 2) metadata analysis, 3) computer vision, 4) image transformation, and 5) visualization.

3.1. Download

Built on geospatial libraries such as GeoPandas and OSMnx (Boeing, 2017), this sub-package implements wrapper functions with parallel processing to acquire SVI from two major platforms: Mapillary and KartaView, which have global coverage. In addition, the City of Amsterdam's SVI API is supported, which showcases ZenSVI's capability to accommodate various data sources, not limited to commercial or crowdsourced platforms.

The package standardizes input types and output formats while enhancing the official APIs' functionalities. Users can provide location data for SVI acquisition in various formats: coordinate pairs, CSV files with longitude and latitude columns, shapefiles, GeoJSON files, or place names as string values. By leveraging GeoPandas for spatial operations and OSMnx for street network analysis, the package efficiently and robustly processes these inputs to generate and filter sampling points for SVI downloads with parallel processing and checkpoints. With these additional features, ZenSVI eliminates the need for custom API integration code by researchers.

3.2. Metadata Analysis

Metadata is important in facilitating exploratory data analyses and for adding more context to each image. This is particularly important for crowdsourced SVI because they are usually captured in diverse ways, unlike those from commercial platforms. Such inconsistency makes it difficult for researchers to know if a specific image is suitable for their research when they acquire numerous SVI (e.g., millions at the city scale). For example, if one is interested in nighttime walkability, it is unsuitable to use SVI from daytime or SVI captured on highways without sidewalks (Hou et al., 2024).

As exemplified in Figure 2, this sub-package for metadata analysis is built upon the basic metadata provided by Mapillary (such as capture time, camera parameters, and spatial coordinates) and further augments it by computing additional features. The package provides solutions for comprehensive metadata analysis at three levels: 1) image-level, 2) grid-level, and 3) street-level analyses.

At the most granular level, users can compute the indicators listed in Table 3 for each of the SVI points (i.e., spatial locations of SVI) by inputting the CSV file outputted from the downloading sub-package after downloading SVI. These indicators extend beyond the basic Mapillary metadata to provide more detailed insights about each image.

For the street- and grid-level analyses, we match each SVI point with the nearest street and the corresponding grid cell to compute indicators at the aggregated level (see Table 4), which are commonly used in use cases (e.g., average sky view factor at the district scale) and could also be useful for data quality analyses such as spatial and temporal coverage (Hou and Biljecki, 2022). This sub-package also contributes to the standardization of data properties by enabling researchers to explicitly query SVI based on both the original Mapillary metadata and our augmented indicators, allowing them to fetch only images suitable for their needs.

3.3. Transformation

Transformation of SVI is also important because different image projections and data formats are required for different purposes (Biljecki et al., 2023). Our sub-package for image transformation offers options to transform panoramic imagery into two further view types: 'perspective' and 'fisheye'. An overview and definitions of different image projections are provided in Figure 3.

The left side of Figure 3 illustrates the four fisheye projection methods included in this sub-package: 'orthographic', 'equisolid angle', 'equidistant', and 'stereographic'. These projection methods have been employed across diverse research fields (Schneider et al., 2009). For example, the orthographic projection has been

Table 3: A list of metadata indicators at the image level.

Attribute	Description
year	Year of the image collection.
month	Month of the image collection.
day	Day of the image collection.
hour	Hour of the image collection.
day_of_week	Day of the week of the image collection.
daytime_nighttime	Daytime or nighttime based on astral package.
season	Season determined by hemisphere and month.
relative_angle	Camera angle relative to nearest street segment.
h3_id	Uber's H3 hexagon IDs (zoom level 0-15).
speed_kmh	Speed in km/h, computed between points in a sequence.

Table 4: A list of metadata indicators at the grid and street level.

Attribute	Description
coverage	Coverage of user-defined buffers from SVI points in percentage.
count	Number of SVI points.
days_elapsed	Days between the oldest and newest SVI.
most_recent_date	Most recent date of SVI.
oldest_date	Oldest date of SVI.
number_of_years	Number of unique years.
number_of_months	Number of unique months.
number_of_days	Number of unique days.
number_of_hours	Number of unique hours.
number_of_days_of_week	Number of unique days of the week.
number_of_daytime	SVIs collected during daytime.
number_of_nighttime	SVIs collected during nighttime.
number_of_spring	SVIs collected during spring.
number_of_summer	SVIs collected during summer.
number_of_autumn	SVIs collected during autumn.
number_of_winter	SVIs collected during winter.
average_compass_angle	Average compass angle of SVI.
average_relative_angle	Average relative angle to nearest street segment.
average_is_pano	Average ratio of panoramic SVIs.
number_of_users	Number of unique users contributing to SVI.
number_of_sequences	Number of unique sequences containing SVI.
number_of_organizations	Number of unique organizations contributing.
average_speed_kmh	Average speed in km/h.
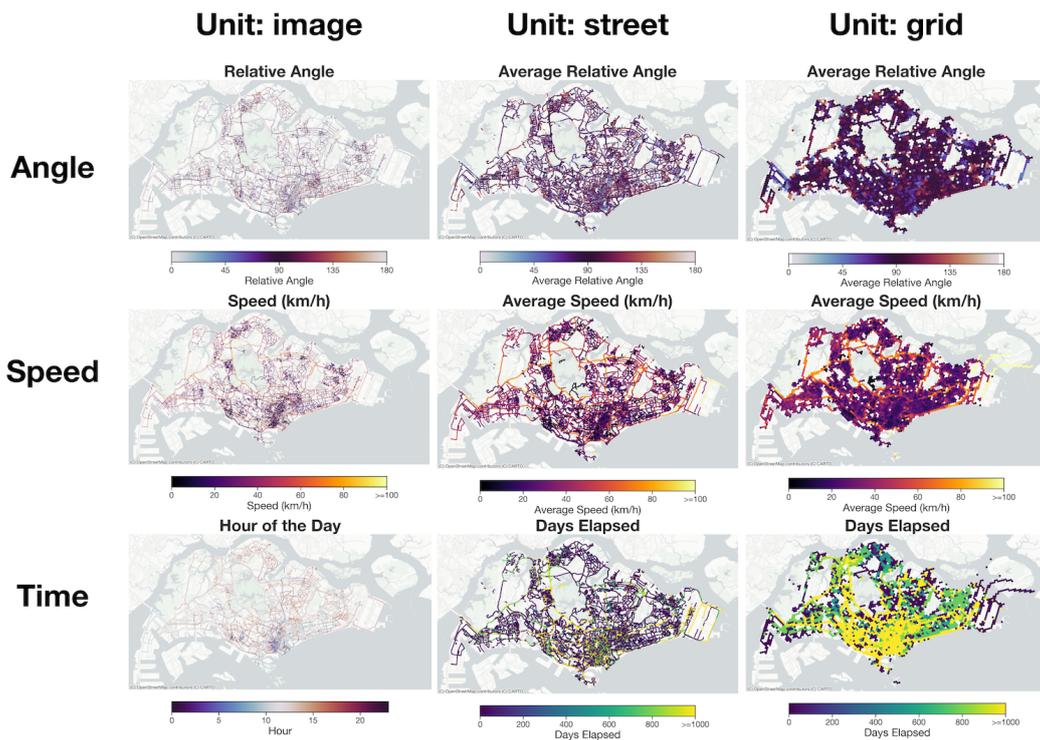
Figure 2: A matrix of maps representing various metadata. Columns show three different types of units of analysis: 1) image, 2) street, and 3) grid, and rows display three exemplary types of metadata: 1) angle, 2) speed, and 3) time.
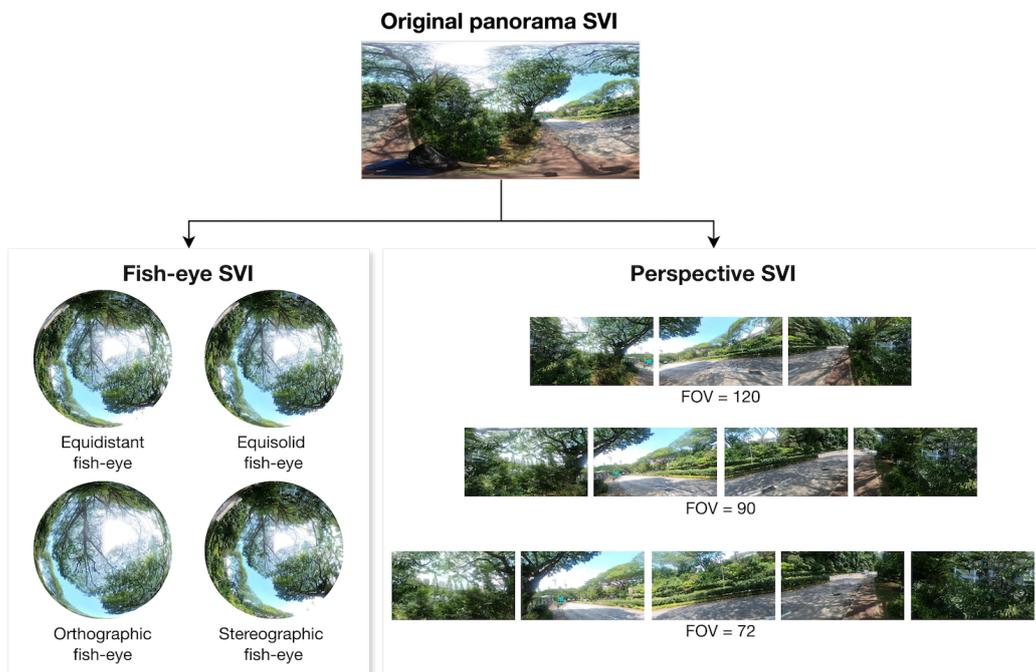
Figure 3: This figure illustrates different configurations of transformation from panorama SVI to fish-eye and perspective SVI. Source of imagery: Mapillary.

instrumental in calculating the sky view factor for urban microclimate studies (Miao et al., 2020; Carrasco-Hernandez et al., 2015), as the ratio of sky pixels in an upper fisheye photo with orthographic projection directly correlates to the diffuse solar irradiance absorbed by a horizontal surface (Ivanova and Gueymard, 2019). This method also facilitates long-wave radiation transfer calculations between surfaces and surrounding objects, as the ratio of sky pixels to a surface's normal vector corresponds to the surface radiation (Watanabe et al., 2013). For studies examining visual perceptions such as openness and oppressiveness (Asgarzadeh et al., 2012; Miyake et al., 2019), the equisolid angle projection proves valuable because its pixel ratio directly corresponds to solid angles from a viewpoint. The equidistant projection finds its primary application in celestial position detection, as the distance between the center point and any other point in the fisheye photo directly corresponds to the angle in three-dimensional space (Garcia-Gil and Ramirez, 2019).

The right side of Figure 3 displays three configurations for perspective SVI with fields of views (FOVs) at 72, 90, and 120 degrees. A perspective view is created by cropping an area with a specific direction and angle of view, which can be set as input arguments. This view type simulates human vision when the angle of view accounts for the real ranges of human sight, making it suitable for both objective measurements of the environment and studies of subjective perception (Kruse et al., 2021; Wang et al., 2019; Zhou et al., 2019; Meng et al., 2020). The FOV setting significantly affects the resulting image: a lower FOV (e.g., 72 degrees in Figure 3) creates less distortion but limits the captured area, whereas a higher FOV (e.g., 120 degrees) captures more area but introduces more distortion.

These diverse projection options enable researchers to select methods that best suit their specific research requirements and fields of study. The flexibility based on the parametrization of key settings contributes to meeting various analytical demands across different disciplines while ensuring reproducibility.

Our transformation sub-package also provides point cloud generation capabilities based on depth estimation in the computer vision sub-package, thus enabling advanced spatial analysis of street-level imagery. These techniques have evolved rapidly, particularly driven by autonomous vehicle applications (Li et al., 2020), and can be utilized for various urban modeling applications (Biljecki et al., 2015).

For point cloud generation, we implemented the methodology of Cavallo (2015), which combines depth estimates with RGB color information. The process projects pixels into spherical coordinates with depth-determined radii and then converts them to Cartesian coordinates for three-dimensional positioning. As demonstrated in Figure 4, this workflow transforms panoramic SVI into detailed point clouds, enabling researchers to conduct advanced spatial analyses such as sidewalk width

(a) A sample panoramic SVI in London.

(b) Depth map estimated via Depth Anything model.

(c) Point cloud generated (top view).
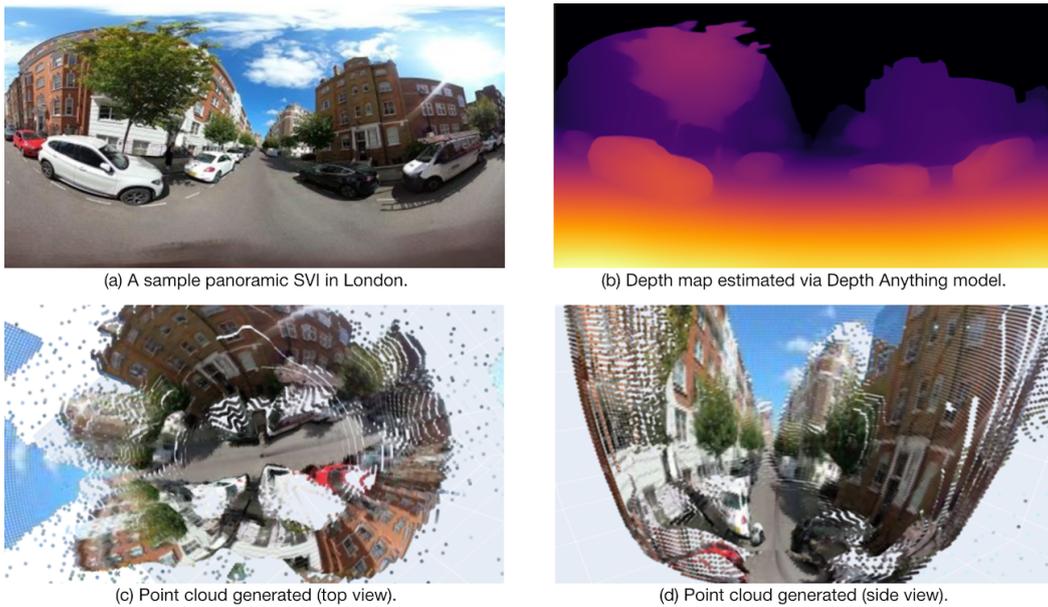
(d) Point cloud generated (side view).

Figure 4: Example of depth estimation for a Mapillary SVI using the Depth Anything model and the resulting point cloud generation. Source of imagery: Mapillary.

measurement and tree volume estimation (Tu et al., 2024; Orozco Carpio et al., 2024). The visualization shows both top and side views, revealing the scene's outline, direction, and relative distances between street objects.

3.4. Computer vision

Computer vision techniques have evolved into being indispensable for SVI data analysis. Our framework facilitates the use of well-established computer vision models by providing them with unified input and output formats. These models include object detection, scene segmentation, and feature extraction algorithms (see Table 5). Figure 5 demonstrates sample outputs from these models, and their practical applications are further discussed in Section 4.

Our package provides both semantic and panoptic segmentation capabilities. While semantic segmentation classifies pixels into categories, panoptic segmentation additionally identifies unique object instances, enabling tasks like counting individual trees or cars — a crucial feature for scalable urban analysis (Sánchez and Labib, 2024). We use the Mask2Former model (Cheng et al., 2022), pre-trained on two key urban datasets. The model achieves 84.5% mIoU for semantic and 66.6% for panoptic segmentation on Cityscapes, 64.7% mIoU for semantic and 45.5% for panoptic segmentation on Mapillary Vistas.

As demonstrated in Figure 6, the segmentation results clearly identify and color-code various street features including buildings, roads, sidewalks, and vegetation. The package outputs pixel ratios for all labels to a CSV/JSON file, facilitating

15

Category	Class	Value
Segmentation	Sky	10.3 / 100
Segmentation	Vegetation	17.7 / 100
Segmentation	Building	0 / 100
Scene	Street	56.5 / 100
Perceptions	Safer	6.2 / 10
Perceptions	Livelier	5.1 / 10
Perceptions	More beautiful	7.0 / 10
Perceptions	Wealthier	5.8 / 10
Perceptions	More boring	5.4 / 10
Perceptions	More depressing	3.2 / 10
Low-level	Hue	133.54 / 360
Low-level	Saturation	22.5 / 100
Low-level	Lightness	44.5 / 100

Figure 5: An example of scores and indicators on the right computed for the image on the left. Source of the image: Mapillary.

Figure 6: Panoptic segmentation results of five Mapillary images using Cityscapes and Mapillary Vistas pre-trained models. The dual-model approach enhances adaptability to diverse urban contexts: Mapillary Vistas offers specialized classes (e.g., bike lanes) for detailed analysis, while Cityscapes provides broader classes for general urban assessment.

16

Table 5: Summary of computer vision models integrated in ZenSVI.

Model Type	Model Purpose	Description
Semantic/Panoptic Segmentation	Semantic Information Quantification	Mask2Former model pre-trained on Cityscapes and Mapillary Vistas datasets. Achieves mean IoU of 84.5% (semantic) and 64.7% (panoptic) for semantic segmentation and 66.6% and 45.5% for panoptic segmentation (Cheng et al., 2022).
Image Classification	Perception Classification	Place Pulse 2.0 model trained using ResNet50 architecture. Measures human perception of streetscapes across six dimensions. Achieves an average top-2 accuracy of 66.03% (Dubey et al., 2016).
Image Classification	Scene Classification	Places365 model trained using ResNet architecture. Classifies 365 scene categories and 102 scene attributes. Achieves 85.08% precision for the top-5 scene categories (Zhou et al., 2017).
Image Classification	Weather Classification	Model trained on Hou et al. (2024) dataset. Classifies weather in SVI as clear, cloudy, foggy, rainy, or snowy (Hou et al., 2024).
Image Classification	Glare Detection	Model trained on Hou et al. (2024) dataset. Detects glare in SVI (True or False) (Hou et al., 2024).
Image Classification	Time of Day Classification	Model trained on Hou et al. (2024) dataset. Classifies time of day in SVI as day, dusk/dawn, or night (Hou et al., 2024).
Image Classification	Panorama Detection	Model trained on Hou et al. (2024) dataset. Detects if SVI is a panorama (True or False) (Hou et al., 2024).
Image Classification	Platform Classification	Model trained on Hou et al. (2024) dataset. Classifies platform of SVI as cycling surface, driving surface, fields, railway, tunnel, or walking surface (Hou et al., 2024).
Image Classification	Quality Classification	Model trained on Hou et al. (2024) dataset. Classifies quality of SVI as good, slightly poor, or very poor (Hou et al., 2024).
Image Classification	Reflection Detection	Model trained on Hou et al. (2024) dataset. Detects reflection in SVI (True or False) (Hou et al., 2024).
Image Classification	View Direction Classification	Model trained on Hou et al. (2024) dataset. Classifies view direction of SVI as front/back or side (Hou et al., 2024).
Low-level Features Extraction	Edge Detection	Canny, SobelX, SobelY, and Laplacian techniques included. Detects edges of objects in images (Canny, 1986; Sobel and Feldman, 1968; Marr et al., 1997).
Low-level Features Extraction	Blob Detection	Identifies regions in the image that differ in properties like brightness or color compared to surrounding regions (Lindeberg, 1998).
Low-level Features Extraction	Blur Detection	Determines the blurriness of an image for quality control (Gonzalez and Woods, 2006).
Low-level Features Extraction	Color Space Analysis	Analyzes color characteristics in hue, saturation, and lightness.
Image Embedding	Feature Extraction	Converts images into vector representation using pre-trained models like ResNet, VGG, DenseNet, AlexNet, and EfficientNet for various tasks such as image retrieval, clustering, and classification (He et al., 2016; Simonyan and Zisserman, 2014; Huang et al., 2017; Krizhevsky et al., 2012; Tan and Le, 2019).

quantitative analysis (Ibrahim et al., 2020). This functionality enables measurement of street elements across different locations and periods (Ye et al., 2019; Ramírez et al., 2021; Ma et al., 2021), supporting applications such as green view index calculation — a measure of visible greenery at eye-level (Li et al., 2015; Yang et al., 2009).

Our package provides several pre-trained image classification models, with particular emphasis on street-level perception analysis. A key model is trained on the Place Pulse 2.0 dataset (Dubey et al., 2016), which comprises 110,988 images from 56 cities across 28 countries. This dataset captures the human perception of streetscapes along six dimensions: depressing, boring, beautiful, safe, lively, and wealthy, based on volunteer-provided labels. Following the methodology of Kang et al. (2023), we implemented a ResNet50 architecture where perceptual scores are categorized into five classes. The models achieved an average top-2 accuracy of 66.03% across all six perceptual dimensions. When processing input images,

the package generates a five-dimensional probability vector for each perceptual category, which is then converted into final dimensional scores ranging from 0 to 10. An existing model used in Global Streetscapes (Hou et al., 2024) was also added for benchmarking purposes. This model has a Vision Transformer (ViT) architecture and also generates perception scores, ranging from 0 to 10, with an average accuracy ranging from 61.6% to 77.1% for all six perceptual dimensions.

This perceptual analysis framework has found widespread applications, from urban change tracking (Liang et al., 2023; Ito et al., 2024b) to public housing evaluation (Wang et al., 2024c). Figure 7 presents ranked image samples across the six perceptual dimensions and reveals that locations featuring diverse urban elements (e.g., buildings, vehicles, vegetation) typically score higher in safety and liveliness metrics. Conversely, construction sites and monotonous street designs tend to receive higher scores for boringness and depressiveness. This human-centric evaluation approach enables researchers to quantitatively assess the built environment's perceptual qualities and integrate these insights into broader urban studies.

ZenSVI incorporates a scene classification model pre-trained on the Places365 dataset by Zhou et al. (2017), which achieves 85.08% top-5 precision in scene category classification. This comprehensive dataset contains 365 distinct scene categories (such as residential neighborhoods, building facades, and field roads) and 102 scene attributes (including features like biking areas, asphalt surfaces, and grass coverage). The model's usefulness has been validated through numerous research applications, from environmental feature extraction (Luo et al., 2023; Yang and Zhang, 2024) to targeted scene filtering for specific urban analyses (Chen et al., 2020a; de Oliveira et al., 2020; Seresinhe et al., 2017).

As demonstrated in Figure 8, our implementation excels at analyzing SVI across diverse urban and rural settings, including highways, residential neighborhoods, slums, campuses, and farms. For each image, the model generates both probability distributions across scene categories and class activation mapping (CAM) heatmaps. These CAM heatmaps visually reveal which regions of the image most significantly contribute to the scene classification (Zhang et al., 2018). For example, the model focuses on road infrastructure and traffic patterns when classifying highways, while for residential neighborhoods, it attends to architectural features, landscaping elements, and the overall spatial organization of buildings and vegetation. This functionality enables researchers to conduct detailed environmental analyses through automated scene recognition (Chen et al., 2020b; Yang and Zhang, 2024).

In our computer vision sub-package, we also integrated image classification models developed by Hou et al. (2024), trained on the manually labeled subset of

Figure 7: Image samples from Mapillary evaluated using human perception models, ranked from lower (left) to higher (right) human perception scores across six dimensions.
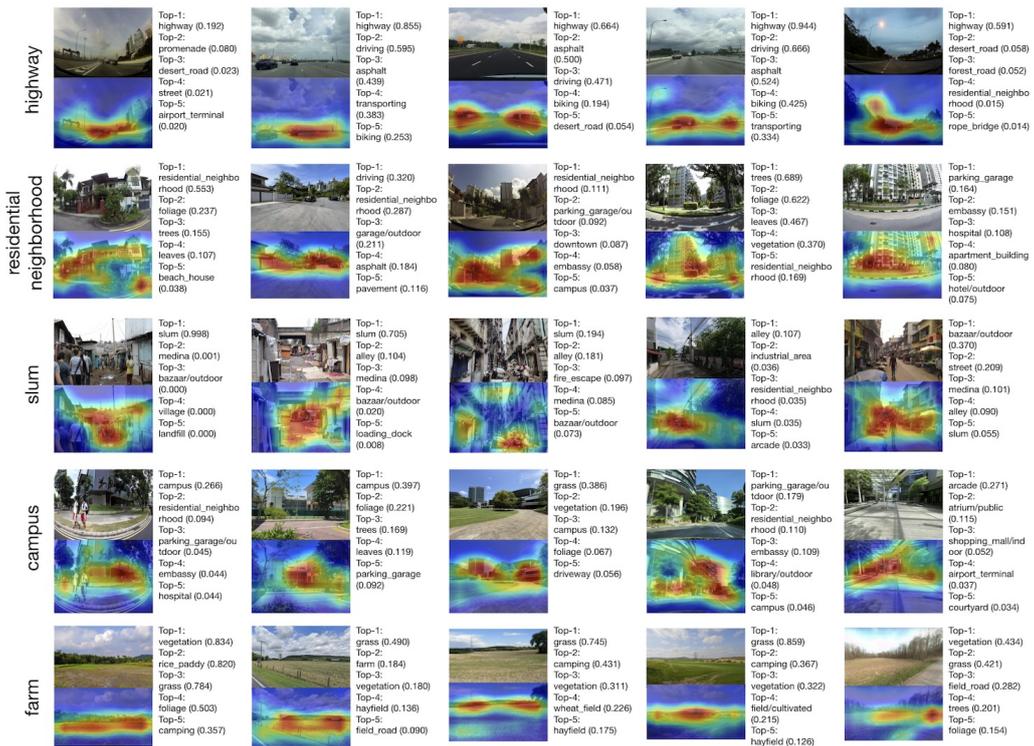
Figure 8: Places365 scene classification examples in five different settings: highway, residential neighborhood, slum, campus, and farm.

the NUS Global Streetscapes dataset comprising more than 10,000 images from over 400 cities across all continents. These images were sourced from Mapillary and KartaView platforms. Our sub-package has eight distinct classification models that analyze the following image attributes:

- Weather conditions (clear, cloudy, foggy, rainy, or snowy)

- Glare presence (True or False)

- Time of day (day, dusk/dawn, or night)

- Panorama status (True or False)

- Platform type (cycling surface, driving surface, fields, railway, tunnel, or walking surface)

- Image quality (good, slightly poor, or very poor)

- Reflection presence (True or False)

- View direction (front/back or side)
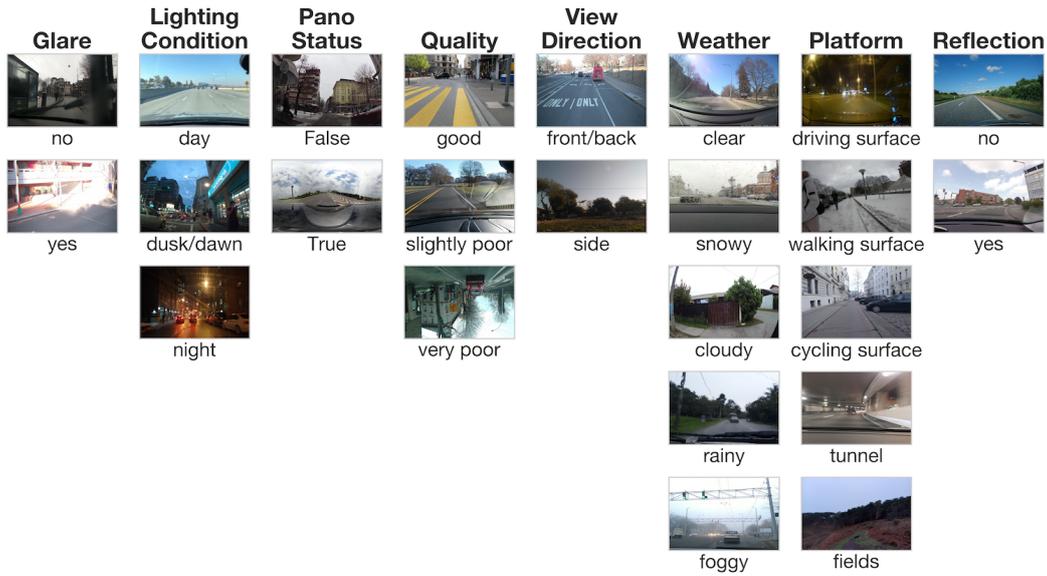
Figure 9: This figure shows examples of SVI classified into different conditions, under which they were captured. Columns from left to right show glare, lighting condition, pano status, quality, view direction, weather, platform, and reflection.

As illustrated in Figure 9, these models enable comprehensive filtering of SVI based on specific research requirements. For instance, researchers studying night-time pedestrian experiences can isolate high-quality images taken from walking surfaces during nighttime conditions (Hou et al., 2024). This systematic filtering capability helps mitigate potential biases in SVI analysis.

For object detection tasks, we integrate GroundingDINO in ZenSVI. This state-of-the-art model can detect arbitrary objects based on human inputs such as category names or referring expressions (Liu et al., 2023). Trained on several widely used object detection datasets — including COCO, O365, and Open Images — GroundingDINO has also been implemented in urban research, enabling large-scale analyses of built environment features from SVI (Liang et al., 2024; Ramalingam and Kumar, 2025). By specifying category names (e.g., 'tree', 'building', or multiple objects such as 'person . car .'), this open-set detector can generate bounding boxes for the identified objects in images. This open-set capability enables the quantification of urban elements and offers flexibility for a wide range of applications.

Our sub-package also offers image embedding capabilities, leveraging various pre-trained models including ResNet, VGG, DenseNet, AlexNet, and EfficientNet (He et al., 2016; Simonyan and Zisserman, 2014; Huang et al., 2017; Krizhevsky et al., 2012; Tan and Le, 2019). These models transform SVI into multi-dimensional vector representations that capture their essential features. The embedding process extracts high-level features from the models' final layers. For instance, ResNet
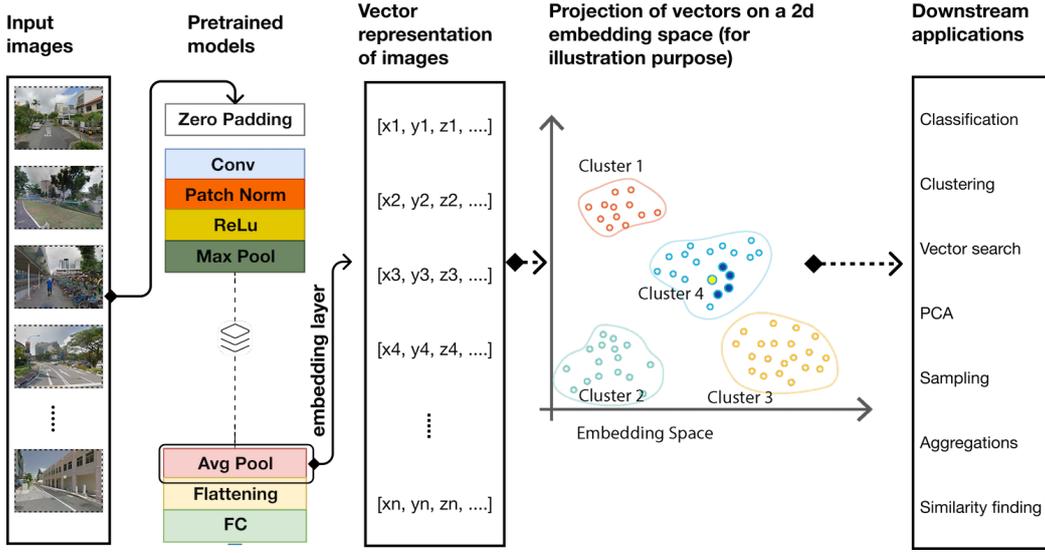
Figure 10: Extracting image embeddings from pre-trained Models. The image embeddings are extracted from the last layer of the pre-trained models, capturing the high-level features of the images. The embeddings are then used to calculate the similarity between images and identify clusters of images with similar features.

generates a 2048-dimensional vector from its average pooling layer, as illustrated in Figure 10. To quantify image similarities, we implemented cosine distance metrics, defined as:

$$\text{Cosine Distance} = 1 - \frac{A \cdot B}{\|A\| \times \|B\|} \tag{1}$$

where A and B represent the embedding vectors being compared. This metric ranges from 0 (identical) to 1 (completely different), enabling efficient image comparison and clustering. As demonstrated in Figure 11, these embeddings can be projected into a two-dimensional space for visualization, where proximity between points indicates feature similarity. This capability facilitates various applications, including image retrieval (Jush et al., 2023), clustering (Abrar et al., 2023), and pattern identification (Ypsilantis et al., 2023) in SVI. The technique proves particularly valuable for efficiently sampling representative images while eliminating redundant ones in downstream applications.

The sub-package also supports the following two types of depth estimation:

- Relative depth estimation: Returns normalized distances (0-255) for each pixel using the Dense Prediction Transformer (Ranftl et al., 2021).

- Absolute depth estimation: Provides distances in meters using the Depth Anything model (Yang et al., 2024).
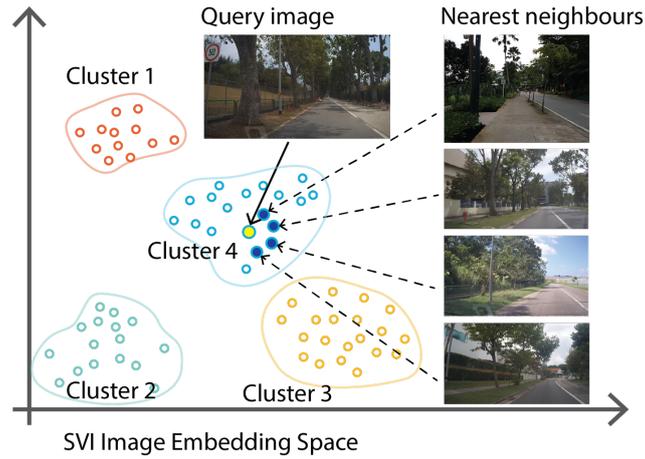
Figure 11: Illustration of image embeddings in embedding space. Each image can be represented as a multi-dimensional vector. Each vector can be projected in the embedding space, the distance between each vector in the space indicates the similarity between features of the images.

Lastly, our computer vision sub-package includes comprehensive capabilities for extracting low-level image features, which are crucial for applications such as human perception prediction (Guan et al., 2021; Dubey et al., 2016). The package implements several key feature detection techniques:

Edge Detection: We provide four distinct edge detection methods:

- Canny detection for precise edge localization (Canny, 1986).

- SobelX and SobelY detection for directional edge identification (horizontal and vertical respectively) (Sobel and Feldman, 1968).

- Laplacian detection for emphasizing rapid intensity changes and identifying blobs or junctions (Marr et al., 1997). The Laplacian operator is defined as:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{2}$$

which is implemented using the following discrete kernel:

$$K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

Additional Feature Analysis:

- Blob detection for identifying regions with distinct properties (brightness, color) relative to their surroundings (Lindeberg, 1998).

- Blur detection for image quality assessment (Gonzalez and Woods, 2006).

- Color space analysis in HSL (Hue, Saturation, Lightness) format, providing a more intuitive representation of color characteristics compared to RGB space.

3.5. Visualization

The visualization sub-package, built with Matplotlib (Hunter, 2007), provides four high-level plotting functions to streamline the presentation of SVI analysis for quick initial data exploration. ZenSVI supports histogram generation for discrete data visualization, such as object detection counts, and kernel density estimation plots for continuous variables, such as semantic segmentation ratios. Users can create image grids to display SVI samples, either randomly selected or sorted by specified variables.

For spatial visualization, the package implements three map formats: point maps showing individual SVI locations, line maps using OpenStreetMap street segments (Boeing, 2017), and hexagonal maps using Uber's H3 grid (Uber, 2018). In both line and hexagonal maps, SVI values are aggregated to the nearest features, defaulting to point counts when no specific variable is specified.

These visualization tools effectively communicate spatial patterns in urban environments. For example, Figure 12 reveals higher building view indices in the central business district and residential areas, elevated vegetation indices in central forested regions, and increased sky view indices along coastal and riverfront areas. The distribution plots highlight the varying patterns of transportation modes and view indices across the study area. All the functions return figure and axis instances from the Matplotlib package so that the users can further modify the plots flexibly.

4. Examples and case study

In this section, we conducted a comprehensive case study to showcase examples of how our package can be used to help researchers in one of the most typical studies by exhibiting each functionality in our package. We used Singapore as the study site and aim to investigate the daytime urban characteristics by conducting k-means clustering based on the features extracted from SVI obtained from Mapillary. This analysis was chosen because this type of analysis with SVI has been implemented for various purposes by researchers (Fan and Biljecki, 2024; Gao et al., 2024; Zhang et al., 2023; Liang et al., 2023; Wang et al., 2024a), and it also includes data engineering processes such as filtering unsuitable imagery. This process was conducted as follows:

1. Download SVI from Mapillary.

2. Apply computer vision models to extract features.

Building in the Study Area
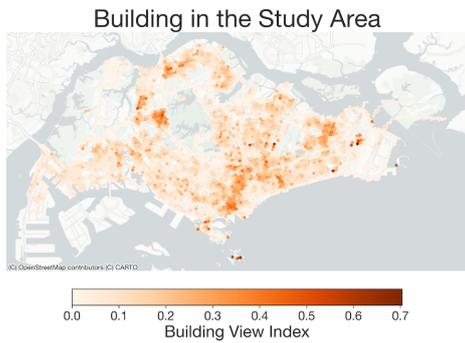
(a) A map of building view index in Singapore.

Vegetation in the Study Area
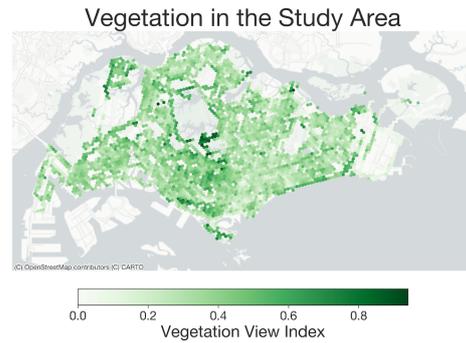
(b) A map of vegetation view index in Singapore.
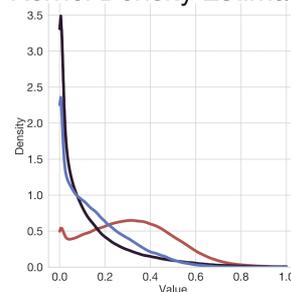

Sky in the Study Area
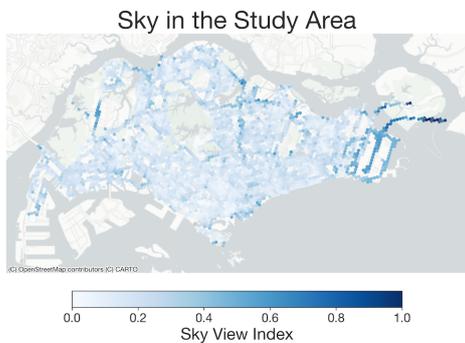
(c) A map of sky view index in Singapore

Histogram
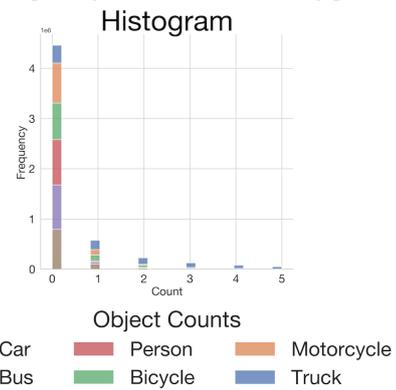
Object Counts

(d) Frequency distribution of detected urban objects (vehicles, people, and bicycles) in Singapore street-level imagery. The histogram shows the count distribution for different object categories, with most observations having 0-1 instances per image.

Kernel Density Estimation

View Index

(e) Probability density distribution of view indices across Singapore. The plot shows the relative frequency of different index values [0, 1] for buildings, vegetation, and sky visibility, highlighting the varying urban morphology patterns. Building views show high density at lower values, while sky views demonstrate a bimodal distribution.

Figure 12: Examples of how our visualization sub-package can be used to plot maps, histograms, and kernel density estimation plots.

3. Conduct quality assessment and remove unsuitable SVI.

4. Run *k*-means clustering analysis.

4.1. Download and quality control

We downloaded SVI from Mapillary by simply inputting 'Singapore' to the `download_svi` function, which was then passed to OSMnx to retrieve the boundary of Singapore. This yielded 932,596 images as of June 2024. Then, we applied the following computer vision models to extract features used for the subsequent cleaning process and analysis:

- Image classification model trained on Places365.

- Image classification model trained on PlacePulse.

- Image classification model trained on NUS Global Streetscape dataset (e.g., image quality, platform, etc).

- Semantic segmentation trained on Mapillary Vistas.

- Low-level image processing (e.g., blur detection, etc).

In addition to these extracted features, we also used metadata from Mapillary for the cleaning process, such as the speed and timestamp of the device when the image was taken. After extracting the necessary features, the dataset was cleaned using a two-stage filtering approach to ensure images were both of high quality and suitable for our specific use case. For this case study focusing on characterizing the urban environment during daytime, we applied the following criteria:

1. **Image Quality Requirements**

   a) **Overall Quality**: Images labeled as "very poor" by the image quality classification model were excluded.

   b) **Sharpness**: Blurry images with Laplacian variance over 100 were removed.

2. **Image Properties for Use Case Suitability**

   a) **Visual Complexity**: A minimum visual complexity score of 1.0 was required, calculated using the Shannon Index:

$$H = -\sum_{i=1}^{n} p_i \ln(p_i) \tag{4}$$

where p_i is the normalized ratio of pixels for segment i, and n is the number of different segments in the image. The normalization ensures $\sum_{i=1}^{n} p_i = 1$. This measure, originally from information theory, is now used to clean up SVI datasets in urban studies (Yap et al., 2023).

b) **Speed Criterion**: The device speed was required to be ≤ 200 km/h when available in the metadata. We applied this step to remove outliers, such as images taken from an airplane, which may also indicate positional inaccuracy due to GPS lag at high velocities.

c) **Lighting Conditions**: Images were filtered to exclude nighttime shots based on the NUS Global Streetscape's lighting model and the metadata.

d) **View Direction**: Only images labeled as "front/back" viewing by the NUS Global Streetscape's image quality classification model were retained.

e) **Platform**: Based on the NUS Global Streetscape's platform classification model, we only kept images labeled as being recorded from the driving surface.

By following these criteria, we targeted the daytime SVI with good quality and sufficient visual information. For each criterion, missing values (`NA`) were treated as suitable to avoid excessive data loss. The final suitability flag (`is_suitable`) was computed as the logical conjunction of all individual criteria, meaning an image needed to pass all applicable filters to be considered suitable. After the cleaning process, 270,864 images — about 30% of the initial dataset — remained. As displayed in Figure 13, our suitable SVI points cover most of Singapore, and the majority of the images were taken within the last 5 years. Some examples of unsuitable images are also shown in Figure 13, such as nighttime images, images with blur, low visual complexity, and low quality, and images taken at extremely high speeds. We applied rather strict criteria in this case by specifying the view direction and platform; however, these criteria can be flexibly adjusted for each use case as required by users. This filtering approach ensures a consistent and high-quality dataset by removing potentially unsuitable images in the crowdsourced images.

4.2. Clustering analysis

For the clustering analysis, we used indicators relevant to urban characteristics from Places365 classification (i.e., highway, residential neighborhood, construction site, tree farm, forest path, and forest road), PlacePulse classification (i.e., beautiful, boring, depressing, lively, safe, and wealthy) and semantic segmentation (i.e., sky,

Figure 13: A map of suitable SVI points in Singapore downloaded from Mapillary. The color represents the year of being recorded. This map also shows some examples of unsuitable images removed from the dataset, including nighttime images, images with blur, low visual complexity, and low quality, and images taken at extremely high speeds.

building, vegetation, road, and sidewalk). These indicators were aggregated by taking the mean values to Uber's H3 grid (Uber, 2018) at a resolution of nine. After the z-score standardization, we applied *k*-means clustering with k=5 clusters. This number was chosen through a combination of quantitative and qualitative considerations: while the elbow method suggested a range of potentially optimal *k* values, we selected five clusters as it provided a balance between statistical validity and human-interpretable differentiation of urban streetscapes.

Figure 14 illustrates the five clusters formed from *k*-means clustering. The ridge plots on the leftmost column display distributions of some relevant indicators in the respective clusters and the images on the right side of the ridge plots are randomly selected examples of SVI for the clusters. As shown in the sample images, Cluster 1 represents open highways with high sky view factors, probability of being classified as highway, as well as depressingness and boringness scores, and low green view index as well as safety and liveliness scores. Cluster 2 includes dense urban areas, characterized by tall buildings, urban infrastructure, and commercial developments. Cluster 3 is distinguished by green highways and corridors with abundant roadside vegetation. Cluster 4 shows mixed urban infrastructure with medium-rise buildings and moderate vegetation, while Cluster 5 represents residential neighborhoods with mature tree cover and lower building density. The spatial distribution of these clusters across Singapore in Figure 15 reveals clear patterns, with Cluster 1 (blue) following major highway networks, Cluster 2 (red) concentrated in urban centers, Cluster 3 (green) along green corridors, and Clusters 4 and 5 (light green and light blue respectively) distributed across residential areas throughout the city.

5. Extensibility and export

5.1. Extending ZenSVI

Our package is designed in a way that allows future development and extensions that suit a specific context such as a data stream or application. For example, the download sub-package can be extended further by including platforms other than Mapillary and KartaView. To demonstrate this extensibility, we added another source of SVI data — an official SVI platform supported by the City of Amsterdam[9]. Other commercial platforms, crowdsourced services, and city-led initiatives can be similarly incorporated into ZenSVI in the future.

The computer vision sub-package can be extended with new models, such as zero-shot segmentation (Kirillov et al., 2023) and various generative models for

---

[9] https://amsterdam.github.io/projects/open-panorama/

Cluster 1

Cluster 2

Cluster 3

Cluster 4

Cluster 5

■ Sky	■ Vegetation	■ Residential Neighborhood	■ More Beautiful	■ More Depressing	■ Safer
■ Building	■ Road	■ Highway	■ More Boring	■ Livelier	■ Wealthier
Semantic Segmentation (Mapillary Vistas)		**Scene Classification** (Places365)	**Human Perception** (Place Pulse 2.0)		

Figure 14: Representative street-level images and their corresponding feature profiles for the five visual clusters identified in Singapore. The stacked plots on the left show the relative intensity of each feature (semantic segmentation based on Mapillary Vistas, scene classification based on Places365, and human perception based on PlacePulse) for each cluster. Cluster 1 represents open highways, Cluster 2 shows dense urban areas, Cluster 3 captures green highways, Cluster 4 represents mixed urban infrastructure, and Cluster 5 shows residential areas with mature vegetation. Source of imagery: Mapillary.
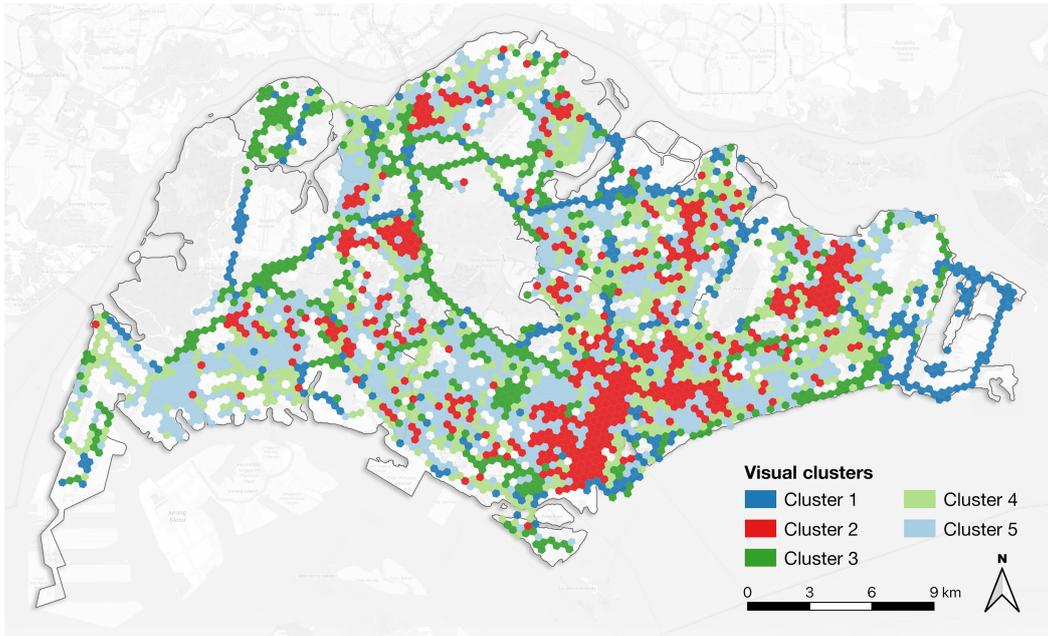
Figure 15: Spatial distribution of the five visual clusters across Singapore. The clusters show distinct spatial patterns: Cluster 1 (blue) primarily follows major highways, Cluster 2 (red) concentrates in dense urban areas and town centers, Cluster 3 (green) follows green corridors, while Clusters 4 (light green) and 5 (light blue) represent different types of residential areas distributed throughout the city.

text-to-image (Saharia et al., 2022), image captioning (Li et al., 2022a), image-to-sound (Zhuang et al., 2024), image-to-video (Singer et al., 2022), image-to-image (Ito et al., 2024c), and inpainting (Lugmayr et al., 2022). The visualization sub-package can also be expanded with additional plot types like alluvial plots for time-series analysis. While leveraging ecosystems like HuggingFace, ZenSVI provides an abstraction layer that simplifies data handling while supporting models from multiple platforms through a consistent interface.

Lastly, ZenSVI's software components are released under the MIT license, which grants users extensive freedom to use, modify, distribute, and sell the software without significant restrictions. This permissive license allows users to modify and incorporate the code into their projects, whether open-source or proprietary, with only minimal requirements of preserving the copyright notice and license terms. However, it is important to note that the utilization of ZenSVI in conjunction with Mapillary or KartaView datasets necessitates adherence to their respective Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licensing protocols. These protocols mandate appropriate attribution and stipulate that any derivative works must be distributed under identical licensing conditions.

5.2. *Import and export of the data, integration, and downstream analyses thanks to ZenSVI*

The modularized structure of ZenSVI, with its independent download and other processing sub-packages, allows users to flexibly import and export data that work with external data sources and software tools. For example, users can apply ZenSVI's transformation and computer vision sub-packages to SVI from other sources not covered in the current version, such as Baidu Street View and Bing Maps. Most of the functions in ZenSVI also allow users to save outputs directly to external files. Thus, it is easy to integrate the analysis in our package with other packages. For example, users can download the SVI and perform semantic segmentation to compute the green view index, which can be saved to a CSV file. This design enables users to perform subsequent analysis and visualization in external software. For example, one can conduct a statistical analysis of the index in R and visualize the results in QGIS. Such a bridge facilitates conducting analyses and visualization in their most familiar environment while efficiently conducting the SVI analysis in ZenSVI.

6. Conclusion

SVI has gained popularity in many research domains for its scalability, spatial resolution, and capability to capture street-level information with high veracity and usability. Despite many applications demonstrated in the literature, the lack of standardized, cohesive, and reproducible software makes it difficult for other researchers to reproduce or expand existing methodologies that rely on such data and develop new ones. To fill this gap, we designed and implemented ZenSVI — an integrated Python package that incorporates a comprehensive set of tools, simplifying the process of SVI analysis and contributing to the acceleration of research relying on this emerging and powerful geospatial dataset.

The introduction of ZenSVI significantly advances the principles of open science and enhances research reproducibility in urban analytics. By offering a unified open-source solution, ZenSVI reduces the barrier to entry for researchers who may lack advanced technical skills, thus democratizing access to SVI analysis. The extensive documentation ensures that users can effectively replicate studies and build upon existing research, while the standardized workflow facilitates comparative studies across different contexts and regions.

The integration of diverse functionalities into a one-stop Python package enables users to conduct end-to-end SVI analysis with minimal code. For example, users can analyze the perception of safety in areas like `Fifth Avenue, New York` by downloading SVIs, converting panoramic images to 90-degree field-of-view

perspectives using `transform_images`, calculating image greenness and openness via the `segment` method, and determining safety perception scores using `calculate_score`. This streamlined process demonstrates ZenSVI's capability to facilitate detailed urban studies and support informed decision-making.

Moving forward, ZenSVI's extensible architecture allows for continuous enhancement through new features and models. Future developments will focus on implementing text-to-image generative models and multi-modal translation capabilities, as well as expanding support for satellite, aerial, and CCTV imagery analysis. We also plan to create ready-to-use open datasets of SVI-derived indicators globally, further supporting users without extensive technical experience. Additionally, ZenSVI will serve as a platform for researchers to publish and share novel models, fostering collaboration and innovation in urban analysis and related fields.

Acknowledgments

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT in order to format tables and proofread. After using this tool/service, the author(s) reviewed

and edited the content as needed and take(s) full responsibility for the content of the published article.

References

Abrar, S., Sekmen, A., Samad, M.D., 2023. Effectiveness of Deep Image Embedding Clustering Methods on Tabular Data. doi:10.48550/arXiv.2212.14111, arXiv:2212.14111.

Ai, D., Jiang, G., Siew Kei, L., Li, C., 2018. Automatic Pixel-Level Pavement Crack Detection Using Information of Multi-Scale Neighborhoods. IEEE Access 6, 24452–24463. doi:10.1109/ACCESS.2018.2829347.

Aikoh, T., Homma, R., Abe, Y., 2023. Comparing conventional manual measurement of the green view index with modern automatic methods using google street view and semantic segmentation. Urban Forestry & Urban Greening 80, 127845.

Alhasoun, F., González, M., 2019. Streetify: using street view imagery and deep learning for urban streets development, in: 2019 IEEE International Conference on Big Data (Big Data), IEEE. pp. 2001–2006.

Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D., Mraz, A., Kashiyama, T., Sekimoto, Y., 2021. Deep learning-based road damage detection and classification for multiple countries. Automation in Construction 132, 103935. doi:10.1016/j.autcon.2021.103935.

Asgarzadeh, M., Lusk, A., Koga, T., Hirate, K., 2012. Measuring oppressiveness of streetscapes. Landsc. Urban Plan. 107, 1–11.

Bahrehdar, A.R., Adams, B., Purves, R.S., 2020. Streets of london: Using flickr and openstreetmap to build an interactive image of the city. Computers, Environment and Urban Systems 84, 101524.

Bansal, M., Sawhney, H.S., Cheng, H., Daniilidis, K., 2011. Geo-localization of street views with aerial image databases, in: Proceedings of the 19th ACM international conference on Multimedia, pp. 1125–1128.

Bendixen, M., Kanhema, T., Iversen, L.L., 2023. Putting africa on the map. Nature Africa doi:10.1038/d44148-023-00204-1.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning 215, 104217. doi:10.1016/j.landurbplan.2021.104217.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A., 2015. Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information 4, 2842–2889. doi:`10.3390/ijgi4042842`.

Biljecki, F., Zhao, T., Liang, X., Hou, Y., 2023. Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. International Journal of Applied Earth Observation and Geoinformation 122, 103385. doi:`10.1016/j.jag.2023.103385`.

Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139. doi:`10.1016/j.compenvurbsys.2017.05.004`.

Bromm, K.N., Lang, I.M., Twardzik, E.E., Antonakos, C.L., Dubowitz, T., Colabianchi, N., 2020. Virtual audits of the urban streetscape: Comparing the inter-rater reliability of GigaPan® to Google Street View. International Journal of Health Geographics 19. doi:`10.1186/s12942-020-00226-0`.

Cai, Q., Abdel-Aty, M., Zheng, O., Wu, Y., 2022. Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety. Transportation research part C: emerging technologies 135, 103541.

Callau, A.À., Albert, M.Y.P., Rota, J.J., Giné, D.S., 2019. Landscape characterization using photographs from crowdsourced platforms: content analysis of social media photographs. Open Geosciences 11, 558–571.

Campbell, A., Both, A., Sun, Q.C., 2019. Detecting and mapping traffic signs from google street view images using deep learning and gis. Computers, Environment and Urban Systems 77, 101350.

Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 679–698. doi:`10.1109/TPAMI.1986.4767851`.

Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., Zhang, Q., Qiu, G., 2018. Integrating Aerial and Street View Images for Urban Land Use Classification. Remote Sensing 10, 1553. doi:`10.3390/rs10101553`.

Carrasco-Hernandez, R., Smedley, A.R.D., Webb, A.R., 2015. Using urban canyon geometries obtained from google street view for atmospheric studies: Potential applications in the calculation of street level total shortwave irradiances. Energy Build. 86, 340–348.

Cavallo, M., 2015. 3d city reconstruction from google street view. URL: https://api.semanticscholar.org/CorpusID:26811584.

Chen, J., Zhou, C., Li, F., 2020a. Quantifying the green view indicator for assessing urban greening quality: An analysis based on Internet-crawling street view data. Ecological Indicators 113, 106192. doi:10.1016/j.ecolind.2020.106192.

Chen, M., Arribas-Bel, D., Singleton, A., 2020b. Quantifying the characteristics of the local urban environment through geotagged flickr photographs and image recognition. ISPRS International Journal of Geo-Information 9, 264.

Chen, S., Biljecki, F., 2023. Automatic assessment of public open spaces using street view imagery. Cities 137, 104329.

Chen, X., Meng, Q., Hu, D., Zhang, L., Yang, J., 2019. Evaluating greenery around streets using baidu panoramic street view images and the panoramic green view index. Forests 10, 1109.

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-attention Mask Transformer for Universal Image Segmentation. doi:10.48550/arXiv.2112.01527, arXiv:2112.01527.

Danish, M., Labib, S.M., Ricker, B., Helbich, M., 2025. A citizen science toolkit to collect human perceptions of urban environments using open street view images. Computers, Environment and Urban Systems 116, 102207. doi:10.1016/j.compenvurbsys.2024.102207.

de Mesquita, R.G., Ren, T.I., Mello, C.A.B., Silva, M.L.P.C., 2022. Street pavement classification based on navigation through street view imagery. AI & SOCIETY doi:10.1007/s00146-022-01520-0.

de Oliveira, W.B., Dorini, L.B., Minetto, R., Silva, T.H., 2020. OutdoorSent: Sentiment Analysis of Urban Outdoor Images by Using Semantic and Deep Features. ACM Transactions on Information Systems 38, 1–28. doi:10.1145/3385186, arXiv:1906.02331.

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C.A., 2016. Deep learning the city: Quantifying urban perception at a global scale, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer. pp. 196–212.

Fan, Z., Biljecki, F., 2024. Nighttime Street View Imagery: A new perspective for sensing urban lighting landscape. Sustainable Cities and Society 116, 105862. doi:10.1016/j.scs.2024.105862.

Fan, Z., Zhang, F., Loo, B.P.Y., Ratti, C., 2023. Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences 120, e2220417120. doi:10.1073/pnas.2220417120.

Fang, F., Zeng, L., Li, S., Zheng, D., Zhang, J., Liu, Y., Wan, B., 2022. Spatial context-aware method for urban land use classification using street view images. ISPRS Journal of Photogrammetry and Remote Sensing 192, 1–12.

Fujiwara, K., Ito, K., Ignatius, M., Biljecki, F., 2024. A panorama-based technique to estimate sky view factor and solar irradiance considering transmittance of tree canopies. Build. Environ. 266, 112071.

Gao, M., Guo, H., Liu, L., Zeng, Y., Liu, W., Liu, Y., Xing, H., 2024. Integrating street view imagery and taxi trajectory for identifying urban function of street space. Geo-spatial Information Science 0, 1–23. doi:10.1080/10095020.2024.2311866.

Garcia-Gil, G., Ramirez, J.M., 2019. Fish-eye camera and image processing for commanding a solar tracker. Heliyon 5, e01398.

Gonzalez, R.C., Woods, R.E., 2006. Digital Image Processing (3rd Edition). Prentice-Hall, Inc.

Gou, A., Zhang, C., Wang, J., 2022. Study on the identification and dynamics of green vision rate in jing'an district, shanghai based on deeplab v3+ model. Earth Science Informatics 15, 163–181.

Graser, A., 2019. MovingPandas: Efficient Structures for Movement Data in Python. GI_Forum – Journal of Geographic Information Science 7, 54–68. URL: https://hw.oeaw.ac.at?arp=0x003aba2b, doi:10.1553/giscience2019_01_s54.

Guan, W., Chen, Z., Feng, F., Liu, W., Nie, L., 2021. Urban perception: Sensing cities via a deep interactive multi-task learning framework. ACM Transactions on Multimedia Computing, Communications and Applications 17, 1–20. doi:10.1145/3424115.

Hara, K., Le, V., Froehlich, J., 2013. Combining crowdsourcing and google street view to identify street-level accessibility problems, in: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 631–640.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

He, N., Li, G., 2021. Urban neighbourhood environment assessment based on street view image processing: A review of research trends. Environmental Challenges 4, 100090.

Helbich, M., Danish, M., Labib, S., Ricker, B., 2024. To use or not to use proprietary street view images in (health and place) research? that is the question. Health & Place 87, 103244. doi:`10.1016/j.healthplace.2024.103244`.

Higgs, C., Liu, S., Boeing, G., Arundel, J., Lowe, M., Adlakha, D., Resendiz, E., Heikinheimo, V., Giles-Corti, B., Delclòs-Alió, X., Puig-Ribera, A., Vargas, J.C.B., Castillo-Riquelme, M., Jafari, A., Molina-García, J., Queralt, A., Cerin, E., Suel, E., Domínguez-Mallafré, M., Hernandez-Alcaraz, C., 2023. Global healthy and sustainable city indicators software (version 1). URL: `https://doi.org/10.25439/rmt.24760260.v1`, doi:`10.25439/rmt.24760260.v1`.

Hou, Y., Biljecki, F., 2022. A comprehensive framework for evaluating the quality of street view imagery. International Journal of Applied Earth Observation and Geoinformation 115, 103094.

Hou, Y., Quintana, M., Khomiakov, M., Ouyang, J., Ito, K., Wang, Z., Zhao, T., Biljecki, F., 2024. Global streetscapes – a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. ISPRS Journal of Photogrammetry and Remote Sensing .

Hu, Y., Qian, F., Yan, H., Middel, A., Wu, R., Zhu, M., Han, Q., Zhao, K., Wang, H., Shao, F., Bao, Z., 2024. Which street is hotter? street morphology may hold clues -thermal environment mapping based on street view imagery. Build. Environ. 262, 111838.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

Huang, J., Fei, T., Kang, Y., Li, J., Liu, Z., Wu, G., 2024. Estimating urban noise along road network from street view imagery. International Journal of Geographical Information Science 38, 128–155.

Hunter, J.D., 2007. Matplotlib: A 2d graphics environment. Computing in Science & Engineering 9, 90–95. doi:`10.1109/MCSE.2007.55`.

Ibrahim, M.R., Haworth, J., Cheng, T., 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. Cities 96, 102481.

Ito, K., Bansal, P., Biljecki, F., 2024a. Examining the causal impacts of the built environment on cycling activities using time-series street view imagery. Transportation Research Part A: Policy and Practice 190, 104286. doi:10.1016/j.tra.2024.104286.

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transportation Research Part C: Emerging Technologies 132, 103371. doi:10.1016/j.trc.2021.103371.

Ito, K., Kang, Y., Zhang, Y., Zhang, F., Biljecki, F., 2024b. Understanding urban perception with visual data: A systematic review. Cities 152, 105169. doi:10.1016/j.cities.2024.105169.

Ito, K., Quintana, M., Han, X., Zimmermann, R., Biljecki, F., 2024c. Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. International Journal of Geographical Information Science 38, 2514–2544. doi:10.1080/13658816.2024.2391969.

Ivanova, S.M., Gueymard, C.A., 2019. Simulation and applications of cumulative anisotropic sky radiance patterns. Solar Energy 178, 278–294.

Jiao, J., Wang, H., 2023. Forecasting Traffic Speed during Daytime from Google Street View Images using Deep Learning. Transportation Research Record 2677, 743–753. doi:10.1177/03611981231169531.

Jordahl, K., den Bossche, J.V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A.G., Farmer, C., Hjelle, G.A., Snow, A.D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L.J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, Leblanc, F., 2020. geopandas/geopandas: v0.8.1. doi:10.5281/zenodo.3946761.

Juhász, L., Hochmair, H.H., 2016. User contribution patterns and completeness evaluation of mapillary, a crowdsourced street level photo service. Transactions in GIS 20, 925–947.

Jush, F.K., Truong, T., Vogler, S., Lenga, M., 2023. Medical Image Retrieval Using Pretrained Embeddings. doi:10.48550/arXiv.2311.13547, arXiv:2311.13547.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., Zhu, X.X., 2018. Building instance classification using street view images. ISPRS Journal of Photogrammetry and Remote Sensing 145, 44–59. doi:10.1016/j.isprsjprs.2018.02.006.

Kang, Y., Abraham, J., Ceccato, V., Duarte, F., Gao, S., Ljungqvist, L., Zhang, F., Näsman, P., Ratti, C., 2023. Assessing differences in safety perceptions using geoai and survey across neighbourhoods in stockholm, sweden. Landscape and Urban Planning 236, 104768.

Kang, Y., Zhang, F., Gao, S., Lin, H., Liu, Y., 2020. A review of urban physical environment sensing using street view imagery in public health studies. Annals of GIS 26, 261–275. doi:10.1080/19475683.2020.1791954.

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. doi:10.48550/arXiv.2304.02643, arXiv:2304.02643.

Koo, B., Guhathakurta, S., Botchwey, N., 2022. How are Neighborhood and Street-Level Walkability Factors Associated with Walking Behaviors? A Big Data Approach Using Street View Images. Environment and Behavior 54, 211–241. doi:10.1177/00139165211014609.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25.

Kruse, J., Kang, Y., Liu, Y.N., Zhang, F., Gao, S., 2021. Places for play: Understanding human perception of playability in cities using street view images and deep learning. Comput. Environ. Urban Syst. 90, 101693.

Krylov, V.A., Kenny, E., Dahyot, R., 2018. Automatic discovery and geotagging of objects from street view imagery. Remote Sensing 10, 661.

Law, S., Paige, B., Russell, C., 2019. Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. ACM Transactions on Intelligent Systems and Technology 10, 54:1–54:19. doi:10.1145/3342240.

Law, S., Seresinhe, C.I., Shen, Y., Gutierrez-Roig, M., 2020. Street-frontage-net: urban image classification using deep convolutional neural networks. International Journal of Geographical Information Science 34, 681–707.

Li, J., Li, D., Xiong, C., Hoi, S., 2022a. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. doi:10.48550/arXiv.2201.12086, arXiv:2201.12086.

Li, S., Ma, S., Tong, D., Jia, Z., Li, P., Long, Y., 2022b. Associations between the quality of street space and the attributes of the built environment using large

volumes of street view pictures. Environment and Planning B: Urban Analytics and City Science 49, 1197–1211.

Li, X., Ning, H., Huang, X., Dadashova, B., Kang, Y., Ma, A., 2022c. Urban infrastructure audit: an effective protocol to digitize signalized intersections by mining street view images. Cartography and Geographic Information Science 49, 32–49.

Li, X., Ratti, C., Seiferling, I., 2018a. Quantifying the shade provision of street trees in urban landscape: A case study in boston, usa, using google street view. Landscape and Urban Planning 169, 81–91.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening 14, 675–685. doi:10.1016/j.ufug.2015.06.006.

Li, Y., Chen, Y., Rajabifard, A., Khoshelham, K., Aleksandrov, M., 2018b. Estimating Building Age from Google Street View Images Using Deep Learning (Short Paper), in: Winter, S., Griffin, A., Sester, M. (Eds.), 10th International Conference on Geographic Information Science (GIScience 2018), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany. pp. 40:1–40:7. doi:10.4230/LIPIcs.GISCIENCE.2018.40.

Li, Y., Ma, L., Zhong, Z., Liu, F., Cao, D., Li, J., Chapman, M.A., 2020. Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review. doi:10.48550/arXiv.2005.09830, arXiv:2005.09830.

Li, Y., Yabuki, N., Fukuda, T., 2022d. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. Sustainable Cities and Society 86, 104140.

Liang, J., Gong, J., Sun, J., Zhou, J., Li, W., Li, Y., Liu, J., Shen, S., 2017. Automatic Sky View Factor Estimation from Street View Photographs—A Big Data Approach. Remote Sensing 9, 411. doi:10.3390/rs9050411.

Liang, X., Chang, J.H., Gao, S., Zhao, T., Biljecki, F., 2024. Evaluating human perception of building exteriors using street view imagery. Building and Environment 263, 111875.

Liang, X., Zhao, T., Biljecki, F., 2023. Revealing spatio-temporal evolution of urban visual environments with street view imagery. Landscape and Urban Planning 237, 104802. doi:10.1016/j.landurbplan.2023.104802.

Lindeberg, T., 1998. Feature Detection with Automatic Scale Selection. International Journal of Computer Vision 30, 79–116. doi:`10.1023/A:1008045108935`.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J., et al., 2023. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. arXiv e-prints , arXiv–2303.

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., Woodcock, J., 2017. The Propensity to Cycle Tool: An open source online system for sustainable transport planning. Journal of Transport and Land Use 10. doi:`10.5198/jtlu.2016.862`.

Lu, B., Hu, Y., Yang, D., Liu, Y., Liao, L., Yin, Z., Xia, T., Dong, Z., Harris, P., Brunsdon, C., et al., 2023a. Gwmodels: A software for geographically weighted models. SoftwareX 21, 101291.

Lu, Y., 2019. Using google street view to investigate the association between street greenery and physical activity. Landscape and Urban Planning 191, 103435.

Lu, Y., Ferranti, E.J.S., Chapman, L., Pfrang, C., 2023b. Assessing urban greenery by harvesting street view data: A review. Urban Forestry & Urban Greening , 127917.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L., 2022. RePaint: Inpainting using Denoising Diffusion Probabilistic Models. doi:`10.48550/arXiv.2201.09865`, `arXiv:2201.09865`.

Luo, Y., Leong, C.T., Jiao, S., Chung, F.L., Li, W., Liu, G., 2023. Geo-Tile2Vec: A Multi-Modal and Multi-Stage Embedding Framework for Urban Analytics. ACM Transactions on Spatial Algorithms and Systems 9, 10:1–10:25. doi:`10.1145/3571741`.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., Ren, F., 2021. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. Cities 110. doi:`10.1016/j.cities.2020.103086`.

Mahabir, R., Schuchard, R., Crooks, A., Croitoru, A., Stefanidis, A., 2020. Crowdsourcing street view imagery: a comparison of mapillary and openstreetcam. ISPRS International Journal of Geo-Information 9, 341.

Mahajan, S., 2024. greenR: An open-source framework for quantifying urban greenness. Ecological Indicators 163, 112108. doi:`10.1016/j.ecolind.2024.112108`.

Maniat, M., Camp, C.V., Kashani, A.R., 2021. Deep learning-based visual crack detection using google street view images. Neural computing and applications 33, 14565–14582.

Marr, D., Hildreth, E., Brenner, S., 1997. Theory of edge detection. Proceedings of the Royal Society of London. Series B. Biological Sciences 207, 187–217. doi:`10.1098/rspb.1980.0020`.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., Raubal, M., 2023a. Trackintel: An open-source Python library for human mobility analysis. Computers, Environment and Urban Systems 101, 101938. doi:`10.1016/j.compenvurbsys.2023.101938`.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., Raubal, M., 2023b. Trackintel: An open-source python library for human mobility analysis. Computers, Environment and Urban Systems 101, 101938.

Meng, Y., Xing, H., Yuan, Y., Wong, M.S., Fan, K., 2020. Sensing urban poverty: From the perspective of human perception-based greenery and open-space landscapes. Comput. Environ. Urban Syst. 84, 101544.

Miao, C., Yu, S., Hu, Y., Zhang, H., He, X., Chen, W., 2020. Review of methods used to estimate the sky view factor in urban street canyons. Build. Environ. 168, 106497.

Mirowski, P., Grimes, M.K., Malinowski, M., Hermann, K.M., Anderson, K., Teplyashin, D., Simonyan, K., Kavukcuoglu, K., Zisserman, A., Hadsell, R., 2018. Learning to navigate in cities without a map, in: Neural Information Processing Systems (NeurIPS).

Miyake, H., Yamamoto, T., Yamaguchi, H., Yoshizawa, N., 2019. Effects of lighting on perception of spaciousness. PROCEEDINGS OF the 29th Quadrennial Session of the CIE .

Muller, E., Gemmell, E., Choudhury, I., Nathvani, R., Metzler, A.B., Bennett, J., Denton, E., Flaxman, S., Ezzati, M., 2022. City-wide perceptions of neighbourhood quality using street view images. arXiv preprint arXiv:2211.12139 .

Ning, H., Li, Z., Wang, C., Hodgson, M.E., Huang, X., Li, X., 2022. Converting street view images to land cover maps for metric mapping: A case study on sidewalk network extraction for the wheelchair users. Computers, Environment and Urban Systems 95, 101808. doi:10.1016/j.compenvurbsys.2022.101808.

Orozco Carpio, P.R., Viñals, M.J., López-González, M.C., 2024. 3D Point Cloud and GIS Approach to Assess Street Physical Attributes. Smart Cities 7, 991–1006. doi:10.3390/smartcities7030042.

Pappalardo, L., Simini, F., Barlacchi, G., Pellungrini, R., 2022. scikit-mobility: A python library for the analysis, generation, and risk assessment of mobility data. Journal of Statistical Software 103, 1–38. URL: https://www.jstatsoft.org/index.php/jss/article/view/v103i04, doi:10.18637/jss.v103.i04.

Pönkänen, M., Tenkanen, H., Mladenović, M., 2025. Spatial accessibility and transport inequity in finland: Open source models and perspectives from planning practice. Computers, Environment and Urban Systems 116, 102218. doi:10.1016/j.compenvurbsys.2024.102218.

Ramalingam, S.P., Kumar, V., 2025. Building usage prediction in complex urban scenes by fusing text and facade features from street view images using deep learning. Building and Environment 267, 112174.

Ramírez, T., Hurtubia, R., Lobel, H., Rossetti, T., 2021. Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. Landscape and Urban Planning 208. doi:10.1016/j.landurbplan.2020.104002.

Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision Transformers for Dense Prediction. doi:10.48550/arXiv.2103.13413, arXiv:2103.13413.

Ren, M., Zhang, X., Chen, X., Zhou, B., Feng, Z., 2023. Yolov5s-m: A deep learning network model for road pavement damage detection from urban street-view imagery. International Journal of Applied Earth Observation and Geoinformation 120, 103335.

Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, A., Wilson, N., 2018. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. Health & Place 52, 240–246. doi:10.1016/j.healthplace.2018.07.001.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S.K.S., Gontijo-Lopes, R., Ayan, B.K., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M., 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, in: Advances in Neural Information Processing Systems.

Sánchez, I.A.V., Labib, S., 2024. Accessing eye-level greenness visibility from open-source street view images: A methodological development and implementation in multi-city and multi-country contexts. Sustainable Cities and Society 103, 105262. doi:10.1016/j.scs.2024.105262.

Sanusi, R., Johnstone, D., May, P., Livesley, S.J., 2016. Street orientation and side of the street greatly influence the microclimatic benefits street trees can provide in summer. Journal of environmental quality 45, 167–174.

Schneider, D., Schwalbe, E., Maas, H.G., 2009. Validation of geometric models for fisheye lenses. ISPRS J. Photogramm. Remote Sens. 64, 259–266.

Seresinhe, C.I., Preis, T., Moat, H.S., 2017. Using deep learning to quantify the beauty of outdoor places. Royal Society Open Science 4, 170170. doi:10.1098/rsos.170170.

Shu, Z., Yan, Z., Xu, X., 2021. Pavement Crack Detection Method of Street View Images Based on Deep Learning. Journal of Physics: Conference Series 1952, 022043. doi:10.1088/1742-6596/1952/2/022043.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., Taigman, Y., 2022. Make-A-Video: Text-to-Video Generation without Text-Video Data. doi:10.48550/arXiv.2209.14792, arXiv:2209.14792.

Sobel, I., Feldman, G., 1968. A 3x3 Isotropic Gradient Operator for Image Processing.

de Souza-Daw, T., Ross, R., Nhan, T.D., Hung, L.A., Trung, N.D.Q., Chau, L.H., Phuong, H.M., Ngoc, L.H., Nkhoma, M., 2015. Design and evaluation of a low-cost street-level image capturing vehicle for south-east asia. Journal of Engineering, Design and Technology 13, 579–595.

Sun, M., Zhang, F., Duarte, F., Ratti, C., 2022. Understanding architecture age and style through deep learning. Cities 128, 103787.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114.

Tu, P., Zhou, X., Wang, M., Yang, X., Peng, B., Chen, P., Su, X., Huang, Y., Zheng, Y., Xu, C., 2024. NeRF2Points: Large-Scale Point Cloud Generation From Street Views' Radiance Field Optimization. doi:10.48550/arXiv.2404.04875, arXiv:2404.04875.

Uber, 2018. H3: Uber's Hexagonal Hierarchical Spatial Index. https://www.uber.com/en-JP/blog/h3/.

Wang, J., Hu, Y., Duolihong, W., 2023. Diagnosis and planning strategies for quality of urban street space based on street view images. ISPRS International Journal of Geo-Information 12, 15.

Wang, J., Liu, W., Gou, A., 2022a. Numerical characteristics and spatial distribution of panoramic street green view index based on segnet semantic segmentation in savannah. Urban Forestry & Urban Greening 69, 127488.

Wang, M., Haworth, J., Chen, H., Liu, Y., Shi, Z., 2024a. Investigating the potential of crowdsourced street-level imagery in understanding the spatiotemporal dynamics of cities: A case study of walkability in Inner London. Cities 153, 105243. doi:10.1016/j.cities.2024.105243.

Wang, M., He, Y., Meng, H., Zhang, Y., Zhu, B., Mango, J., Li, X., 2022b. Assessing Street Space Quality Using Street View Imagery and Function-Driven Method: The Case of Xiamen, China. ISPRS International Journal of Geo-Information 11, 282. doi:10.3390/ijgi11050282.

Wang, R., Liu, Y., Lu, Y., Zhang, J., Liu, P., Yao, Y., Grekousis, G., 2019. Perceptions of built environment and health outcomes for older chinese in beijing: A big data approach with street view images and deep learning technique. Comput. Environ. Urban Syst. 78, 101386.

Wang, S., Huang, X., Liu, P., Zhang, M., Biljecki, F., Hu, T., Fu, X., Liu, L., Liu, X., Wang, R., Huang, Y., Yan, J., Jiang, J., Chukwu, M., Reza Naghedi, S., Hemmati, M., Shao, Y., Jia, N., Xiao, Z., Tian, T., Hu, Y., Yu, L., Yap, W., Macatulad, E., Chen, Z., Cui, Y., Ito, K., Ye, M., Fan, Z., Lei, B., Bao, S., 2024b. Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review. International

Journal of Applied Earth Observation and Geoinformation 128, 103734. doi:`10.1016/j.jag.2024.103734`.

Wang, Z., Ito, K., Biljecki, F., 2024c. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. Cities 145, 104704. doi:`10.1016/j.cities.2023.104704`.

Watanabe, S., Horikoshi, T., Ishii, J., Tomita, A., 2013. The measurement of the solar absorptance of the clothed human body – the case of japanese, college-aged male subjects. Build. Environ. 59, 492–500.

Weld, G., Jang, E., Li, A., Zeng, A., Heimerl, K., Froehlich, J.E., 2019. Deep Learning for Automatically Detecting Sidewalk Accessibility Problems Using Streetscape Imagery, in: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ACM. pp. 196–209. doi:`10.1145/3308561.3353798`.

Wojke, N., Bewley, A., 2018. Deep cosine metric learning for person re-identification, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 748–756. doi:`10.1109/WACV.2018.00087`.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 3645–3649. doi:`10.1109/ICIP.2017.8296962`.

Xia, Y., Yabuki, N., Fukuda, T., 2021a. Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. Urban Forestry & Urban Greening 59, 126995.

Xia, Y., Yabuki, N., Fukuda, T., 2021b. Sky view factor estimation from street view images based on semantic segmentation. Urban Climate 40, 100999.

Yang, C., Zhang, Y., 2024. Public emotions and visual perception of the East Coast Park in Singapore: A deep learning method using social media data. Urban Forestry & Urban Greening 94, 128285. doi:`10.1016/j.ufug.2024.128285`.

Yang, J., Zhao, L., Mcbride, J., Gong, P., 2009. Can you see green? Assessing the visibility of urban forests in cities. Landscape and Urban Planning 91, 97–104. doi:`10.1016/j.landurbplan.2008.12.004`.

Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. https://arxiv.org/abs/2401.10891v2.

Yang, X., Lindquist, M., Van Berkel, D., 2025. "streetscape" package in r: A reproducible method for analyzing open-source street view datasets and facilitating research for urban analytics. SoftwareX 29, 101981.

Yap, W., Janssen, P., Biljecki, F., 2022. Free and open source urbanism: Software for urban planning practice. Computers, Environment and Urban Systems 96, 101825. doi:10.1016/j.compenvurbsys.2022.101825.

Yap, W., Stouffs, R., Biljecki, F., 2023. Urbanity: Automated modelling and analysis of multidimensional networks in cities. npj Urban Sustainability 3, 1–11. doi:10.1038/s42949-023-00125-w.

Ye, Y., Zeng, W., Shen, Q., Zhang, X., Lu, Y., 2019. The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. Environment and Planning B: Urban Analytics and City Science 46, 1439–1457. doi:10.1177/2399808319828734.

Yin, L., Wang, Z., 2016. Measuring visual enclosure for street walkability: Using machine learning algorithms and google street view imagery. Applied geography 76, 147–153.

Ypsilantis, N.A., Chen, K., Cao, B., Lipovský, M., Dogan-Schönberger, P., Makosa, G., Bluntschli, B., Seyedhosseini, M., Chum, O., Araujo, A., 2023. Towards Universal Image Embeddings: A Large-Scale Dataset and Challenge for Generic Image Representations. doi:10.48550/arXiv.2309.01858, arXiv:2309.01858.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H.H., Lin, H., Ratti, C., 2018. Measuring human perceptions of a large-scale urban region using machine learning. Landscape and Urban Planning 180, 148–160. doi:10.1016/j.landurbplan.2018.08.020.

Zhang, G., Yi, J., Yuan, J., Li, Y., Jin, D., 2023. DAS: Efficient Street View Image Sampling for Urban Prediction. ACM Trans. Intell. Syst. Technol. 14, 35:1–35:20. doi:10.1145/3576902.

Zhang, L., Pei, T., Wang, X., Wu, M., Song, C., Guo, S., Chen, Y., 2020. Quantifying the urban visual perception of chinese traditional-style building with street view images. applied sciences 10, 5963.

Zhang, Y., Dong, R., 2018. Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in beijing. ISPRS International Journal of Geo-Information 7, 104.

Zhao, C., Ogawa, Y., Chen, S., Oki, T., Sekimoto, Y., 2023a. Quantitative land price analysis via computer vision from street view images. Engineering Applications of Artificial Intelligence 123, 106294. doi:10.1016/j.engappai.2023.106294.

Zhao, K., Liu, Y., Hao, S., Lu, S., Liu, H., Zhou, L., 2021. Bounding boxes are all we need: Street view image classification via context encoding of detected buildings. IEEE Transactions on Geoscience and Remote Sensing , 1–17doi:10.1109/TGRS.2021.3064316.

Zhao, T., Liang, X., Tu, W., Huang, Z., Biljecki, F., 2023b. Sensing urban soundscapes from street view imagery. Computers, Environment and Urban Systems 99, 101915.

Zheng, X., Amemiya, M., 2023. Method for applying crowdsourced street-level imagery data to evaluate street-level greenness. ISPRS International Journal of Geo-Information 12, 108.

Zhou, B., Lapedriza, À., Khosla, A., Oliva, A., Torralba, A., 2017. Places: A 10 Million Image Database for Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence PP, 1–1. doi:10.1109/TPAMI.2017.2723009.

Zhou, H., He, S., Cai, Y., Wang, M., Su, S., 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. Sustainable Cities and Society 50, 101605.

Zhu, Y., Zhang, Y., Biljecki, F., 2025. Understanding the user perspective on urban public spaces: A systematic review and opportunities for machine learning. Cities 156, 105535.

Zhuang, Y., Kang, Y., Fei, T., Bian, M., Du, Y., 2024. From hearing to seeing: Linking auditory and visual place perceptions with soundscape-to-image generative artificial intelligence. Computers, Environment and Urban Systems 110, 102122. doi:10.1016/j.compenvurbsys.2024.102122.