Seasonal Perception and Sentiment in Urban Scenes

A multimodal time-series study using social media data

Yan Lan[1], Jinghan Liu[2], Zhengyang Li[3], Han Tu[4]
[1]Tongji University [2, 3, 4]Independent Researcher
[1]lany082801@gmail.com [2]GINLJH@163.com [3]lzyqwe2500@outlook.com
[4]hantu@alum.mit.edu

Street scene data has been widely used to assess urban visual characteristics, but its low temporal resolution limits the exploration of dynamic human perception. To address this, this study proposes a dual-path multimodal analytical framework leveraging the real-time and multimodal nature of social media images. Focusing on Beijing's Second Ring Road, Weibo data from four representative months in 2023 were collected for analysis. At the scene level, the BLIP model was used to generate image captions, followed by K-Means clustering and manual semantic alignment to derive nine scene categories, later grouped into three macro types. This process revealed seasonal shifts in public spatial perception. At the object-sentiment level, YOLO extracted visual objects, while SNLP quantified sentiment from accompanying texts. Statistical and modeling approaches, including SHAP-based interpretation, were used to explore seasonal object-sentiment relationships. Findings show that natural scenes exhibit strong seasonal variation, while architectural and cultural scenes remain more stable. Vegetation and buildings consistently evoke positive sentiment, whereas elements like water bodies, animals, entertainment, and furniture show more seasonal fluctuation. The study demonstrates the utility of social media–based methods for spatiotemporal perception analysis and offers a novel approach to understanding the emotional dimensions of urban experience.

Keywords: *Urban Perception, Seasonal Fluctuation, Social Media Data, Multimodal Analysis, Image Scene Semantics, Sentiment Reactions.*

INTRODUCTION

In recent years, the rapid proliferation of social media platforms has profoundly reshaped how people perceive, experience, and express urban environments. The exponential growth of geotagged content has enabled platforms such as Weibo, Douyin, and Tencent Street View to emerge as important media channels for capturing the emotional dimensions of urban life . These platforms reflect two primary dimensions of urban emotional responses: on the one hand, perceptual responses derived from images, such as spatial impressions formed by building morphology, color saturation, and green coverage; on the other hand, subjective emotional expressions conveyed through text, such as user narratives about urban safety, comfort, or cultural belonging (Yao et al., 2019; Pipitone et al, 2022). For example, Yao et al. (2019) introduced the DeepSentiStreet framework, which leverages deep learning techniques to extract visual features from street-level imagery and generate spatialized sentiment maps. Pipitone et al (2022) combined photographs of neighborhood natural features taken by participants with their narrative descriptions, conducting both quantitative and qualitative analyses of the visual and textual content in conjunction with geospatial data, found a positive correlation between greening rates and residents' sense of security and belonging.

Although multimodal data acquisition has become increasingly accessible, current research on multimodal urban sentiment remains limited. On the one hand, existing studies have largely focused on unimodal modeling, such as visual recognition of images or semantic analysis of text (Oliveira et al., 2020), thereby neglecting potential mismatches between spatial perception and subjective expression. On the other hand, existing models show limited adaptability across spatial and temporal contexts. Daniel Nüst and Edzer Pebesma (2020) pointed out that spatial heterogeneity and geographic dependence constrain the replicability of GeoAI models across different cities, while Kedron et al. (2019) emphasized the lack of generalizability across temporal scales in current geospatial analytical methods.

Meanwhile, public preferences for urban spaces are not static but exhibit significant variations across seasons. From a spatial perception perspective, seasonal environmental factors—such as city colors, vegetation condition, and street activity—can significantly shape people's preferences for specific types of urban spaces (Geng et al. 2021; Peng et al., 2023). Geng et al. (2021) found that natural landscapes within parks were more satisfying to visitors in winter and autumn, and that human activities within parks also played an important positive role in visitor satisfaction during winter. Liang et al. (2022) also found that parks in Shanghai provided visitors with higher levels of happiness during the Spring Festival and autumn. Peng et al. (2023) further revealed the seasonal association between urban parks and activities, finding that areas with higher levels of shade attract more people in summer but deter visitor activities in winter.

Although prior research has highlighted the influence of seasonal changes on urban space usage, there remains a significant gap in multimodal analysis. Most existing studies have focused either on detecting changes in greenery through image recognition or on capturing mobility patterns through location-based social media data, making it difficult to reveal the mechanisms underlying both perceptual and emotional responses to spatial features. Consequently, potential links between "seasonal public preferences " and "seasonal spatial characteristics" remain largely unexplored.

METHODOLOGY

Data Construction

The data used in this study were sourced from the Chinese social media platform Weibo. The filtering criteria included publicly accessible posts published within the Beijing metropolitan area throughout the year 2023, which contained both geo-tagged location data and multimodal content (i.e., images and text). To ensure seasonal coverage, four representative months—April, July, October, and December, corresponding to spring, summer, autumn, and winter—were selected as analytical nodes. A total of 212,350 publicly accessible Weibo posts meeting the criteria were retrieved for analysis.

To enhance spatial consistency, we first excluded all check-in posts located outside of Beijing's Second Ring Road, focusing specifically on the urban core and its associated streetscape perceptions. Next, by combining an image classification model with manual screening, we removed image content irrelevant to urban spatial environments, such as selfies, food photography, or posters. Simultaneously, the textual content used for sentiment analysis was filtered to exclude posts located indoors or those whose narratives deviated from urban spatial contexts. This step ensured the retention of image-text pairs directly related to urban public spaces and streetscapes. However, the reliance on social media data introduces inherent sampling biases due to user behavior, platform usage patterns, and demographic disparities.

Image scene-level characteristics semantic recognition method

To achieve the scene-level characteristics of public perception of urban space through social media images, this study addresses the limitations of traditional visual recognition models when applied to non-standardized and complex image content. Therefore, we employed the BLIP (Bootstrapped Language Image Pretraining) model to generate text descriptions for the social media images of four seasons, which transforms the difficult-to-interpret visual information of the images into semantically decipherable text descriptions. And then a K-Means clustering algorithm with a fixed cluster number of six was applied to the image captions for each month to identify major semantic groupings. To assist in evaluating the coherence and separability of clusters, we further applied t-SNE for dimensionality reduction and two-dimensional visualization (see Figure 1).

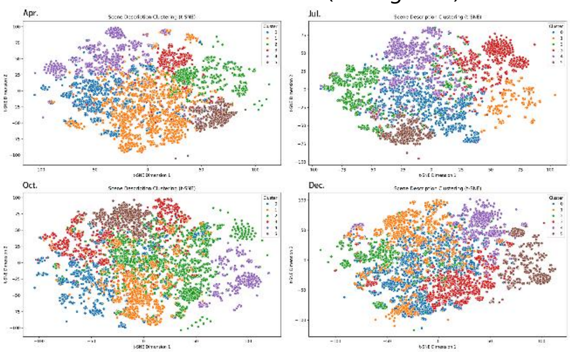

Figure 1
t-SNE visualization of scene description clusters for four seasons

Then, in order to realize the comparison of scene features on spatio-temporal sequences, we performed a two-step generalization of the clustering labels:

1. Temporal alignment: A total of 24 clustering labels for the four seasons were synthesized to manually semantically align the representative descriptions and keywords of the clustered images, and finally 9 categories of unified scene labels were generalized: historic building, street environment, tourist landmark, urban waterfront, natural flower view, human subject scene, religious building, seasonal tree, and snow scene.
2. Spatial generalization: We integrate the 9 categories of semantic labels into three macro-scenes: architectural (including historic building, street environment, religious building), natural (including natural flower view, seasonal tree, urban waterfront and tree snow scene), and humanities (including tourist landmark, human subject scene).

Utilizing the geo-location metadata embedded in the images, we visualized the spatial distribution of the three macro scene types using the Kepler.gl platform, mapping them onto the urban coordinate system to explore their seasonal spatial patterns across the city.

Text Sentiment Reaction Analysis

This study adopts a two-stage approach to analyze user-generated content on the Weibo platform, focusing on discussions related to urban space and their evaluative tendencies.

In the preprocessing phase, Weibo posts from four representative months—April, July, October, and December—were manually filtered using location metadata to exclude irrelevant content. We then applied semantic vectorization using the embedding layer of a pre-trained Transformer model, which maps linguistic units into a continuous high-dimensional space based on semantic proximity (Vaswani et al., 2017). To visualize these clusters and support interpretation, we used the UMAP algorithm for dimensionality reduction, which effectively preserves both local and global semantic structures (McInnes et al., 2018).

For sentiment analysis, we employed the SnowNLP library, which applies a Naive Bayes classifier to assign a polarity score ranging from 0 (negative) to 1 (positive) to each post. Widely used in Chinese social media research, SnowNLP handles informal expressions and mixed emotions with reasonable accuracy (Zhang et al., 2020). The resulting sentiment scores were integrated into a spatial-temporal framework to examine seasonal differences in public perceptions of urban spaces.

Although SnowNLP was selected for its efficiency and ease of deployment, we also tested BERT-based models (e.g., Chinese-BERT, ERNIE, RoBERTa, Mengzi) on a manually labeled subset. While models like ERNIE showed ~3% higher accuracy, their heavy computational demands made them less practical for large-scale analysis. Future work will explore lightweight alternatives such as distilled BERT to enhance semantic sensitivity without compromising scalability.

Analysis of the Relationship between Object level Features and Sentiment Reaction

This study constructs object-level semantic features of images based on YOLOv5 and combines the emotional values of Weibo images to conduct multi-layered analyses, including correlation analysis and variance analysis. The aim is to explore the relationship between urban semantic elements and emotional perception. The study covers four seasonal periods—spring, summer, autumn, and winter—focusing on revealing how this perception mechanism varies across seasons. Additionally, we trained emotion prediction models for each of the four seasons and introduced SHAP methods to analyze the contribution of different categories in the model, strengthening the mutual validation between statistical analysis and model-based interpretation.

To adapt to the distribution of city street scene images posted by social media users, we used an image segmentation algorithm suitable for street object recognition to preprocess all sample images. The labeling was manually adjusted to fine-tune the YOLOv5 model. A total of 22 target categories were defined, covering typical visible elements in urban spaces, such as Building, Vegetation, and Waterbody. The model achieved a mAP@0.5 of 0.229 on the validation set, with stable recognition performance for mainstream categories. However, a few categories exhibited weaker recognition due to limited sample size.

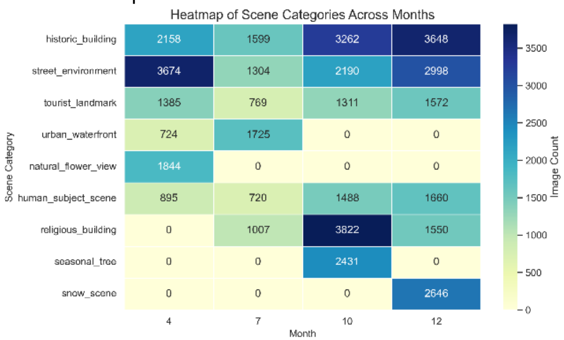

Figure 2
Heatmap showing the monthly distribution of nine scene categories

The fine-tuned YOLOv5 model was used to detect objects in 46,293 images from four seasonal periods, extracting 22-dimensional object frequency vectors. Each row represents the detection results for one image, with undetected categories assigned a value of zero. The data was then matched with emotional scores extracted from the corresponding Weibo texts, forming a structured feature set for multi-seasonal urban perception analysis.

This study employs the following three analysis methods to investigate the relationship between object-level features and emotional perception:

Correlation Analysis: Pearson correlation coefficients were used to measure the linear relationships between various semantic objects in the images and their emotional values.

Analysis of Variance (ANOVA): Single-factor and two-factor variance analyses were applied to examine the statistical significance of different object categories and their combinations on emotional scores, with a focus on seasonal variations.

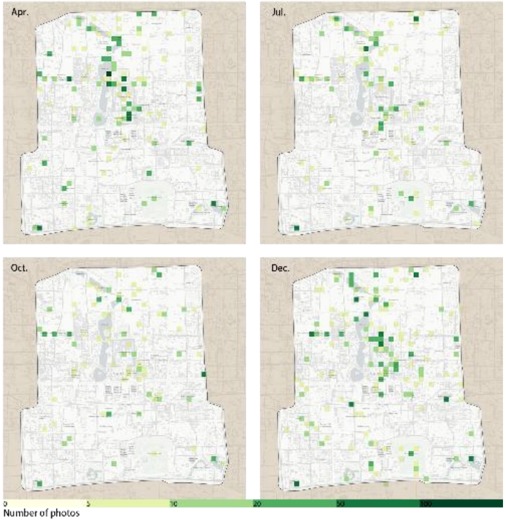

Figure 3
Seasonal spatial distribution maps of natural scene perceptions

Emotion Prediction Model and SHAP Analysis: We built a binary emotion prediction model (high/low emotion) based on object features and used SHAP to assess the contribution of each category to the model's predictions, highlighting the influence of semantic objects on emotional responses.

RESULTS & ANALYSIS

Seasonal Variation of Perceived Urban Scenes

As shown in the heatmap (Fig. 2), public perception of the nine urban scene categories exhibits clear seasonal variation across the four selected months. Two built environment types—historic buildings and street environments—remain consistently prominent throughout the year, attracting particularly strong attention outside of summer. Tourist landmarks and human subject scenes also show relative seasonal stability, with only a slight dip in summer. This may be due to high temperatures discouraging outdoor activities, leading people to favor indoor environments during this period.

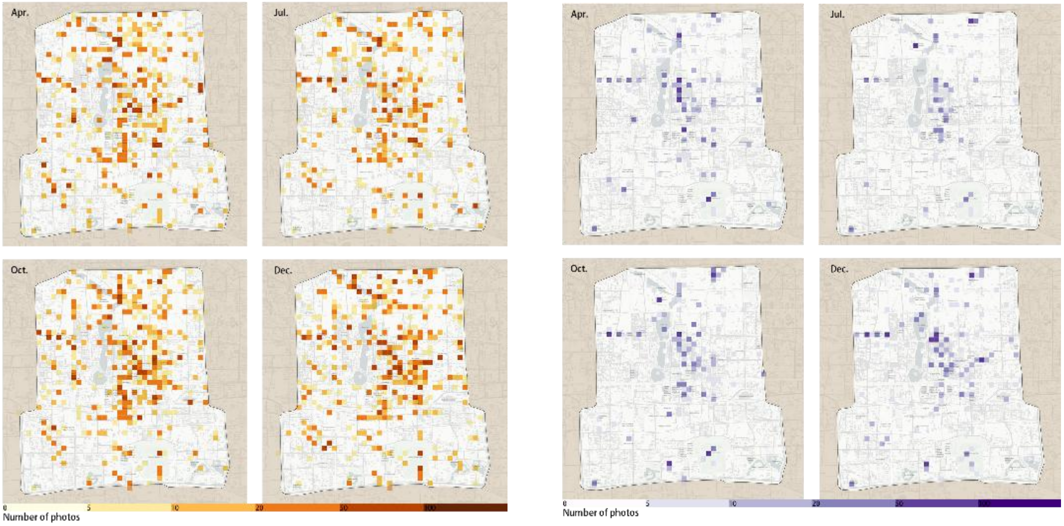

Figure 4
left to right, Seasonal spatial distribution maps of humanistic scene perceptions; Seasonal spatial distribution maps of architectural scene perceptions

In contrast, scene types associated with natural environments demonstrate more pronounced seasonal variation. For instance, seasonal trees and snow scenes are typical seasonal features that appear only in autumn and winter, respectively. Natural flower views are concentrated in spring, while urban waterfronts dominate summer, making them the most prominent scene type during that season. This may reflect a general public preference for water-related environments in hot weather.

It is worth noting that religious buildings show a significant peak in October. This can be attributed to the location of the study area—the Second Ring Road in Beijing—which includes Tiananmen, a highly symbolic national site. The spike in attention coincides with China's National Day on October 1st, when people tend to share photos of the national flag and Tiananmen. Unlike seasonal changes in the physical environment, this surge reflects a temporally specific social and cultural event, underscoring the sentimental attachment to symbolic urban spaces. These findings suggest that public perception of urban scenes is shaped not only by seasonal environmental features, but also by temporally anchored sentimental and cultural recognition.

Spatiotemporal Distribution of Perceived Urban Scenes

As shown in Fig. 3, natural scenes also display seasonal fluctuation in their spatial distribution, while the distribution of architectural and human-centered scenes remains more stable, with relatively minor seasonal variation. In spring and winter, public attention to natural scenes is concentrated in well-known open green spaces along Beijing's central axis, such as Shichahai Park, Jingshan Park, Baita Temple, and southern sites like Taoranting Park and Temple of Heaven. In summer and autumn, natural scene perception becomes more scattered: summer shows localized attention near waterfronts, while autumn reveals no significant hotspots. This may be explained by the presence of flower blooms and snowfall during spring and winter, which tend to trigger collective public attention, whereas in summer and autumn, more uniform weather and landscape conditions lead to a dispersed perception pattern.

In comparison, the perception of architectural scenes (see Fig. 4) remains consistently focused on areas such as the Forbidden City, the central axis, and the Temple of Heaven across all four seasons. Human-centered scenes (see Fig. 4) similarly show strong spatial consistency, concentrating around Tiananmen Square, Nanluoguxiang, Wangfujing, and Shichahai—all major tourist and cultural destinations. These spatial patterns may result from the strong influence of tourism and symbolic cultural identity, which are less affected by seasonal changes.

Semantic Object Features and Seasonal Patterns of Sentiment Response

Our study analyzes the seasonal relationship between semantic objects detected by YOLO and sentiment scores extracted from social media posts, aiming to understand how different visual elements in urban scenes influence sentimental responses. This includes both correlation and variance analysis, as well as a heatmap visualization showing object-sentiment distributions across seasons. (Fig. 5)

Spring: Categories such as Building, Vegetation, and Dining Furniture show strong positive correlations with sentiment, suggesting that scenes combining openness, greenery, and seating tend to evoke positive sentiments. In contrast, Light Vehicles and Vehicles show negative associations, indicating a dampening effect on mood.

Summer: The positive influence of Vegetation and Sidewalk is even more pronounced, highlighting the importance of shaded, walkable environments during hot seasons. While Waterbody contributes positively, traffic-related objects continue to exhibit negative effects.

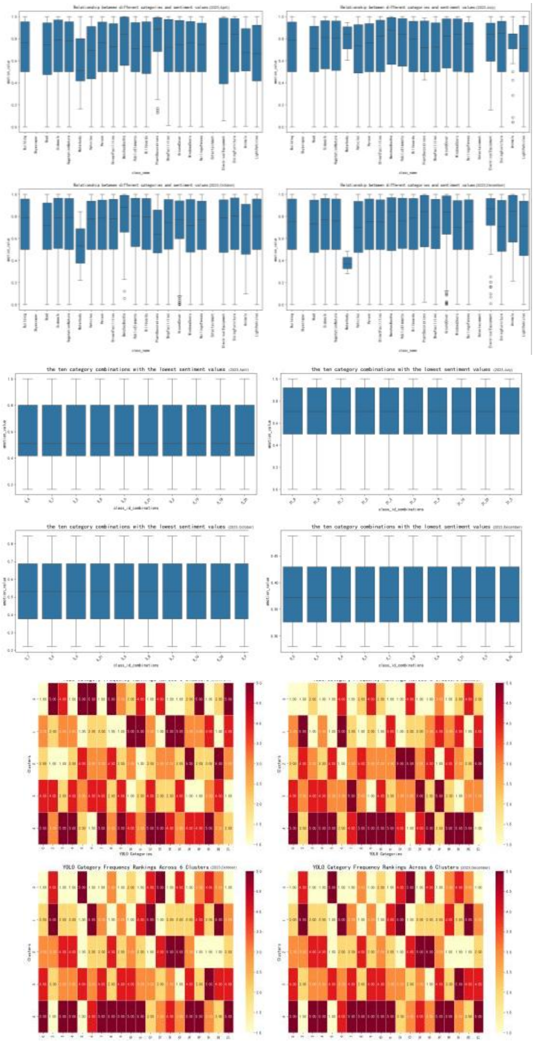

Figure 5
top to bottom
Seasonal relationship between individual object categories and sentiment values;
object category combinations with the highest sentiment values across seasons;
heatmap of object-sentiment distributions across seasons

Autumn: Building and Vegetation maintain their sentimental consistency, while Waterbody sees a notable drop in sentimental value. Combinations like Waterbody + Person or Waterbody + Building tend to be linked with feelings of emptiness or coldness, suggesting a seasonal shift in the perception of static natural elements.

Winter: Ground Cover emerges as the most positively associated category, underscoring the sentimental value of tactile and close-contact surfaces in cold seasons. Animals and Entertainment elements also show increased sentimental appeal, while Waterbody nearly loses its positive influence altogether.

Sentiment Prediction Modeling and Feature Attribution

To examine how object composition influences sentiment perception, we built seasonal binary classifiers using frequencies of 22 YOLO-detected object categories. Sentiment scores were binarized by median split. To interpret model behavior, we applied SHAP (Shapley Additive Explanations), which reveals each feature's marginal contribution to predictions, capturing non-linear effects beyond traditional correlations or ANOVA.

While overall model accuracy was moderate, consistent explanatory patterns supported the viability of object-level sentiment modeling. Notably, the December model performed best (accuracy 57.78%, recall 0.78, F1 0.69 for positive class), suggesting stronger affective signals in winter scenes.

SHAP summary plots (Fig. 6) showed that Buildings, Vegetation, and Street Facilities were positively associated with sentiment across seasons. Spring emphasized greenery and architecture; summer highlighted people and plants; autumn featured signs and retail fronts; winter emphasized ground elements, animals, and entertainment facilities.

Interestingly, features like Animals and Retail-related objects—previously showing weak statistical correlation—emerged as influential in the SHAP framework, underscoring the model's capacity to capture complex emotional cues through non-linear feature interactions.

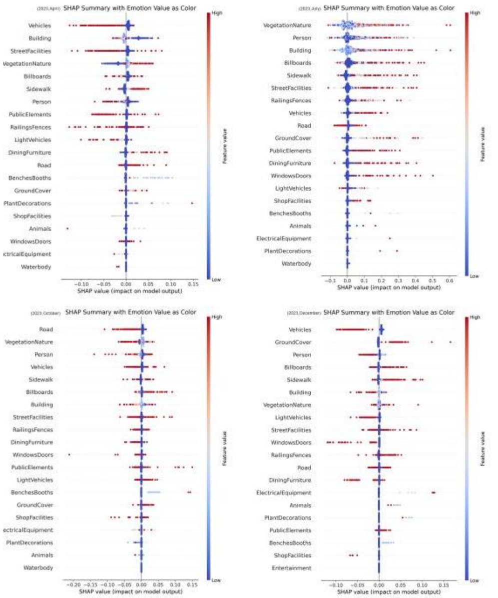

Figure 6
SHAP Summary Plots Showing Feature Contributions to Sentiment Prediction Across Four Seasons

CONCLUSION

Through semantic captioning and clustering of social media images, this study identified that public perception of urban natural scenes exhibits strong alignment with seasonal environmental changes. On the one hand, this finding confirms the effectiveness of our semantic extraction and clustering methodology for analyzing large-scale social media imagery; on the other hand, it underscores the significant influence of seasonal variations in the natural environment on public spatial attention. In contrast, built environment scenes, including architectural and landmark-oriented categories, demonstrated greater spatiotemporal stability, showing less sensitivity to seasonal physical changes but heightened responsiveness to sociocultural and symbolic factors. This suggests that public perception of urban scenes is shaped not only by spatiotemporal attributes but also by sentimentally driven cultural associations.

Our dual-track analysis—integrating object recognition in images with sentiment analysis of accompanying text—revealed pronounced seasonal mechanisms in object-sentiment relationships. Two key patterns emerged:

1. Persistent positive sentimental influence of architectural and vegetative objects: These object categories consistently ranked high in SHAP feature importance across seasonal models, indicating their robust role in enhancing affective perception. Such "stable-background" elements appear to constitute the foundational layer of urban spatial perception. These results validate the capacity of our object-level analytical framework to effectively capture and model the sentimental impact of urban form on public sentiment.
2. Season-specific sentimental modulation by water, animals, entertainment, and furniture-related objects: For instance, water bodies displayed strong affective activation in summer but exhibited inverse sentimental effects in winter, functioning as negative triggers due to perceived inaccessibility or discomfort. Similarly, objects like animals, entertainment facilities, and seating furniture were found to elevate sentimental scores in winter, suggesting the affective significance of thermal comfort and social interaction under low-temperature conditions. These seasonal dynamics demonstrate the context-dependent sentimental valence of urban features, revealing their divergent functions as affective stimulators or suppressors across different climatic contexts.

Based on these findings, we propose a dual-path framework to interpret urban sentiment: one path based on semantic scene perception, the other on object-level emotional cues. This framework not only contributes to urban perception theory, but also offers practical value for public space planning—for instance, by informing context-sensitive interventions like winter-appropriate infrastructure or enhanced summer water access.

Several limitations remain. Manual semantic alignment may limit reproducibility; the temporal design lacks continuity for fine-grained analysis; and the dataset, while large, is context-specific to one city and platform. Lastly, since the study focuses solely on Beijing, future research should explore the applicability of this framework across different cultural, climatic, and geographic settings to evaluate its broader robustness.

To address these limitations, future research could advance in several practical directions. First, to improve the reproducibility and adaptability of the methodology, semi-automated approaches to semantic alignment could be explored—such as clustering based on sentence embeddings or utilizing large language models to assist in label harmonization—reducing the need for manual intervention. Second, enhancing temporal continuity by shifting from seasonal snapshots to finer temporal resolutions, such as monthly or bi-weekly intervals, may help capture more nuanced transitions and short-term emotional fluctuations. Third, applying the proposed framework to a broader range of urban regions and social media platforms could help test its generalizability. In addition, incorporating user demographic information and contextual variables—such as local events, holidays, or policy shifts—may reveal how these factors interact with seasonal changes to influence public sentiment in subtle ways, thereby supporting the development of a more comprehensive model of urban emotional dynamics.

REFERENCES

Geng, D. C., Innes, J. L., Wu, W., Wang, W. and Wang, G. (2021) 'Seasonal variation in visitor satisfaction and its management implications in Banff National Park', *Sustainability*, 13(4), 1681. doi: 10.3390/su13041681

Kedron, P., Trgovac, A., Frazier, A. E., Nelson, T., and Fotheringham, A. S. (2019) 'Reproducibility and replicability in geographical analysis', *Geographical Analysis*, 53(3), pp. 304–318. https://doi.org/10.1111/gean.12221

Li, J., Li, D., Xiong, C. and Hoi, S. C. H. (2022) 'BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation', in *Proceedings of the 39th International Conference on Machine Learning (ICML)*. Baltimore MD. Available at: https://doi.org/10.48550/arXiv.2.201.12086

Liang, H., Yan, Q., Yan, Y., Zhang, L. and Zhang, Q. (2022) 'Spatiotemporal study of park sentiments at metropolitan scale using multiple social media data', *Land*, 11(9), 1497. doi: 10.3390/land11091497

MacQueen, J. (1967) 'Some methods for classification and analysis of multivariate observations' in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 5(1), pp. 281–297. Available at: http://projecteuclid.org/euclid.bsmsp/1200512992

McInnes, L., Healy, J. and Melville, J. (2018) *UMAP: Uniform manifold approximation and projection for dimension reduction*. [online]. [preprint]. Available at: https://arxiv.org/abs/1802.03426

Nüst, D. and Pebesma, E. (2020) 'Practical reproducibility in geography and geosciences', *Annals of the American Association of Geographers*, 110(5), pp. 1300-1310. https://doi.org/10.1080/24694452.2020.1806028

Oliveira, W. B. D., Dorini, L. B., Minetto, R. and Silva, T. H. (2020) 'Outdoorsent: Sentiment analysis of urban outdoor images by using semantic and deep features', *ACM Transactions on Information Systems (TOIS)*, 38(3), pp. 1-28. doi: 10.1145/3385186

Peng, Y., Gan, D., Cai, Z., Xiao, M., Shu, D., Wu, C., Yu, X. and Li, X. (2023) 'Landscape features impact the spatial heterogeneity of visitation density within a comprehensive park: what are the seasonal and diurnal variations?', *Forests*, 14(8), 1627. Available at: https://doi.org/10.3390/f14081627

Pipitone, J. M. and Jović, S. (2022) 'Urban nature and sense of belonging: Photo-narrative exploration of socio-spatial disparities in New York City', *Ecopsychology*, 14(4), pp. 235-252. doi: 10.1089/eco.2021.0056

Reimers, N. and Gurevych, I. (2019) 'Sentence-BERT: Sentence embeddings using Siamese BERT-networks', in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pp. 3982–3992. https://doi.org/10.18653/v1/D19-1410

Uber Technologies Inc. (2018) *Kepler.gl: Geospatial analysis tool for large-scale data visualization*. [Online] Available at: https://github.com/keplergl/kepler.gl [Accessed 1 Apr. 2025].

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017) 'Attention is all you need', *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 30, pp. 6000-6010. Available at: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., et al. (2019) 'A human-machine adversarial scoring framework for urban perception assessment using street-view images', *International Journal of Geographical Information Science*, 33(12), pp. 2363-2384. Available at: https://doi.org/10.1080/13658816.2019.1643024