Visual Engagement: Quantifying Campus Experiences in Urban Open Spaces Using a Computer Vision Model

Nabil Mohareb

nabil.mohareb@aucegypt.edu

American University in Cairo

Abdelaziz Ashraf

October University of Modern Sciences and Arts

Research Article

Keywords: Computer Vision, instance Segmentation, Convolutional Neural Networks, Spatial Analysis, Navigation Behavior, University open spaces

Posted Date: May 10th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4339232/v1

Additional Declarations: No competing interests reported.

Abstract

Introduction

Addressing the gap in quantitative analysis of spatial experiences within academic environments, this study introduces a groundbreaking framework designed to measure and quantify the visual experiences of individuals in academic campus settings. Focused on analyzing the visual composition of the built environment—including aspects such as visible sky, greenery, and spatial enclosure—our framework aims to provide a quantitative reflection of the subjective spatial experiences of campus users.

Methods

The methodology involves using mobile phones with digital cameras and GPS sensors to capture first-person visual data and track movements as they freely traverse campus open spaces. Computer vision techniques, including Instance segmentation and convolutional neural networks, will categorize architectural and natural elements within each frame image extracted from a recorded video, quantify proportional compositions and analyze relative amounts of greenery, open sky, walkways, buildings, and other built structures that participants visually experienced. The framework is translated into a Python model capable of producing quantitative outcomes.

The analysis will be further enriched by integrating Geographic Information Systems (GIS) for spatial analysis to identify navigation and visual engagement patterns. This comprehensive methodology quantifies the visual attributes of spaces and interprets their impact on the behavior and experiences of campus users.

Results and conclusions

The study outcomes reveal relationships between student’s navigation choices, visual experiences, and scene types. The results aim to guide urban designers in understanding university students’ open space needs based on their natural movement and viewing preferences and complement other qualitative approaches.

Introduction

University campus open spaces offer visually rich environments, yet little is known about how specific details and elements influence pedestrians' visual focus, spatial cognition, and movement patterns. This research aims to develop an analytical framework to systematically investigate the relationship between the visual information encountered by pedestrians navigating open spaces and their spatial awareness, attention, and navigation choices. The study examines how the built environment, including components such as buildings, vegetation, pathways, and architectural features, shapes users' visual engagement by analyzing the composition and characteristics of these visual elements within the open spaces.

The research integrates innovative computer vision techniques, such as instance segmentation and convolutional neural networks, to quantify visual attention from pedestrians' perspectives. This visual analysis offers insights into the experiential aspects of perceiving and interacting with the built environment. The proposed analytical framework and its Python code implementation will be tested in a pilot study focused on a single student's navigation experience on a university campus, assessing the feasibility of the methodology and refining it for large-scale applications.

The findings contribute key principles for designing engaging open spaces that enhance visual interest, spatial awareness, and sense of place for university communities. The analytical framework is translated into a Python code to facilitate systematic application in other cases. By unraveling the impact of the visually rich open space realm on pedestrians through this analytical framework, the research provides a comprehensive understanding of sensitively sculpted environments that elevate the journey on foot and foster cohesive, user-centered university precincts.

The study progresses from a Literature Review on the shift towards human-centric urban design via CV/AI, to Technological Advancements in Urban Studies. It details CV's role in Evaluating Urban Environments and Urban Design Applications, outlining the Methods and Models. A pilot Data Analysis and Results demonstrate CV's impact, with Conclusions tying theory to practice in advancing urban design.

Literature Review

This literature review examines how computer vision (CV) and instance image analysis have impacted urban design, shifting the focus from traditional aesthetic approaches to human-centric urban spaces. 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.

Urban design theories have evolved from an aesthetic focus to a deeper understanding of human interactions within urban spaces [1]. 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 [2, 3].

Traditional data collection methods in urban studies, such as images, videos, and direct observation, faced limitations in labor intensity and scalability. The advent of advanced sensing technologies and geotagged imagery enabled more efficient and extensive urban analysis [1]. 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 [4], 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 [1]. The application of computer vision technologies in urban management represents a progression in employing image-based AI for data collection [4]. However, these objective approaches potentially overlook the complexities of human perceptions [5].

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 [2]. However, traditional methods for collecting perception data often lack consistency, and reliability, are time-consuming, expensive, and challenging to interpret [5].

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 [1]. 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 [4]. Semantic image segmentation (SiS), as described by Csurka et al. [6], is crucial in computer vision, where each pixel in an image is assigned to a specific semantic class to understand different image parts.

Implementing activity surveys with computer vision 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 [7]. 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 Computer Vision Methods

The "Urban Visual Intelligence" framework, elucidated by Zhang et al. [1], 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. [8] 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 non-linear 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. [8] utilized semantically segmented street environmental features with a fully convolutional neural network (CNN), specifically the Xception-71 CNN, pre-trained 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. [9] 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 & Ratti [10] emphasize using specialized urban cameras designed to collect visual data about cities. This shift from user-generated, geotagged photographs used in previous studies to assess urban attractiveness or aesthetic appeal represents a move towards more objective urban data collection [10]. Additionally, Lee et al. [11] highlight using computer vision technologies, such as semantic segmentation and edge detection, to evaluate urban design 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 design elements and their impact on urban life.

2.3 Applications of Computer Vision Segmentation in Urban Design 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 Fig. 1. Walking shapes urban experiences and community dynamics [11], 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 [12], measurable through segmentation.

Segmentation enables virtual assessments of pedestrian volumes, a key walkability indicator [13]. Street View Imagery (SVI) and segmentation algorithms like the Visual Walkability Index [14] evaluate aspects like crowdedness and obstacles, revealing walkability variations across locations. Urban design qualities

like area ratio, enclosure, openness, and complexity are extracted through semantic segmentation [11]. 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 [14], 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 [14]. 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 [13], traffic indicators, and cycling pattern exploration based on greenery and architecture [14]. 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 [4], necessitating ongoing policy discussions for responsible and ethical implementation in urban studies.

2.4 Methods and models of segmentation:

Computer vision is a crucial component of artificial intelligence. It enables the extraction of valuable insights from visual data, which is indispensable for urban analysis. This section discusses the design of our model, which leverages established computer vision models and algorithms. A pilot approach was adopted to rigorously evaluate the novel computer vision framework developed for this study, centering on an in-depth analysis of one student's navigational experience through the campus. This methodological choice was driven by the intent to closely monitor and adapt the analytical process in real-time, ensuring the comprehensive testing of our software's capabilities and the framework's analytical precision.

Advancements in hardware and research into convolutional neural networks (CNNs) have led to sophisticated deep-learning models for image analysis. Notable examples include VGGNet [15], the YOLO (You Only Look Once) object detection framework [16, 17], and the Fast R-CNN [18, 19] for object detection and classification.

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.

Our model represents an advancement in computer vision 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.

Our 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," "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 algorithm determines the object closest to the video frame's center by calculating the frame's center coordinates (center_x, center_y) and iterating over the detected object bounding boxes. Each bounding box computes the Euclidean distance between the box's center and the frame's center, verifying if the frame's center is within the bounding box for accuracy. If an object is found closer than the closest distance, the code updates the closest distance and stores the object's label and frame number. This process enables the detailed extraction of object coverage percentages and dynamic tracking of prominent objects over time, enriching the analysis of urban open spaces.

Integrating GPS location data from data loggers or mobile applications enhances each video frame with movement location. This comprehensive extraction of visual elements, including classified objects, pedestrian and vehicle counts, and focus of attention, is compiled into a CSV file and integrated into a GIS platform for spatial analysis within urban studies. The model's utility lies in analyzing visual attention patterns within open urban spaces, employing a corpus of systematically documented videos to identify areas that capture visual attention during navigation.

The proposed framework quantifies observable urban elements, such as the proportion of sky, landscape features, architectural structures, and human presence, offering insights into the visual structure and built environment quality. It governs a code for assessing visual engagement and focus within urban spaces by capturing focal points and gaze duration. Integrating computer vision analysis with geospatial data enhances the understanding of visual attention dynamics; see Fig. 2 for the full Python model process. The production of quantified outcomes in CSV files, geo-referencing these data, enables the correlation of visual attention data with other spatial datasets. This multidimensional analysis provides a holistic understanding of sensory experiences in urban environments, advancing urban studies by facilitating informed decision-making based on a multifaceted understanding of urban open spaces through both temporal dynamics and geographic context.

This model's design and functionality represent a significant computer vision application to urban space analysis. It affords researchers and urban planners a powerful tool for understanding the intricate

dynamics of urban open spaces, facilitating informed decision-making and contributing to advancing urban studies.

2.4.1 Model Methodology

We use the following dataset for this research to detect and segment in open urban spaces. The dataset comprises 2,763 images across 16 distinct classes, with an average image resolution of 2048 x 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 Figs. 3 and 4.

The dataset underwent preprocessing to address class imbalances and resize requirements. Automatic orientation adjustments were applied, and images were resized to 320x224 pixels. Data augmentation techniques were employed to achieve class balance, including horizontal and vertical flips and slight rotations within ± 5 degrees. 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 5,000 images across five distinct classes, divided into training (80%), validation (10%), and testing (10%) sets. The training involved 100 epochs with images resized to 640x640 pixels, a batch size of 16, and using ‘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 Fig. 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 Figs. 5 and 6.

Data Analysis of the case study

The research focuses on the main walking spine at the American University in Cairo's (AUC) new campus in New Cairo, Egypt, as a case study. 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 was conducted to evaluate the model's functionality and identify potential technical issues, as the primary focus is on developing and testing the Python model itself. While using a single user limits the generalizability of findings due to individual characteristics and biases, this concern is less critical given the emphasis on evaluating the analytical model rather than generalizing it to a broader population, which is part of our future intention. The study investigates the prevalence and visual prominence of various environmental elements encountered along this principal walking route, demonstrating the potential of the Python model's approach and refining the methodologies before wider application.

Visual data is processed to elaborate on the classification and segmentation of images from video to quantify environmental elements; see Fig. 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 7 (A) 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 7 (B) 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 Fig. 7 (C) 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. Panel D’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 design for optimal visual accessibility.

Together, these visual data sets corroborate the intricate nature of pedestrian visual experiences, highlighting the necessity for urban designs 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 Fig. 8, which captures the dynamic interplay between architectural presence and natural elements along a delineated path. The upper graph (Figure (A)) quantitatively tracks the visibility of various elements—buildings, roads (pedestrian pathway), sidewalks, sky, trees, and vegetation—as the passerby perceives. Substantial fluctuations 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 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 computer vision with the spatial narrative of the plaza, Fig. 8 (C) 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 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.

Results

The study's application of computer vision to analyze 2,763 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 computer vision 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 design research and practice.

The experiment evaluated the computer vision model's ability to segment and detect the participant's visual focus areas. If applied to diverse groups, the model could potentially identify distinct personas based on preferences for greenery versus architectural features. This finding would highlight the need for inclusive design strategies that accommodate varied visual attractions and user experiences. By categorizing personas according to their visual preferences and navigation patterns, the model could provide a framework for designing urban spaces that cater to different user needs and priorities. Urban designers could then create environments balancing efficiency, visual appeal, and overall satisfaction by considering personas' preferences for navigational clarity, aesthetics, or a combination thereof.

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 design should foster physical and emotional connectivity between individuals and the urban space.

Conclusion

This research offers a methodologically innovative approach, employing computer vision 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 computer vision techniques to urban design research. By focusing on a singular user experience, we have identified key areas for methodological refinement and gained valuable insights that will guide the expansion of this research into larger, more diverse participant groups.

While the computer vision approach provides valuable insights, it's 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 design 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 computer vision 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.

Declarations

Funding Declaration

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

Competing Interest declaration

no Competing Interests

Author Contributions Statement

Nabil Mohareb has contributed the following: The research main idea; 70 % of the text; 60 % of the images; and follow-up with the Python code.

Abdelaziz Ashraf has contributed the following: The Python code; 30% of the text; 40% of the images

References

1. Zhang F, Miranda AS, Duarte F, Vale L, Hack G, Chen M, et al. Urban visual intelligence: Studying cities with AI and street-level imagery. arXiv [preprint]. 2023 Jan 2 [cited 2023 Jun 28]. Available from: https://doi.org/10.48550/arXiv.2301.00580
2. Lynch K. The image of the city. Cambridge: MIT Press; 1960.
3. Whyte WH. The social life of small urban spaces. New York: Project for Public Spaces; 1980.
4. Vanky A, Le R. Urban-semantic computer vision: A framework for contextual understanding of people in urban spaces. AI & Society. 2023; 38(3):1193–207.
5. Qiu W, Li W, Liu X, Huang X. Subjectively measured streetscape qualities for Shanghai with large-scale application of computer vision and machine learning. In: Yuan PF, Chai H, Yan C, Leach N, editors. Proceedings of the 2021 DigitalFUTURES. Singapore: Springer; 2022. p. 242–51.
6. Csurka G, Volpi R, Chidlovskii B. Semantic image segmentation: Two decades of research. Foundations and Trends® in Computer Graphics and Vision. 2022; 14(1–2):1–162.
7. Ong G, Zhang Y, Jin Z, Seah C, Chua T. Observations of urban activities with computer vision. arXiv [preprint]. 2020 Oct 10 [cited 2023 Jun 28]. Available from: https://doi.org/10.48550/arXiv.2010.05080
8. Liu J, Ettema D, Helbich M. Street view environments are associated with the walking duration of pedestrians: The case of Amsterdam, the Netherlands. Landscape and Urban Planning. 2023; 235:104752.
9. Yan L, Chen Y, Zheng L, Zhang Y, Zhu C. Research on the quantification of historical street space based on image semantic segmentation. In: Proceedings of SPIE 12215, MIPPR 2022: Pattern Recognition and Computer Vision. 2022 Mar 18. 122151F.
10. Duarte F, Ratti C. What urban cameras reveal about the city: The work of the Senseable City Lab. In: Hawken S, Han H, Pettit C, editors. Open Cities | Open Data: Collaborative Cities in the Information Era. Singapore: Springer; 2021. p. 491–502.

11. Lee J, Kim D, Park J. A machine learning and computer vision study of the environmental characteristics of streetscapes that affect pedestrian satisfaction. Sustainability. 2022; 14(9):5730.
12. Ernawati J, Adhitama MS, Sudarmo BS. Urban design qualities related walkability in a commercial neighbourhood. Environment-Behaviour Proceedings Journal. 2016; 1(4):242-50.
13. Chen L, Lu Y, Sheng Q, Ye Y, Wang R, Liu Y. Estimating pedestrian volume using Street View images: A large-scale validation test. Computers, Environment and Urban Systems. 2020; 81:101481.
14. Biljecki F, Ito K. Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning. 2021; 215:104217.
15. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, ICLR 2015; 2015 May 7-9; San Diego, CA, USA.
16. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas, NV, USA. IEEE; 2016. p. 779–88.
17. Krishna NM, Reddy RY, Reddy MSC, Madhav KP, Sudham G. Object detection and tracking using YOLO. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA); 2021 Jul 2-4; Coimbatore, India. IEEE; 2021. p. 1310–5.
18. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R, editors. Advances in Neural Information Processing Systems. Vol. 28. Red Hook: Curran Associates, Inc.; 2015.
19. Sahin ME, Ulutas H, Yuce E, Erkoc MF. Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images. Neural Computing and Applications. 2023;35(18):13597–611.
20. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022; 44(7):3523–42.

Figures

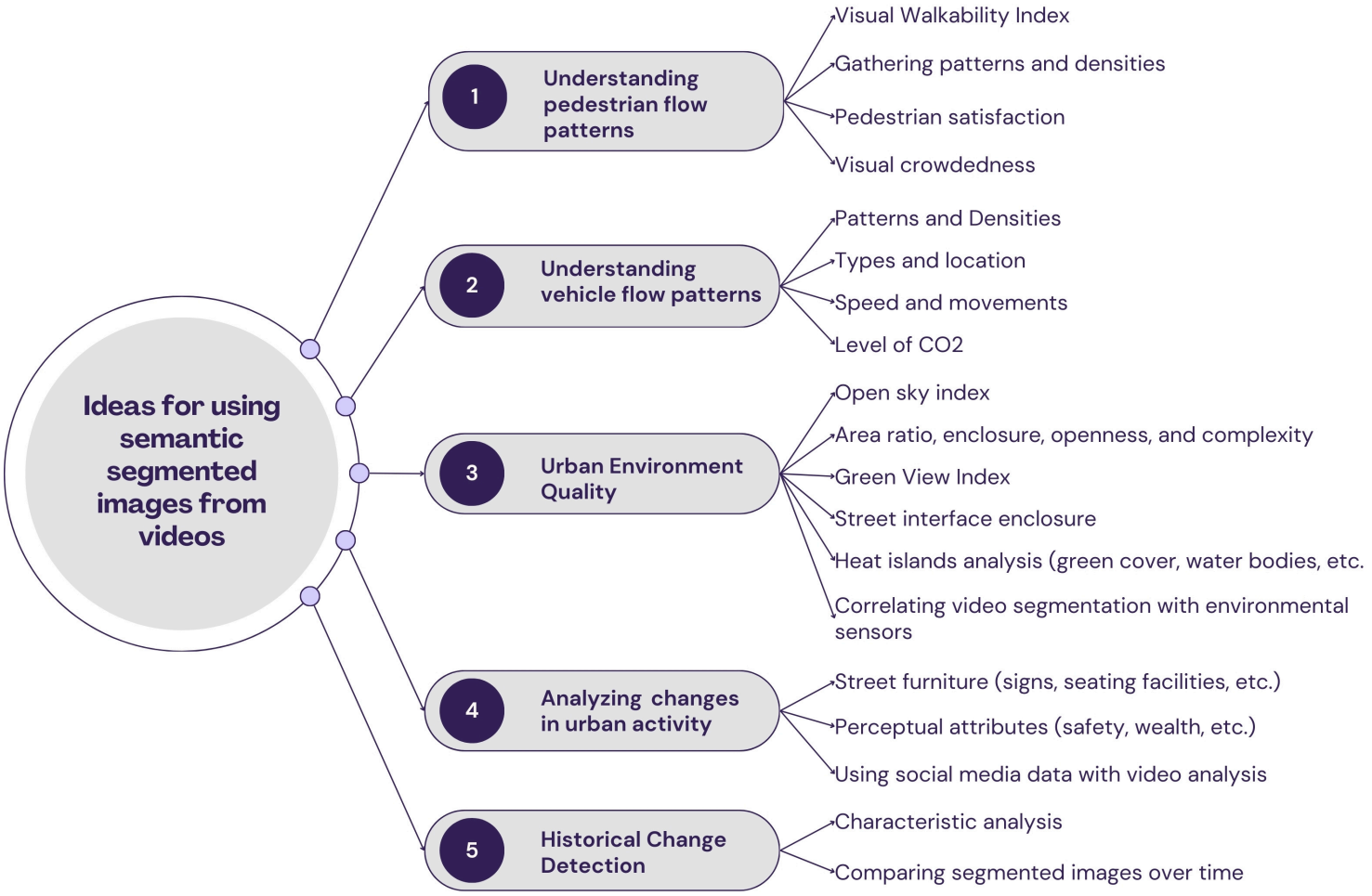

Figure 1

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

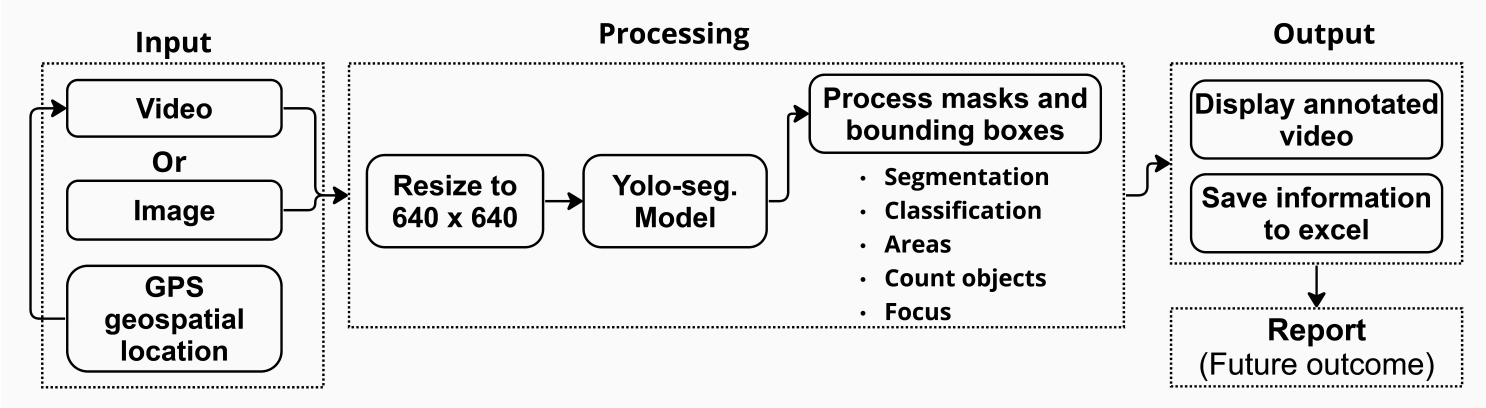

Figure 2

illustrates the framework of the proposed model, translated to Python code, delineating the stages of input, processing, and output.

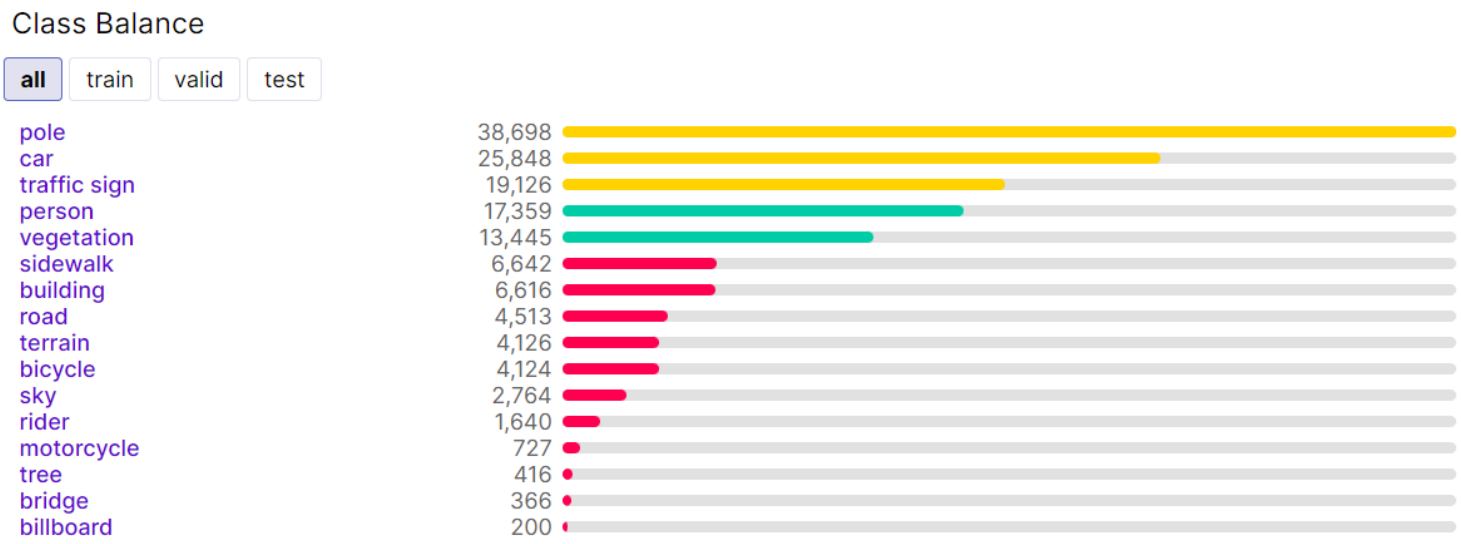

Figure 3

illustrates the class distributions and classes.

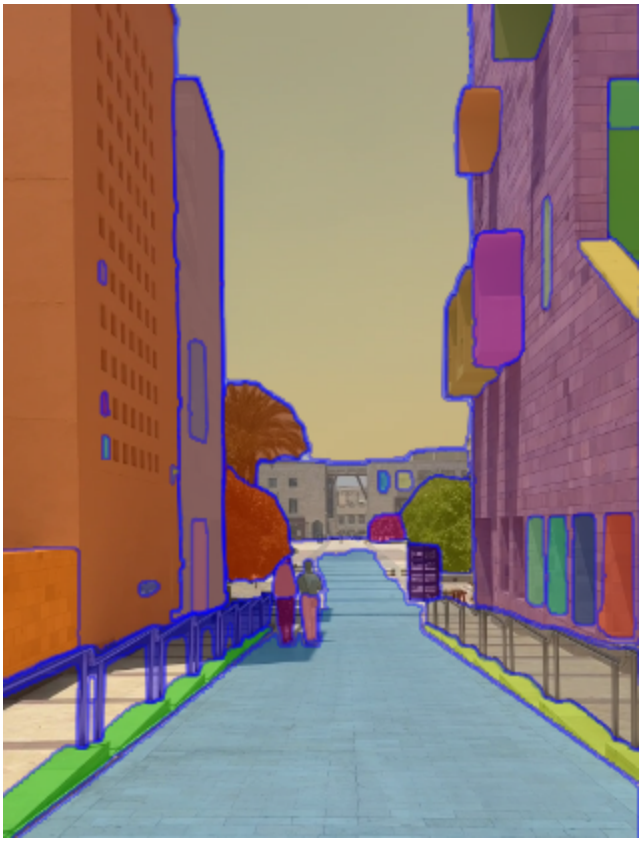

Figure 4

shows the segmentation masks in an image from the dataset.

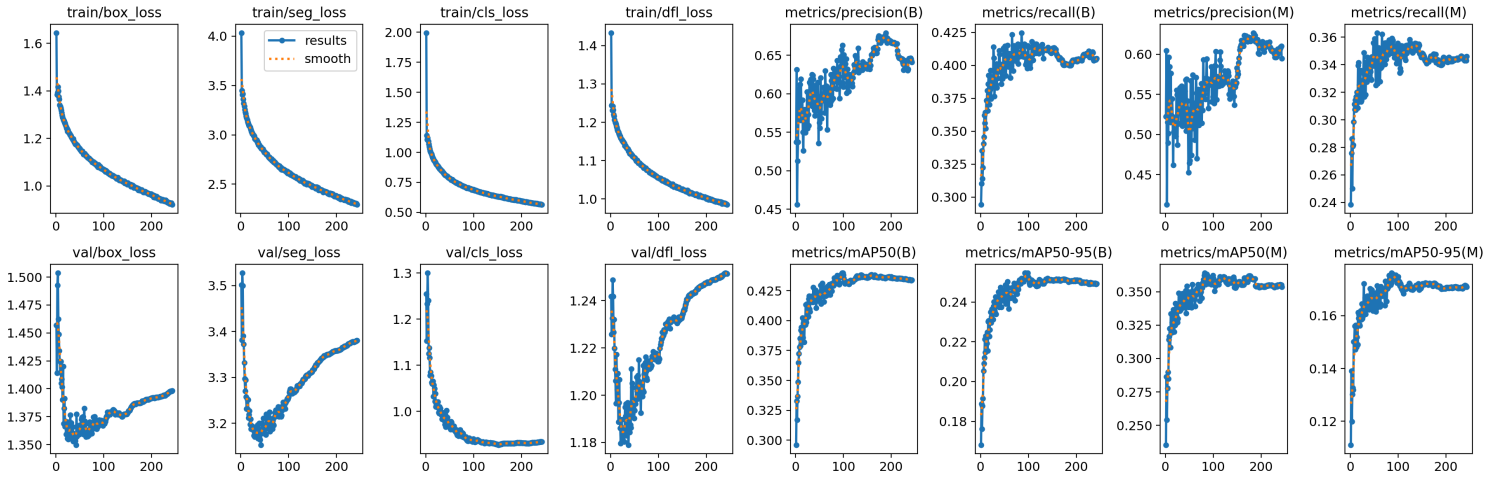

Figure 5

illustrates the model's losses in different metrics for each train and validation dataset, such as precision, recall, and mAP50.

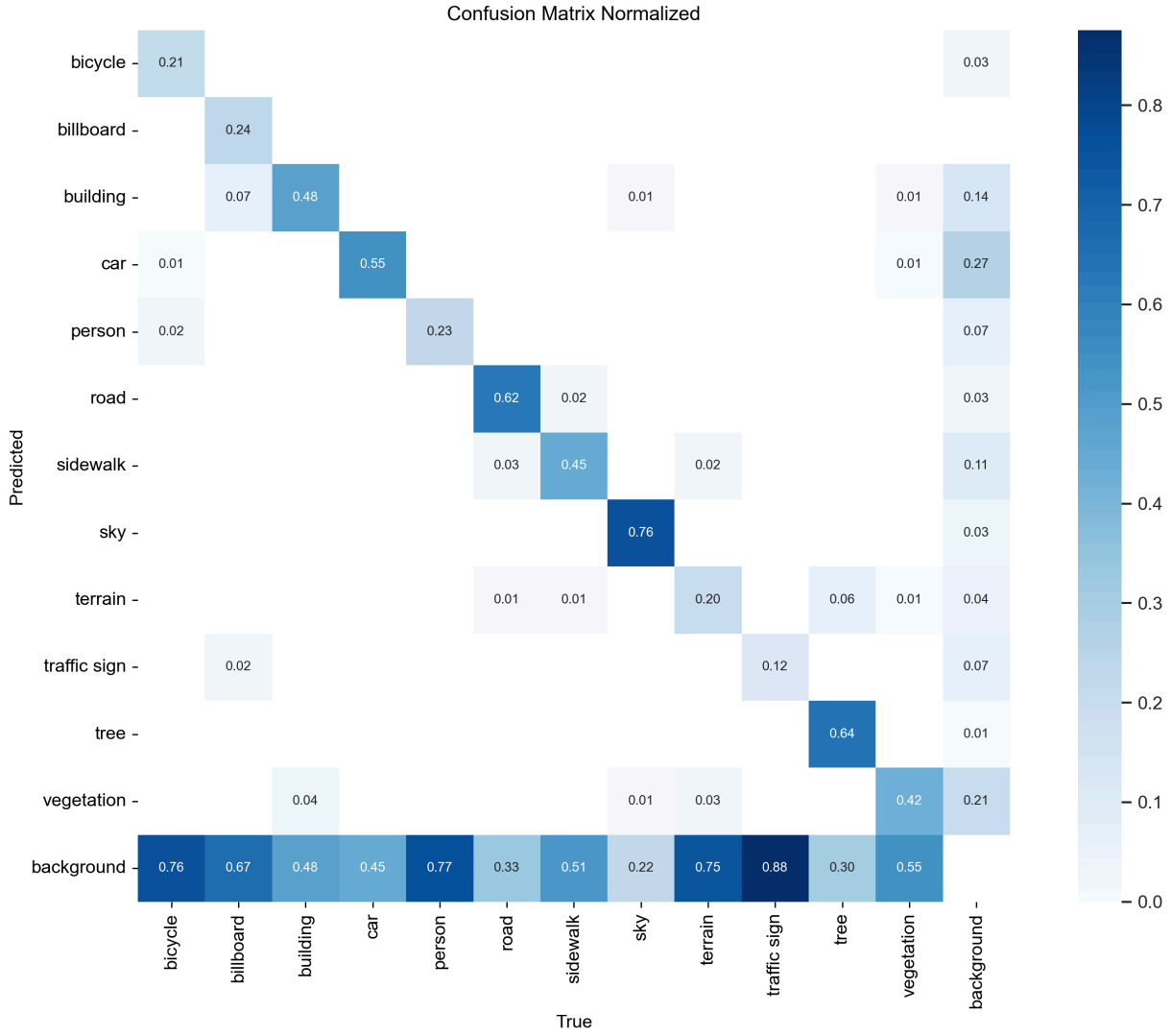

Figure 6

shows the confusion matrix of the YOLO V8 model.

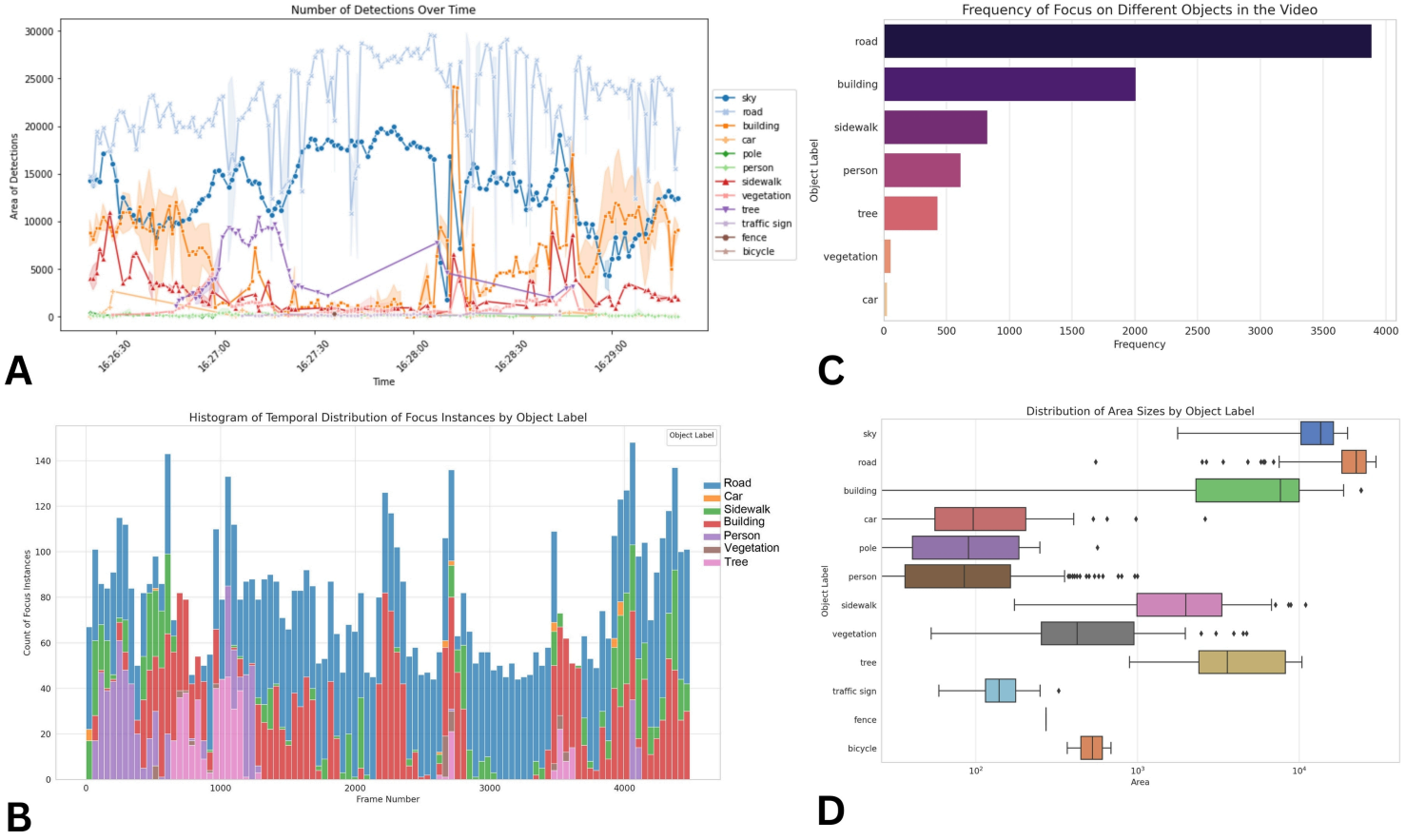

Figure 7

derived from computer vision analysis of a pedestrian navigation video, captures fluctuating object visibility (A), the temporal focus on different elements (B), the frequency of object focus (C), and the scale of visual components encountered (D), charting the dynamics of urban visual engagement.

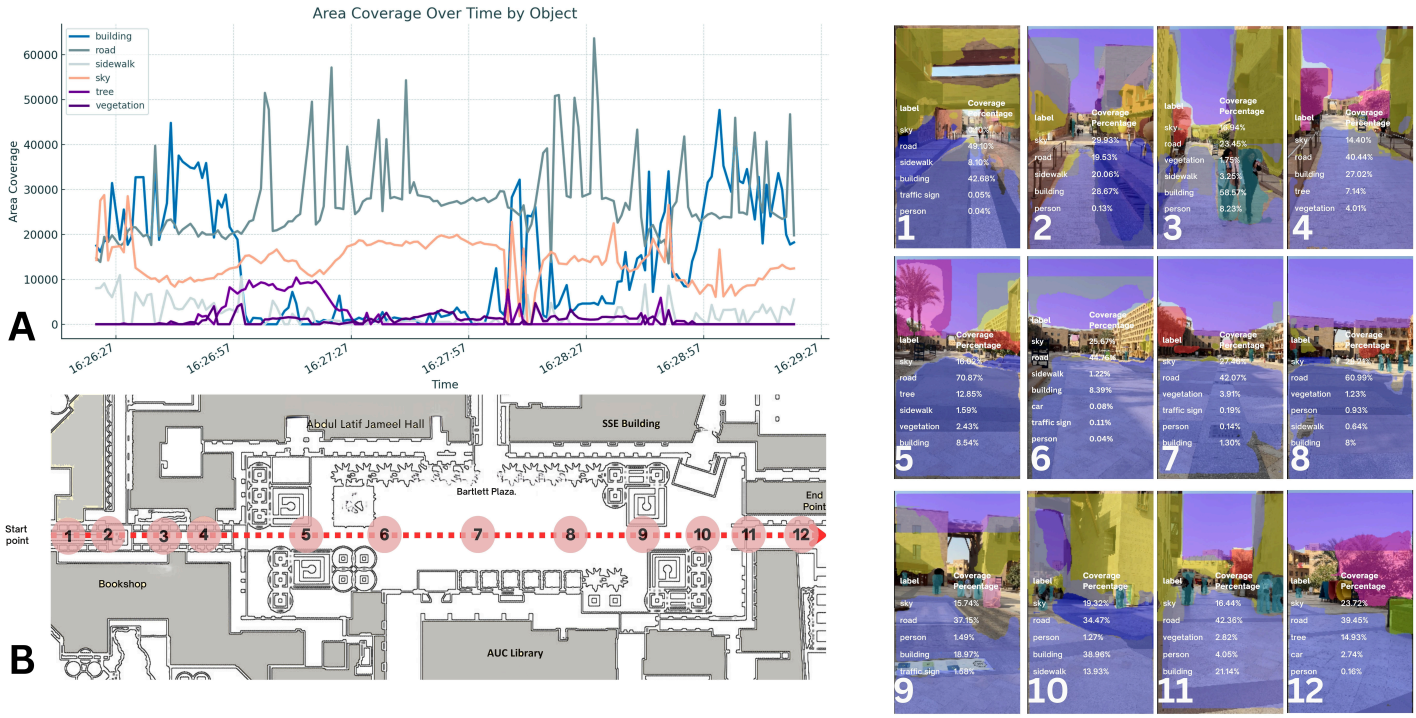

Figure 8

(A & B) analyzes the visual experience of a pedestrian walking through the library plaza at the American University in Cairo (AUC). Figure 8 (A) displays the area coverage over time for different objects or elements encountered along the walking path, such as buildings, roads, sidewalks, sky, trees, and vegetation. Figure 8 (B) shows 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) comprises 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.