


Content-based vs. Similarity-based Deep Learning Approaches for Walkability Assessment

Ankita Kadam¹, Felix Vu¹, Siddheshwari Bankar¹, Alivia Zhao¹, Garland Lau¹, Wan D. Bae ¹, and Shayma Alkobaisi²

¹ Seattle University, Seattle, WA, USA

{akadam1, mvu1, sbankar, azhao4, glau, baew}@seattleu.edu

² United Arab Emirates University, Al Ain, UAE

shayma.alkobaisi@uaeu.ac.ae

Abstract. Urban environments significantly influence residents’ health, well-being, and community engagement. However, current methods for assessing walkability often lack efficiency and accuracy, especially given the challenges posed by urban expansion and environmental shifts. This study presents an automated framework that integrates deep learning techniques with geographic data to evaluate walkability and recommend optimal pedestrian paths. We explore two approaches: context-based learning using GPT-4o mini and similarity-based learning with Contrastive Language-Image Pretraining (CLIP). By analyzing Google Street View images, we generate walkability scores for specific locations and employ a ranking system that combines these scores with geographic features, such as elevation and slope, to identify the most accessible paths. Our comprehensive evaluation across several Seattle neighborhoods, compared with human perception data from surveys, demonstrates the effectiveness of our approach. This scalable framework enhances walkability assessments and supports the creation of safer, more inclusive urban environments.

Keywords: walkability assessment, deep learning, pedestrian path construction

1 Introduction

In response to rapid urbanization, enhancing walkability has become crucial for public safety and health, local economies, and community cohesion, ultimately fostering livable and sustainable cities. Improved walkability enriches the urban experience for both residents and tourists while it assists urban planners in optimizing city landscapes and maintaining the infrastructure.

Research on walkability has traditionally focused on the physical characteristics of neighborhoods. Early studies identified key mesoscale factors, such as residential density and land use, as significant determinants of walking behavior [1]. Recent technological advancements, including streetscape imagery and semantic segmentation, have facilitated the measurement of street-scale features

[6]. However, it is increasingly recognized that perceived walkability is as equally important as physical environmental metrics. Several studies have explored how people perceive urban settings and reported that a comprehensive assessment of walkability must consider both objective features and subjective perceptions [3].

Despite these advancements, assessing walkability remains challenging. Physical walkability is often evaluated through manual or semi-manual inspections, which are resource-intensive and impractical for large urban areas. While GIS techniques facilitate large-scale evaluations, perceived walkability still depends on costly, geographically limited questionnaire-based studies. Pairwise comparison datasets require extensive data collection, and training deep learning models via web research is costly and may overlook key perception factors. Pretrained models struggle to scale, and questionnaires remain constrained by geographic and financial limitations. Addressing these issues is essential for improving walkability assessments.

Street view images (SVIs) and computer vision have introduced a new paradigm for walkability measurement, enabling automated extraction of physical features and real-time assessment of perceived walkability. However, these approaches face limitations. Semantic segmentation accuracy is underexplored, and models often fail to generalize across cities with different walkability characteristics. These accuracy constraints underscore the need for more robust methodologies beyond current deep learning techniques.

To address these challenges, this paper introduces a framework that integrates deep learning models with geographic data for comprehensive urban walkability assessment. We compare content-based and similarity-based models for estimating walkability from SVIs and evaluate their alignment with human perception in a Greater Seattle case study. The framework incorporates a path-ranking method with slope penalties to optimize pedestrian routes. By combining automated, scalable analysis with perceptual insights, our approach enhances urban planning and accessibility.

2 Related Work

Our approach builds on existing deep learning models in computer vision and natural language processing to develop a novel walkability assessment system. In this section, we summarize key related works in image and text analysis for walkability evaluation and path construction.

In [10], a system enhancing route diversity for tourism was proposed. The system utilized publicly available data from FourSquare and Google Street View to generate diverse routes in cities like Kyoto and San Francisco. While innovative, it focuses on tourism rather than daily walkability. The Drop-and-Spin method [7] used 360-degree imagery for virtual neighborhood audits. This method offers a systematic way to assess environment features linked to health outcomes and hence evaluates the built environment’s impact on public health.

Recent studies have integrated advanced AI models to tackle urban walkability challenges. For instance, [5] used OpenAI’s CLIP to compare physical

and perceived walkability via street view imagery and text prompts and uncovered notable discrepancies. Using data from downtown Amsterdam, the study highlighted differences in walkability aspects and offered insights into pedestrian environments. Similarly, UrbanCLIP [8] employed contrastive pretraining on large-scale multimodal data, aligning textual and visual features to enhance urban area profiling, improving classification of diverse urban settings. In another study, [4] utilized street-view images and multimodal large language models to assess changes in the visual quality of urban street spaces. Their deep learning framework, capable of processing both visual and textual data, aimed to analyze and quantify streetscape transformations over time.

[11] developed a deep learning framework to assess walkability from Baidu Map Street View images in Shenzhen, China. The study shows that their Visual Walkability Index outperformed K-means and SVMs and revealed social inequalities. Similarly, [2] used a CNN to evaluate perceived and physical walkability across 196,624 images and suggested infrastructure improvements.

3 Methods

3.1 Deep Learning for Walkability Assessment

3.2 System Overview

The proposed framework consists of three phases: (1) data preprocessing, (2) walkability assessment, and (3) pedestrian path selection, as shown in Figure 1.

Phase 1 Data preprocessing: Street view images (SVIs) are obtained via the Google Street View API and processed using a CLIP model. The model filters duplicates based on cosine similarity and removes irrelevant images (e.g., indoor or non-street views) using text-based classification.

Phase 2 Walkability assessment: Filtered SVIs are fed into a walkability assessment model, which estimates walkability scores using deep learning and text prompts. These scores are added to the SVI metadata. Figure 1 outlines two deep learning approaches used in this phase, detailed in subsequent sections.

Phase 3 Pedestrian path selection and ranking: Walkability scores are mapped to the road network, and segment weights are computed. A path selection algorithm generates and ranks k paths ($k = 100$ in our experiments) between a given start and endpoint, incorporating geographic constraints via penalty functions. The final output consists of the top k pedestrian paths.

Similarity-Based Learning We used CLIP, an OpenAI model that links images and text via contrastive learning. Its zero-shot capability enables classification without task-specific training. Following the approach in [5], our method involves the following steps:

1. Input SVIs and text prompts into CLIP.
2. Adjust probabilities using positive/negative label entropy.
3. Compute walkability scores.

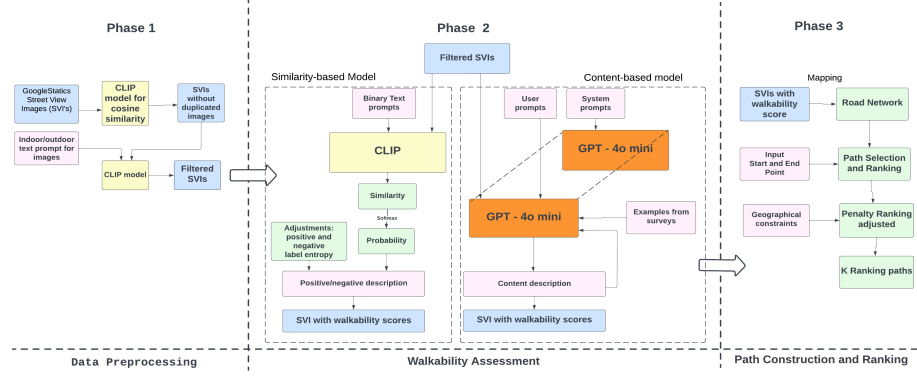


Fig. 1. An overview of system framework

CLIP was implemented using CLIPProcessor and CLIPModel libraries. It processes 25 prompts categorized into two types: safety and appeal. These prompts assess features like sidewalks, intersections, and greenery, each assigned a positive and negative score (e.g., “Paved sidewalk” vs. “No sidewalk”). Scores are weighted based on their significance and combined into an overall walkability score, scaled from 1 to 10.

Indicator weights were derived from literature [8] and survey data. Essential factors like sidewalk condition were weighted higher than amenities like shops, reflecting their greater impact on walkability. Our survey reinforced these priorities, emphasizing safety, sidewalk quality, and elevation over aesthetics. These weights were fine-tuned accordingly, with future refinements expected to enhance accuracy.

Content-based Learning We used GPT-4o Mini, a cost-efficient OpenAI model, to assess SVIs based on pedestrian safety and accessibility. Unlike CLIP, it employs guided prompts, including system and user messages, along with example images and corresponding walkability scores. Our experiments used 100 reference images representing scores from 1 to 10. The process involves the following steps:

1. Provide a system message outlining walkability evaluation criteria.
2. Supply selected example images with corresponding model responses.
3. input the SVI and a user query.
4. Retrieve image descriptions and walkability scores.

In addition to the system message and user query, the model is trained with a set of examples of (image, model response) pairs in range of walkability score 1 through 10. This process is applied across multiple images, improving response quality by processing SVIs in batches. The model outputs both description of the image and its walkability score. A selected example is shown in Figure 2.



Fig. 2. An example of user query response

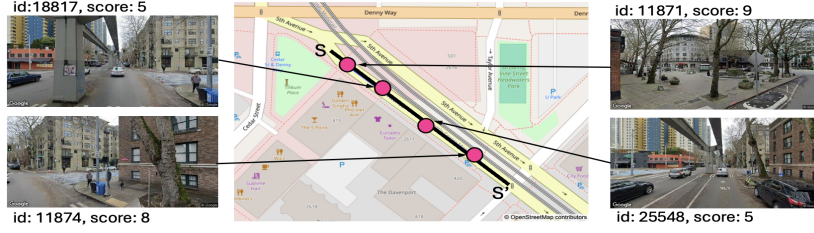


Fig. 3. An example of mapping walkability on road segments

3.3 Pedestrian Path Construction

Mapping Walkability Score to the Road Network Pedestrian paths are constructed using the OpenStreetMap Network (OSMnx) API, which provides geospatial data, street networks, and routing information. The OSM Network includes 11,659 nodes and 37,796 segments with coordinates, path lengths, and labels (e.g., highway, footway). Walkability scores for SVIs are mapped to this network based on their location. Each road segment's score is calculated as the average walkability score of the SVIs located between its nodes. For example, in Figure 3, the score for segment $\overline{ss'}$ is the average of four SVIs: $\overline{ss'} = \frac{9+5+5+8}{4} = 6.25$.

To identify the most walkable paths, we use a shortest loopless path method [9] with OSMnx. This method finds multiple shortest paths using Dijkstra's algorithm, ensuring no loops and generating distinct alternatives. Path scores are the sum of the segment scores along the path, adjusted by penalty functions based on geographic constraints.

Penalty Functions for Geographic Constraints Two penalty functions refine path rankings: (1) extra altitude gain/loss and (2) maximum slope. The altitude-based penalty function is defined as:

$$\Delta h(start, end) = \begin{cases} \sum_{a_{i+1} > a_i} (a_{i+1} - a_i), & \text{if } a_{end} \geq a_{start} \\ \sum_{a_{i+1} < a_i} (a_{i+1} - a_i), & \text{otherwise} \end{cases}$$

, where $\Delta h(start, end)$ is the gain (or loss) in altitude along the path from location $start$ to end , and the path consists of n number of segments, a_i and a_{i+1} are the altitude values of a road segment $\overline{x_i x_{i+1}}$. Then we have a penalty

function $P_{elevation}$ for the extra gain or loss in elevation:

$$P_{elevation} = \Delta h(start, end) - |a_{end} - a_{start}| \quad (1)$$

The second function that measures the maximum slope:

$$P_{slope} = \max\{\theta(i, i+1) | i = 1, 2, \dots, n\} \quad (2)$$

, where $\theta(i, i+1)$ represents the slope in degrees between locations x_i and x_{i+1} and $\theta(x_i, x_{i+1}) = |\arctan\left(\frac{a_i - a_{i+1}}{d(x_i - x_{i+1})}\right)|$. By combining (1) and (2), we have:

$$P(start, end) = \alpha P_{elevation} + \beta P_{slope} \quad (3)$$

, where α and β are scaling factor that adjust the impact of the extra gain (loss) in altitude and slope on the penalty.

4 Experiments

4.1 Datasets and System Setup

We collected 26,322 SVIs from six Seattle neighborhoods using the Google Street View API. Each 600x300 pixel image includes metadata such as location, pano ID, and capture date (median: July 2021). After filtering out 5,225 duplicate or non-street view images, 21,097 remained. Experiments were conducted using Python 3.11.7 with Jupyter Notebook 7.0.8.

Figure 4 (a) displays the six neighborhoods: Seattle Center, South Lake Union, Capitol Hill, Downtown, First Hill, and Chinatown-International District. Figures 4 (b) and (c) show images with low (red) and high (green) walkability scores, respectively.

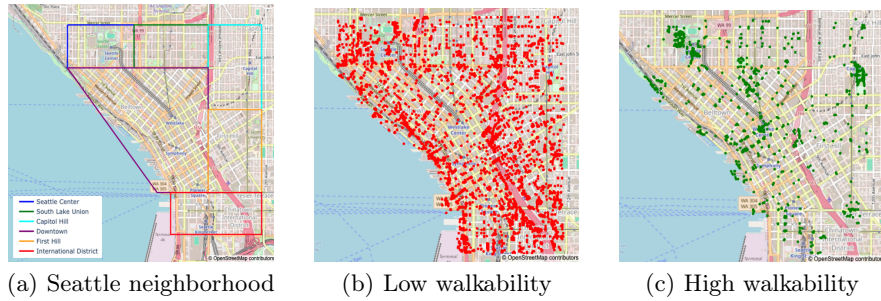




Fig. 4. Datasets and overall walkability trends assessed by the content-based moel

Table 1. Statistical comparison of walkability scores across six neighborhoods

Neighborhood	# SVIs	Content-based model				Similarity-based model			
		Mean	STD	Min	Max	Mean	STD	Min	Max
Seattle Center	2,056	5.26	2.62	1.00	10.00	6.34	1.17	2.44	9.72
South Lake Union	1,927	5.51	2.37	1.00	10.00	6.21	1.05	1.94	9.34
Capitol Hill	3,375	5.44	2.28	1.00	10.00	5.82	0.98	1.98	9.70
Downtown	7,658	5.42	2.54	1.00	10.00	5.91	1.10	1.23	10.00
First Hill	3,447	4.79	2.28	1.00	10.00	6.17	1.05	1.00	9.55
International District	2,634	4.92	2.44	1.00	10.00	5.81	1.12	1.18	9.92

Table 2. Similarity-Based Model vs. Content-Based Model (South Lake Union)

Lowest walkability by CLIP	Highest walkability by CLIP
 <p>CLIP: 1.94 GPT-4o mini: 2.0, "Construction site, barriers obstructing the view, limited pedestrian access, debris potential, urban environment, no sidewalks visible."</p>	 <p>CLIP: 9.34 GPT-4o mini: 2.0, "Indoors, exit signs present, no visible walking paths, enclosed space, unclear destination, limited visibility, potential safety concerns."</p>

4.2 Walkability Assessment Evaluation

Table 1 compares walkability scores across six neighborhoods. The similarity-based model produces higher mean scores (5.81–6.34) with lower variability (0.98–1.17) than the content-based model (mean: 4.79–5.51, SD: 2.28–2.62). Both models identified the least walkable neighborhood (International District) but differed on the most walkable. The similarity-based model shows lower variability, while the content-based model captures a wider range of walkability conditions.

Table 2 examines the lowest and highest walkable images in South Lake Union. Both models agree on the lowest walkable image, but the similarity-based model rated an image with an exit sign 9.34, while the content-based model scored it 2.0, revealing a significant disparity. Similar trends appear across other neighborhoods. The content-based model, which better captures walkability-related features, proves more effective in interpreting urban environments, highlighting the influence of model choice.

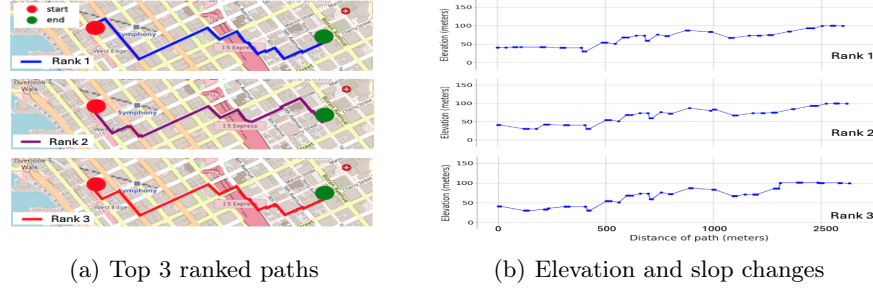
4.3 Pedestrian Path Evaluation

Table 3 shows how geographic penalties affect path rankings. While initial walkability scores reflect pedestrian appeal, penalties adjust these rankings. In both

Table 3. Analysis on the top 3 ranked paths

Model	Path	Initial score	Adjusted	Penalty	Travel time	Distance
Content-based model using GPT-4o mini	Rank 1	313.43	298.51	14.92	19.99	1,598.84
	Rank 2	293.85	279.33	14.52	20.42	1,633.31
	Rank 3	296.22	268.44	27.78	20.23	1,618.44
Similarity-based model using CLIP	Rank 1	341.62	319.30	22.32	19.94	1,594.98
	Rank 2	321.90	311.78	10.12	19.99	1,599.53
	Rank 3	330.08	298.76	31.32	19.85	1,588.00

Notes: travel time in minute and distance in meter

**Fig. 5.** Top 3 ranked paths based on walkability scores by the similarity-based model

models, paths initially ranked 2 had lower scores than those ranked 3 after penalties were applied. Figures 5 (a) and (b) show the top 3 ranked paths before and after adjustments. Although travel time and distance factor into path construction, they do not directly influence rankings. The adjusted scores reveal how penalties can lower perceived walkability, making some paths less attractive despite high initial scores, highlighting the importance of accounting for penalties in walkability assessments.





4.4 Comparisons with Perceived Walkability

We conducted two surveys to evaluate perceived walkability and validate our models: (1) *Survey A*: 50 participants compared 20 pairs of SVIs and selected the more walkable image and (2) *Survey B*: 35 participants rated 100 sampled images based on known walkability factors.

In Survey A, participants compared image walkability using a binary scoring system. Table 4 shows two selected questions alongside model scores. In Q1, both models aligned with survey results, where 49 participants favored option 2. In Q5, the similarity-based model rated option 1 higher, while the content-based model matched 35 participants in selecting option 2. These findings highlight the need to consider additional factors influencing perceived walkability.

Figure 6 presents Survey B compared with walkability score distributions of the two models on 100 images. The content-based model shows greater variability, while the similarity-based model has a narrower range. These trends align

Table 4. Model comparisons with perceived walkability

	Method	Option 1	Option 2
Q1			
	Survey A	1 selected option 1	49 selected option 2
	Content-based model	8	9
	Similarity-based model	5.52	7.44
Q5			
	Survey A	15 selected option 1	35 selected option 2
	Content-based model	3	8
	Similarity-based model	5.84	5.42

with Section 4.2. Figure 6 (a) shows a slight right-skew in distribution, while Figure 6 (b) highlights the survey’s higher mean score (6.21) compared to the content-based (4.68) and similarity-based models (5.91). Despite this, the survey’s variability mirrors the content-based model, with several outliers.

5 Conclusion

This study proposed a deep learning framework using Google SVIs for walkability assessment. We compared two models, integrating walkability scores with geographic data to recommend pedestrian-friendly routes. A case study in Seattle showed the content-based model aligned better with human perception than the similarity-based approach. Challenges remain in improving accuracy, expanding to diverse urban contexts, and refining subjective perception datasets. Increasing survey participants will enhance model validation. Future work should improve model precision, broaden geographic analysis, and incorporate perceptual factors for better urban planning and route optimization.

References

1. Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., Saelens, B.E.: Linking objectively measured physical activity with objectively measured urban form: findings from smartraq. *American journal of preventive medicine* **28**(2), 117–125 (2005)
2. Kang, Y., Kim, J., Park, J., Lee, J.: Assessment of perceived and physical walkability using street view images and deep learning technology. *ISPRS International Journal of Geo-Information* **12**(5), 186 (2023)

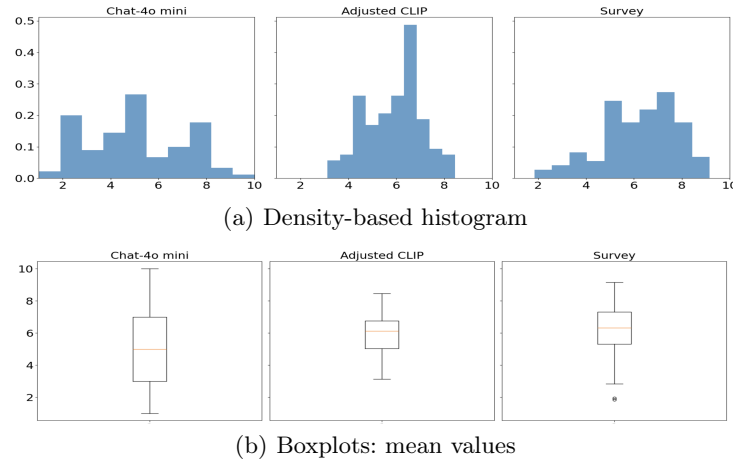


Fig. 6. Survey C: Walkability score distribution

3. Lee, E., Dean, J.: Perceptions of walkability and determinants of walking behaviour among urban seniors in toronto, canada. *Journal of transport & health* **9**, 309–320 (2018)
4. Liang, H., Zhang, J., Li, Y., Wang, B., Huang, J.: Automatic estimation for visual quality changes of street space via street-view images and multimodal large language models. *IEEE Access* (2024)
5. Liu, X., Haworth, J., Wang, M.: A new approach to assessing perceived walkability: Combining street view imagery with multimodal contrastive learning model. In: *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications*. pp. 16–21 (2023)
6. Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., Inoue, S.: Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of google street view images. *Health & Place* **66**, 102428 (2020)
7. Plascak, J.J., Rundle, A.G., Babel, R.A., Llanos, A.A., LaBelle, C.M., Stroup, A.M., Mooney, S.J.: Drop-and-spin virtual neighborhood auditing: assessing built environment for linkage to health studies. *American journal of preventive medicine* **58**(1), 152–160 (2020)
8. Yan, Y., Wen, H., Zhong, S., Chen, W., Chen, H., Wen, Q., Zimmermann, R., Liang, Y.: Urbanclip: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web. In: *Proceedings of the ACM on Web Conference 2024*. pp. 4006–4017 (2024)
9. Yen, J.Y.: Finding the k shortest loopless paths in a network. *management Science* **17**(11), 712–716 (1971)
10. Zhang, Y., Siriaraya, P., Wang, Y., Wakamiya, S., Kawai, Y., Jatowt, A.: Walking down a different path: route recommendation based on visual and facility based diversity. In: *Companion Proceedings of the The Web Conference 2018*. pp. 171–174 (2018)
11. Zhou, H., He, S., Cai, Y., Wang, M., Su, S.: Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable cities and society* **50**, 101605 (2019)