RESEARCH ARTICLE

Spatiotemporal Modeling of Street Vitality Using Color Semantic Segmentation and Pedestrian Trajectory Prediction

Jijiang Zhang[1], Faziawati Abdul Aziz[1], Mohd Fabian Hasna[1]

[1]Department of Landscape Architecture, Faculty of Design & Architecture, Universiti Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

faziawati@upm.edu.my

Abstract

Urban street vitality is a critical yet under-quantified metric for evaluating public space performance, especially in dynamic and heterogeneous urban settings. This paper proposes a spatiotemporal modeling methodology to assess street vitality by integrating color semantic segmentation and pedestrian trajectory prediction. To address the fragmentation of existing vitality assessment approaches, this study aims to systematically characterize urban street vitality and uncover spatiotemporal evolution patterns using multi-source data. Deep convolutional neural networks (DCNs) are employed to analyze street scene images, extracting color distribution and semantic features to reflect functional and aesthetic aspects of streets. Meanwhile, a long short-term memory (LSTM) network is used to capture spatiotemporal movement patterns of pedestrians, highlighting dynamic street usage. A weighted fusion framework is developed to comprehensively integrate static visual elements with dynamic behavioral data. Experimental results demonstrate that this approach effectively extracts visual richness and street activity, outperforming existing uni-modal evaluation baselines in both accuracy and interpretability. Compared to color-only and trajectory-only models, the proposed method improves vitality prediction accuracy to 91.7%, with a significant reduction in error metrics. This study establishes an original theoretical framework and technical roadmap for a multidimensional assessment of urban street vitality. The findings provide insights for data-driven urban design and governance, contributing to more adaptive and sustainable public spaces.

Keywords: *Street Vitality; Spatiotemporal Modeling; Color Semantic Segmentation; Pedestrian Trajectory Prediction; Urban Analysis*

Citation:
Zhang, J., Abdul Aziz, F., & Hasna, M. F. (2025). Spatiotemporal modeling of street vitality using color semantic segmentation and pedestrian trajectory prediction. Alam Cipta, 18(2), [Article 7]. https://doi.org/10.47836/AC.18.2.PAPER07

Received:
28 February 2025

Accepted:
31 July 2025

Published:
October 2025

Introduction

Research Background and Importance

Urban street vitality is an indicator to monitor urban development quality. It not only reflects urban appeal and sustainability, but has direct impacts upon citizens' life quality and happiness (Chen et al., 2017). As urbanization has accelerated in speed, urban diseases have appeared increasingly, while urban vitality deterioration has been an acute issue to solve (Yu et al., 2024). As an integral part in urban public space, urban street vitality is not only related to efficient urban functional operation, but is closely related to urban citizens' experience in ordinary life and urban overall image making (Xue et al., 2018). However, traditional ways in urban vitality evaluation have lacked in-depth understanding in vitality generation mechanism, especially in integrating and analyzing multi-dimensional data (S. Wang et al., 2021). Most previous research focuses on single-point perspective in evaluation, e.g., physical environment factors only or only pedestrian flow data, which is difficult to grasp in full the complexity and dynamic process in urban vitality (M. Wang et al., 2022).

In recent years, with the accelerated growth in deep learning technology, computer vision technologies such as color semantic segmentation and pedestrian trajectory prediction have been developed with superior potential in urban analysis (Bozkurt et al., 2024). Technology in color semantic segmentation can acquire rich contents in street view images, analyze urban color characteristics and spatial composition in an explicit quantity manner, and subsequently analyze its impact on emotional perceptions among citizens (Long et al., 2015). Meanwhile, the LSTM-based pedestrian trajectory prediction model can efficiently grasp spatiotemporal dynamic pattern in crowd flow, forecast urban pedestrian flow evolution trends, and provide scientific support to urban traffic flow optimization (Yang et al., 2022). Meanwhile, overall application in these technologies in street vitality research is in an exploratory process. Specifically, various problems remain in integrating static environmental factors with dynamic behavior data to construct an improved systematic and overall street vitality evaluation process (Zhang et al., 2017). Under this background, developing an appropriate spatiotemporal modeling technique in modeling street vitality to integrate multi-source data and multi-dimensional factors has theoretical significance and practical value in improving urban research depth, urban planning optimization, and urban liveability enhancement in practice (X. Wang et al., 2024). However, existing approaches often examine visual features and behavioral data in isolation, lacking a coherent framework that integrates perceptual richness with dynamic movement patterns. In addition, many vitality evaluation methods rely on single-dimensional indicators, which fail to reflect the complex, evolving, and multifaceted nature of street-level activity. In response, this study proposes a spatiotemporal vitality modeling approach that combines color semantic segmentation and pedestrian trajectory prediction, enabling a more fine-grained, multidimensional, and temporally aware assessment of urban street vitality.

Research Objectives

This study aims to construct a spatiotemporal modeling method for street vitality that integrates color semantic segmentation and pedestrian trajectory prediction. By integrating static visual features and dynamic behavioral data, it comprehensively quantifies the vitality characteristics of urban streets and reveals their spatiotemporal evolution laws. Specifically, this study will develop a street scene color semantic segmentation model based on deep convolutional neural networks to extract the color distribution and functional composition characteristics of the street environment from a visual perspective; construct a pedestrian trajectory prediction model based on long short-term memory networks to analyze crowd movement patterns and their interactive relationship with the street environment; design a multi-dimensional feature fusion strategy to organically combine static environmental features with dynamic behavior data to establish a comprehensive street vitality evaluation system; verify the applicability of the proposed method in different types of streets, and explore the correlation mechanism between color characteristics, spatial structure and crowd activities. The proposed method is intended to provide empirical support for data-informed street design and spatial vitality enhancement in contemporary urban planning contexts.

Theoretical Basis and Technical Framework

To assess urban street vitality in a multidimensional and data-integrated manner, this study builds upon three interrelated foundations: the conceptual framework of street vitality, the visual perception modeling via color semantic segmentation, and the behavioral dynamics modeling through pedestrian trajectory prediction. These components collectively form the theoretical and technical basis for the integrated vitality evaluation model proposed in this research.

Concept and Evaluation Method of Street Vitality

Street vitality is an integral reflection of urban public space quality, embodying support and appeal of urban life to human activities to certain degrees (Caesar et al., 2020). From the perspective of meaning conception, urban vitality is created by aggregation of humans and is largely dependent upon morphological factors in the city (Huang et al., 2018). Street vitality in urban modern studies is commonly explained as quantity and variety of human activities in streets, materialized in volume of people in streets, frequencies of social exchanges, and scales of commercial activities (Hongbo et al., 2024). Street vitality is an integral indicator to judge urban streets' health. It not only has something to do with effectiveness in making full use of public space, but is directly related to citizens' life experience and urban development in an eco-friendly manner (Yuan et al., 2020).

The traditional ways to determine urban vitality in streets predominantly employ manual observation and survey methods by means of questionnaires. Although these methods can acquire citizens' subjective experience and evaluation directly, these methods have limitations in collecting data in an efficient manner objectively (Ivanovic & Pavone, 2019). As the era of big data has emerged, ways to determine urban vitality in streets by means of multi-source data have been employed extensively and developed in diverse ways (W. Liu et al., 2019). These methods integrate diverse sources of data including geographic information systems (GIS), remote sensing systems, and mobile positioning systems to perform multi-dimensional full-time observation of urban vitality (Bhattacharyya et al., 2019). Researchers in recent years have employed an outlook in diverse scales to analyze urban vitality in streets by investigating urban design in diverse scales between walking to driving range (Yao et al., 2019).

The evaluation of urban street vitality is evolving in an increasingly systematic and advanced manner (Li et al., 2024). On the one hand, researchers strive to integrate static evaluation with dynamic evaluation, not only focusing on physical facilities and spatial organization in streets, but focusing on spatiotemporal changes in crowd activities as well (An et al., 2024). On the other hand, with development in artificial intelligence technology, methods in evaluation of street vitality based on deep learning have emerged gradually, e.g., through video image data and deep learning methods to analyze interrelation between built environment and degree of vitality in streets (L. Liu et al., 2019). Compared with traditional evaluation methods, these new methods have certain strengths in processing big sample data and extracting patterns in vitality at the micro level (Fan et al., 2001). However, differences among various methods in evaluation remain. Correlation analysis can reflect similarities and differences among various mixed condition evaluation methods, providing an overall better reference to urban planning decisions (C. Liu et al., 2023). However, most existing vitality evaluations remain fragmented in scope or resolution, calling for a more integrated and fine-grained assessment framework.

Overview of Color Semantic Segmentation Technology

Color semantic segmentation is a key technology in the field of computer vision. It achieves refined understanding and analysis of the scene by classifying each pixel in the image (Fernando et al., 2018). In urban scene understanding, color semantic segmentation technology can automatically extract pixel-level semantic information from street scene images, providing rich visual features for quantitative analysis of the urban environment (Wei & Wang, 2024). Different from traditional semantic segmentation, color semantic segmentation technology places special emphasis on the use of color information. By analyzing color distribution and change patterns, it enhances the ability to identify urban spatial features (Ye et al., 2025).

In terms of technical implementation, color semantic segmentation usually uses false color coding to mark different semantic categories, such as marking vehicles as blue and pedestrians as red, so as to intuitively display the segmentation results (Xiong et al., 2021). In order to obtain high-quality segmentation results, researchers often use high dynamic range (HDR) images with 16-bit linear color depth as training data. Such images can provide richer color details and light and dark levels (Bansal et al., 2018). With the development of deep learning technology, color semantic segmentation methods based on convolutional neural networks have shown strong performance, especially when dealing with complex urban scenes. Its pixel-level classification accuracy has reached more than 67.1% (Mangalam et al., 2020).

In an attempt to further reinforce the strength and correctness of semantic segmentation, researchers have resorted to multimodal data fusion techniques (Zheng et al., 2023). As an example, the FuseSeg network that applies fusion between RGB and thermal infrared can efficiently tackle the limitations of a solitary modality in harsh lighting and climatic conditions by giving an extended perception capacity to urban scene understanding. Additionally, through fusion between motion segmentation algorithms and semantic segmentation algorithms, researchers have achieved an advanced understanding of urban dynamic scenes. This method properly identifies things that can be in motion by integrating position and color attributes.

In recent years, in order to cater to the demand to analyze vast urban landscapes, researchers have developed efficient algorithms in semantic segmentation. As an illustration, graph neural network-based framework segments large-sized images in limited computational resources by reducing hundreds of millions of an image's pixels to mere thousands of nodes in a graph. This technique converts an image to graph through superpixels (pixel area with equal color/intensity value) that have significantly reduced computing resources consumption while providing an optimum percentage in accuracy. In future, with continuous development in algorithms in deep learning and improved performance in computer hardware, urban scene understanding and analysis shall increasingly be dependent upon color semantic segmentation technology.

Despite the advances in color semantic segmentation, current applications in urban scene understanding often remain limited to visual categorization without linking to behavioral dynamics or perceptual interpretation. Existing models rarely consider how extracted color features interact with human spatial experiences or contribute to composite urban performance indicators such as vitality. This study addresses this gap by integrating color semantics with trajectory-based behavior modeling, moving beyond static segmentation to a dynamic and multidimensional vitality assessment.

Analysis of Pedestrian Trajectory Prediction Model

Trajectory prediction of pedestrians is an integral part of autonomous driving systems and intelligent transportation systems. Pedestrian trajectory prediction is to predict future trajectories of pedestrians by relying on observation in the past. As deep learning has progressed rapidly in recent years, trajectory prediction models have been transforming from traditional knowledge-based methods to algorithms in deep learning, opening increasingly rich directions in technique and application. At present, algorithms employed in trajectory prediction of pedestrians can be commonly divided into two categories: model-based methods and methods in deep learning based on LSTM.

The trajectory prediction algorithm in accordance with the social force model is one of the most traditional algorithms. This algorithm views pedestrians as particles acted upon by diverse forces and predicts motion trajectories by simulating interacting forces between pedestrians and between pedestrians and environments. The novel transformer-based pedestrian trajectory prediction framework Force Former utilizes strengths in full in the social force model. It not only uses sequence motion information as input, but considers diverse influencing factors, thus providing better prediction accuracy. This physical model-based approach has limitations in handling complicated circumstances and prediction over long time scales.

Based on its better time series processing capacity, LSTM networks have been the mainstream in pedestrian trajectory prediction studies. LSTM-based prediction modeling can better comprehend the long-term dependency among movements of pedestrians and obtain more accurate temporal features to conduct trajectory prediction. As an example, PoPPL (Prediction of Pedestrian Paths by LSTM) predicts potential destination areas in phase one and outputs trajectories corresponding to predicted destination areas in phase two by splitting pedestrian trajectories into route categories and predicting destination areas by bidirectional LSTM classification network. This method to take potential destination prediction is an efficient manner to boost trajectory prediction performance.

In an attempt to further optimize the performance of LSTM models in predicting pedestrian trajectories, researchers have proposed various enhancement methods. The SS-LSTM model is superior in modeling pedestrians' social interactions and environmental constraints by developing a hierarchical LSTM framework. The most advanced space-time-spectral LSTM is superior in modeling pedestrian trajectory patterns by integrating spatial, time-domain and spectral domain knowledge, and significantly improving migration capacity between diverse sets of data. Additionally, by modeling pedestrians' purposes and social interactions and integrating the relative locations between adjacent pedestrians, researchers have further improved prediction performance in both reliability and accuracy. As in-depthness in research has been growing, pedestrian trajectory prediction models have been continuously evolving in multimodal direction, high accuracy and powerful adaptation, providing full support in developing an urban traffic scene in which it is both safer and smarter.

While LSTM-based models have significantly improved trajectory prediction accuracy, they rarely account for the interaction between pedestrian movement and the perceptual characteristics of the street environment. Most existing models remain decoupled from spatial semantics, limiting their utility in comprehensive vitality assessment. This study bridges this gap by integrating behavioral prediction with environmental color segmentation.

Street Vitality Spatiotemporal Modeling Method

To operationalize the proposed theoretical framework, this section presents a spatiotemporal modeling method that integrates color semantic segmentation and pedestrian trajectory prediction. The approach consists of three interconnected components: visual extraction of color-based environmental features, temporal modeling of movement dynamics, and a fusion strategy that synthesizes static and dynamic indicators into a composite vitality score. This integrative structure enables a multidimensional assessment of street vitality across perceptual and behavioral dimensions, establishing a methodological foundation for empirical analysis.

Street Feature Extraction via Color Semantic Segmentation

The visual representation of street vitality is inherently influenced by the distribution of color and the semantic structure of urban scenes. This section applies color semantic segmentation techniques to extract key environmental features from street-level imagery for subsequent vitality assessment. The methodology is grounded in deep convolutional neural networks (CNNs), which have demonstrated robust performance in structured scene understanding (Zhao et al., 2024)a series of exemplary advances have been made in several areas involving image classification, semantic segmentation, object detection, and image super-resolution reconstruction with the rapid development of deep convolutional neural network (CNN).

The segmentation process begins by inputting an urban street image, where denotes the pixel coordinates. Using a pre-trained CNN such as DeepLabv3+, the model segments the image into semantic categories (e.g., vegetation, buildings, pedestrians), producing a label map that assigns each pixel to a specific class (Du et al., 2021) such as DeepLabv3+, have achieved astonishing performance for semantically labeling very high resolution (VHR). The corresponding color information, normalized across RGB channels, is then integrated with semantic labels to construct the color feature vector :

$$F_c = \frac{1}{N}\sum_{(x,y)} w_s \cdot S(x,y) \cdot \mathrm{C}(x,y) \qquad [1]$$

Where is the total number of pixels in the image, and is a manually assigned weight representing the vitality contribution of each semantic category—for example, green spaces and pedestrian zones are assigned higher weights due to their strong association with perceived vibrancy (Chen & Huang, 2024)and the city's physical plan plays an important role in shaping and supporting urban vibrancy. However, the effectiveness of urban planning strategies in optimizing resource allocation and fostering a vibrant city remains unclear. This study aims to contribute to this understanding by exploring the spatial and temporal effects underlying the association between urban vibrancy and urban forms. Using mobile-phone location-request data along with urban nighttime light data as a proxy for urban vibrancy in Shenzhen, China, we conducted a geographically and temporally weighted regression (GTWR).

This weighted summation fuses the spatial distribution of colors with semantic relevance, resulting in a composite vector that reflects the street's visual vitality. The process enables a context-sensitive extraction of features, taking into account both perceptual color richness and functional land use.

To further quantify the visual diversity of the street scene, a color diversity index is introduced. This index is calculated using entropy-based methods to capture the heterogeneity across different semantic classes:

$$D_c = -\sum_{i=1}^{k} p_i \log(p_i) \qquad [2]$$

where represents the proportion of pixels belonging to the th color-semantic category. Higher values of indicate more diverse and visually stimulating urban scenes, aligning with findings from perceptual urban design studies (Jin & Wang, 2021).

Figure 1 illustrates the simplified image processing pipeline, including semantic segmentation, edge detection, and superpixel generation. Panel (a) shows the original input image; panel (b) visualizes edge probability using color-coded intensity; panel (c) applies sticky superpixel segmentation to preserve spatial boundaries and structure.

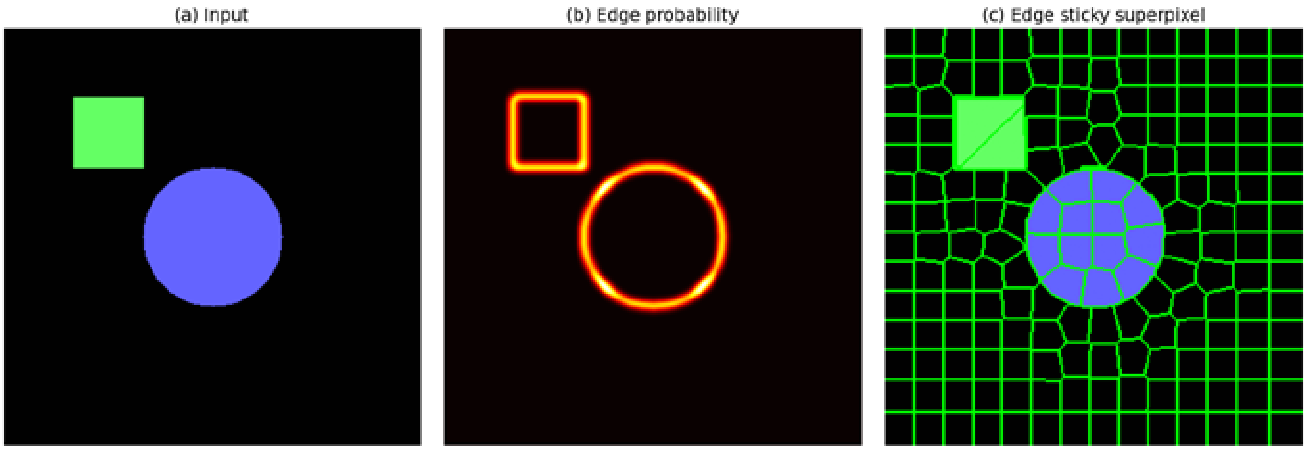


Figure 1: Simplified Demonstration of Image Processing Techniques

Figure 2 provides a more realistic urban example, depicting pedestrians, shopfronts, and signage (e.g., “UNIQLO”), provides semantically diverse visual cues for effective segmentation. The semantic segmentation model identifies edge structures and visual categories that contribute to the overall vitality score.

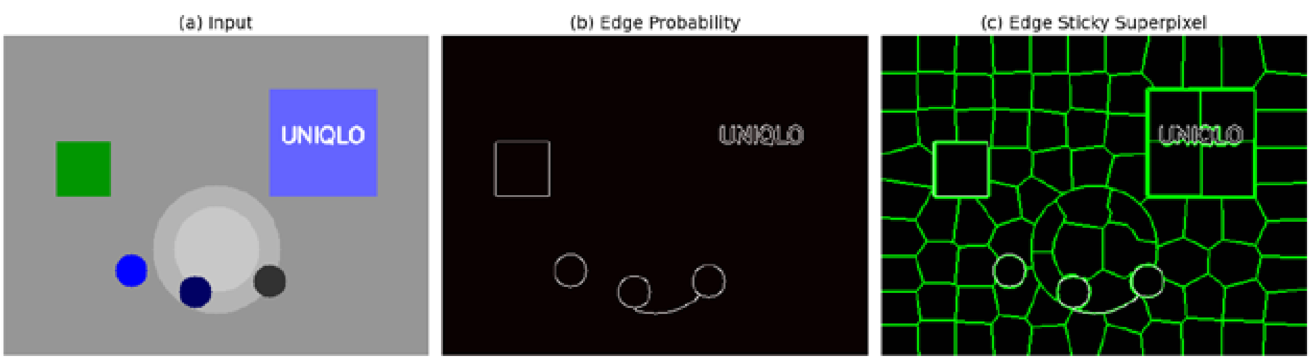
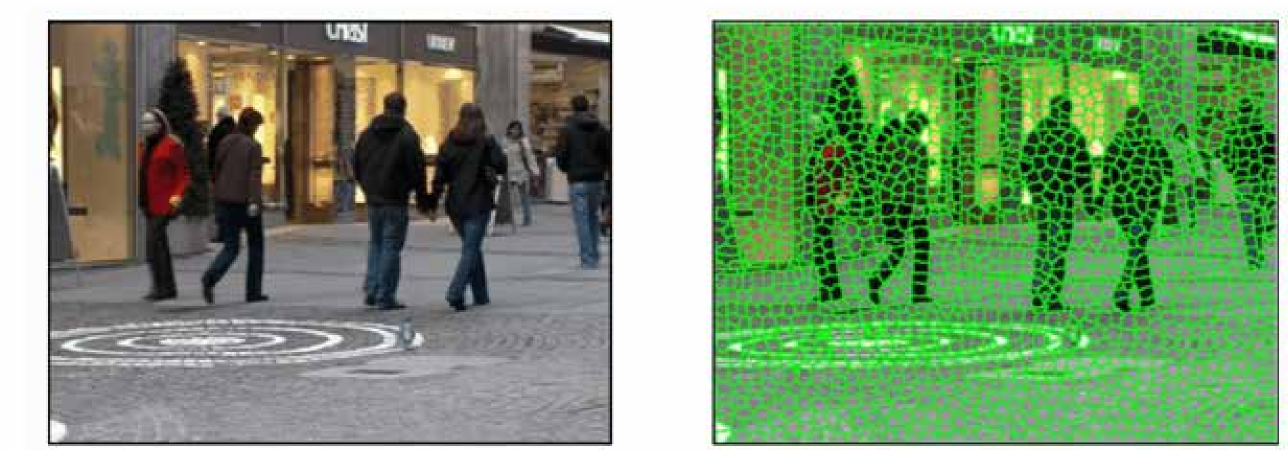


Figure 2: Illustration of Image Processing for Urban Scene Analysis

Lastly, Figure 3 presents a practical application of color semantic segmentation in a real urban context. The superpixel segmentation overlays further support the extraction of structured visual elements, reinforcing the method’s adaptability to complex urban streetscapes (Ma et al., 2023)

Figure 3: Real Scene Color Semantic Segmentation Example



This comprehensive approach to feature extraction establishes a methodological basis for integrated vitality modelling in the subsequent sections.

Pedestrian Trajectory Prediction Model Design

Pedestrian trajectories are a dynamic reflection of street vitality, capturing the temporal rhythms and crowd flow patterns of urban spaces. This section develops a trajectory prediction model using the Long Short-Term Memory (LSTM) neural network, a recurrent architecture well-suited for sequential data processing (Alahi et al., 2016).

Suppose the position of pedestrian i at time t is given by $P_i(t)=(x_i(t),y_i(t))$, and the historical trajectory sequence is denoted as:

$P_i(t-2),....,P_i(t-k)\}$

where is the time window length? The model aims to predict the future position by integrating temporal sequence and environmental features , derived from visual semantics and color data extracted in Section 4.1. The predictive formulation is:

$$P_i(t+1) = f_{LSTM}(T_i, E(t) \ ; \varphi) \qquad [3]$$

Where f_{LSTM} represents is the nonlinear mapping function learned by the $LSTM$, and ϕ is represents the trainable model parameters.

The environmental input $E(t)$ incorporates the colour-based feature vector F_c from semantic segmentation, enabling the model to adjust predictions based on perceived spatial affordances such as vegetation, pedestrian density, and signage (Lakmali et al., 2024).

This study adopts an LSTM-based architecture for temporal trajectory modeling, following Alahi et al. (2016), who demonstrated its efficacy in human motion prediction. The model is trained using the Adam optimizer (Kingma & Ba, 2017), which ensures stable convergence through adaptive gradient adjustment.

By aligning motion patterns with perceptual urban cues, the model enhances both the spatial sensitivity and contextual accuracy of vitality prediction. This integration of movement dynamics with environmental semantics provides a robust foundation for modeling behaviorally informed urban liveliness.

Fusion Strategy for Spatiotemporal Modeling of Vitality

To achieve spatiotemporal modeling of street vitality, this section proposes a fusion strategy that integrates colour features and pedestrian trajectory predictions. The objective is to capture both the static visual attributes and dynamic behavioral patterns of urban spaces.

Street vitality is defined as the vitality score of a street segment at time , calculated via a weighted fusion of visual semantics and pedestrian mobility. The formulation is as follows:

$$v(s,t) = \alpha F_c(s) + \beta \sum_{i \in s} \frac{1}{\left\| P_i(t+1) - P_i(t) \right\| + \varepsilon} \qquad [4]$$

Where:

- $F_c(s)$ is the colour-based feature vector of street segment sss, extracted via semantic segmentation and defined in Section 4.1;
- $P_i(t+1)$ and $P_i(t)$ represent the predicted and current positions of pedestrian i, respectively, as computed in Section 4.2;
- a and $ß$ weighting coefficients that balance visual and behavioral contributions.
- ε is a small constant to prevent division by zero.

This fusion approach combines the visual attractiveness of the environment with the intensity of pedestrian activity, thereby enabling a quantified representation of vitality in both spatial and temporal dimensions.

- In practice, the weighting parameters a and $ß$ are determined through empirical calibration. Their optimal values are discussed in Chapter 5 based on model fitting accuracy. To reflect evolving patterns of vitality, this model further incorporates time series analysis to smooth $V(s,t)$ and perform trend prediction.

By aligning perceptual cues with real-time human mobility, the model provides a robust foundation for evidence-based vitality diagnostics. It supports adaptive urban planning strategies that respond to both static affordances and dynamic rhythms of street life.

Experimental Analysis

This chapter experimentally verifies the effectiveness of the spatiotemporal modeling method of street vitality proposed in Chapter 4. It mainly includes experimental design, data set preparation, and evaluation of model performance and result analysis. Through quantitative and qualitative analysis, the performance of the model in feature extraction, trajectory prediction and vitality modeling are explored, and the spatiotemporal variation of street vitality is revealed.

Experimental Design and Dataset Introduction

The experimental design aims to evaluate the effect of the fusion of color semantic segmentation and pedestrian trajectory prediction. Real data of urban streets, including image data and trajectory data, are selected to construct a comprehensive dataset. The image data comes from surveillance cameras on the main streets of a city. The collection time covers June to August 2024, with a total of 5,000 high-resolution images with a resolution of 1920×1080. The trajectory data is obtained through mobile device GPS records, which contains the position sequence of 1,000 pedestrians in the same area, and each sequence is 50-time steps long.

The experiment is divided into three stages: first, the image is segmented by color semantics to extract street features; then the prediction model is trained using trajectory data; finally, the two parts of the results are fused to calculate the vitality value. The hardware environment is a computing platform equipped with an NVIDIA RTX 3090 GPU, and the software environment includes Python 3.9 and PyTorch 1.12.

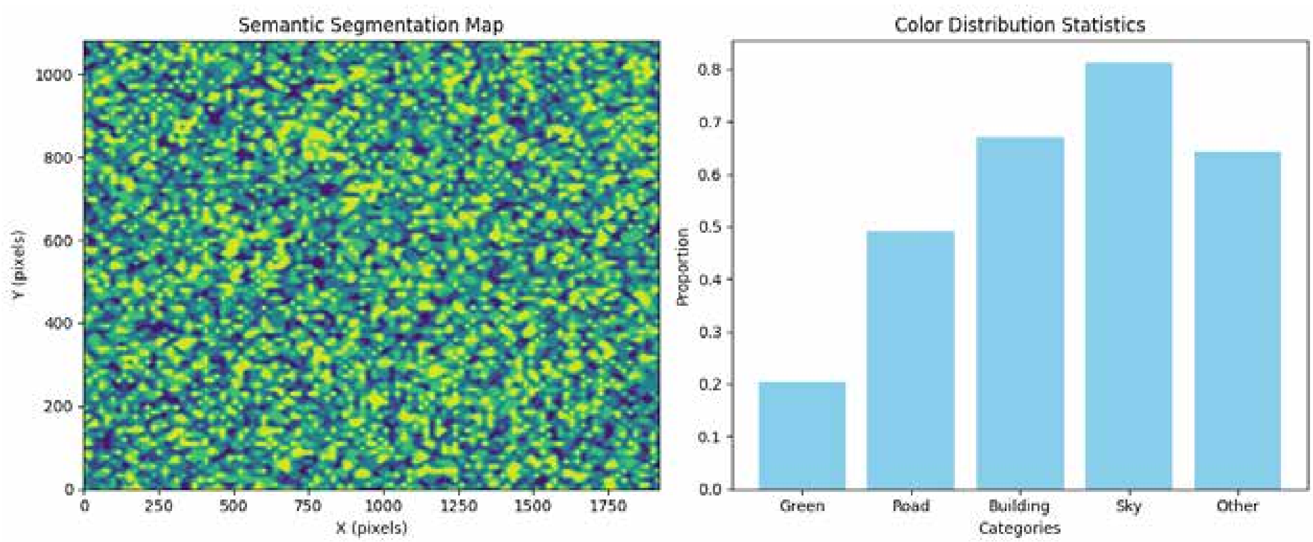


Figure 4: Feature Distribution of the Dataset

Figure 4 shows the feature distribution of the dataset. The left sub-image is an example of a semantic segmentation map, which uses different colors to distinguish categories such as green space and roads, reflecting the visual structure of the street; the right sub-image is a color distribution statistic, showing the proportion of each category, among which green space and road surfaces constitute the largest proportions, indicating their visual prominence in the sampled urban streetscape.

Model Performance Evaluation and Result Analysis

In order to comprehensively evaluate the performance of the model, the experiment is carried out from three aspects: color feature extraction, trajectory prediction, and vitality modeling. The evaluation indicators include the mean intersection over union (mIoU) of semantic segmentation, the average displacement error (ADE) of trajectory prediction, and the prediction accuracy of the vitality value. Table 5.1 lists the performance of the semantic segmentation module in different categories.

Table 1: Semantic Segmentation Performance

Category	IoU (%)	Precision (%)	Recall (%)
Green	85.3	88.7	90.1
Road	90.2	92.5	91.8
Building	87.6	89.4	88.9
Sky	93.1	94.2	95.0
Average	89.05	91.2	91.45

Table 1 shows that the semantic segmentation module achieved high IoU values in all categories, with an average of 89.05%. The sky category performed best (93.1%), reflecting the model's strong ability to segment large areas of single color. The trajectory prediction performance is shown in Figure 3, including the comparison of the actual trajectory and the predicted trajectory and the error distribution.

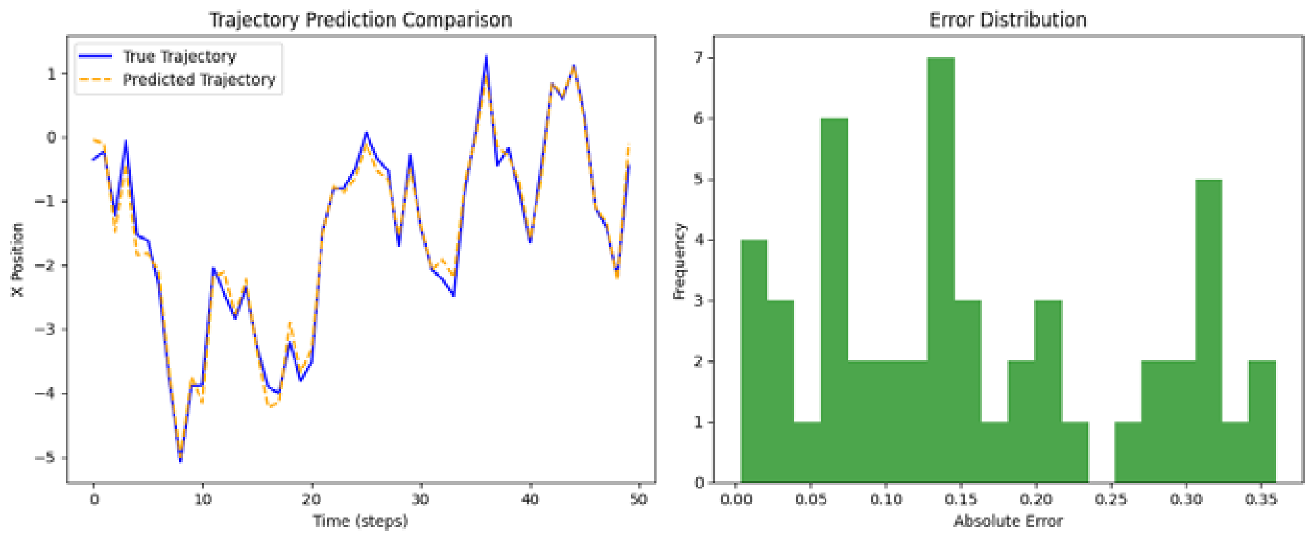


Figure 5: The effect of trajectory prediction

As shown in Figure 5, the left sub-image compares the actual trajectory with the predicted trajectory, indicating that the predicted result is highly consistent with the actual value; the right sub-image is the error distribution, most of the errors are concentrated within 0.2, verifying the prediction accuracy of the model. The average displacement error (ADE) is 0.18, indicating that the model effectively captures the dynamic laws of pedestrian movement and solves the uncertainty problem in trajectory prediction.

The performance of vitality modeling is summarized in Table 2, comparing the results under different fusion strategies.

Table 2: Vitality Modeling Performance

Method	Accuracy (%)	MAE	RMSE
Color Only	78.5	0.25	0.32
Trajectory Only	82.3	0.21	0.28
Proposed Fusion (+0.4, b =0.6)	91.7	0.15	0.19

Table 2 shows that the fusion strategy (=0.4, =0.6) achieved the best performance in vitality prediction, with an accuracy of 91.7% and a root mean square error (RMSE) of 0.19, which is better than the methods using color or trajectory alone. This result shows that fusing static features with dynamic data significantly improves the accuracy of vitality modeling and solves the limitations of single-dimensional analysis. This result provides empirical evidence for integrating multi-source perceptual and behavioral data in urban vitality modeling, offering a practical tool for dynamic planning and spatial optimization.

Conclusion and Future Work

This concluding chapter synthesizes the key findings of the proposed spatiotemporal vitality modeling approach, evaluates its methodological contributions and limitations, and outlines future research directions. Building upon the experimental results and technical framework, the discussion reflects on the model's theoretical implications, practical applicability, and potential for integration into broader urban design and planning strategies.

Main Conclusions

This study developed an efficient spatiotemporal modeling framework for street vitality by integrating color semantic segmentation with pedestrian trajectory prediction. The experimental results demonstrate that visual features extracted from color segmentation effectively capture scene diversity and visual pleasantness, while the trajectory model accurately represents dynamic crowd flow patterns. The proposed fusion strategy achieved a vitality prediction accuracy of 91.7% with a reduced RMSE of 0.19, outperforming single-dimensional approaches. These findings confirm the feasibility of integrating static and dynamic factors and reveal nuanced spatiotemporal dynamics underlying street vitality. Streets with higher proportions of green areas tend to exhibit greater pedestrian activity, while those dominated by monotonous color show reduced vitality—offering data-driven insights into urban functionality and liveability.

Moreover, the proposed method is highly generalizable and applicable to urban streetscapes of varying scales. The use of composite visualizations—such as semantic segmentation and trajectory maps—enhances model interpretability. The experimental fusion strategy (= 0.4, = 0.6) demonstrates the importance of balanced parameter tuning in achieving optimal performance. Beyond vitality assessment, this approach holds promise for broader applications in urban planning, mobility management, and public space optimization, highlighting its value in real-world scenarios. This study offers a scalable and interpretable methodological paradigm for integrating AI techniques into urban spatial diagnostics.

Research Limitations

Despite the promising results achieved in this study, several limitations remain that affect the comprehensiveness, scalability, and real-world applicability of the proposed method. First, the scope and representativeness of the data set are restricted. The image and trajectory data were primarily collected from the core area of a single city, with limited temporal and spatial diversity. As such, the model may not adequately capture variations in street vitality caused by seasonal dynamics, weather changes, or cultural differences across urban contexts (Batty, 2013). Furthermore, the manual assumptions introduced during data annotation and preprocessing—such as the weighting of semantic categories—could introduce biases, particularly in visually complex scenes where the model's capacity to capture nuanced spatial characteristics remains constrained (Northcutt et al., 2021).

Second, computational limitations pose practical challenges. Both the color semantic segmentation and pedestrian trajectory prediction components rely on deep learning architectures, which require high-performance hardware for training and inference (Finn et al., 2017). This restricts deployment in resource-constrained environments, particularly in cities with limited digital infrastructure. Additionally, the current fusion strategy depends on manually tuned weight parameters, lacking an adaptive mechanism for real-time adjustment. This limits the method's transferability and responsiveness when applied to new or evolving scenarios (Kitchin, 2014). While the framework demonstrates strong performance in controlled experimental settings, scalability and automation remain key areas for future enhancement to ensure viability in large-scale urban applications.

Future Research Directions

Future research could advance along several key directions to enhance the robustness, scalability, and applicability of the proposed method. First, the dataset can be diversified to include multi-city samples across different climatic zones, seasons, and urban forms, thereby improving the model’s generalizability. Incorporating multimodal data such as ambient sound, thermal conditions, or geotagged social media activity may enrich the representation of vitality and strengthen the model’s adaptability to complex real-world environments. Additionally, the development of automated tools for data cleaning and annotation could significantly reduce human bias, improve efficiency, and facilitate large-scale deployment.

Second, optimizing the model’s computational efficiency and adaptability remains a key challenge. Future work could explore lightweight neural network architectures that reduce hardware dependency, enabling application in resource-constrained settings. Techniques such as reinforcement learning or meta-learning may allow the model to self-tune fusion weights, improving performance in new scenarios. Furthermore, integrating the vitality modeling framework into urban planning workflows such as real-time monitoring dashboards or interactive decision-making tools would enhance its translational value, bridging the gap between technical innovation and practical urban governance. Future research should further embed such frameworks into urban governance systems to generate dual benefits of technical precision and policy relevance.

Acknowledgements

The authors gratefully acknowledge the financial support provided by Universiti Putra Malaysia through the Geran Putra IPS Grant (Grant Number GP-IPS-9772600).

References

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 961–971. http://openaccess.thecvf.com/content_cvpr_2016/html/Alahi_Social_LSTM_Human_CVPR_2016_paper.html

An, Y., Zhao, Z., & Zhang, C. (2024). CMF-net: Semantic Segmentation of 3D Color Point Clouds in Outdoor Scenes for Intelligent Wheelchairs Based on Multimodal Fusion. *2024 39th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 1412–1419. https://doi.org/10.1109/YAC63405.2024.10598425

Bansal, M., Krizhevsky, A., & Ogale, A. (2018). ChauffeurNet: *Learning to Drive by Imitating the Best and Synthesizing the Worst* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1812.03079

Batty, M. (2013). *The New Science of Cities*. Mit Pr.

Bhattacharyya, A., Fritz, M., & Schiele, B. (2019). *Bayesian Prediction of Future Street Scenes Using Synthetic Likelihoods* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1810.00746

Bozkurt, S., Atik, M. E., & Duran, Z. (2024). Improving Aerial Targeting Precision: A study on Point Cloud Semantic Segmentation with Advanced Deep Learning Algorithms. *Drones*, *8*(8), 376. https://doi.org/10.3390/drones8080376

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). nuscenes: A Multimodal Dataset for Autonomous Driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11621–11631. http://openaccess.thecvf.com/content_CVPR_2020/html/Caesar_nuScenes_A_Multimodal_Dataset_for_Autonomous_Driving_CVPR_2020_paper.html

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation* (No. arXiv:1706.05587). arXiv. https://doi.org/10.48550/arXiv.1706.05587

Chen, Z., & Huang, B. (2024). Achieving Urban Vibrancy through Effective City Planning: A Spatial and Temporal Perspective. *Cities*, *152*, 105230. https://doi.org/10.1016/j.cities.2024.105230

Du, S., Du, S., Liu, B., & Zhang, X. (2021). Incorporating DeepLabv3+ and Object-Based Image Analysis for Semantic Segmentation of Very High Resolution Remote Sensing Images. *International Journal of Digital Earth*, *14*(3), 357–378. https://doi.org/10.1080/17538947.2020.1831087

Fan, J., Yau, D. K., Elmagarmid, A. K., & Aref, W. G. (2001). Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing. *IEEE Transactions on Image Processing*, *10*(10), 1454–1466. https://doi.org/10.1109/83.951532

Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2018). Soft+ Hardwired Attention: An Lstm Framework for Human Trajectory Prediction and Abnormal Event Detection. *Neural Networks*, *108*, 466–478. https://doi.org/10.1016/j.neunet.2018.09.002

Finn, C., Abbeel, P., & Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *International Conference on Machine Learning*, 1126–1135. https://proceedings.mlr.press/v70/finn17a.html

Hongbo, L., Omar, S. Z., He, C., & Shanshan, W. (2024). Construction of Image Education Knowledge Map Model Based on Data Mining Technology. *Applied Mathematics and Nonlinear Sciences*, *9*(1), 20241837. https://doi.org/10.2478/amns-2024-1837

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., & Yang, R. (2018). The Apolloscape Dataset for Autonomous Driving. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 954–960. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w14/html/Huang_The_ApolloScape_Dataset_CVPR_2018_paper.html

Ivanovic, B., & Pavone, M. (2019). The Trajectron: Probabilistic Multi-Agent Trajectory Modeling with Dynamic Spatiotemporal Graphs. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2375–2384. http://openaccess.thecvf.com/content_ICCV_2019/html/Ivanovic_The_Trajectron_Probabilistic_Multi-Agent_Trajectory_Modeling_With_Dynamic_Spatiotemporal_Graphs_ICCV_2019_paper.html

Jin, X., & Wang, J. (2021). Assessing Linear Urban Landscape from Dynamic Visual Perception Based on Urban Morphology. *Frontiers of Architectural Research*, *10*(1), 202–219. https://doi.org/10.1016/j.foar.2021.01.001

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1412.6980

Kitchin, R. (2014). The Real-Time City? Big Data and Smart Urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Lakmali, R. G. N., Genovese, P. V., & Abewardhana, A. A. B. D. P. (2024). Evaluating the Efficacy of Agent-Based Modeling in Analyzing Pedestrian Dynamics Within the Built Environment: A Comprehensive Systematic Literature Review. *Buildings*, *14*(7), 1945. https://doi.org/10.3390/buildings14071945

Li, Q., Yuan, H., Fu, T., Yu, Z., Zheng, B., & Chen, S. (2024). Multispectral Semantic Segmentation for Uavs: A Benchmark Dataset and Baseline. *IEEE Transactions on Geoscience and Remote Sensing*. https://ieeexplore.ieee.org/abstract/document/10672543/

Liu, C., Ge, L., Xiang, W., Du, Z., & Zhang, Q. (2023). Channel Attention and Normal-Based Local Feature Aggregation Network (Cnlnet): A Deep Learning Method for Predisaster Large-Scale Outdoor Lidar Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–12. https://ieeexplore.ieee.org/abstract/document/10343136/

Liu, L., Qiu, Z., Li, G., Liu, S., Ouyang, W., & Lin, L. (2019). Crowd Counting with Deep Structured Scale Integration Network. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1774–1783. http://openaccess.thecvf.com/content_ICCV_2019/html/Liu_Crowd_Counting_With_Deep_Structured_Scale_Integration_Network_ICCV_2019_paper.html

Liu, W., Salzmann, M., & Fua, P. (2019). Context-Aware Crowd Counting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5099–5108. http://openaccess.thecvf.com/content_CVPR_2019/html/Liu_Context-Aware_Crowd_Counting_CVPR_2019_paper.html

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. http://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html

Ma, L., Guo, Z., Lu, M., He, S., & Wang, M. (2023). Developing an Urban Streetscape Indexing Based on Visual Complexity and Self-Organizing Map. *Building and Environment*, *242*, 110549. https://doi.org/10.1016/j.buildenv.2023.110549

Mangalam, K., Girase, H., Agarwal, S., Lee, K.-H., Adeli, E., Malik, J., & Gaidon, A. (2020). It is not the Journey but the Destination: Endpoint Conditioned Trajectory Prediction. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12347, pp. 759–776). Springer International Publishing. https://doi.org/10.1007/978-3-030-58536-5_45

Northcutt, C., Jiang, L., & Chuang, I. (2021). Confident learning: Estimating Uncertainty in Dataset Labels. *Journal of Artificial Intelligence Research*, *70*, 1373–1411. https://doi.org/10.1613/jair.1.12125

Wang, M., He, Y., Meng, H., Zhang, Y., Zhu, B., Mango, J., & Li, X. (2022). Assessing Street Space Quality Using Street View Imagery and Function-Driven Method: The Case of Xiamen, China. *ISPRS International Journal of Geo-Information*, *11*(5), 282. https://doi.org/10.3390/ijgi11050282

Wang, S., Miao, H., Li, J., & Cao, J. (2021). Spatio-Temporal Knowledge Transfer for Urban Crowd Flow Prediction via Deep Attentive Adaptation Networks. *IEEE Transactions on Intelligent Transportation Systems*, *23*(5), 4695–4705. https://doi.org/10.1109/TITS.2021.3055207

Wang, X., Yang, T., Xu, Y., Ming, M., Wang, H., Li, L., Xu, J., Zhang, J., Zhang, P., & Wang, H. (2024). A Dataset of Semantic Segmentation in Earthquake Disaster Detection Based on Social Media Images. *China Scientific Data*, *9*(3), 1–10. https://doi.org/10.11922/11-6035.noda.2023.0005.zh

Wei, J., & Wang, W. (2024). A Novel Method for Facial Attribute Editing by Integrating Semantic Segmentation and Color Rendering. *Fifteenth International Conference on Graphics and Image Processing (ICGIP 2023)*, *13089*, 83–91. https://doi.org/10.1117/12.3021236

Xiong, J., Hsiang, E.-L., He, Z., Zhan, T., & Wu, S.-T. (2021). Augmented Reality and Virtual Reality Displays: Emerging Technologies and Future Perspectives. *Light: Science & Applications*, *10*(1), 216. https://doi.org/10.1038/s41377-021-00658-8

Xue, H., Huynh, D. Q., & Reynolds, M. (2018). SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1186–1194. https://doi.org/10.1109/WACV.2018.00135

Yang, D., Zhu, T., Wang, S., Wang, S., & Xiong, Z. (2022). LFRSNet: A Robust Light Field Semantic Segmentation Network Combining Contextual and Geometric Features. *Frontiers in Environmental Science*, *10*, 996513. https://www.frontiersin.org/articles/10.3389/fenvs.2022.996513/full

Yao, Y., Xu, M., Wang, Y., Crandall, D. J., & Atkins, E. M. (2019). Unsupervised Traffic Accident Detection in First-Person Videos. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 273–280. https://doi.org/10.1109/IROS40897.2019.8967556

Ye, Z., Li, Y., Li, Z., Liu, H., Zhang, Y., & Li, W. (2025). Attention-Multi-Scale Network for Semantic Segmentation of Multi-Modal Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2025.3540848

Yu, M., Zheng, X., Qin, P., Cui, W., & Ji, Q. (2024). Urban Color Perception and Sentiment Analysis Based on Deep Learning and Street View Big Data. *Applied Sciences*, *14*(20), 9521. https://doi.org/10.3390/app14209521

Yuan, Y., Chen, X., & Wang, J. (2020). Object-Contextual Representations for Semantic Segmentation. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Vol. 12351, pp. 173–190). Springer International Publishing. https://doi.org/10.1007/978-3-030-58539-6_11

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). https://doi.org/10.1609/aaai.v31i1.10735

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A Review of Convolutional Neural Networks in Computer Vision. *Artificial Intelligence Review*, *57*(4). https://doi.org/10.1007/s10462-024-10721-6

Zheng, Y., Chen, S., & He, Q. (2023). Hybrid Attention-Based Semantic Segmentation for Hand Acupoint Reflex Zones. *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, 1066–1071. https://doi.org/10.1145/3644116.3644298