



Implications of Artificial Intelligence for Assessing the Built Environment

Lily Moradi & Nimish Bioria

To cite this article: Lily Moradi & Nimish Bioria (2025) Implications of Artificial Intelligence for Assessing the Built Environment, *Journal of Urban Technology*, 32:3, 163-191, DOI: [10.1080/10630732.2025.2468142](https://doi.org/10.1080/10630732.2025.2468142)

To link to this article: <https://doi.org/10.1080/10630732.2025.2468142>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 May 2025.



Submit your article to this journal [↗](#)



Article views: 643




View related articles [↗](#)



View Crossmark data [↗](#)

Implications of Artificial Intelligence for Assessing the Built Environment

Lily Moradi  and Nimish Bioria

University of Technology Sydney, Australia

ABSTRACT



The exponential rate of global urban growth makes evaluating numerous socio-spatial attributes of our everyday environment an increasingly daunting task. Consequently, many modern urban studies have taken advantage of artificial intelligence (AI), specifically machine learning and computer vision techniques, as preferred techniques for image-based assessment of the built environment. This article investigates the critical role of AI in urban studies, associated urban data collection methods, and preferred machine learning algorithms. Several implemented computer vision models are assessed for their accuracy, study area, and the urban attributes they evaluate. These attributes include neighborhood perception, child-friendliness, walkability, safety and security, aesthetics, urban forestry and greenery, land use, and transportation. Thereafter, the deficiencies and potential of the future applications of AI for assessing the built environment are discussed within the current context, which lacks comprehensive assessment models for evaluating critical aspects of urban life.

KEYWORDS

built environment; machine learning; image processing; object detection; Google Street View (GSV)

Introduction

Rapid urbanization and population growth present new challenges for people-centric sustainable development. However, the extensive size and variations of urban environments make this a daunting task (Hipp et al., 2017). As a result, numerous researchers have considered using artificial intelligence-based (AI) methods to evaluate cities from diverse viewpoints. AI has been proven to be a revolutionary field of study in the past decade (Goodfellow et al., 2016; Norvig and Russell, 1995) and has found numerous urban planning and urban design applications due to its scalability and generalizability. An extensive amount of data is also being generated by citizens, and IoT devices are specifically developed for data collection from within urban areas. Access to unprecedented data has also initiated better interdisciplinary research opportunities, including applying machine learning in urban design (Liu et al., 2017). Besides this, utilizing

CONTACT Lily Moradi  f.moradi.k@gmail.com  University of Technology Sydney, Department of Design, Architecture, and Building.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

non-human judges, such as machine learning-based models, could result in more accurate interpretations of urban areas as they will be free from personal judgment and biases (Ye et al., 2019). The fusion of machine-learning techniques with modern computer vision methods using artificial neural networks has also proven to be less time-consuming and scalable for studying multi-scalar urban regions across towns and countries (Acosta and Camargo, 2019).

The intersection of AI and urban studies has given rise to a new paradigm in urban development and governance, which (Cugurullo et al., 2023) term “AI urbanism.” This emerging field represents a significant shift from traditional smart city approaches, as AI technologies move beyond mere data collection and analysis to actively shaping urban spaces and decision-making processes. Cugurullo et al. (2023) argue that AI urbanism is characterized by the deployment of urban AIs, including autonomous vehicles, robots, city brains, and urban software agents, which are increasingly capable of making independent decisions that affect urban life. This transition towards AI-driven urbanism raises critical questions about the future of cities, including issues of sustainability, governance, and the changing nature of human-machine relationships in urban environments (Luusua et al., 2023). As AI becomes more deeply integrated into urban systems, it challenges existing notions of smart urbanism and paves the way for what could potentially become “autonomous cities”—urban spaces where AI performs crucial social and managerial functions with minimal human supervision (Cugurullo, 2020). However, this transition also brings forth concerns about the ethical implications and potential biases of AI systems in urban contexts (Yan et al., 2023). Our study contributes to this evolving discourse by examining the multifaceted applications of AI within smart cities, focusing on key urban attributes such as neighborhood perception, child-friendliness, walkability, safety and security, aesthetics, urban forestry and greenery, land use, and transportation. Through a comprehensive analysis of machine learning and computer vision techniques applied to urban data collection and management, our research sheds light on how AI is transforming various aspects of urban life, from intelligent transportation systems to public safety and healthcare. By investigating both the potential benefits and challenges of integrating AI into urban environments, our work adds to the growing body of knowledge on how AI is reshaping urban landscapes and the implications this holds for future urban development and planning.

This paper explores how AI approaches have revolutionized urban planning studies. The next section explains the methods and materials we used for conducting the research. After that, the following sections offer a literature review and an overview of machine learning methods, data collection, data management procedures for urban planning research, multiple urban attributes, and associated AI-based methods for evaluating them. Finally, we highlight the results, limitations, and potential future AI applications for assessing the urban environment.

Methods and Materials

Study Selection

For the literature review, we adopted Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), developed by Liberati et al. (2009). PRISMA proposes a

checklist of 27 sections for a literature review and a three-phase flow diagram for state-of-the-art paper collection. These three phases have been adapted for the literature analysis. In the first phase—identification, a search strategy, and article sources to collect a set of potential papers are used. The articles selected in this phase are based on the following urban design and machine learning keywords: urban planning, urban design, urban environment, built environment, public policy, perception, risk perception, age-friendly, walkability, street view, aesthetics, forestry, greenery, flora, land use, traffic, tracking system, flow network, security, and safety. The following computer-science keywords were also taken into consideration: AI, machine learning, computer vision, artificial neural network, deep learning, convolutional neural networks (CNN), support vector machines (SVMs), image processing, object detection, image classification, scene perception, semantic segmentation, action recognition, data collection, and dataset.

Using these keywords, the following databases were mined: Google Scholar (GS), Web of Science (WoS), Xplore by Institute of Electrical and Electronics Engineers (IEEE), Microsoft Academic by Microsoft (MA), ScienceDirect by Elsevier (SD), Association for Computing Machinery (ACM) Digital Library, ProQuest, and Sage. This first phase resulted in identifying more than 300 papers from urban planning and computer science domains.

The second phase involved screening the chosen papers and excluding irrelevant items according to our research purposes. Since the use of machine learning tools that use computer vision for urban analysis only appeared during the last two decades, we, therefore, removed articles published before the year 2000 during this screening phase. Secondly, we excluded duplicate papers across the eight selected repositories (GS, WoS, IEEE, MA, SD, ACM, ProQuest, and Sage). Finally, we screened all documents to distinguish items non-relevant to computer vision or urban planning concepts. This phase resulted in more than 150 related articles being taken into account for the final phase. The third phase—eligibility, involved accurately reading and selecting papers relevant to the use cases of machine learning in urban design studies. The selected papers were subsequently categorized based on the considered urban characteristics and by their adopted algorithms.

Information Structuring

In order to extract structured information from the selected articles, the following three steps were followed: First, relevant data items and their corresponding values were extracted from the articles. Subsequently, these data items and values were categorized based on the urban features that have been assessed in the research. This acquired information was structured as a table in which rows represent each paper, and columns consist of processed information items such as the authors' names, year of publication, study area, and scale, considered urban features, and utilized algorithms. [Table 1](#) presents these data items classified by urban characteristics. The distribution of the urban characteristics used in this study and the distribution of the study regions and scales are also depicted in [Figure 1](#). These graphs demonstrate that urban safety and security is the researchers' most popular analysis topic. The majority of research studies employed datasets from one or more cities within a single nation to propose their approaches, and among all locations, the United States has produced the most research articles.

Table 1. An overview of the selected papers with information about the urban feature, location, date, and the algorithms used

Urban Feature Category	Authors and Date	Location/Space		Sampling Period	Specific Urban Features	Algorithm/Task
		Study Area	Scale			
Perceived Neighborhood Environment	Porzi et al., 2015	New York, Boston, Linz, and Salzburg	Cities	2015	Perceived safety of urban scenes	Convolutional Neural Network
	Ordonez and Berg 2014	56 cities worldwide	City	2010	uniqueness, safety, and wealth	Convolutional Neural Network
	Cheng et al., 2014	-	Worldwide	2010-2011	Perceived safety of urban scenes	Support Vector Machine
	Dubey et al., 2016	56 cities worldwide	City	2010-2014	Safety, Liveliness, Beauty, and Wealth	Convolutional Neural Network (CNN)
	Verma et al., 2019	Mumbai, India	City	2018	liveliness, safety, and health	Convolutional Neural Network/ Long Short-Term Memory (LSTM)
	Ibrahim et al., 2019	-	Worldwide	2019	informal areas, pedestrians, transport modes	Convolutional Neural Network
Tian et al., 2021	Berlin, Germany	City	2021	enclosure, aesthetics, richness, accessibility, etc.	Six ML models including KNN, ADA, Random Forest, and Decision Tree	
Freitas et al., 2022	Raleigh, North Carolina, US	Neighborhoods	2008-2020	Safety, maintenance, subjective metrics, objective metrics	You Only Look Once (YOLOv3)	
Child-Friendly Features	Lee et al., 2016	Ulsan, South Korea	City	2010-2014	pedestrians, walkability, crash risk	negative binomial (NB) regression and zero-inflated negative binomial (ZINB)
Walkability	Kwon et al., 2020	Ulsan, South Korea	City	2010-2014	pedestrians, walkability, crash risk	K-means clustering
	Hara Kotaro et al., 2013	Honolulu, Hawaii, USA	Country	2014	Cross walking, Accessibility	linear Support Vector Machine
	Dragan Ahmetovic, 2017	Washington, DC	City	2017	sidewalk accessibility	Convolutional Neural Network (ResNet)
	Yin et al., 2016	Buffalo, USA	City	2016	visual enclosure, accessibility	Support Vector Machine
	Law et al., 2019	London, UK	City	2018	Pedestrians, frontage, aesthetics	Convolutional Neural Network (AlexNet)
	Ito and Biljecki, 2021	Singapore and Tokyo	Cities	N/A	Bikeability and Walkability using 34 indicators	GlouCV and YOLOv3
	Wong et al., 2021	Multiple cities in the UK, USA, etc.	Cities	N/A	Pedestrian tracking, pedestrian flow statistics, human-centric planning and design	FairMOT, a multi-object tracking (MOT) method
	Hosseini et al., 2022	Multiple cities in USA	Cities	2018-2019	Mapping sidewalk networks	Hierarchical Multi-Scale Attention semantic segmentation
	Koo et al., 2022	Atlanta, Georgia, USA	City	2017	Quantifying street-level factors, association with walking mode choice	Logistic regression models
Li et al. 2023	US transit stations	National	2007-2020	Changes in intersection-level marked crosswalks, analysis of geographic variations and associations with built environment	Custom deep learning CNN object detection model	

Security and Safety	Wang et al., 2019	Guangzhou, China	City	N/A	Perception of neighborhood safety, mental health	fully convolutional neural network (FCN-8s)
	Acosta and Camargo, 2019	Bogota, Colombia	City	2018	Safety perception	Support Vector Machine
	Dubey et al., 2016	56 world-wide cities	Country	2016	Safe, lively, boring, wealthy, depressing, beautiful	Convolutional Neural Network, TrueSkill
	Ordonez, and Berg, 2014	NY, Chicago, Boston, Baltimore	Cities	2014	Safety, uniqueness, wealthy	Generic image features-support vector regression
	Naik et al., 2014	21 US Cities	Cities	2014	Safety	Support vector regression
	Suel et al., 2019	London, UK	City	2018	health, crime	DNN
	De Nadai et al., 2016	Milan and Rome, Italy	Cities	2016	Safety, liveliness, activity	Convolutional Neural Network
	Salesses et al., 2013	Boston, NY Linz, Salzburg	Cities	2013	Safety, class, uniqueness	Descriptive Statistics
	Dixit et al., 2021	N/A	N/A	2021	Lane detection, vehicle crash predictability, trajectory predictability of neighboring vehicles	Slice-based Convolutional Neural Networks (SCNNs)
	Larkin et al., 2021	56 cities (global)	Cities	2010-2014	Safety, lively and beauty perceptions, population density, traffic air pollution, tree cover	GIS/remote sensing, deep learning image segmentation
Aesthetics and Upkeep	Bustos et al., 2021	Barcelona, Madrid, and San Francisco	Cities	2010-2018	Pedestrian and vehicle safety, hazard index determination, image segmentation, class activation mapping	Residual Convolutional Neural Network, image segmentation,
	Zhang et al., 2021	Houston, USA	City	2018	Perception of safety score, crime rates, urban liveliness	Pre-trained scene understanding ResNet50 model
	Naik, 2017	21 American cities	Cities	2007-2014	Visual appearance, urban upkeep	RankSVM, Streetscore- Convolutional Neural Network
Urban Forestry and Green Space	Law et al., 2020	London, UK	City	2018	Pedestrians, frontage, aesthetics	Convolutional Neural Network (AlexNet)
	Ye et al., 2019	Shanghai, China	City	2019	street greenery, sky views, building frontage, pedestrian space	Convolutional Neural Network (SegNet)
	Li et al., 2015	New York City	City	2014	Greenery, urban management	Green View Index (GVI) classification
	Song et al., 2022	New York City and Atlanta, USA	Cities	2017	Urban greenways	YOLOv3 Object detection
	Bai et al., 2023	Shenzhen, China	City	2018	Urban greenways	FCNN-based segmentation

(Continued)

Table 1. Continued.

Urban Feature Category	Authors and Date	Location/Space		Sampling Period	Specific Urban Features	Algorithm/Task
		Study Area	Scale			
Land Use and Population Density	Audebert et al., 2017	Vaihingen and Potsdam, Germany			Semantic labeling of very high-resolution multi-modal remote sensing data	Convolutional Neural Network (SegNet and ResNet)
	Chew et al., 2018	Nigeria and Guatemala	Countries	2017	building pattern, community size, closeness	DNN
	Srivastava et al., 2019	Ile-de-France and Nantes, France	Cities	2018	Automated landuse mapping at the urban-object level	DNN
	Chaturvedi and de Vries, 2021	-	Global	2020	A Review on Urban Land Use Planning articles	Random forest (RF), CNN, support vector machine (SVM), GAN
	Zhang et al., 2021	Beijing, China	City	2020	Land cover classification with satellite imagery	Random forest (RF) with Bayesian parameter optimization, support vector machine (SVM)
Traffic Flow, and Transportation	Boguszewski et al., 2021	Rural areas across Poland	Country	2015-2018	Land cover mapping using aerial imagery	Deep learning semantic segmentation with DeepLabv3 and Xception71
	Bottino et al., 2016	Turin, Italy	City	2014-2015	flow network, traffic monitoring	Multi-layer model
	Yu et al., 2017	Beijing, China	City	2016	transportation network	Convolutional Neural Network/ Long Short-Term Memory (LSTM)
	Yang et al., 2018	-	Worldwide	2012-2014	traffic surveillance	BPNN (backpropagational Neural Network)
	Xie et al., 2021	N/A	Stanford Cars and CompCars datasets	N/A	Visual surveillance system (VSS) and Intelligent transportation systems (ITS)	CNN and variations for comparison (FCNN, RCNN, etc.)
	Hasibuan et al., 2021	N/A	N/A	N/A	Traffic surveillance, vehicle detection, tracking, and classification	YOLOv4, Deep Sort
	Umair et al., 2021	Toronto, Canada	City	2020	Vehicle queue length estimation, low-resolution traffic videos	CNN-based approaches, YOLOv4, ResNet, Inception
	Wael et al., 2022	Cairo, Egypt	City	N/A	User experiences in city contexts, pedestrian flow, destination proximity	GoodVision Video Insights software

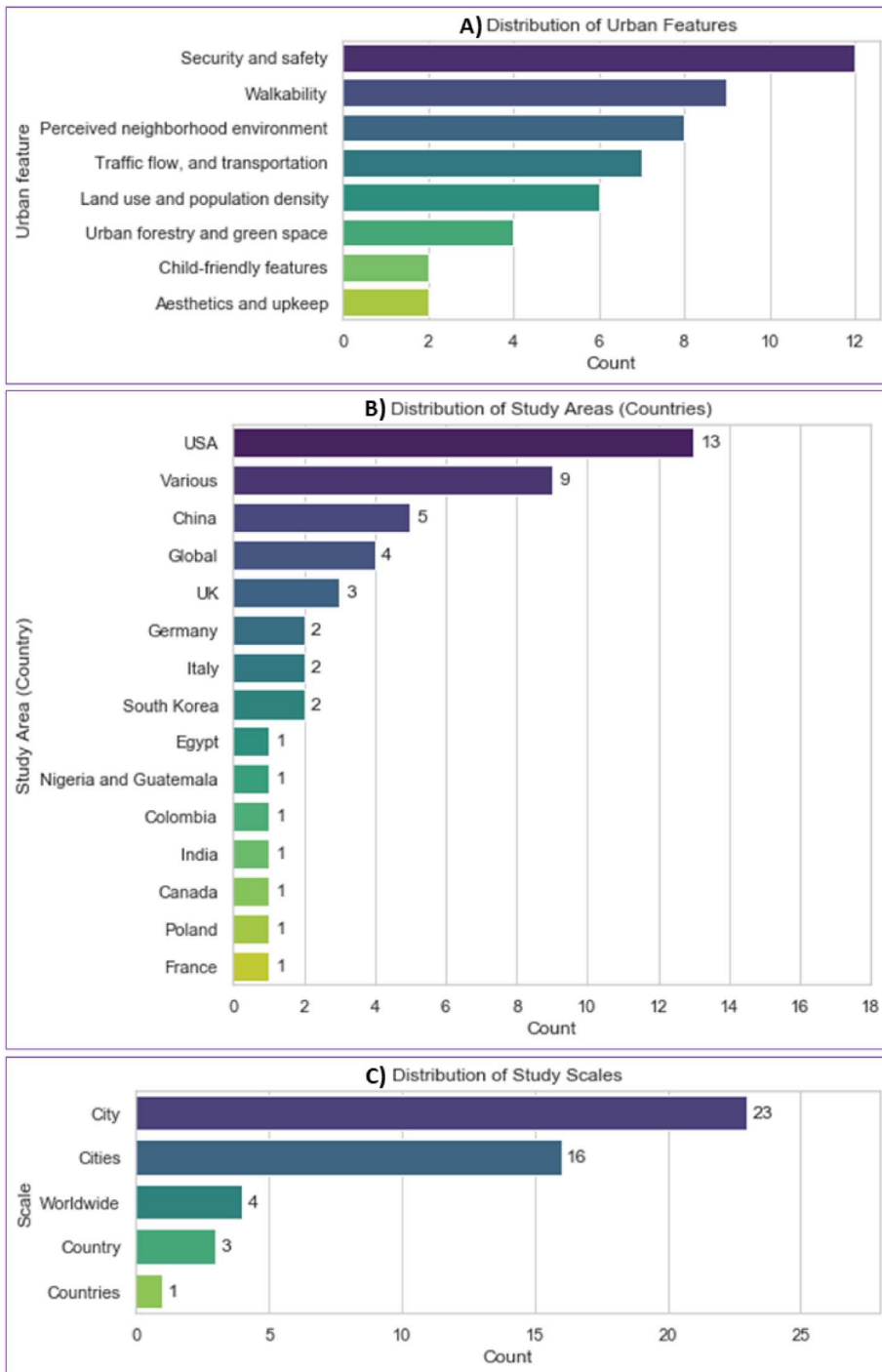


Figure 1. The distribution of research articles analyzed in this study based on: (A) the urban characteristics; (B) study areas (by countries); and (C) the study scales

It was observed that the quantum of urban studies assessing urban attributes using imagery datasets, as well as computer vision solutions, were conducted after 2010, perhaps due to advances in “deep learning” models post 2010 (Harikandeh et al., 2023; Taheri and Aliakbary, 2022), which improved the accuracy of AI-based solutions to a great extent. For instance, exploring the Science Direct repository reveals the progress made in Urban Planning studies through the utilization of AI-based techniques. Figure 2 showcases the quantity of articles that mention both urban planning concepts and AI-related keywords, encompassing implications such as computer vision, machine learning, and deep learning. These research works mainly studied North American, European, and East Asian cities. However, it must be noted that machine learning models are often good at generalizing, and the models trained to analyze urban attributes for a specific region are not necessarily applicable in other regions due to inherent socio-cultural and spatial diversities.

Machine learning algorithms have also been successfully applied in the field of image analysis within urban studies. These algorithms are primarily classified as supervised learning algorithms, in which the model is provided with input data and expected results during the training phase and is subsequently expected to find patterns between them (Ben-David, 2014). Table 1 outlines the most popular supervised learning algorithms used for urban studies, namely: SVMs, Feed-Forward Neural Networks, CNNs, and combinations of CNNs and Long Short-Term Memories (LSTMs). CNNs have proved to be the most reliable among these algorithms and have been successful in various applications, including image classification and segmentation (Krizhevsky et al., 2017).

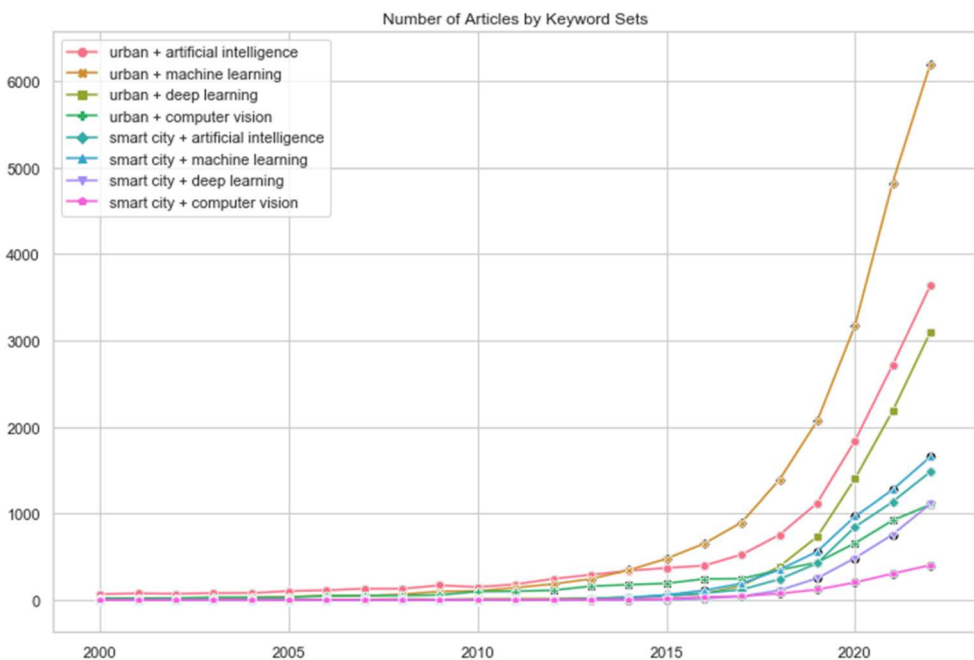


Figure 2. The trend of using AI-based approaches in urban analysis studies from 2000 to 2022 on the Scienec Direct repository. Each line shows the number of co-occurrence of two keywords.

Discussion of Research Findings

Data Collection and Data Management Procedures in Urban Computation

Cities are spatial structures that are too large to be experienced as a single entity. Therefore, to apply computer vision approaches to predict the human perception of places, several images are required (Salesses et al., 2013). The following sub-sections outline data collection tools and techniques successfully employed in the reviewed literature.

Application Programming Interfaces. Google Street View (GSV) provides a visual recording of panoramic 360-degree views of urban areas, allowing for easy observation of spatial features, and is thus considered a reliable tool for conducting in-person audits in urban regions (Anguelov et al., 2010). Using GSV images proves to be an efficient, safe, and low-cost option that also provides historical images of the studied area as well as its current appearance (Liu et al., 2017). Photos are updated every two to three years by autonomous or human-driven Street View cars, three-wheelers, snowmobiles, and other mobile devices (Anguelov et al., 2010). GSV provides an application programming interface (API) for area-specific data collection using geotagging and is one of the most prevalent data collection methods among scholars for numerous urban planning and urban design studies (Bader et al., 2015).

Acosta et al. (2019) crawled GSV API based on a file containing geographical information about target zones from Bogota, Colombia. They then adopted a crowdsourced surveying approach to finding safe areas and built their dataset to analyze city safety perception over the gathered imagery. Bader et al. (2015) presented an online application known as the Computer Assisted Neighborhood Visual Assessment System (CANVAS), which employs GSV to assess neighborhood scenes concerning walkability and physical disorder. They built their dataset by sampling 150 street sections in the United States, wherein high school interns were asked to rate them through a simple computer gateway interface (CGI) as a proof of concept. Ordóñez and Berg (2014) collected a dataset of geotagged photos of New York, Boston, Baltimore, and Chicago using the GSV API from random spots inside each city's boundaries. They combined these images with an available dataset to generate denser urban perception maps for depicting wealth, uniqueness, and safety assessments.

In another study, the reliability of a neighborhood audit instrument was assessed by Clarke et al. (2010) using the data obtained from the Chicago Community Adult Health Study (CCAHS). Neighborhoods' impacts on residents' health were investigated in 60 residential blocks on Google Earth's Street View images. Yin and Wang (2016) also obtained panoramic photographs sampled from more than 300 street blocks across Buffalo, US, to evaluate the streets' walkability measures. Law et al. (2019) collected city-level data for estimating street frontage quality from the GSV API and street images captured from the CityEngine software by using a graph from London's route network obtained from the OS Meridian dataset. Likewise, Odgers et al. (2012) observed a local neighborhood's essential features through a systematic social observation (SSO) study using over 1,000 neighborhood GSV data from England and Wales.

OpenStreetMap (OSM) is another community-driven map-providing API. Geolocational data can be obtained from the OSM API by defining a specific area and downloading the included XML information from OpenStreetMap or using websites

such as Cloudmade and Geofabrik, that allow downloading OSM data in different formats, including XML and Shapefiles. Chirkin and König (2016) also took advantage of OSM's existing spatial configurations as the primary dataset for unsupervised learning to let the algorithm investigate a design space using clustering techniques. Forghani et al. (2014) performed a qualitative assessment of OpenStreetMap data as one of the most reliable crowdsourced products for volunteered geospatial information (VGI). They proposed a novel quality metric for assessing the VGI accuracy and performing quality analysis on the OSM dataset by judging its consistency in Tehran, Iran.

Crowdsourced Applications. Crowdsourced services, such as *Mapillary* and *KartaView*, provide valuable street view imagery (SVI) data for urban analysis research (Zheng and Amemiya, 2023). These services offer advantages over GSV, including higher temporal resolution and broader coverage in areas where GSV is unavailable. Mapillary (Neuhold et al., 2017), with its extensive repository of crowdsourced street-level images, enables the study of transportation patterns, land use, and architectural characteristics in diverse urban environments (Biljecki and Ito, 2021). It also provides advanced computer vision algorithms for automated feature extraction. Similarly, KartaView focuses on urban areas, offering high-resolution images captured from imaging vehicles and facilitating detailed studies of urban morphology, environmental quality, and urban design (Crooks and See, 2022). Both services contribute significantly to various aspects of urban analysis research, empowering researchers to extract valuable information and generate insights for urban policy and planning. The spatial sampling and ability to download imagery from different users and equipment provide finer temporal resolution and higher image resolution compared to GSV (Yang et al., 2022).

Therefore, besides the GSV programming interface, APIs from such mobile phone applications are also used for urban data collection. For example, Remigio et al. (2019) took advantage of a mobile phone-based neighborhood auditing approach to assess street-level factors such as physical disorder and physical deterioration in Rio de Janeiro, Brazil. Data for each street section were gathered using Fulcrum's API, and a customized data entry system was developed through the mobile application's platform. For this purpose, auditors were asked to take photos of the street segment's potentially health-relevant characteristics for each sampled street segment. The completed audit was then saved and geotagged as a unique data point onto the area's map with locations obtained from the global positioning systems (GPS) embedded in smartphones.

In other research, Verma et al. (2019) proposed a temporal data collection method demonstrating a smartphone-based camera and handheld video recorder to collect visual and audio datasets simultaneously from urban streets. The visual data are captured using the Mapillary mobile application designed to continuously record snapshots at specific intervals according to the duration and the distance covered by the user. A smartphone with the active Mapillary app was attached to a bicycle handlebar, driven by the surveyors on a predefined path. The application also captured the geo-coordinates during this data collection phase. An audio dataset is also captured simultaneously with an audio recorder worn by the surveyors.

Social Networks. The large number of geotagged images being posted on social networks creates an opportunity for scholars to extract image datasets. For example, IM2GPS

(Hays and Efros, 2008) proposes an algorithm for estimating the geographic locations from an image using a scene matching approach by leveraging a dataset of more than six million GPS-tagged photos from *Flickr* and representing the predicted location as a probability distribution over the earth's surface. Likewise, PlaNet (Weyand et al., 2016) resolved this problem as a classification problem by subdividing the earth's surface into multiscale geographic cells and training an LSTM deep network using millions of geotagged images collected from Flickr photos across the world. The model only learns from the geotagged images, including a range of 1 to 5 textual tags demonstrating the image's location. Zamir and Shah (2014) also presented an image geo-locating framework using multiple nearest neighbor feature matching schemes with Generalized Minimum Clique Graphs (GMCP) and evaluated the proposed algorithm using a reference dataset from GSV images of Pittsburgh, Orlando, and Manhattan. This test set consisted of unconstrained user-uploaded GPS-tagged images from three social networks: Flickr, Panoramio, and Picasa.

Available Datasets. Many scholars practice publishing their image datasets for further use by other researchers. One excellent representation of such a dataset is the Place Pulse dataset collected by Salesses et al. (2013) to measure the perception of uniqueness, safety, and class in the cities of Linz and Salzburg in Austria as well as Boston and New York City in the United States. They collected data on city perception by utilizing more than four thousand geotagged images from these cities. Images from multiple towns in the United States were taken from GSV, while the rest of the dataset was collected manually onsite. Multiple users were subsequently asked to select images from a pair to assess people's perceptions of urban environments. This dataset is still available for further use as Place Pulse v1.0. The Place Pulse dataset is now upgraded as Place Pulse v2.0, containing more than a 110,000 pictures and over one million crowd-sourced pairwise comparisons from 56 cities worldwide. The new dataset is used in several urban studies: Dubey et al. (2016) studied the correlation between each city's appearance and its citizens' health and behavior. City perception was analyzed concerning safety, liveliness, and wealth from a computer vision perspective. Naik (2018) also used the Place Pulse 2.0 to train CNN to forecast the pairwise comparisons for these attributes by taking a pair of images as input data, which outperforms similar network architectures.

Mirowski et al. (2019) analyzed that although existing navigation benchmarks generally rely on static datasets, they cannot be adopted for decision-making in urban studies as well as reinforcement learning approaches. Therefore, they presented StreetLearn, an interactive, partially observed visual environment employing approximately 143 thousand GSV photos for its imagery content and full coverage. StreetLearn's data is now accessible in Manhattan and Pittsburgh.

Erwin and Sweetkind-singer (2010) proposed a collecting, maintaining, and access-providing method for more than ten terabytes of imagery and geospatial data. They combined Google map's imagery data with the California Spatial Information Library (CASIL), National Map, and National Atlas dataset. These datasets include high-resolution imagery, layers for school districts, transportation networks, boundaries, and thematic data on geology, climate, and water. This repository is available as "The National Geospatial Digital Archive."

Likewise, other available datasets are employed by scholars such as Youssofi et al. (2020) who, applied the data from the French National Geographical Institute database to identify grasslands, croplands, forests, artificialized spaces, residential open spaces, and main transport infrastructures such as highways and railways to investigate the residents' neighborhood perception considering their neighborhood's spatial features and their socio-economic status. Suel et al. (2019) combined GSV Images with data available from government statistics for the entire population on a subset of images from the four largest urban areas in England to measure environmental, social, and health inequalities. Introducing KITTI-360 (Liao et al., 2023), a successor to the KITTI dataset, brings a wealth of input modalities, semantic annotations, and accurate localization for research in computer vision, graphics, and robotics. Although designed for mobile perception, this dataset can also be utilized for urban planning tasks, allowing researchers to gain insights into urban scenes and address various urban planning problems. Additionally, several datasets intended initially for training autonomous driving vehicles, such as nuScenes (Caesar et al., 2020), ApolloScape (Huang et al., 2020), A2D2 (Geyer et al., 2020), and ViP-DeepLab (Qiao et al., 2021), offer valuable resources for understanding urban environments through machine learning and computer vision methods. These datasets provide high-resolution images, comprehensive annotations, and synchronized sensor data, enabling tasks like object detection, semantic segmentation, and spatial distribution analysis, which are crucial for urban analysis and planning. Leveraging these datasets expands the potential for gaining valuable insights into urban scenes and advancing urban planning research (Biljecki and Ito, 2021; Cinnamon and Jahiu, 2021).

Evaluation of Urban Attributes Using Computer Vision Methods

Machine Learning and deep learning methods, specifically ConvNets (CNNs), have been employed to detect, segment, and recognize regions and objects in images. The relatively abundant labeled data in these tasks are used for problems such as traffic sign recognition and the detection of human bodies and faces, pedestrians, and text in common images (Lecun et al., 2015). These computer vision algorithms perform three fundamental tasks: Image Classification, Segmentation, and Object Detection.

Several trained neural networks are available for performing image classifications and are used to solve a wide variety of research questions. These operate on the principle of perception (regarded as a classification task) for predicting information not necessarily visualized but information that can still be deduced from an image's overall appearance (Goodfellow et al., 2016; Lecun et al., 2015). Models such as VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2014), ResNet (He, 2016), and CondenseNet (Huang and Weinberger, 2018) are among the best-performing image classifiers. These models are primarily trained on the ImageNet dataset (Krizhevsky et al., 2017), which contains more than 15 million images from 22 thousand distinct classes. There are over 30 types and versions of pre-trained classification models with various characteristics and architectures (Stančić et al., 2022). Apart from image classifiers, some computer vision models developed explicitly for scene perception have demonstrated more use cases in urban investigations. Cao et al.'s (2018) perceptual network based on the ResNet, He et al.'s (2018) work focuses on scene depth for perceptual analysis, and Eslami et al.'s (2018) unsupervised perceptual learning model are among the most

well-known image perception research works employed in urban studies. Object detection models are also used as base models for feature extraction to analyze scenes. Faster R-CNN by Ren et al. (2017), YOLO by Redmon et al. (2015), and RefineNet by Lin et al. (2017) are known to be the most reliable among such models.

Image segmentation is another computer vision application employed in urban analysis research works as well as a wide variety of different uses, from scene perception to autonomous driving (Badrinarayanan et al., 2017; He et al., 2016). SegNet (Badrinarayanan et al., 2017), RefineNet (Lin et al., 2017), LinkNet (Chaurasia and Culurciello, 2017), and DenseASPP (Yang et al., 2018) take advantage of CNN to partition images into various regions based on their pixels' characteristics. Moreover, modern semantic segmentation models have significantly contributed to urban analysis and planning by augmenting real-world datasets with synthetic images, resulting in improved accuracy for urban scene segmentation (Ivanovs et al., 2022), or by incorporating Transformer-based decoders which offer faster inference speeds and higher accuracy, such as UNetFormer (Wang et al., 2022); and by leveraging the advantages of geometric information from light field cameras, leading to the creation of a comprehensive dataset and the development of tailored baseline models that outperform other methods (Sheng et al., 2022).

Most of the pre-trained deep learning architectures introduced in this section are available freely for researchers to use for various purposes. These pre-trained models make it possible to take advantage of limited datasets for addressing more specific urban problems and have been used in several urban studies, as outlined in the following sections.

Perception of Neighborhood Environments. Globalization is transforming urban areas rapidly, prompting several researchers to identify the most meaningful indicators for the perception of urban regions (Hanák et al., 2015). There are several classification pieces of research explicitly conducted in the field of urban design. For example, Ordóñez and Berg (2014) developed a visual model that could determine human perceptions of places with classification and regression predictions on the uniqueness, safety, and wealth of urban areas. Adopting state-of-the-art image classifiers, namely DeCAF (Donahue et al., 2014) convolutional network-based characteristics and Fisher Vector (FV) (Kawano and Yanai, 2013) encodings, they also demonstrated this model's use cases for predicting perceptual properties at the city-scale and reported their classification performance on the Place Pulse 1.0 dataset. Cheng et al. (2014) developed a rotation-invariant model for geographic image classification as well as multi-class geospatial object detection. The model is based on the Collection of Part Detectors (COPD), made of many representative part detectors; each is a binary SVM classifier with a linear kernel used to detect a specific range of orientation objects. By treating the selected class's photos as positive instances and the remaining images as negative instances, each simple one-vs-all scheme SVM image classifier is trained separately, enabling the model to classify images between 21 distinct categories. Dubey et al. (2016) also trained a convolutional neural architecture to learn from a ranking loss and joint classification to forecast human interpretations of pairwise picture comparisons and produce global-scale urban perception data. They propose two network architectures, namely the Street-score-CNN (SS-CNN) and the Ranking SS-CNN (RSS-CNN), to simultaneously

minimize the classification loss while predicting pairwise comparisons and ordinal ranking throughout the dataset. The models are tested and evaluated on the Place Pulse 2.0 dataset and can outperform the pre-trained AlexNet (Krizhevsky et al., 2017), VGGNet (Dubey et al., 2016; Simonyan and Zisserman, 2014), and PlaceNet (Zhou et al., 2014).

Several articles have concentrated on image perception methods rather than classifications. Verma et al. (2019) employed state-of-the-art algorithms for feature extraction from the visual and auditory data to establish comparisons between various urban regions on liveliness, safety, and health. Deep learning classifiers are employed to gather information from the collected audio and image datasets, demonstrating innovative multi-modal data gathering and interpretation methods. A recurrent neural network (RNN) with an LSTM architecture and a convolutional neural net implementation named Faster-RCNN-NAS (Neural Architecture Search) structure were trained to perform the audio and imagery city perception tasks. Acosta et al. (2019b) presented a method to use machine learning techniques to model the perception of safety within urban residences by processing city images. This model combines five SVMs trained to predict how safe a given street image is, based on crowdsourced visual survey votes on pairs of urban street photographs. Images with negative perceptions of safety properties include garbage, empty streets, and dirty houses. In contrast, the positive perceptions of safety might be caused by green zones, clean streets, and proper lighting. Phillips et al. (2021) question the assumption of accurate localization against a high-definition map in state-of-the-art autonomy stacks and proposes a system that jointly performs perception, prediction, and localization. The architecture can reuse computation between the three tasks and correct localization errors efficiently. Borkman et al. (2021) introduce the Unity Perception package, an open-source toolset that simplifies and accelerates the generation of synthetic datasets for computer vision tasks by extending the Unity Editor and engine components to generate perfectly annotated examples for several computer vision tasks, including urban perception. The article shows the value of the generated synthetic datasets by training a 2D object detection model, which outperforms the model trained using only actual data.

Zhao et al. (2018) deduced the existence of a correlation between a neighborhood's environment and the value of its properties and land. This work's main contribution is presenting an accurate perspective to urban judgment by integrating multiple information sources, utilizing Deep CNN architectures to interpret low-resolution street view images. One of these Deep CNN models uses semantic segmentation to learn features from images, and the other trains the model to identify urban regions' physical attributes. Ibrahim et al. (2019) offer a multipurpose realistic-dynamic urban modeling framework using deep CNNs to detect slums and informal areas as well as pedestrian and transport modes in urban scenes from street-level images of cities globally. Tian et al. (2021) propose a quantitative framework to measure eye-level urban landscape perceptions, using computer vision and machine learning technologies to evaluate eight perceptual qualities, including ecology and enclosure, in seven residential districts in Berlin. Finally, Freitas et al. (2022) present a machine learning and computer vision-based approach to provide a street-level characterization of urban spaces, such as maintenance, in urban neighborhoods. The approach collects and annotates street-view images, extracts objective metrics through computer vision techniques, and models the

perception of subjective metrics through crowdsourcing. Overall, these studies demonstrate the versatility and potential of computer vision for perceiving and analyzing urban environments and provide valuable frameworks and tools for future research.

Child-Friendly Urban Features. Apart from large-scale urban investigations, some researchers have also studied the urban environment with the intent to study children's needs as a target population group (Hassen and Kaufman, 2016). Numerous studies examine how urban areas throughout the globe might be more child-friendly by examining how they perceive their surroundings (Andal, 2022) or attempting to modify this sentiment through workshops (Elkhouly et al., 2022). However, ML-based investigations have also produced promising outcomes when analyzing child-friendliness as more modern methods. For instance, Lee et al. (2016) examined the impact of environmental attributes on crash risk among school-aged children, providing information on the likelihood of pedestrian crashes in a sample of 800 children aged between 10- to 12-years-old. Their outcomes revealed that the number of student crossings, the path width, the availability of crosswalks, and four-way intersections are associated with crash risk among school-aged children. Kwon et al. (2020) subsequently presented a semantic scene analysis method employing street-level images to investigate further the relationship between environmental features and crash risk at 533 intersections in South Korea. These attributes were quantified, applying object detection and scene segmentation algorithms and k-means clustering method to classify all these intersections into various types based on their similarities. The data on perceived crash risk were obtained from the questionnaires of Lee et al.'s research (2016). Their reported results showed that some of the intersection features, such as the sky and roadways, were correlated with children's perception of crash risk.

Walkability. Walkability refers to the physical environment's conduciveness to walking and can be used to measure a residents' physical activity levels. It can also be interpreted as the environment's degree of friendliness to people walking to work, accessing services, traveling, or exercising (Wang and Yang, 2019). Many studies have been performed to measure walkability with various approaches, including Duncan et al.'s (2013) walkability regression analysis on the Boston Youth Survey Geospatial dataset, which is called CANVAS (Bader et al., 2015). CANVAS employs GSV images for a virtual audition of neighborhood scenes to evaluate walkability-related items' inter-rater reliability in US streets with statistical measures.

Researchers have also used SVI to assess walkability measures. For example, Yin et al. (2016) explore machine learning and big data analytics' use-cases on GSV imagery to measure streetscape characteristics by generating three visual enclosure measures (in three directions) in each studied area. They proved the correlation between these visual enclosure measures with a walk score and pedestrian volume, as well as their universal applicability to examine the association with walking behaviors in urban regions of all countries. The Street-Frontage-Net by Law et al. (2020) also indicates the effects of active frontages on walkability, stating that the active frontages might be more salient in places where pedestrians tend to walk more. Wong et al. (2021) focus on pedestrian tracking and attribute recognition, proposing a robust methodology incorporating high-level pedestrian attributes, a similarity measure, and a probation mechanism.

This approach enhances pedestrian tracking accuracy, leading to more reliable pedestrian flow statistics. Similarly, Hosseini et al. (2022) address data gaps in pedestrian mapping by presenting a computer vision approach for generating sidewalk network datasets using machine learning techniques. This low-cost and scalable methodology creates equal data opportunities for non-motorized street users. On the other hand, Ito and Biljecki (2021) and Li et al. (2023) examine the use of SVI and computer vision to assess bike-ability and crosswalk visibility, respectively. Li et al. (2023) introduces a deep learning-based computer vision model for identifying changes in marked crosswalks. The study reveals an increase in high-visibility crosswalks, indicating improvements in walkability. Lastly, Koo et al. (2022) explores the use of computer vision to quantify street-level factors contributing to walkability. By analyzing street view images, this study identifies highly correlated streetscape factors that significantly impact walking mode choice, emphasizing the importance of considering street-level characteristics in walkability assessments.

Security and Safety. Street-level urban features influence citizens' perception of health and safety, thus impacting their wellbeing. Park and Garcia (2020) analyzed the correlation between pedestrian safety perception and street situations in Alabama, US, using an online survey containing data about residents' characteristics, walking behaviors, and anxiety levels while exposing them to specific street images. The results demonstrate that the primary features to consider for improving the perception of street safety are proper street illumination, the morphologic complexity of streets, outdoor dining, street performances, and various pedestrian-heavy activities. Bustos et al. (2021) also present an automated data-processing scheme combining computer vision and large-scale street imagery to improve pedestrian and vehicle safety, while Dixit et al. (2021) develop a dynamic, safe assessment system for autonomous vehicles, focusing on lane detection and extending the operational design domain for enhanced safety.

Suel et al. (2019) applied a deep neural network architecture trained on a collection of images from London to measure income distribution and evaluate education, unemployment rate, housing, health, and crime in urban regions. They proved that street imagery could help identify attributes, such as safety, linked with visual elements and population density. Remigio et al. (2019) also adapted urban auditory approaches to characterize a sizeable informal community's environment in a part of Rio de Janeiro, Brazil, using a smartphone-based observation system. The system was used to collect street-level information and conduct a SSO. Naik and Philipoom (2014) developed Streetscore to understand urban scenes and predict the streetscape's safety by applying a training dataset obtained from an online survey. The model predicts the image features of approximately one million pictures from GSV, applying SVR to estimate a streetscape's perceived safety and creates high-quality maps of perceived safety for 21 cities in the United States. De Nadai et al. (2016) explored the relationship between safety perception and activity levels of the residents of two Italian cities. They consolidated estimations of perceived safety by using a Convolutional Neural Network (CNN) with citizens' mobile phone data pertaining to liveliness and activity. The CNN was trained with crowdsourced rated GSV images, resulting in a better understanding of reasons underlying safety perception, namely, employee density, distance to the city center, and residents' activity levels. They also demonstrate a significant correlation between safety and activity for

people over 50 and females and a negative correlation for people under the age of 30. More recently, research articles such as F. Zhang et al. (2021) quantitatively examine the relationship between perceived safety and actual crime rates using a scene understanding algorithm, identifying perception bias categories. Larkin et al. (2021) also explore the associations between built environment factors and safety, liveliness, and beauty perceptions across multiple cities, highlighting the importance of considering perceptions alongside physical features.

Urban Forestry and Green Space. An interesting area of research is focused on understanding how spatial characteristics such as urban forestry, urban greenery, and flora and fauna affect city perceptions. Rundle et al. (2011) confirmed that it is feasible to use GSV and similar dataset images to evaluate such urban attributes with computer vision techniques. Naik (2018) concluded that the variance of green areas is the second-best measure to evaluate urban scenes after building structures. Ye et al. (2019) applied SegNet (Badrinarayanan et al., 2017) to extract the pixels representing design elements concerning the streets' visual quality, such as sky view and greenery. Notably, street greenery demonstrates the highest importance for the in-depth perception of urban visual quality. For instance, to increase the visual quality of a street with enough structure diversity and pedestrian space, adding street greenery is considered a potentially successful policy. Li et al. (2015) investigated GSV as an urban greenery evaluation tool by modifying an existing formula named Green View Index (GVI) and assessed street greenery utilizing street images of some sections of New York City. After classifying these photos with Gonzalez method (2009), the results confirm GSV images as a reliable source for street-level greenery assessments and urban planning and management combined with the Green View Index (GVI). The Semantic Naturalness metric was also calculated to assess street greenery by adjusting the Green View Index in GSV images in Rzotkiewicz et al.'s work (2018). Other data sources have also been used in this field of urban analysis. For instance, Song et al. (2022) used Computer Vision and Instagram postings to analyze urban greenway usage. Results showed high user engagement, with different activities and temporal patterns between The High Line and the Atlanta Beltline. This study highlights the potential of social media images combined with AI for understanding greenway usage in landscape architecture and urban design. Bai et al. (2023) examined the relationship between urban greenways and cycling behavior using bike-sharing data and street-view images. Findings revealed positive correlations between greenness and cycling on weekdays and weekends, but the effects of greenway openness varied daily.

Land Use and Population Density. Land-use planning is a set of methods to enhance the use of land to support the more efficient use of resources and more favorable environmental outcomes (Corcoran et al., 2018). Numerous studies in this field used satellite imagery data to assess the land use of urban zones. For instance, Chaturvedi (2021) discusses the use of AI and ML algorithms, such as random forest (RF), CNNs, support vector machine (SVM), and generative adversarial networks (GANs), for land use planning and urban growth analysis using earth observation-based data. Another research introduces the LandCover.ai dataset, which combines aerial imagery with computer vision and deep learning for precise land cover assessment and change detection

(Boguszewski et al., 2021). The dataset includes high-resolution images of rural areas in Poland, annotated with four classes: buildings, woodlands, water, and roads. T. Zhang et al. (2021) presents a systematic method for automatically tuning the hyperparameters of the random forest classifier with Bayesian parameter optimization for land cover classification using Sentinel-2 satellite imagery. Audebert et al. (2017) examined semantic segmentation methods to deal with high-resolution urban remote sensing data with deep fully convolutional networks to assess this urban attribute. They also implemented an efficient CNN using SegNet (Badrinarayanan et al., 2017) and ResNet (He, 2016) for land cover classification and semantic labeling of remote sensing map data from urban areas. Chew et al. (2018) developed a deep learning classification model to foretell if the given aerial images are residential or non-residential, considering several geospatial features such as building patterns, community size, and closeness to other land uses. However, some studies utilize street-view images for this purpose as well. For example, Srivastava et al. (2019) automated the urban-level land-use mapping with an end-to-end trainable deep learning method based on two data sources containing Google Maps overhead images and street-level pictures' ensembles per urban object from the GSV images. The land-use labels are also obtained from OpenStreetMap annotations. The model's performance was tested over the areas of Île-de-France and Nantes in France and provided higher accuracy than methods using a single image modality.

Traffic Flow and Transportation. The expansion of vehicle availability in urban areas is accompanied by accidents and traffic congestion (Yu et al., 2017). Various research projects have proposed solutions to such issues by building new infrastructure, developing intelligent traffic lights systems, and rescheduling transportation itineraries. The Street Viewer (Bottino et al., 2016) is one such real-time system operating on embedded hardware architectures to analyze traffic behavior and observe an area's traffic flow patterns from images. This system can automatically identify traffic flow changes due to accidents or construction works and produces a complete graph representation of the area's traffic flow, enabling various statistical analyses on the flow network with small error rates. The Street Viewer is trained to learn the traffic model autonomously and update it in case of a shift detection in the traffic scheme using object detection and object tracking methods.

Yu et al. (2017) also propose a network grid representation system to retain a fine-scale transportation network structure using motion prediction. Their traffic forecasting is performed by a combination of LSTM and CNNs, which is referred to as spatio-temporal recurrent convolutional networks (SRCNs). To validate the model's accuracy, snapshots from the Beijing transportation network were analyzed by the SRCNs, proving its ability to learn the complex road network topology, containing intersections, interchanges, and ramps. Umair et al. (2021) proposes a CNN-based approach for estimating vehicle queue length in urban traffic using low-resolution videos. It achieves 83 percent and 93 percent accuracy for vehicle-based and pixel-based estimation, respectively, using a pre-trained YOLOv4 model. Hasibuan et al. (2021) also combines YOLOv4 and Deep Sort algorithms to detect and track vehicles on CCTV footage, achieving 87.98 percent accuracy for 13 vehicle types. Xie et al. discuss deep learning-based computer vision methods for visual-based Intelligent Transportation System (ITS) services. It compares state-of-the-art methods in terms of performance

and identifies challenges for future research (Xie et al., 2021). Similarly, Wael et al. (2022) examine factors influencing user experiences around Cairo Metro stops using GIS and video analysis, providing insights for urban planners. Finally, Li et al. (2023) use deep learning and Street View images to monitor changes in marked crosswalks around US transit stations in a recent study, revealing an increase in their number and a shift towards high-visibility designs.

Discussion

The utilization of AI techniques, particularly machine learning, and computer vision, holds significant implications for assessing the urban environment. The findings of this research shed light on the critical role of AI in urban studies and its potential to address the challenges posed by the exponential rate of global urban growth. By leveraging AI algorithms, researchers have been able to analyze image data and extract valuable insights related to various socio-spatial attributes of the built environment (Larkin et al., 2021; Tian et al., 2021). Integrating machine learning algorithms with urban data collection methods has facilitated more accurate and efficient assessment processes. Notably, deep learning algorithms, including CNNs and recurrent neural networks (RNNs), have emerged as highly efficient methods for evaluating urban imagery, surpassing other computer vision techniques (Ibrahim et al., 2019). This efficiency can be attributed to advancements in algorithm development, the availability of diverse imagery datasets, and the utilization of pre-trained models. The dominance of deep-learning architectures in the analyzed articles signifies their potential to shape the future of urban studies. However, despite these advancements, several implications and areas for further exploration exist. First, the research community would greatly benefit from adopting comprehensive and accessible data collection and storage approaches for city-level imagery data worldwide. Such an approach would support and enhance urban research endeavors. Additionally, there is a need for more comprehensive models that assess the correlation and interconnections between different urban attributes. Developing holistic frameworks and models would enable a deeper understanding of the relationships between urban conditions and human wellbeing, fostering inclusive and sustainable urban development.

Furthermore, the limitations associated with computational power and the lack of overarching image datasets pose challenges to conducting large-scale, multi-variate research within the urban domain. Researchers are encouraged to develop computationally efficient computer vision approaches that enable the creation of applicable urban perception models. This would facilitate comprehensive assessments of urban environments, including specific age groups, disadvantaged populations, and individuals with different abilities, ensuring urban inclusiveness.

While these findings highlight the potential of AI in urban assessment, it is important to consider the broader implications of AI integration in urban environments and its impact on the evolution of smart cities. The integration of AI into urban environments represents a significant evolution of the smart city concept, pushing it beyond its initial focus on data collection and connectivity towards more autonomous and anticipatory forms of urban governance. This shift, which can be termed “AI urbanism” or “urban

AI,” presents both opportunities and challenges for the future development of cities. As Cugurullo et al. (2023) argue, AI urbanism differs significantly from smart urbanism in both theory and practice. While smart cities primarily focused on data collection and analysis to improve efficiency, AI urbanism introduces autonomous decision-making capabilities that can fundamentally alter urban governance processes.

The emergence of “city brains” and other large-scale AI systems designed to manage multiple aspects of urban life exemplifies this evolution. For instance, the Haidian City Brain in Beijing, as described by Xu et al. (2024), demonstrates how AI can be employed to integrate diverse data sources and provide real-time insights for urban management. Such systems represent a move towards what (Cugurullo, 2020) terms “anticipatory governance”, where AI technologies are used to predict and preemptively address urban challenges. This capability allows for more proactive approaches to urban issues, potentially improving efficiency and responsiveness in areas such as traffic management, environmental monitoring, and public service delivery (Murshid Reza, 2023).

The implementation of AI-driven systems in urban governance raises critical questions about democracy, transparency, and citizen participation. Scholars have noted that the deployment of such technologies often follows a top-down approach, potentially marginalizing citizen input in favor of technocratic decision-making (Cardullo and Kitchin, 2019; Cugurullo, 2020). The autonomous nature of urban AI systems introduces complex moral implications, particularly evident in scenarios involving autonomous vehicles and predictive policing. These developments necessitate a thorough examination of accountability mechanisms and the ethical frameworks embedded within these systems (Yigitcanlar et al., 2020).

Urban AI systems have the potential to either mitigate or exacerbate existing social inequalities. Research on data-driven governance in various contexts has shown that algorithmic decision-making can reinforce existing biases and create new forms of digital discrimination (O’Neil, 2017). The risk of differential treatment based on algorithmically determined profiles is a growing concern, highlighting the need for robust safeguards and ethical guidelines in the development and deployment of urban AI (Cardullo and Kitchin, 2019).

The development and implementation of urban AI systems often involve complex relationships between state actors and private corporations. This trend towards public-private partnerships in smart city initiatives has been observed globally, from North America to Asia (Szipilko et al., 2023). These collaborations, while potentially driving innovation, also blur the lines between state and corporate interests, raising questions about data ownership, privacy, and the potential for corporate influence over urban governance (Sadowski and Bendor, 2019).

It is crucial to recognize that the implementation and impact of urban AI systems are not uniform across different geographical and socio-economic contexts. Studies across various regions have demonstrated significant variations in how urban AI is conceptualized, implemented, and regulated (Karvonen et al., 2020). This diversity underscores the importance of considering local contexts, existing socioeconomic disparities, and cultural factors when studying or implementing urban AI initiatives.

As the field of urban AI continues to evolve, several key areas require further research and consideration. These include:

1. Meaningful citizen participation in the design and governance of AI systems
2. Development of ethical AI frameworks tailored to urban contexts
3. Interdisciplinary approaches combining insights from urban planning, computer science, and social sciences
4. Comparative studies of urban AI implementations across diverse global contexts
5. Assessment of long-term social, economic, and environmental impacts of AI-driven urban governance

While AI urbanism presents significant opportunities for enhancing urban management and addressing complex challenges, it also introduces new risks and ethical dilemmas. As cities increasingly adopt AI-driven technologies, it is crucial to critically examine these developments, ensuring they align with broader goals of sustainability, equity, and democratic governance (Cugurullo et al., 2023).

In conclusion, the integration of AI into urban assessment and governance presents both profound opportunities and significant challenges. While AI enhances our understanding of urban attributes and supports evidence-based decision-making, it also raises critical questions about democracy, equity, and ethics in urban development. The potential of AI to revolutionize urban studies and create comprehensive assessment models is tempered by concerns over privacy, algorithmic bias, and the changing dynamics of public-private partnerships in urban governance. As AI technologies advance and data collection methods improve, there is a pressing need for interdisciplinary collaboration to address these multifaceted challenges. Researchers, policymakers, and stakeholders must work together to develop ethical frameworks, ensure meaningful citizen participation, and critically examine the long-term impacts of AI-driven urban systems. By embracing these complex implications and actively shaping the development of urban AI, we can strive towards more inclusive, sustainable, and equitable cities that harness the benefits of technology while safeguarding democratic values and social justice.

Conclusion

This paper reviewed the use of AI for analyzing image data within the built environment. The literature review process proposed by Liberati et al. (2009), namely PRISMA, was followed to analyze more than 300 research works related to urban studies which use machine learning algorithms. Moreover, since all machine learning algorithms depend on the quality and quantum of fed data, widely used data collection methods, as well as the applied algorithms in this field, were discussed. Additionally, several implemented computer vision models used in urban studies were categorized: classification and perception models, object detection, and image segmentation. Finally, the most prominent urban attributes and AI methods used to analyze them were discussed. The integrated use of computer vision and machine learning algorithms within these studies was also elaborated upon. Deep learning algorithms were found to be the most efficient methods for evaluating urban imagery among all other computer vision techniques available. This efficiency can be attributed to the advent of algorithms such as CNN and RNNs alongside the availability of larger, more diverse imagery datasets and APIs. [Figure 3](#) demonstrates that

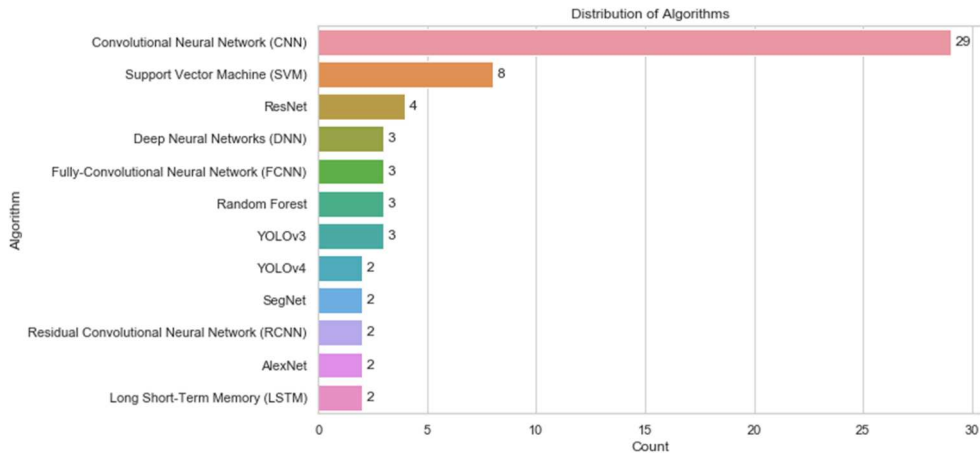


Figure 3. Top learning algorithms used in the studies analyzed in this research. All algorithms utilized in an article have been counted; therefore, each article might have more than one representative in this figure.

CNNs have been used in over half of all the analyzed articles in various forms, such as custom-designed and pre-trained architectures, and are followed by SVMs. Pre-trained models, such as different variations of ResNet and YOLO, are also among the top models used in urban studies. This figure reveals that deep-learning architectures have surpassed other AI-based methods recently and are more likely to influence this field of study in the future.

However, despite the remarkable advancement of state-of-the-art machine learning methods, the research community would greatly benefit by adopting a comprehensive and accessible data collection and storage approach for city-level imagery data of urban areas worldwide for supporting urban research. Besides this, predominantly urban attributes have been analyzed separately; there still remains a need for more comprehensive models to assess the correlation and interconnections between these features. Furthermore, no generic model or framework exists to relationally evaluate all types of urban conditions and characteristics to assess human wellbeing in urban areas holistically. For instance, a lack of a pervasive model for specific age groups, disadvantaged, or differently abled population still exists, thus negating aspects of urban inclusiveness. It is also noteworthy that computational power, as well as lack of overarching image datasets, cause hindrances to conduct large-scale multi-variate research within the urban domain. Thus, research scholars are encouraged to develop further computationally efficient computer vision approaches that enable the development of applicable accumulative urban perception models.

Notes on Contributors

Lily Moradi completed her PhD in architecture from the University of Technology Sydney (UTS), with expertise in urban planning and design. Fereshteh's research focuses on the inclusive built environment and emergent technologies. Her research interest includes Inclusive Cities, Sustainable Cities, Smart Cities, Spatial Analysis and Urban Informatics.

Nimish Biloria is an Associate Professor at the Faculty of Design Architecture Building at the University of Technology Sydney, Australia and a Visiting Associate Professor at the University of Johannesburg, South Africa. His research area: Empathic Environments, involves the study of human-environment interaction. Under this research umbrella, he has amassed extensive research and design experience in leading multi-scalar transdisciplinary projects spanning the areas of Architecture, Product innovation, Smart Cities, Urban Informatics and Spatial Data Science, Sustainable Mobility, Social Robotics, and Tangible and Embedded Interaction with a primary focus on enhancing urban health and wellbeing.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Lily Moradi  <http://orcid.org/0000-0002-6584-3282>

References

- S. F. Acosta and J. E. Camargo, “City Safety Perception Model Based on Visual Content of Street Images,” paper presented at the 2018 IEEE International Smart Cities Conference (2018) <https://doi.org/10.1109/ISC2.2018.8656949> Accessed February 5, 2025.
- S. F. Acosta and J. Camargo, *Predicting City Safety Perception Based on Visual Image Content* (Springer International Publishing, 2019) <https://doi.org/10.1007/978-3-030-13469-3> Accessed February 5, 2025.
- D. Ahmetovic, R. Manduchi, J. M. Coughlan, and S. Mascetti, “Mind Your Crossings: Mining GIS Imagery for Crosswalk Localization,” *ACM Transactions on Accessible Computing* 9: 4 (2017) 1–25.
- A. G. Andal, “Children’s Spaces in Coastal Cities: Challenges to Conventional Urban Understandings and Prospects for Child-Friendly Blue Urbanism,” *Children’s Geographies* 20: 5 (2022) 688–700 <https://doi.org/10.1080/14733285.2021.1962515> Accessed February 5, 2025.
- D. Anguelov et al., “Google the World at Capturing Street View: Street Level” (2010).
- N. Audebert, B. Le, and S. Lefèvre, “Beyond RGB: Very High Resolution Urban Remote Sensing with Multimodal Deep Networks,” *ISPRS Journal of Photogrammetry and Remote Sensing* (2017) <https://doi.org/10.1016/j.isprsjprs.2017.11.011> Accessed February 5, 2025.
- M. D. M. Bader et al., “Development and Deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to Measure Health-Related Neighborhood Conditions,” *Health & Place* 31 (2015) 163–172 <https://doi.org/10.1016/j.healthplace.2014.10.012> Accessed February 5, 2025.
- V. Badrinarayanan, A. Kendall, R. Cipolla, and S. Member, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39: 12 (2017) 2481–2495 <https://doi.org/10.1109/TPAMI.2017.2691124> Accessed February 5, 2025.
- Y. Bai et al., “Exploring Associations Between the Built Environment and Cycling Behaviour Around Urban Greenways from a Human-Scale Perspective,” *Land* 12: 3 (2023) 1–17.
- S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge: Cambridge University Press, 2014).
- F. Biljecki and K. Ito, “Street View Imagery in Urban Analytics and GIS: A Review,” *Landscape and Urban Planning* 215 (2021) 104217.
- A. Boguszewski, D. Batorski, N. Ziemia-Jankowska, T. Dziedzic, and A. Zambrzycka, “LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and

- Roads From Aerial Imagery,” paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Virtual Conference, June 19–25, 2021) 1102–1110.
- S. Borkman, et al., “Unity Perception: Generate Synthetic Data for Computer Vision,” *ArXiv* (2021) <http://arxiv.org/abs/2107.04259> Accessed February 5, 2025.
- A. Bottino et al., “Street Viewer: An Autonomous Vision Based Traffic Tracking System,” *Sensors* 16: 6 (2016) 813.
- C. Bustos, D. Rhoads, A. Solé-Ribalta, D. Masip, A. Arenas, A. Lapedriza, and J. Borge-Holthoefer, “Explainable, Automated Urban Interventions to Improve Pedestrian and Vehicle Safety,” *Transportation Research Part C: Emerging Technologies* 125 (2021) 103018.
- H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nuScenes: A Multimodal Dataset for Autonomous Driving,” paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Seattle, WA, June 13–19, 2020) 11618–11628.
- Y. Cao, Z. Wu, and C. Shen, “Estimating Depth from Monocular Images as Classification Using Deep Fully Convolutional Residual Networks,” *IEEE Transactions on Circuits and Systems for Video Technology* 28: 11 (2018) 3174–3182.
- P. Cardullo and R. Kitchin, “Smart Urbanism and Smart Citizenship: The Neoliberal Logic of ‘Citizen-focused’ Smart Cities in Europe,” *Environment and Planning C: Politics and Space* 37: 5 (2019) 813–830.
- V. Chaturvedi and W. T. de Vries, “Machine Learning Algorithms for Urban Land Use Planning: A Review,” *Urban Science* 5: 3 (2021) 68.
- A. Chaurasia and E. Culurciello, “LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation,” paper presented at the Visual Communications and Image Processing Conference (St. Petersburg, FL, December 10–13, 2017) 1–4.
- G. Cheng et al., “Multi-class Geospatial Object Detection and Geographic Image Classification Based on Collection of Part Detectors,” *ISPRS Journal of Photogrammetry and Remote Sensing* 98 (2014) 119–132.
- R. F. Chew et al., “Residential Scene Classification for Gridded Population Sampling in Developing Countries Using Deep Convolutional Neural Networks on Satellite Imagery,” *International Journal of Health Geographics* 17 (2018) 1–17.
- A. M. Chirkin and R. König, “Concept of Interactive Machine Learning in Urban Design Problems,” paper presented at the SEACHI 2016 on Smart Cities for Better Living with HCI and UX (San Jose, CA, May 7–12, 2016) 10–13.
- J. Cinnamon and L. Jahiu, “Panoramic Street-Level Imagery in Data-Driven Urban Research: A Comprehensive Global Review of Applications, Techniques, and Practical Considerations,” *ISPRS International Journal of Geo-Information* 10: 7 (2021) 471.
- P. Clarke et al., “Using Google Earth to Conduct a Neighborhood Audit: Reliability of a Virtual Audit Instrument,” *Health & Place* 16: 6 (2010) 1224–1229.
- J. Corcoran et al., “Neighbourhood Land Use Features, Collective Efficacy and Local Civic Actions,” *Urban Studies* 55: 11 (2018) 2372–2390.
- A. Crooks and L. See, “Leveraging Street Level Imagery for Urban Planning,” *Environment and Planning B: Urban Analytics and City Science* 49: 3 (2022) 773–776.
- F. Cugurullo, “Urban Artificial Intelligence: From Automation to Autonomy in the Smart City,” *Frontiers in Sustainable Cities* 2 (2020) 1–14.
- F. Cugurullo et al., “The Rise of AI Urbanism in Post-Smart Cities: A Critical Commentary on Urban Artificial Intelligence,” *Urban Studies* 61: 6 (2023) 1168–1182.
- M. De Nadai, R. L. Vieriu, G. Zen, S. Dragicevic, N. Naik, and B. Lepri, “Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life,” paper presented at the 25th ACM International Conference on Multimedia (Mountain View, CA, October 23–27, 2017) 1123–1131.
- A. Dixit, R. Kumar Chidambaram, and Z. Allam, “Safety and Risk Analysis of Autonomous Vehicles Using Computer Vision and Neural Networks,” *Vehicles* 3: 3 (2021) 595–617.

- J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," paper presented at the 31st International Conference on Machine Learning (Beijing, China, June 22–24, 2014) 647–655.
- A. Dubey et al., "Deep Learning the City: Quantifying Urban Perception at a Global Scale," in *Computer Vision – ECCV 2016* (Cham: Springer International Publishing, 2016) 398–413.
- D. T. Duncan et al., "The Built Environment and Depressive Symptoms Among Urban Youth: A Spatial Regression Study," *Spatial and Spatio-Temporal Epidemiology* 5 (2013) 11–25.
- A. P. D. A. Elkhoully, A. P. D. I. M. Shawket, and A. P. D. M. M. Ahmed, "Child Friendly Cities Between Theoretical Conceptualization and Real Perception: Capturing the Children's Perspectives to the City; a Pilot Case Study in GCR," *Journal of Architecture, Arts and Humanities Sciences* 5 (2022) 136–153.
- T. Erwin and J. Sweetkind-singer, "The National Geospatial Digital Archive: A Collaborative Project to Archive Geospatial Data," *Journal of Map & Geography Libraries* 5: 1 (2010) 6–25.
- S. M. A. Eslami et al., "Neural Scene Representation and Rendering," *Science* 360: 6394 (2018) 1204–1210.
- M. Forghani and M. R. Delavar, "A Quality Study of the OpenStreetMap Dataset for Tehran Mohammad," *ISPRS International Journal of Geo-Information* 3: 2 (2014) 750–763.
- F. Freitas, T. Berreth, Y.-C. Chen, and A. Jhala, "Characterizing the Perception of Urban Spaces From Visual Analytics of Street-Level Imagery," *AI & Society* 38: 4 (2023) 1361–1371.
- J. Geyer, et al., "A2D2: Audi Autonomous Driving Dataset," *ArXiv* (2020) <http://arxiv.org/abs/2004.06320> Accessed February 5, 2025.
- I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning* (Cambridge, MA: MIT Press, 2016).
- R. C. Gonzalez, *Digital Image Processing* (New Delhi: Pearson Education India, 2009).
- T. Hanák, I. Marović, and P. Aigela, "Perception of Residential Environment in Cities: A Comparative Study," *Procedia Engineering* 117: 1 (2015) 495–501.
- K. Hara, V. Le, and J. Froehlich, "Combining Crowdsourcing and Google Street View to Identify Street-Level Accessibility Problems," paper presented at the SIGCHI Conference on Human Factors in Computing Systems (Paris, France, April 27–May 2, 2013).
- S. R. T. Harikandeh, S. Aliakbary, and S. Taheri, "An Embedding Approach for Analyzing the Evolution of Research Topics With a Case Study on Computer Science Subdomains," *Scientometrics* 128: 3 (2023) 1567–1582.
- N. N. Hasibuan, M. Zarlis, and S. Efendi, "Detection and Tracking Different Type of Cars with YOLO Model Combination and Deep Sort Algorithm Based on Computer Vision of Traffic Controlling," *Jurnal Dan Penelitian Teknik Informatika* 6: 1 (2021) 210–220.
- N. Hassen and P. Kaufman, "Examining the Role of Urban Street Design in Enhancing Community Engagement: A Literature Review," *Health and Place* 41 (2016) 119–132.
- J. Hays and A. A. Efros, "IM2GPS: Estimating Geographic Information From a Single Image," paper presented at the 26th IEEE Conference on Computer Vision and Pattern Recognition (Anchorage, AK, June 23–28, 2008) 1–8.
- K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," paper presented at the Computer Vision and Pattern Recognition Conference (Las Vegas, NV, June 27–30, 2016) 770–778.
- L. He, G. Wang, and Z. Hu, "Learning Depth From Single Images With Deep Neural Network Embedding Focal Length," *IEEE Transactions on Image Processing* 27: 9 (2018) 4676–4689.
- J. R. Hipp, K. Kane, and J. H. Kim, "Recipes for Neighborhood Development: A Machine Learning Approach Toward Understanding the Impact of Mixing in Neighborhoods," *Landscape and Urban Planning* 164 (2017) 1–12.
- M. Hosseini et al., "Mapping the Walk: A Scalable Computer Vision Approach for Generating Sidewalk Network Datasets from Aerial Imagery," *SSRN Electronic Journal* 101 (2022) <https://doi.org/10.2139/ssrn.4086624> Accessed February 5, 2025.

- G. Huang and K. Q. Weinberger, "CondenseNet: An Efficient DenseNet using Learned Group Convolutions," paper presented at the Computer Vision and Pattern Recognition Conference (Salt Lake City, UT, June 18–22, 2018) 2752–2761.
- X. Huang et al., "The ApolloScape Open Dataset for Autonomous Driving and its Application," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42: 10 (2020) 2702–2719.
- M. R. Ibrahim, J. Haworth, and T. Cheng, "URBAN-i: From Urban Scenes to Mapping Slums, Transport Modes, and Pedestrians in Cities Using Deep Learning and Computer Vision," *Environment and Planning B: Urban Analytics and City Science* 48: 1 (2019) 76–93.
- K. Ito and F. Biljecki, "Assessing Bikeability With Street View Imagery and Computer Vision," *Transportation Research Part C: Emerging Technologies* 132 (2021) 103371.
- M. Ivanovs et al., "Improving Semantic Segmentation of Urban Scenes for Self-Driving Cars with Synthetic Images," *Sensors* 22: 6 (2022) 1–13.
- A. Karvonen, M. Cook, and H. Haarstad, "Urban Planning and the Smart City: Projects, Practices and Politics," *Urban Planning* 5: 1 (2020) 65–68.
- Y. Kawano and K. Yanai, "Rapid Mobile Object Recognition Using Fisher Vector," paper presented at the IAPR Asian Conference on Pattern Recognition (Naha, November 2013).
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification With Deep Convolutional Neural Networks," *Communications of the ACM* 60: 6 (2017) 84–90.
- J. H. Kwon and G. H. Cho, "An Examination of the Intersection Environment Associated with Perceived Crash Risk Among School-Aged Children: Using Street-Level Imagery and Computer Vision," *Accident Analysis and Prevention* 146 (2020) 105716.
- B. W. Koo, S. Guhathakurta, and N. Botchwey, "How Are Neighborhood and Street-Level Walkability Factors Associated With Walking Behaviors? A Big Data Approach Using Street View Images," *Environment and Behavior* 54: 1 (2022) 211–241.
- A. Larkin, X. Gu, L. Chen, and P. Hystad, "Predicting Perceptions of the Built Environment Using GIS, Satellite and Street View Image Approaches," *Landscape and Urban Planning* 216 (2021) 104257.
- S. Law, C. I. Seresinhe, Y. Shen, and M. Gutierrez-Roig, "Street-Frontage-Net: Urban Image Classification Using Deep Convolutional Neural Networks," *International Journal of Geographical Information Science* 34: 4 (2020) 681–707.
- S. Law, B. Paige, and C. Russell, "Take a Look Around: Using Street View and Satellite Images to Estimate House Prices," *ACM Transactions on Intelligent Systems and Technology* 10: 5 (2019) 1–19.
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature* 521: 7553 (2015) 436–444.
- J. Lee, B. Nam, and M. Abdel-Aty, "Effects of Pavement Surface Conditions on Traffic Crash Severity," *Journal of Transportation Engineering* 141: 10 (2015) 4015020.
- M. Li, H. Sheng, J. Irvin, H. Chung, A. Ying, T. Sun, A. Y. Ng, and D. A. Rodriguez, "Marked Crosswalks in US Transit-Oriented Station Areas, 2007–2020: A Computer Vision Approach Using Street View Imagery," *Environment and Planning B: Urban Analytics and City Science* 50: 2 (2023) 350–369.
- X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing Street-Level Urban Greenery Using Google Street View and a Modified Green View Index," *Urban Forestry & Urban Greening* 14: 3 (2015) 675–685.
- Y. Liao, J. Xie, and A. Geiger, "KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45: 3 (2023) 3292–3310.
- A. Liberati et al., "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration," *Journal of Clinical Epidemiology* 62: 10 (2009) e1–e34.
- G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation," paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (Honolulu, HI, July 21–26, 2017).
- T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," paper presented at the IEEE International Conference on Computer Vision (ICCV) (Venice, Italy, October 22–29, 2017) 2980–2988.

- L. Liu, E. A. Silva, C. Wu, and H. Wang, "A Machine Learning-Based Method for the Large-Scale Evaluation of the Qualities of the Urban Environment," *Computers, Environment and Urban Systems* 65 (2017) 113–125.
- A. Luusua, J. Ylipulli, M. Foth, and A. Aurigi, "Urban AI: Understanding the Emerging Role of Artificial Intelligence in Smart Cities," *AI and Society* 38: 3 (2023) 1039–1044.
- P. Mirowski et al., "The StreetLearn Environment and Dataset," *arXiv preprint arXiv: 1903.01292* (2019) 1.
- M. Reza and M. Hassan, "AI-Driven Solutions for Enhanced Waste Management and Recycling in Urban Areas," *International Journal of Sustainable Infrastructure for Cities and Societies* 8: 2 (2023) 1–13.
- N. Naik, *Visual Urban Sensing: Understanding Cities through Computer Vision* (Cambridge, MA: Massachusetts Institute of Technology, 2017).
- N. Naik, J. Philipoom, R. Raskar, and C. A. Hidalgo, "Streetscore - Predicting the Perceived Safety of One Million Streetscapes," paper presented at the IEEE Conference on Computer Vision and Pattern Recognition Workshops (Columbus, OH, June 23–28, 2014) 1.
- G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kotschieder, "The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes," paper presented at the IEEE International Conference on Computer Vision (Venice, October 22–29, 2017) 1.
- S. J. Norvig and P. Russell, *Artificial Intelligence: A Modern Approach* (New Jersey: Prentice Hall, 1995).
- C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2017).
- C. L. Odgers et al., "Systematic Social Observation of Children's Neighborhoods Using Google Street View: A Reliable and Cost-Effective Method," *Journal of Child Psychology and Psychiatry* (2012) <https://doi.org/10.1111/j.1469-7610.2012.02565.x> Accessed February 5, 2025 1.
- V. Ordonez and T. L. Berg, "Learning High-Level Judgments of Urban Perception," in D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds., *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VI* (Cham: Springer, 2014).
- Y. Park and M. Garcia, "Pedestrian Safety Perception and Urban Street Settings," *International Journal of Sustainable Transportation* 14: 11 (2020) 860–871.
- J. Phillips et al., "Deep Multi-Task Learning for Joint Localization, Perception, and Prediction," paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Virtual, June 19–25, 2021) 1.
- L. Porzi, S. Rota Bulò, B. Lepri, and E. Ricci, "Predicting and Understanding Urban Perception with Convolutional Neural Networks," paper presented at the 23rd ACM International Conference on Multimedia (Brisbane, Australia, October 26–30, 2015).
- S. Qiao et al., "VIP-DeepLab: Learning Visual Perception with Depth-aware Video Panoptic Segmentation," paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Virtual, June 19–25, 2021) 3997–4008.
- J. Redmon S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *ArXiv:1506.02640* (2015) 1.
- R. V. Remigio et al., "A Local View of Informal Urban Environments: A Mobile Phone-Based Neighborhood Audit of Street-Level Factors in a Brazilian Informal Community," *Journal of Urban Health* 96 (2019) 537–548.
- S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39: 6 (2017) 1137–1149.
- A. Roshan Zamir and M. Shah, "Image Geo-Localization Based on Multiple Nearest Neighbor Feature Matching Using Generalized Graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36: 8 (2014) 1546–1558.
- A. G. Rundle, M. D. M. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler, "Using Google Street View to Audit Neighborhood Environments," *American Journal of Preventive Medicine* 40: 1 (2011) 94–100.

- A. Rzotkiewicz, A. L. Pearson, B. V. Dougherty, A. Shortridge, and N. Wilson, "Systematic Review of the Use of Google Street View in Health Research: Major Themes, Strengths, Weaknesses and Possibilities for Future Research," *Health and Place* 52 (2018) 240–246.
- J. Sadowski and R. Bendor, "Selling Smartness: Corporate Narratives and the Smart City as a Sociotechnical Imaginary," *Science Technology and Human Values* 44: 3 (2019) 540–563.
- P. Salesses, K. Schechtner, and A. Hidalgo, "The Collaborative Image of The City: Mapping the Inequality of Urban Perception," *PLOS ONE* 8: 7 (2013) e68400.
- H. Sheng, R. Cong, D. Yang, R. Chen, S. Wang, and Z. Cui, "UrbanLF: A Comprehensive Light Field Dataset for Semantic Segmentation of Urban Scenes," *IEEE Transactions on Circuits and Systems for Video Technology* 32: 11 (2022) 7880–7893.
- K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556 (2014).
- Y. Song, H. Ning, X. Ye, D. Chandana, and S. Wang, "Analyze the Usage of Urban Greenways Through Social Media Images and Computer Vision," *Environment and Planning B: Urban Analytics and City Science* 49: 6 (2022) 1682–1696.
- S. Srivastava, J. E. Vargas-muñoz, and D. Tuia, "Understanding Urban Landuse from the Above and Ground Perspectives: A Deep Learning, Multimodal Solution," *Remote Sensing of Environment* 228 (2019) 129–143.
- A. Stančić, V. Vyroubal, and V. Slijepčević, "Classification Efficiency of Pre-Trained Deep CNN Models on Camera Trap Images," *Journal of Imaging* 8: 2 (2022).
- E. Suel, J. W. Polak, J. E. Bennett, and M. Ezzati, "Measuring Social, Environmental and Health Inequalities Using Deep Learning and Street Imagery," *Scientific Reports* (2019) 1–10.
- C. Szegedy et al., "Going Deeper with Convolutions" (2014).
- D. Szpilko, F. J. Naharro, G. Lăzăroi, E. Nica, and A. de la Torre Gallegos, "Artificial Intelligence in the Smart City - A Literature Review," *Engineering Management in Production and Services* 15: 4 (2023) 53–75.
- S. Taheri and S. Aliakbary, "Research Trend Prediction in Computer Science Publications: A Deep Neural Network Approach," *Scientometrics* (2022).
- H. Tian, Z. Han, W. Xu, X. Liu, W. Qiu, and W. Li, "Evolution of Historical Urban Landscape with Computer Vision and Machine Learning: A Case Study of Berlin," *Journal of Digital Landscape Architecture* 2021: 6 (2021) 436–451.
- M. Umair, M. U. Farooq, R. H. Raza, Q. Chen, and B. Abdulhai, "Efficient Video-Based Vehicle Queue Length Estimation Using Computer Vision and Deep Learning for an Urban Traffic Scenario," *Processes* 9: 10 (2021) 1–19.
- D. Verma, A. Jana, and K. Ramaritham, "Machine-Based Understanding of Manually Collected Visual and Auditory Datasets for Urban Perception Studies," *Landscape and Urban Planning* 190 (2019) 103604.
- S. Wael, A. Elshater, and S. Afifi, "Mapping User Experiences around Transit Stops Using Computer Vision Technology: Action Priorities from Cairo," *Sustainability* 14: 17 (2022).
- H. Wang and Y. Yang, "Neighbourhood Walkability: A Review and Bibliometric Analysis," *Cities* 93 (2019) 43–61.
- L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "UNetFormer: A UNet-Like Transformer for Efficient Semantic Segmentation of Remote Sensing Urban Scene Imagery," *ArXiv* (2022) 1–49.
- T. Weyand, I. Kostrikov, and J. Philbin, "Planet - Photo Geolocation with Convolutional Neural Networks," in *Computer Vision - ECCV 2016* (Cham: Springer International Publishing, 2016).
- P. K. Y. Wong, H. Luo, M. Wang, P. H. Leung, and J. C. P. Cheng, "Recognition of Pedestrian Trajectories and Attributes with Computer Vision and Deep Learning Techniques," *Advanced Engineering Informatics* 49 (2021) 101356.
- J. Xie, Y. Zheng, R. Du, W. Xiong, Y. Cao, Z. Ma, D. Cao, and J. Guo, "Deep Learning-Based Computer Vision for Surveillance in ITS: Evaluation of State-of-the-Art Methods," *IEEE Transactions on Vehicular Technology* 70: 4 (2021) 3027–3042.

- Y. Xu, F. Cugurullo, H. Zhang, A. Gaio, and W. Zhang, "The Emergence of Artificial Intelligence in Anticipatory Urban Governance: Multi-Scalar Evidence of China's Transition to City Brains," *Journal of Urban Technology* (2024) 1–25.
- Z. Yan, L. Jiang, X. Huang, L. Zhang, and X. Zhou, "Intelligent Urbanism with Artificial Intelligence in Shaping Tomorrow's Smart Cities: Current Developments, Trends, and Future Directions," *Journal of Cloud Computing* 12: 1 (2023).
- M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for Semantic Segmentation in Street Scenes," paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Salt Lake City, UT, June 18–22, 2018) 3684–3692.
- Z. Yang, C. Zhao, H. Maeda, and Y. Sekimoto, "Development of a Large-Scale Roadside Facility Detection Model Based on the Mapillary Dataset," *Sensors* 22: 24 (2022) 9992.
- Y. Ye, W. Zeng, Q. Shen, X. Zhang, and Y. Lu, "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: 8 (2019) 1439–1457.
- T. Yigitcanlar, K. C. Desouza, L. Butler, and F. Roozkhosh, "Contributions and Risks of Artificial Intelligence (AI) in Building Smarter Cities: Insights from a Systematic Review of the Literature," *Energies* 13: 6 (2020).
- L. Yin and Z. Wang, "Measuring Visual Enclosure for Street Walkability: Using Machine Learning Algorithms and Google Street View Imagery," *Applied Geography* 76 (2016) 147–153.
- S. Youssoufi, H. Houot, G. Vuidel, S. Pujol, and F. Mauny, "Combining Visual and Noise Characteristics of a Neighborhood Environment to Model Residential Satisfaction: An Application Using GIS-Based Metrics," *Landscape and Urban Planning* 204 (2020) 103932.
- H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks," *Sensors* 17: 7 (2017) 1–16.
- F. Zhang, Z. Fan, Y. Kang, Y. Hu, and C. Ratti, "Perception Bias: Deciphering a Mismatch Between Urban Crime and Perception of Safety," *Landscape and Urban Planning* 207 (2021) 104003.
- T. Zhang, J. Su, Z. Xu, Y. Luo, and J. Li, "Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier," *Applied Sciences* 11: 2 (2021) 1–17.
- J. Zhao, X. Liu, Y. Kuang, Y. V. Chen, and B. Yang, "Deep CNN-Based Methods to Evaluate Neighborhood-Scale Urban Valuation Through Street Scenes Perception," paper presented at the 2018 IEEE 3rd International Conference on Data Science in Cyberspace (Guangzhou, June 18–20, 2018).
- X. Zheng and M. Amemiya, "Method for Applying Crowdsourced Street-Level Imagery Data to Evaluate Street-Level Greenness," *ISPRS International Journal of Geo-Information* 12: 3 (2023) 108.
- B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition Using Places Database," paper presented at the Advances in Neural Information Processing Systems 27 (NIPS) (Montreal, Canada, December 8–13, 2014) 487–495.