Wiley
Human Behavior and Emerging Technologies
Volume 2025, Article ID 8537833, 14 pages
https://doi.org/10.1155/hbe2/8537833

Research Article

Quantifying Visual Navigation in Campus Open Spaces Using a Computer Vision Model

Nabil Mohareb[1] **and Abdelaziz Ashraf**[2]

[1]*Department of Architecture, The American University in Cairo (AUC), New Cairo, Egypt*
[2]*Computer Science Department, October University of Modern Sciences and Arts, Cairo, Egypt*

Correspondence should be addressed to Nabil Mohareb; nabil.mohareb@aucegypt.edu

Received 16 August 2024; Accepted 11 March 2025

Academic Editor: Amine Moussa

Copyright © 2025 Nabil Mohareb and Abdelaziz Ashraf. Human Behavior and Emerging Technologies published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

This study presents a framework specifically designed to measure and quantify visual experiences within academic campus environments. The framework addresses the need for quantitative methods to analyze spatial experiences, focusing on key elements of the built environment, such as visible sky, greenery, and spatial enclosure. While the framework emphasizes visual components, it does not aim to analyze broader sensory or emotional experiences. Instead, it establishes a foundation for future research to explore these dimensions comprehensively. The methodology utilizes mobile phones equipped with digital cameras and GPS sensors to capture first-person visual data while participants freely navigate through campus open spaces. Computer vision techniques, including instance segmentation and convolutional neural networks, are employed to categorize architectural and natural elements within each video frame. This process quantifies the proportional composition of visual elements such as greenery, open sky, walkways, buildings, and other structures that participants encounter. The framework is implemented as a Python model that is capable of generating quantitative outcomes. Additionally, the analysis is enhanced by integrating geographic information systems (GISs) for spatial analysis, allowing us to identify navigation and visual engagement patterns. This comprehensive methodology not only quantifies the visual attributes of spaces but also interprets their impact on the behavior and experiences of campus users. This framework offers insights into how navigation choices, visual experiences, and the types of scenes encountered on campus can be understood and analyzed. The results aim to guide urban designers in better understanding university students' open space needs by exploring the connections between natural movement patterns and visual preferences. This research complements other qualitative approaches, providing a more comprehensive perspective on campus space utilization.

Keywords: computer vision; convolutional neural networks; instance segmentation; navigation behavior; spatial analysis; university open spaces

1. Introduction

As universities worldwide face increasing challenges in creating engaging and sustainable campus environments, a gap exists in understanding how visual elements influence user experience and navigation behavior. While university campus open spaces are visually rich environments, current research primarily emphasizes their role in enhancing students' quality of life through restorative settings that improve the overall campus experience [1]. These spaces are essential for fostering social engagement, happiness, and well-being [2] while also contributing to ecological sustainability by enhancing biodiversity and environmental education [3]. The importance of sensory design is particularly noted in hot–humid climates, where comfort is critical for improving outdoor experiences [3]. Additionally, spatial layouts and physical features—such as seating, shade, and visibility—significantly influence student behavior and social interactions, with informal adaptations emerging when formal setups are unavailable [4]. Campus landscapes, including green and blue spaces, positively affect mental health and well-being, fostering happiness, inclusiveness [5], and relaxation while reducing feelings of disappointment and depression [6, 7].

However, existing research primarily focuses on these qualities in isolation or as outcomes of designed environments, with limited attention to how users directly perceive urban scene elements during navigation. This study addresses this gap by developing an analytical framework to systematically quantify visual information from pedestrians' direct visual perspectives. By examining components of the built environment—such as buildings, vegetation, pathways, and architectural features—this research lays the groundwork for future correlation of these visual factors with emotional and spatial awareness, as well as environmental influences.

To achieve this, the research integrates computer vision (CV) techniques, such as instance segmentation and convolutional neural networks (CNNs), to analyze visual attention. Unlike traditional methods relying on static or panoramic online images [8–10], this approach captures real-time visual experiences using images extracted from videos recorded at eye level during walking scenarios. This application, combined with georeferenced data, provides actionable insights into how users perceive their surroundings, offering a new perspective on campus design that aligns with user visual behavior and spatial preferences.

The proposed framework, implementing a developed Python application, will be tested in a pilot study focused on a user navigation experience. This proof-of-concept study evaluates the feasibility of the methodology and refines it for broader applications. Findings from this research will provide valuable insights into designing open spaces that enhance visual engagement and spatial navigation, advancing evidence-based practices for creating functional and appealing university campuses.

2. Literature Review

This literature review examines how CV and instance image analysis have influenced urban studies. While urban theories often emphasized aesthetic and formal concerns, subsequent critiques by scholars and practitioners prompted a broader human- and behavior-centered approach. Today, advanced technologies, such as CV, enable more detailed analyses of how people interact with urban environments, enhancing our understanding of both physical design and social dynamics. It highlights the evolution from basic data collection to advanced technologies like CV for understanding cities, blending technological objectivity with insights into human behavior. The review also addresses these technological advances' social and ethical implications in shaping urban environments. Lucas and Romice [11] emphasize that urban design often prioritizes visual sensory modalities while underrepresenting other sensory dimensions such as sound, touch, and thermal perception. They advocate for a multisensory approach to urban representation and design, which more accurately reflects human experiences in urban spaces.

Urban study theories have evolved from an aesthetic focus to a deeper understanding of human interactions within urban spaces [12]. This shift, notably in the mid-20th century, moved from prescriptive theories to empirical studies, driven by concerns over the alienating effects of modernist designs. This change underscores the importance of human-centric approaches in urban planning, exemplified by Lynch and Whyte's work on mental mapping and public space usage, which significantly influences contemporary urban design by emphasizing space vitality and usage [5, 13, 14].

Traditional data collection methods in urban studies, such as images, videos, and direct observation, faced labor intensity and scalability limitations. The advent of advanced sensing technologies and geotagged imagery enabled more efficient and extensive urban analysis [12]. Whyte's 1980 Street Life Project, utilizing time-lapse filming of pedestrian behavior, marked an early integration of technology and detailed observation in urban studies [15], revealing key behavioral trends.

Modern urban studies have evolved by incorporating hybrid sensing, big data, and AI, revolutionizing the analysis of physical and socioeconomic conditions and human dynamics [12]. The application of CV technologies in urban management represents a progression in employing image-based AI for data collection [15]. However, these objective approaches potentially overlook the complexities of human perceptions [16].

A balanced approach integrating objective and subjective measures is necessary. Subjective measures, derived from interviews and surveys, offer deeper insights into human behavior by considering the cognitive mapping of environments [13]. However, traditional methods for collecting perception data often lack consistency and reliability and are time-consuming, expensive, and challenging to interpret [16].

By combining advanced technologies like CV with subjective insights, urban studies can comprehensively understand urban spaces, informing design decisions that cater to human needs and experiences while addressing social and ethical implications.

2.1. Technological Advancements in Urban Studies: Sensing, Big Data, and AI Integration. Street-level imagery has become vital in various research areas, including urban planning, public health, and real estate, due to its accessibility, coverage, and objective views [12]. Computer vision techniques, which convert images and videos into numerical matrices by analyzing red, green, and blue (RGB) pixel values, require extensive data to train models for accurately differentiating human actions from background noise [15]. Semantic image segmentation (SiS), as described by Csurka et al. [17], is crucial in CV, where each pixel in an image is assigned to a specific semantic class to understand different image parts.

Implementing activity surveys with CV involves data collection and processing handling distinct video and GPS data streams. Adherence to guidelines ensuring that videos reflect human perception and are consistent for accurate tracking is crucial [18]. Integrating these data streams facilitates accurate mapping of activity data in physical space, enhancing understanding of human interactions in urban environments.

2.2. Evaluating Urban Environments: Semantic Segmentation and CV Methods. The "Urban Visual Intelligence" framework, elucidated by Zhang et al. [12], integrates AI with imagery to analyze urban environments, addressing physical and socioeconomic dimensions. This framework overcomes the limitations of traditional methods, providing a nuanced perspective that includes observing urban environments at a human scale, deriving semantic information from imagery, quantifying physical environments, and exploring their physical and socioeconomic interplay.

Challenging the assumption of linear relationships between the built environment and walking behavior, Liu et al. [19] introduced an alternate perspective. Their research suggests that intrinsic motivations and utilitarian travel needs may influence walking desires, indicating a potential saturation point in the built environment's influence on walking. Addressing the geographical gap in research, the study explores nonlinear associations between street view–derived environmental characteristics and pedestrian walking duration in Amsterdam, focusing on identifying influencing features and understanding their variations across different times, including weekdays and weekends.

To support this analysis, Liu et al. [19] utilized semantically segmented street environmental features with a fully CNN, specifically the Xception-71 CNN, pretrained on the Cityscapes dataset comprising pixel-level annotated street scenes from 50 different cities, demonstrating favorable performance compared to alternative CNN architectures. Yan et al. [20] employed an urban perception evaluation framework to complement this approach. This framework analyzes a vast collection of old city landscape street images, focusing on image semantic segmentation to categorize data based on landscape spatial elements. The study quantifies elements such as building area, road area, green viewing rate, human and vehicle flow, and sky area, filtering out extreme proportions of certain elements to maintain accuracy. Their findings reveal an average greenness rate of 30.14% in the analyzed images.

Shifting the focus from social media imagery, Duarte and Ratti [21] emphasize using specialized urban cameras designed to collect visual data about cities. This approach has also been effectively applied in Cairo, where Wael et al. [22] used CV technology to monitor movement behavior from above, providing actionable insights into pedestrian dynamics and urban configuration. This shift from user-generated, geotagged photographs used in previous studies to assess urban attractiveness or aesthetic appeal represents a move toward more objective urban data collection [21]. Additionally, Lee et al. [23] highlight using CV technologies, such as semantic segmentation and edge detection, to evaluate urban space quality. These technologies assess aspects like enclosure, openness, greenery, and the ratio of feature areas, while edge detection quantifies complexity. The data obtained is processed and interpreted using a machine learning model and the SHAP algorithm, contributing to a comprehensive understanding of urban space elements and their impact on urban life.

2.3. Applications of CV Segmentation in Urban Spaces and Pedestrian Behavior Analysis. Computer vision segmentation techniques are invaluable for understanding and optimizing urban spaces for pedestrians by assessing factors that influence walkability and pedestrian behavior and various other aspects (see Figure 1). Walking shapes urban experiences and community dynamics [23], with street environment qualities like green spaces, building layouts, and open areas directly impacting walking appeal. Pedestrian satisfaction relates to perceptions of imageability, enclosure, human scale, openness, and complexity [24], measurable through segmentation.

Segmentation enables virtual assessments of pedestrian volumes, a key walkability indicator [25]. Street View Imagery (SVI) and segmentation algorithms like the visual walkability index [26] evaluate aspects like crowdedness and obstacles, revealing walkability variations across locations. Urban morphology qualities like area ratio, enclosure, openness, and complexity are extracted through semantic segmentation [23]. Enclosure (D:H ratio) and openness (visible sky proportion) impact pedestrian comfort, while complexity enhances satisfaction. Greenery analysis, such as the green view index (GVI) from SVI [26], assesses urban greenery's pedestrian perspective, shading, and aesthetic value. Health studies utilize SVI to analyze links between physical activity, neighborhood greenery, sidewalk quality, and recreational facilities [26]. Urban perception research leverages SVI's human-centered street characterization for assessing safety, wealth, and vibrancy, integrating surveys and audio data. In transportation, SVI enables virtual street audits, road analysis, pedestrian volume assessment [25], traffic indicators, and cycling pattern exploration based on greenery and architecture [26]. Monitoring pedestrian and vehicle traffic informs planning decisions. Historical change detection through image comparisons supports urban heritage conservation.

However, using SVI raises ethical concerns regarding transparency, accuracy, and potential profiling [15], necessitating ongoing policy discussions for responsible and ethical implementation in urban studies.

3. Research Framework and Methodology

A pilot approach was adopted to rigorously evaluate the suggested CV framework developed for this study, centering on an in-depth analysis of one student's navigational experience through the campus. This methodological choice was not intended to generalize findings across user groups but to demonstrate the feasibility and robustness of the framework in capturing and quantifying visual experiences. The primary focus was ensuring the framework's analytical precision and validating CV and geospatial data integration. Future studies may expand to include multiple users to explore variations in visual experiences and navigation patterns; however, this was beyond the scope of the current research.

Advancements in hardware and research into CNNs have led to sophisticated deep learning models for image analysis. Notable examples include the VGGNet [27], the YOLO (You Only Look Once) object detection framework [28, 29], and the Fast R-CNN [30, 31] for object detection and classification.

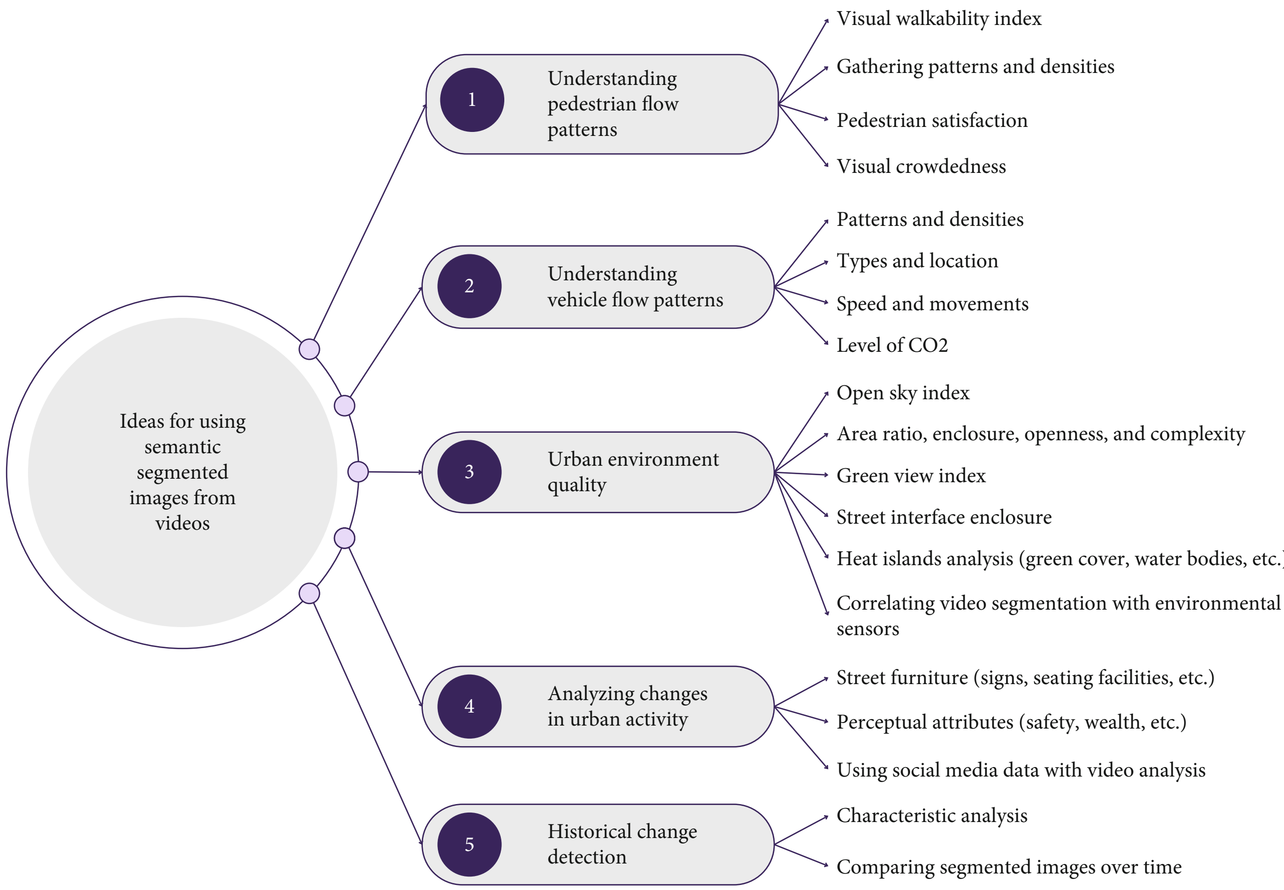


Figure 1: These points summarize the literature review findings on using images with segmented content extracted from videos.

Object detection involves identifying and localizing objects within an image. Classification assigns labels to entire images or objects, which is essential for interpreting visual content. Segmentation offers a granular approach by assigning labels to individual pixels, providing detailed analysis crucial for applications like land cover delineation and augmented reality. Segmentation divides an image into meaningful regions, grouping pixels into semantically coherent clusters. Semantic segmentation classifies each pixel into predefined categories, while instance segmentation differentiates between individual object instances [32].

The suggested model represents an advancement in CV applications for urban space analysis, blending the YOLO framework's object detection and segmentation capabilities with geospatial data integration. Utilizing the Ultralytics library for using the YOLO segmentation model for prediction, the framework processes video frames to extract masks, calculate segmented object areas, and identify focal objects through bounding boxes.

The algorithm calculates the total frame area, retrieves object labels, finds contours defining object shapes within masks, and computes the percentage of frame coverage by objects. It also determines the object closest to the frame's center using Euclidean distance calculations and bounding box center coordinates. The algorithm calculates the total area of the image frame (area = height × width) and retrieves the label of the detected object (e.g., "car" and "person"). It then finds the contours defining the object's shape within the mask using the green formula and calculates the total area enclosed by those contours (sum of each contour object area). Utilizing this object area and the total frame area, it computes the percentage of the frame covered by the object (percentage of coverage = (object area/total area) × 100).

The proposed framework captures and quantifies what user (s) see and where they see it as they navigate urban spaces, providing a detailed analysis of their visual experience. This is achieved by integrating a Cloud-GIS-based mobile application platform with CV analysis. The mobile application records the user's movement in space by tracking x, y, and z coordinates and synchronizing these spatial data points with the timestamps of video and image frames. This synchronization enables precise geolocation of visual data, allowing each video frame to be accurately linked to the specific location where it was captured during navigation (see Figure 2).

The algorithm processes each video frame to determine the object closest to its center by calculating the Euclidean distance between the frame's center coordinates (center_x and center_y) and the centers of detected object bounding boxes. If the frame's center falls within a bounding box, the closest object is identified, and the object's label and frame number are recorded. This process enables detailed extraction of object coverage percentages and dynamic tracking of prominent objects over time, enriching the understanding of urban elements encountered along the user's path.

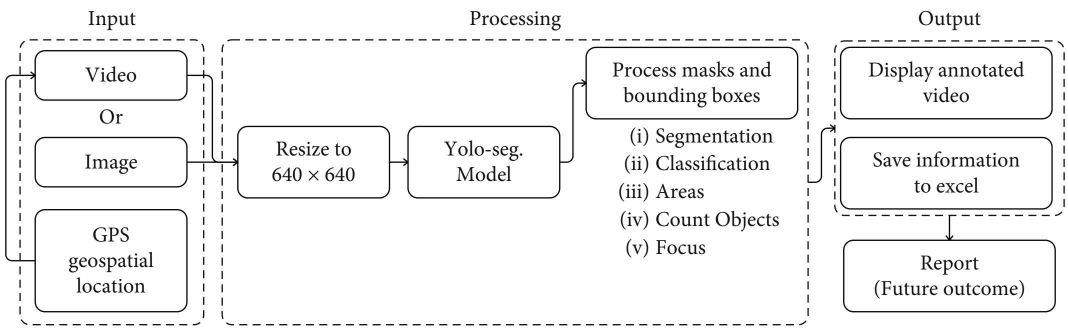


Figure 2: Framework of the proposed model, translated to Python code, delineating the stages of input, processing, and output.

The model generates a comprehensive dataset that links spatial and visual information. Data on classified objects, pedestrian and vehicle counts, and areas of attention focus are compiled into a CSV file. This file is an intermediary for integrating CV outputs with the GIS platform. The georeferenced data enables spatial analysis of visual attention patterns, revealing how different urban elements, such as architectural features, landscape components, open sky, and human presence, are perceived and distributed in urban open spaces.

The case study used a simple straight-line movement to validate the synchronization between spatial and visual data. The results demonstrated accurate mapping of visual elements to their corresponding spatial locations, confirming the robustness of the integration. This quantified approach provides insights into urban spaces' visual structure and design quality, offering a foundation for future applications. While this study focuses on visual perception, future research could expand the framework to incorporate other sensory dimensions, such as auditory inputs, thermal comfort, or emotional responses. This would enable a holistic analysis of sensory experiences during navigation, advancing the understanding of user interactions with urban environments.

This framework represents a new approach to urban studies by quantifying and mapping visual experiences through dynamic movement in urban contexts. By combining CV analysis with geospatial data, it facilitates a multidimensional understanding of visual attention and spatial context, supporting more informed urban planning and design decisions.

This framework leverages CV and geospatial data to map visual experiences in urban spaces, providing researchers and planners with a tool to analyze spatial dynamics and support informed decision-making.

4. Developing the Digital Application

The following datasets for this research are used to detect and segment open urban spaces. The dataset comprises 2763 images across 16 distinct classes, with an average image resolution of 2048 × 1024 pixels. These images were collected from various sources, including "Cityscapes" and other online repositories via "Roboflow" datasets, to create a comprehensive dataset tailored to the research needs. Specific classes required consolidation and preprocessing from diverse sources to ensure representativeness for the study (see Figures 3 and 4).

The dataset underwent preprocessing to address class imbalances and resize requirements. Automatic orientation adjustments were applied, and images were resized to 320 × 224 pixels. Data augmentation techniques were employed to achieve class balance, including horizontal and vertical flips and slight rotations within ±5°. The figure illustrates augmentation and represents the model's perception of classes like sky, road, and persons.

The YOLOv8 segmentation model was fine-tuned using a custom dataset of 5000 images across five distinct classes, divided into training (80%), validation (10%), and testing (10%) sets. The training involved 100 epochs with images resized to 640 × 640 pixels, a batch size of 16, and using the "Adamsw" optimizer. The model achieved a mean average precision (mAP@0.5) of 43.8 on the held-out test set. Class-specific mAP analysis demonstrated the model's segmentation proficiency across various object categories.

Comparative assessment against state-of-the-art approaches emphasized the proposed method's competitiveness. Qualitative inspection through visual examples showcased the model's accuracy in segmenting objects across diverse scenarios. Analysis of the loss curve during training (see Figure 5) provided insights into convergence behavior, with no significant signs of overfitting or underfitting observed. The results contextualized the achieved mAP score, outlining the strengths and limitations of the YOLO model for segmentation tasks and suggesting avenues for future research and enhancement (see Figures 5 and 6).

5. Data Analysis of the Case Study

As a case study, the research focuses on the main walking spine at the American University in Cairo's (AUC) new campus in New Cairo, Egypt. This primary pedestrian corridor, connecting most campus buildings, was carefully selected for its controlled environment, rich environmental elements, and diverse morphological cross-sections along the spine. The main intention is to test the developed Python model by analyzing video footage captured while navigating through this spine.

A pilot study with a single user evaluated the model's functionality and refined its design. While a single-user approach inherently limits the ability to generalize findings,

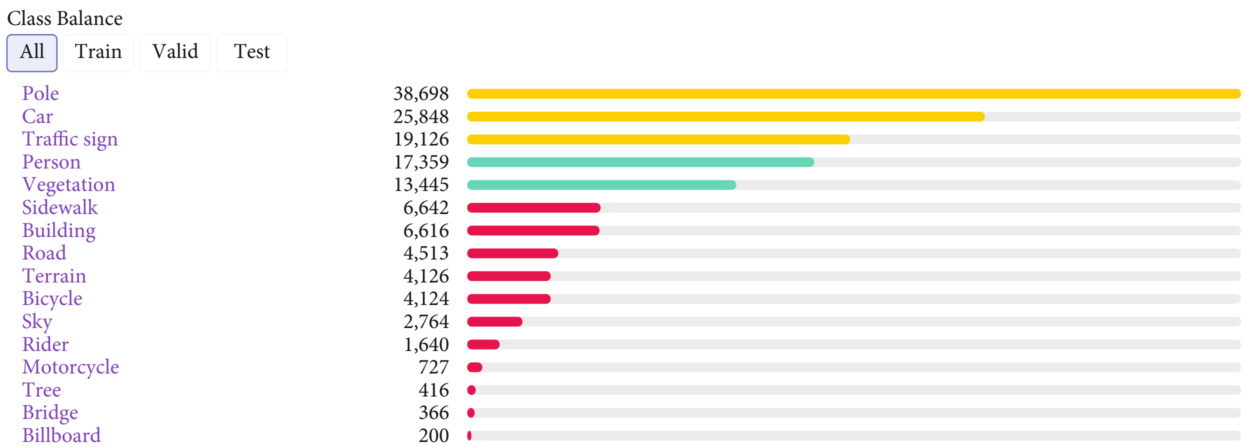


Figure 3: Class distributions and classes.

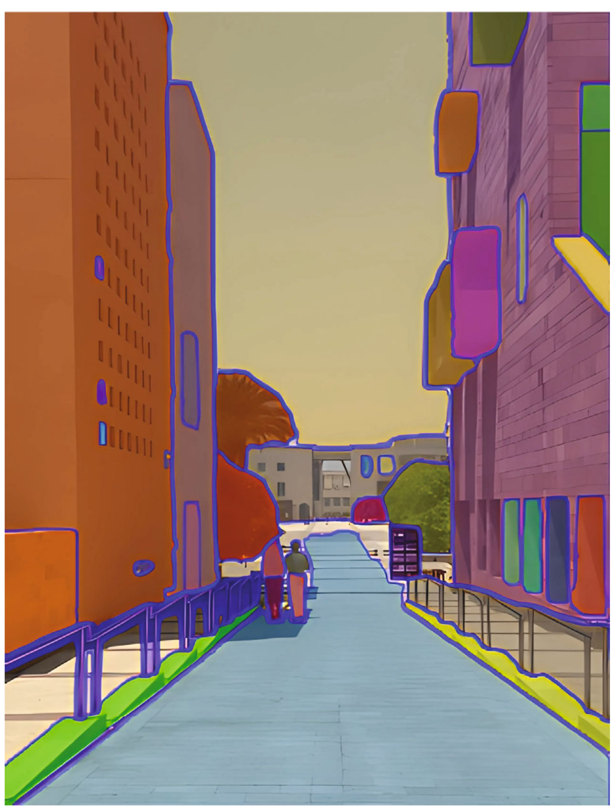


Figure 4: Segmentation masks in an image from the dataset.

the primary aim of this study is to build and validate a replicable framework capable of capturing and quantifying visual experiences. The objective is not to analyze user differences but to establish a methodology that can be applied to study such variations in future research. The findings from this pilot test serve as a proof of concept, demonstrating the Python model's potential for broader applications in urban study research.

Visual data is processed to elaborate on the classification and segmentation of images from video to quantify environmental elements (see Figure 7). It dissects the fluctuating patterns of visual engagement in an urban setting, which is pivotal for decoding the spatial configuration's influence on pedestrian behavior. Figure 7a maps out the instances of object detection over time, highlighting the dominance of buildings and sky in the visual field—a reflection of their physical and visual prominence. Figure 7b complements this with a histogram connecting the visual focus to spatial progression, indicating selective visual attention influenced by environmental cues. Regular detections of vegetation suggest its role as a visual constant and navigational aid. The bar chart in Figure 7c prioritizes visual elements by frequency, underscoring roads as pivotal navigation aids and reducing electric cars and poles to less-noticed visual noise unless they become obstructions or hold significance. Figure 7d's box plots expose the size distribution of detected objects, underscoring the scale's impact on perception. The variability of areas covered by each object, especially the outliers, can inform proportional balance in urban spaces for optimal visual accessibility.

Together, these visual data sets corroborate the intricate nature of pedestrian visual experiences, highlighting the necessity for urban spaces that cater to pedestrians' dynamic visual stimuli, which affect their navigation and psychological well-being.

The AUC Library Plaza offers a captivating pedestrian experience, as evidenced by the integrative visual and quantitative analysis presented in Figure 8, which captures the dynamic interplay between architectural presence and natural elements along a delineated path. The upper graph (Figure 8a) quantitatively tracks the visibility of various elements—buildings, roads (pedestrian pathway), sidewalks, sky, trees, and vegetation—as the passerby perceives. Substantial fluctuation in-building coverage, observed at Waypoints 2, 6, and 9, correspond to the dominance of prominent architectural forms such as the Abdul Latif Jameel Hall and the AUC Library. These points of architectural prominence contrast with more open spatial arrangements in Zones 4 and 8, where the sky and vegetation are more visible.

The graphical portrayal illuminates the visual rhythm dictated by the built environment and highlights the consistent punctuation of the pedestrian's view by greenery, implying a sustained visual dialogue with nature amidst the urban backdrop. The coverage of roads and sidewalks ebbs and

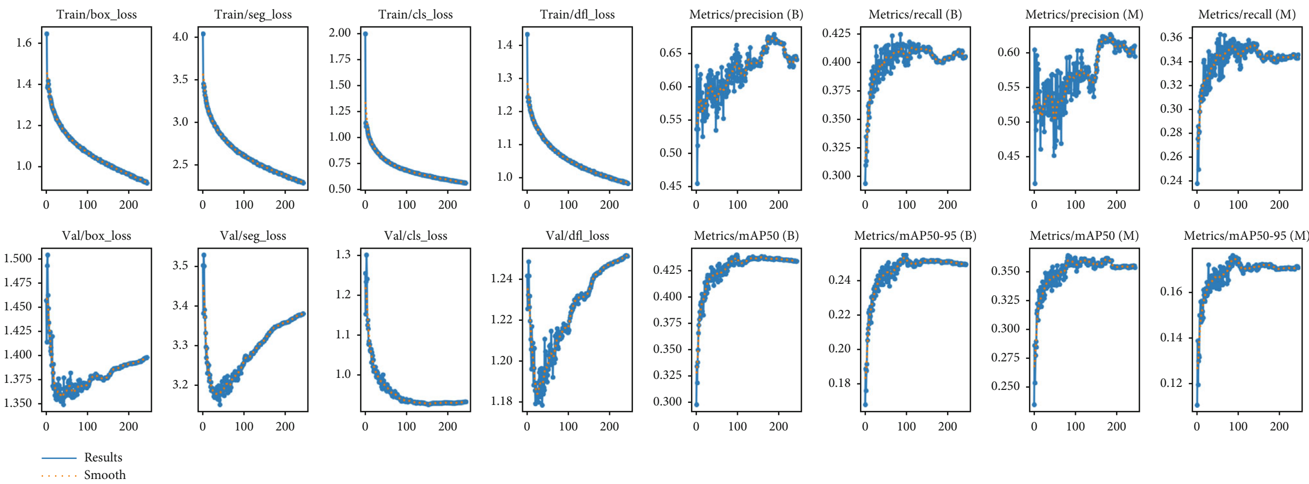


Figure 5: Model's losses in different metrics for each train and validation dataset, such as precision, recall, and mAP50.

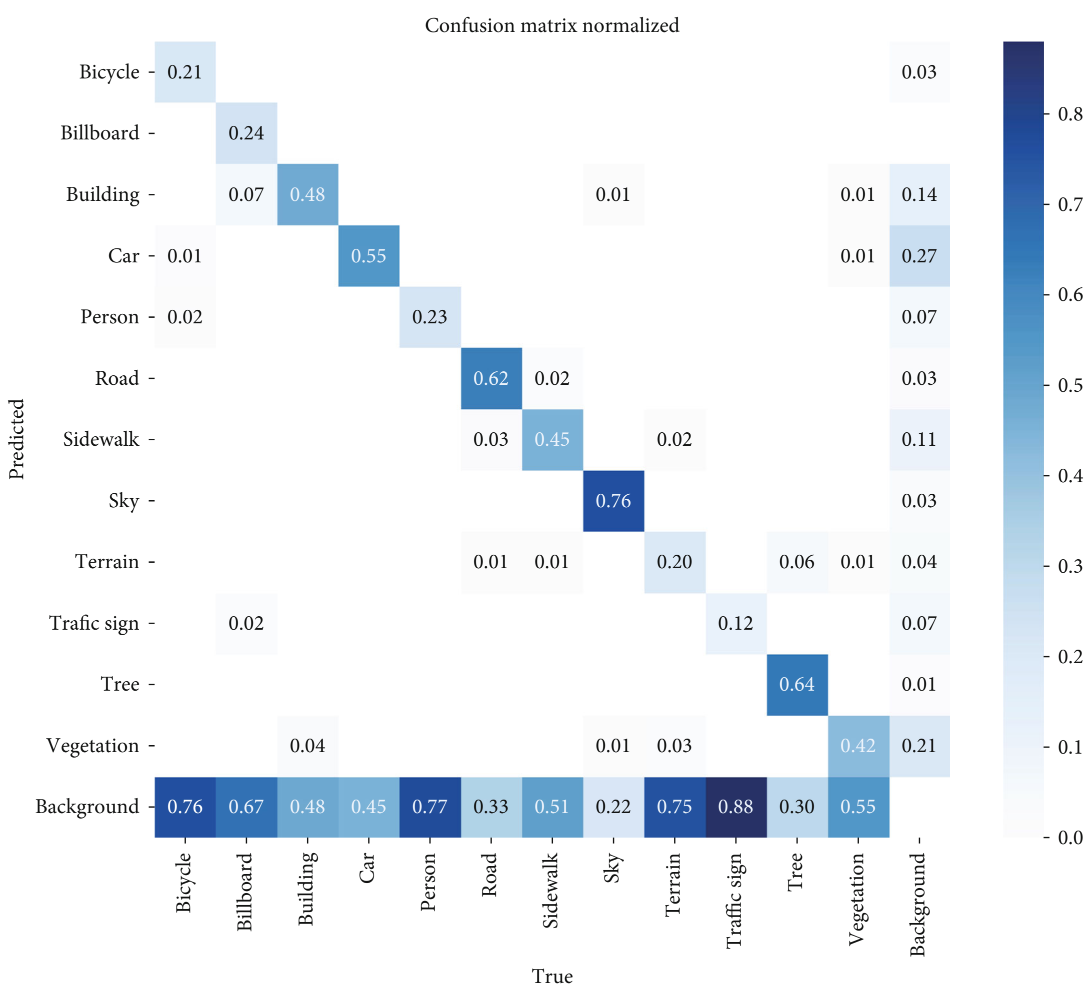
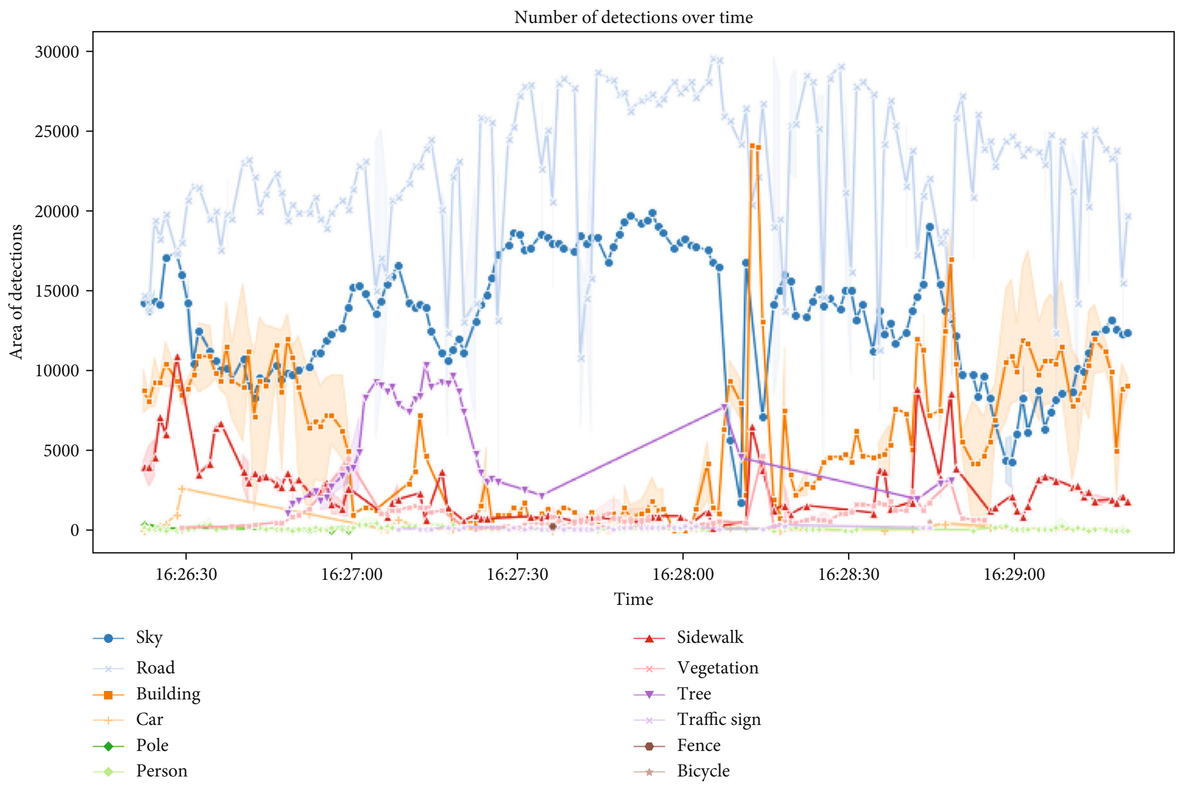


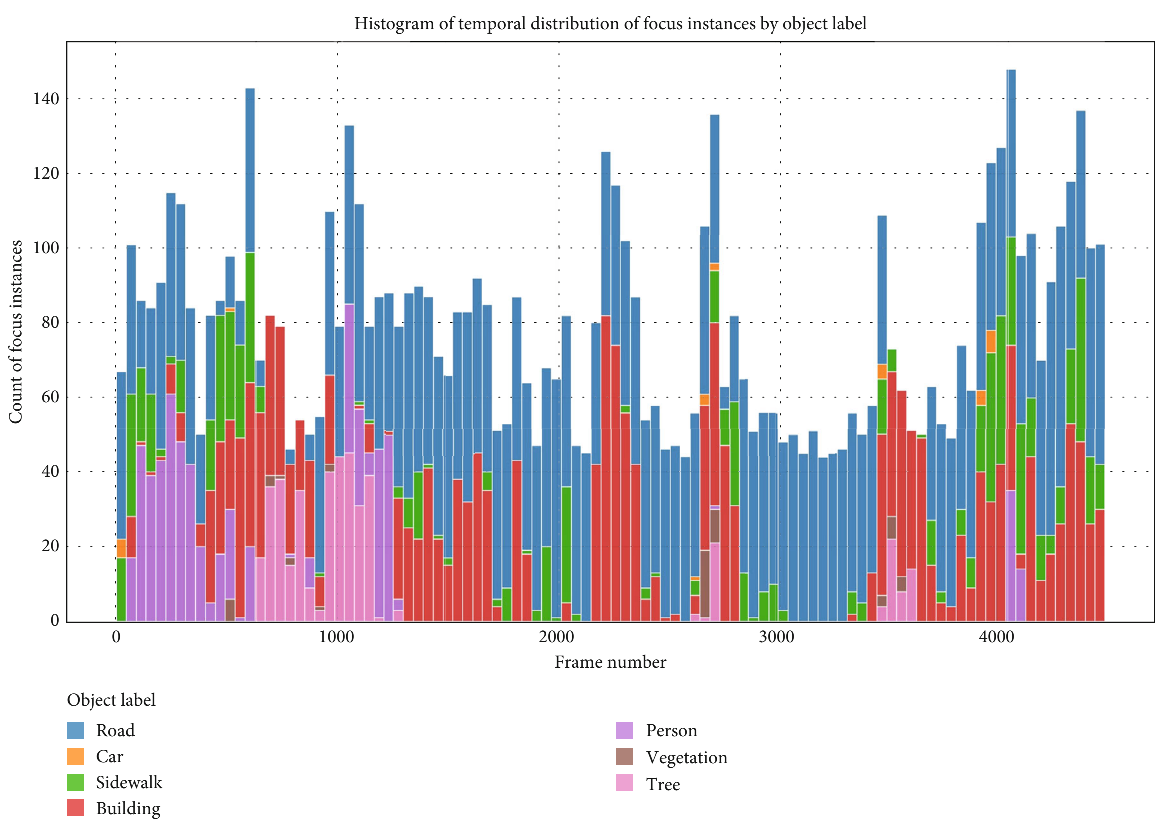
Figure 6: Confusion matrix of the YOLO V8 model.

flows in concert with the built form, underscoring their complementary roles in the spatial experience. For instance, the juxtaposition of decreased building coverage with enhanced road and sidewalk visibility at Location 7 illustrates the nuanced balance of the plaza's design.

By merging the precise metrics of CV with the spatial narrative of the plaza, Figure 8c vividly depicts the pedestrian's journey, marked by varying degrees of visual engagement with both the built and natural environments. This analysis advances our understanding of the spatial

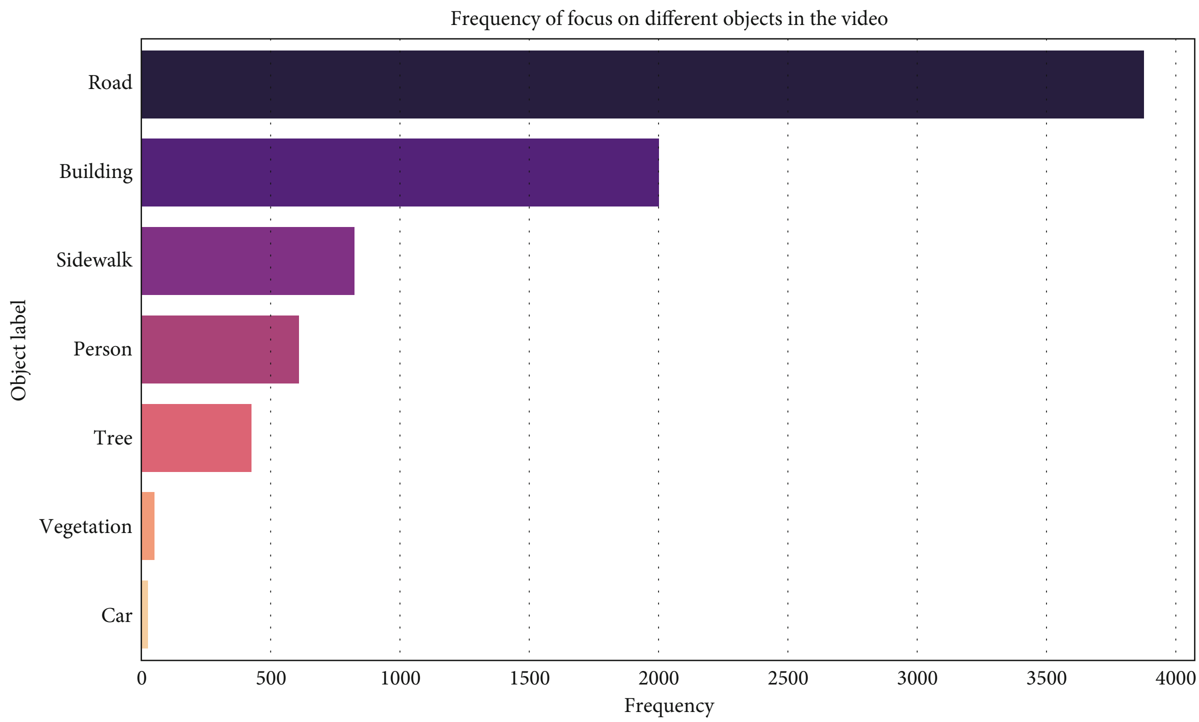


(a)

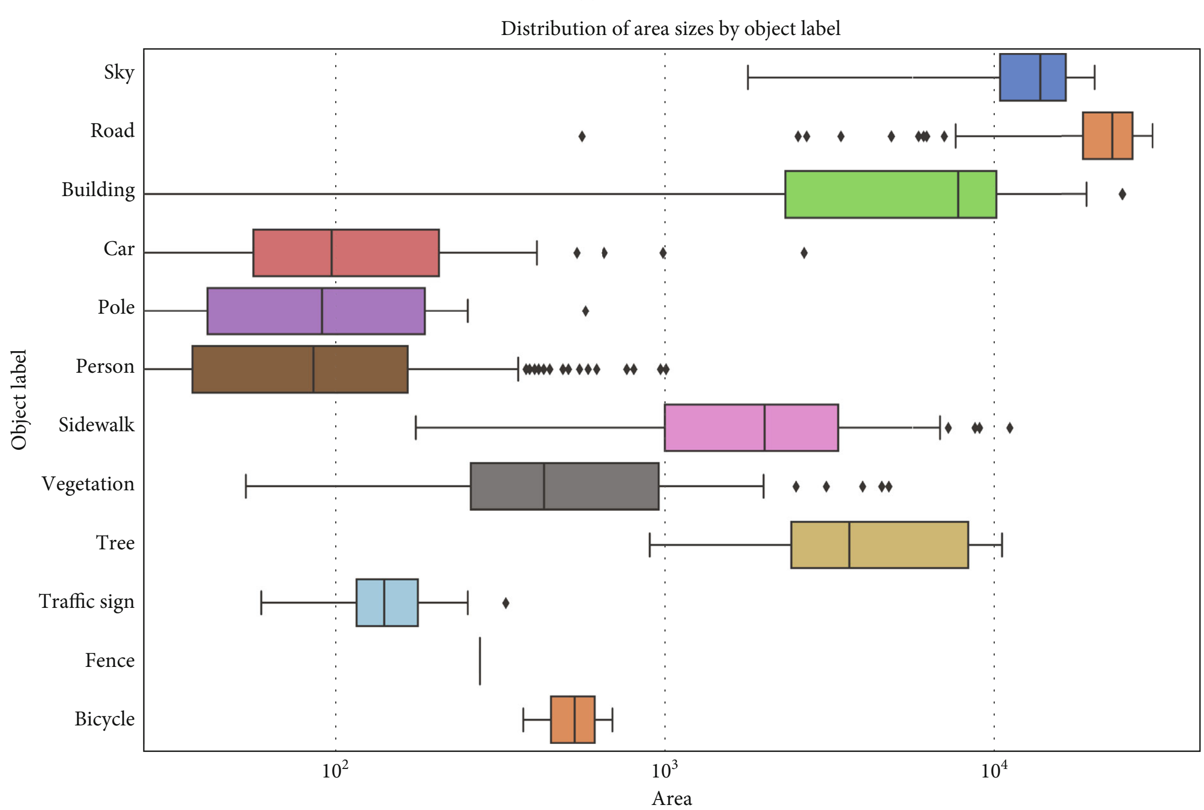


(b)

Figure 7: Continued.

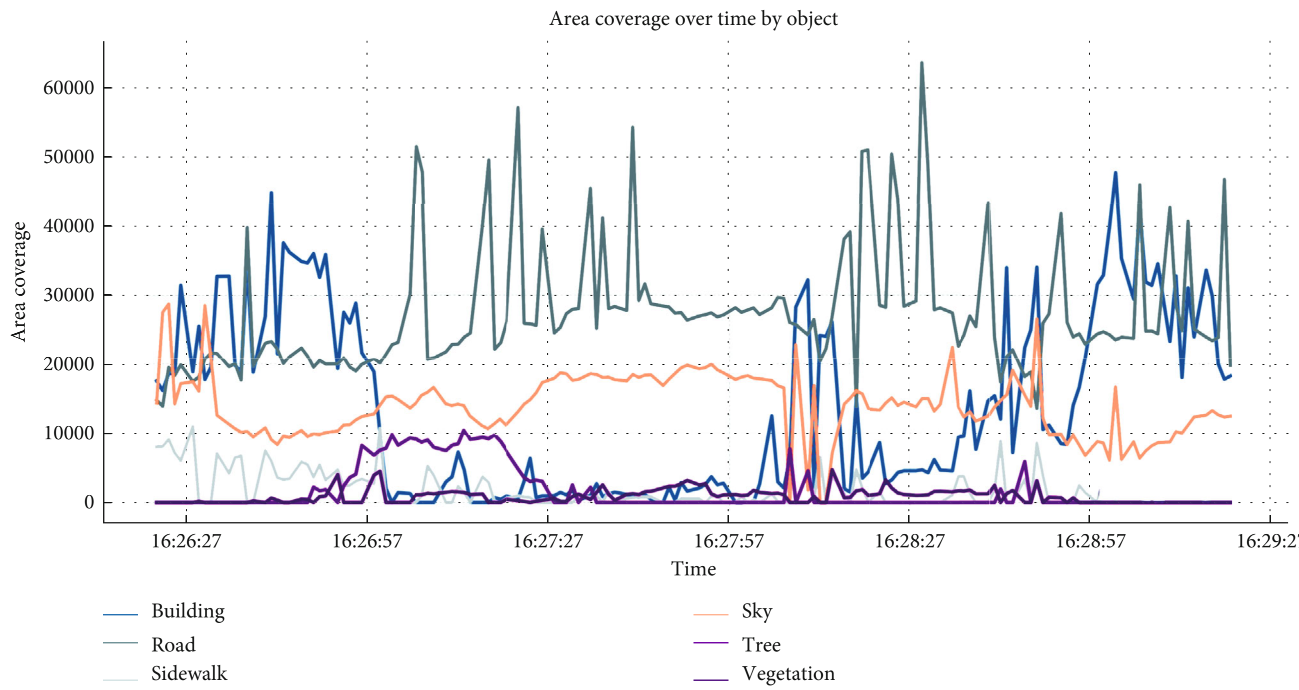


(c)

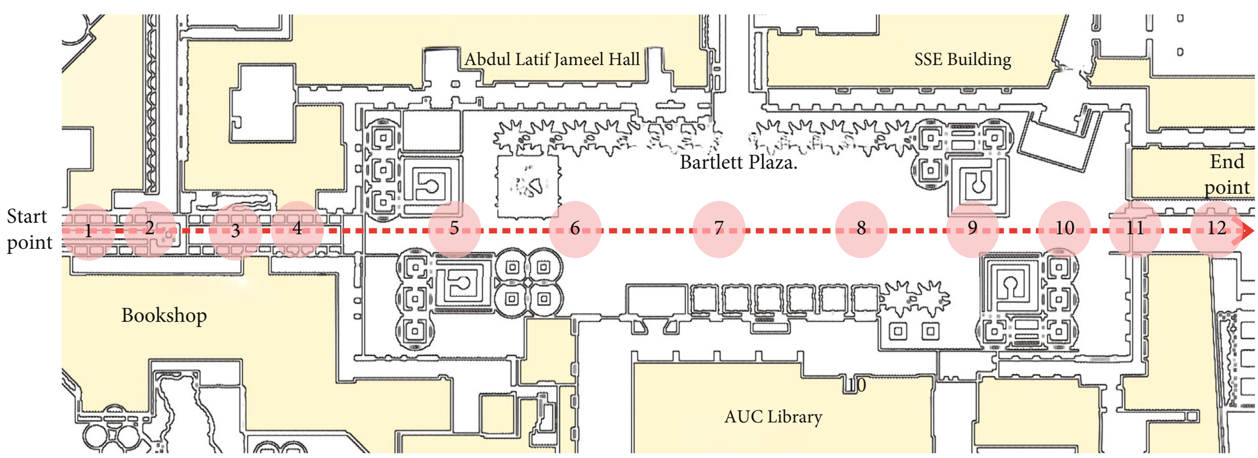


(d)

Figure 7: Derived from computer vision analysis of a pedestrian navigation video, captures (a) fluctuating object visibility, (b) the temporal focus on different elements, (c) the frequency of object focus, and (d) the scale of visual components encountered, charting the dynamics of urban visual engagement.

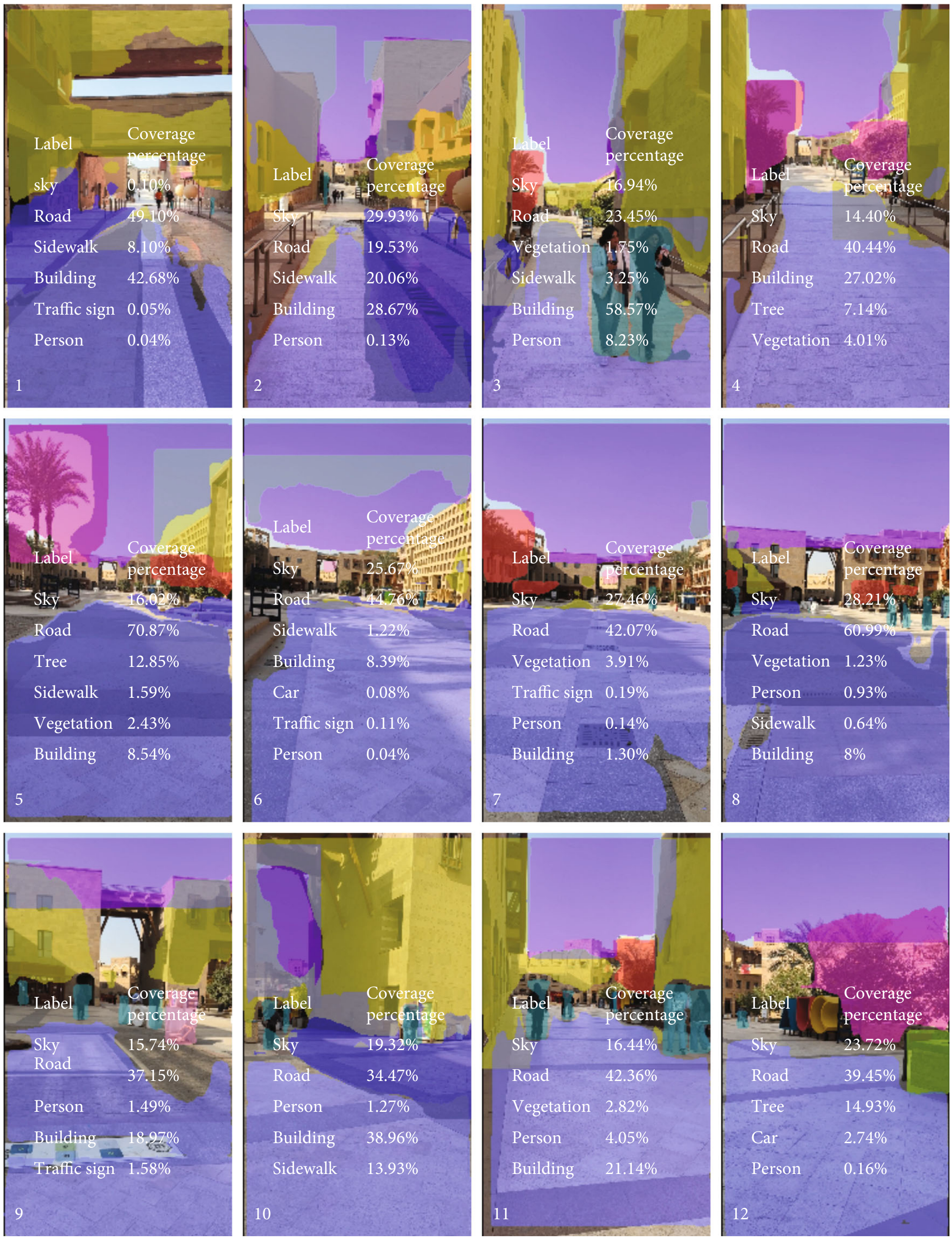


(a)



(b)

Figure 8: Continued.



(c)

FIGURE 8: (a, b) Analysis of the visual experience of a pedestrian walking through the library plaza at the American University in Cairo (AUC). (a) The area coverage over time for different objects or elements encountered along the walking path, such as buildings, roads, sidewalks, sky, trees, and vegetation. (b) The floor plan of the library plaza area, with the walking path marked in red dots, indicating the specific locations (1 to 12) corresponding to the area coverage data in the graph. (c) A series of 12 visual snapshots corresponding to the waypoints on the schematic layout, each accompanied by a coverage percentage legend. These images, processed through a computer vision segmentation algorithm, categorize and color-code elements within the user's field of vision. The legend specifies the percentage coverage of each category (sky, road, person, building, traffic sign, vegetation, sidewalk, and car) within that frame, offering a quantitative insight into the visual experience of the pedestrian at specific locations along the path.

configurations within an urban campus and posits a framework for considering how such configurations might influence the experiential quality of pedestrian movement and space perception. Integrating quantitative data with visual representations provides a comprehensive understanding of the complex interplay between architectural elements and natural features in shaping the pedestrian experience.

6. Results and Discussion

This section presents the study's findings and discusses their implications in the context of campus design and visual engagement. The study's application of CV to analyze 2763 campus images revealed a rich tapestry of visual engagement characterized by different experiences. While 35% of visual engagement focused on green spaces, the remainder concentrated on architectural elements and open skies. This signals the importance of a harmonious balance between nature and architecture in creating stimulating urban spaces. The findings from this pilot study, though based on the visual engagement patterns of a single participant, offer preliminary insights into the potential of the applied CV framework to enhance our understanding of pedestrian experiences in urban spaces. The nuanced analysis of the participant's interaction with the built and natural environment underscores the framework's capacity to capture and quantify complex visual engagement behaviors, suggesting promising avenues for future urban study research and practice.

The experiment assessed the CV model's ability to segment and detect the participant's visual focus areas, distinguishing it from approaches that rely on aerial views, such as those used by Wael et al. [22], or static and panoramic images analyzed in prior studies [8–10]. If applied to diverse groups—such as individuals with special needs, varying age groups, genders, or different urban settings—the model has the potential to identify distinct personas based on preferences for urban elements, landscape components, and architectural features.

This capability highlights the importance of inclusive design strategies that account for diverse visual preferences and user experiences. By categorizing personas according to their visual focus and navigation patterns, the model provides a framework for designing urban spaces that cater to varied user needs and priorities. These insights could guide urban designers in creating environments that balance navigational clarity, aesthetic quality, and overall user satisfaction, ensuring that spaces are both functional and engaging for a wide range of users.

The correlation between visual stimuli and navigation choices extends beyond mere aesthetics, influencing feelings of safety, belonging, and identity within the campus environment. The study suggests that effective urban spaces should foster physical and emotional connectivity between individuals and the urban space.

7. Conclusion

This research offers a methodologically innovative approach, employing CV and semantic segmentation to uncover the intricate relationship between visual stimuli and navigation behavior in understanding the visual engagement of university students with their campus environments.

This pilot study serves as a foundational step in exploring the applicability of CV techniques to urban study research. By focusing on a singular user experience, key areas for methodological refinement were identified, providing valuable insights to guide the expansion of this research into larger, more diverse participant groups.

While the CV approach provides valuable insights, it is essential to recognize its limitations, such as potential data collection and interpretation biases. Using this technology in public spaces raises important ethical questions regarding privacy and consent. Preserving privacy is crucial and requires appropriate measures to prevent unintentional identification of individuals. To address this concern, the model should improve an automated process of blurring pedestrians' faces. Future research must carefully navigate these privacy and ethical considerations, ensuring that technological applications in urban studies adhere to appropriate standards and respect individual privacy rights.

Furthermore, the current model works better for analyzing individual experiences; however, capturing group or crowded scenarios requires further adjustments or different techniques. One potential enhancement could be incorporating higher camera angles, providing a wider view beyond an individual's perspective, and enabling tracking and analysis of movements and interactions within larger groups or crowds rather than relying solely on a person's perspective.

This study lays the groundwork for future research exploring the relationship between urban spaces and pedestrian visual engagement. Ultimately, the aim is to create more inclusive and engaging public spaces. Future work could expand the methodology by incorporating varied data sources and analysis methods to deepen our understanding of the navigation experience in open or indoor spaces.

The initial case study was instrumental in validating the proposed model. However, its application in varied urban contexts will require careful consideration of contextual differences and potential adaptations to maintain its relevance and effectiveness. The model's future applications should also consider the impact of seasonal changes and cultural contexts on visual preferences and navigation patterns, offering richer insights into the complex dynamics of urban spaces.

In conclusion, this study employs CV to significantly enhance our understanding of urban environments. By providing a nuanced analysis using qualitative data, critically evaluating methodologies, and considering ethical implications, the research opens new pathways for creating truly engaging, inclusive, and responsive urban spaces. The model's readiness for broader application holds immense promise to enrich urban design and planning discourse, advocating for a data-driven and deeply attuned approach to the human experience.

Data Availability Statement

No data is available; however, if requested by email, the author will provide it.

Disclosure

An earlier preprint version of this paper, which was incomplete, can be accessed at https://www.researchsquare.com/article/rs-4339232/v1.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This paper is part of a grant funded by the American University in Cairo (AUC), SSE-ARCH-N.M-FY22-RG-2022-Sep-19-23-53-26.

Acknowledgments

This paper is part of a grant funded by the American University in Cairo, SSE-ARCH-N.M-FY22-RG-2022-Sep-19-23-53-26.

References

[1] C. Tudorie, M. Vallés-Planells, E. Gielen, R. Arroyo, and F. Galiana, "Towards a Greener University: Perceptions of Landscape Services in Campus Open Space," *Sustainability* 12, no. 15 (2020): 6047, https://doi.org/10.3390/su12156047.

[2] O. Kwon, I. Hong, J. Yang, D. Wohn, W. Jung, and M. Cha, "Urban Green Space and Happiness in Developed Countries," *Science* 10, no. 1 (2021): 28, https://doi.org/10.1140/epjds/s13688-021-00278-7.

[3] F. Mengjia, H. Hussein, and A. Jamaludin, "Perceiving the Transition of Urban Campus Open Spaces Under the Influence of COVID-19," *Planning Malaysia* 20 (2022): https://doi.org/10.21837/pm.v20i24.1183.

[4] N. Hasan, I. Hijazi, D. Enab, et al., "Influence of Campus Outdoor Spaces on Students Behavior: Enhancing Social Interaction and Learning at An-Najah University," *An-Najah University Journal for Research-a (Natural Sciences)* 39, no. 2 (2024).

[5] A. M. Elshater, "Urban Design Redux: Redefining a Professional Practice of Specialization," *Ain Shams Engineering Journal* 6, no. 1 (2015): 25–39, https://doi.org/10.1016/j.asej.2014.08.004.

[6] Z. Peng, R. Zhang, Y. Dong, and Z. Liang, "A Study on the Relationship Between Campus Environment and College Students' Emotional Perception: A Case Study of Yuelu Mountain National University Science and Technology City," *Buildings* 14, no. 9 (2024): 2849, https://doi.org/10.3390/buildings14092849.

[7] K. Aghabozorgi, A. van der Jagt, S. Bell, and H. Smith, "How University Blue and Green Space affect Students' Mental Health: A Scoping Review," *Urban Forestry & Urban Greening* 97 (2024): 128394, https://doi.org/10.1016/j.ufug.2024.128394.

[8] J. Wei, W. Yue, M. Li, and J. Gao, "Mapping Human Perception of Urban Landscape From Street-View Images: A Deep-Learning Approach," *International Journal of Applied Earth Observation and Geoinformation* 112 (2022): 102886, https://doi.org/10.1016/j.jag.2022.102886.

[9] 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, no. 8 (2019): 1439–1457.

[10] Y. Meng, Q. Li, X. Ji, et al., "Research on Campus Space Features and Visual Quality Based on Street View Images: A Case Study on the Chongshan Campus of Liaoning University," *Buildings* 13, no. 5 (2023): 1332, https://doi.org/10.3390/buildings13051332.

[11] R. Lucas and O. Romice, "Representing Sensory Experience in Urban Design," *Design Principles and Practices* 2, no. 4 (2008): 83–94, https://doi.org/10.18848/1833-1874/CGP/v02i04/37579.

[12] F. Zhang, A. S. Miranda, F. Duarte, et al., "Urban Visual Intelligence: Studying Cities With AI and Street-Level Imagery" 2023, https://arxiv.org/abs/2301.00580.

[13] K. Lynch, *The Image of the City* (MIT Press, 1960).

[14] W. H. Whyte, *The Social Life of Small Urban Spaces* (Project for Public Spaces, 1980).

[15] A. Vanky and R. Le, "Urban-Semantic Computer Vision: A Framework for Contextual Understanding of People in Urban Spaces," *AI & Society* 38, no. 3 (2023): 1193–1207, https://doi.org/10.1007/s00146-022-01625-6.

[16] W. Qiu, W. Li, X. Liu, and X. Huang, "Subjectively Measured Streetscape Qualities for Shanghai With Large-Scale Application of Computer Vision and Machine Learning," in *Proceedings of the 2021 DigitalFUTURES*, eds. P. F. Yuan, H. Chai, C. Yan, and N. Leach (Springer, 2022), 242–251, https://doi.org/10.1007/978-981-16-5983-6_23.

[17] G. Csurka, R. Volpi, and B. Chidlovskii, "Semantic Image Segmentation: Two Decades of Research," *Foundations and Trends in Computer Graphics and Vision* 14, no. 1-2 (2022): 1–162, https://doi.org/10.1561/0600000095.

[18] G. Ong, Y. Zhang, Z. Jin, C. Seah, and T. Chua, "Observations of Urban Activities With Computer Vision," in *The 13th Conference of the International Forum on Urbanism: Urbanism in the Mobile Internet Era* (Nanjing University, 2020).

[19] J. Liu, D. Ettema, and M. Helbich, "Street View Environments Are Associated With the Walking Duration of Pedestrians: The Case of Amsterdam, the Netherlands," *Landscape and Urban Planning* 235 (2023): 104752, https://doi.org/10.1016/j.landurbplan.2023.104752.

[20] L. Yan, Y. Chen, L. Zheng, Y. Zhang, and C. Zhu, "Research on the Quantification of Historical Street Space Based on Image Semantic Segmentation," in *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2021)* 12168, SPIE, 2022), https://doi.org/10.1117/12.2630998.

[21] F. Duarte and C. Ratti, "What Urban Cameras Reveal About the City: The Work of the Senseable City Lab," in *Open Cities|Open Data: Collaborative Cities in the Information Era*, eds. S. Hawken, H. Han, and C. Pettit (Springer, 2021), 491–502, https://doi.org/10.1007/978-981-15-8983-6_27.

[22] S. Wael, A. Elshater, and S. Afifi, "Mapping User Experiences Around Transit Stops Using Computer Vision Technology: Action Priorities From Cairo," *Sustainability* 14, no. 17 (2022): 11008, https://doi.org/10.3390/su141711008.

[23] J. Lee, D. Kim, and J. Park, "A Machine Learning and Computer Vision Study of the Environmental Characteristics of Streetscapes That Affect Pedestrian Satisfaction," *Sustainability* 14, no. 9 (2022): 5730, https://doi.org/10.3390/su14095730.

[24] J. Ernawati, M. S. Adhitama, and B. S. Sudarmo, "Urban Design Qualities Related Walkability in a Commercial Neighbourhood," *Environment-Behaviour Proceedings Journal* 1, no. 4 (2016): 242–250, https://doi.org/10.21834/e-bpj.v1i4.385.

[25] L. Chen, Y. Lu, Q. Sheng, Y. Ye, R. Wang, and Y. Liu, "Estimating Pedestrian Volume Using Street View Images: A Large-

Scale Validation Test," *Computers, Environment and Urban Systems* 81 (2020): 101481, https://doi.org/10.1016/j.compenvurbsys.2020.101481.

[26] F. Biljecki and K. Ito, "Street View Imagery in Urban Analytics and GIS: A Review," *Landscape and Urban Planning* 215 (2021): 104217, https://doi.org/10.1016/j.landurbplan.2021.104217.

[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of the 3rd International Conference on Learning Representations* (ICLR, 2015).

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), 779–788.

[29] N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav, and G. Sudham, "Object Detection and Tracking Using YOLO," in *Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (IEEE, 2021).

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks," in *Advances in Neural Information Processing Systems*, eds. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (28, Curran Associates, Inc., 2015).

[31] M. E. Sahin, H. Ulutas, E. Yuce, and M. F. Erkoc, "Detection and Classification of COVID-19 by Using Faster R-CNN and mask R-CNN on CT Images," *Neural Computing and Applications* 35, no. 18 (2023): 13597–13611, https://doi.org/10.1007/s00521-023-08450-y.

[32] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 7 (2022): 3523–3542, https://doi.org/10.1109/TPAMI.2021.3059968.