SENSING THE CITY

Leveraging geotagged social media posts and street view imagery to model urban streetscapes using deep neural networks

JAYEDI AMAN[1*], TIMOTHY C MATISZIW[2*], JONG BUM KIM[3] and DAN LUO[4]
[1,2,3]University of Missouri Columbia
[4]The University of Queensland
[1]jayediaman@mail.missouri.edu, 0000-0002-6128-8293
[2]matisziwt@missouri.edu, 0000-0003-2159-4904
[3]kimjongb@missouri.edu, 0000-0001-8762-9554
[4]d.luo@uq.edu.au, 0000-0003-1760-0451
[*]Contributed equally

Abstract. Understanding the relationships between individuals and the urban streetscape is an essential component of sustainable city planning. However, analysis of these relationships involves accounting for a complex mix of human behaviour, perception, as well as geospatial context. In this context, a comprehensive framework for predicting preferred streetscape characteristics utilizing deep learning and geospatial techniques is proposed. Geotagged social media posts and street view imagery are employed to account for individual sentiment and geospatial context. Natural Language Processing (NLP) and computer vision (CV) are then used to infer sentiment and model the visual environment within which individuals make posts to social media. An application of the developed framework is provided using Instagram posts and Google Street View imagery of the urban environment. A spatial analysis is conducted to assess the extent to which urban attributes correlate with the sentiment of social media postings. The results shed light on sustainable streetscape planning by focusing on the relationship between users and the built environment in a complex urban setting. Finally, limitations of the developed methodology as well as future directions are discussed.

Keywords. Urban Sustainability; Data Mining; Pedestrian Sentiments; Transportation Behaviour, Street Level Imagery; Transformers; SDG 11.

1. Introduction

Two-thirds of the world's population, 6.5 billion people will live in urbanized areas by 2050. As communities have become more global, dynamic, and transitory, sustainable urban development can be difficult to achieve unless the way cities are planned and managed changes dramatically. Urban streetscapes are thought to play an essential role in influencing social interactions (Salesses et al., 2013). The design, condition, and spatial configuration of urban features such as buildings, traffic infrastructure, and parks can trigger a variety of collective and individual human experiences in this sense. Thus, urban planning and regulatory agencies are investing extensively in enhancing the condition of the streets in order to improve how they are experienced by individuals. To this end, it is increasingly recognized that the linkage between individuals' sentiment and the location(s) associated with sentiment is an important consideration. The spatial context explicitly enables and stimulates human activity. In return, these activities impact the form and perception of streetscapes (Batty et al., 2010). As a result, assessing people's perceptions of urban streetscapes could provide important insights on street semantics, leading to more informed community planning decisions.

Various approaches for gauging perception of streetscapes in urban studies have been proposed. Traditionally, data on humanistic experience of urban environments have been by way of field surveys, interviews, and questionnaires (Montello et al., 2003). Such modes of data collection often entail complex research protocols to ensure use/applicability over wide geographic areas or time periods. Smart city technologies such as drones, street view imagery and social media data have recently facilitated the ability to sense the urban streetscape in the field of urban sustainability research (Ilieva and McPhearson, 2018). In particular, Guerrero et al. (2016) explored how social media posts can be used to learn about how people interact with and perceive urban green spaces, and how this can be used to inform future sustainable infrastructure planning. More recently, Marti et al. (2017) used social media posts and urban cartographies to show what features may make open public spaces, such as urban plazas, more successful than others.

These newer approaches provide information on a variety of facets of urban processes and the dynamics thereof. In order to make full use of more robust streams of data though, there is a need to identify mechanisms for linking them with indicators of human perception of urban streetscapes. Furthermore, detailed accounts of the use of geolocated spatiotemporal social media data and street view imagery for urban perception analysis are relatively rare. To this end, a novel analysis framework is proposed to better integrate individual indicators of urban sentiment with geospatial context in predicting preferred streetscape and identifying linkages between streetscape features and individual sentiment. Open-source geospatial datasets such as images of urban features (Google Street View), roads, and public social media (Instagram) postings are analysed using deep learning algorithms to quantify individual sentiment relative to geospatial context. The developed framework is applied to a medium sized urban area to demonstrate its applicability.

2. Analytical Framework

The proposed analytical framework is depicted in Figure 1. In Phase 1, information that can serve as indicators of individual sentiment is collected and processed. Social media postings are a rich source of data in this respect given that they can contain texture and tone reflective of an individual's sentiment. In many instances, social media postings also contain indicators of the geographic location (latitude, longitude) as well as temporal references (e.g., date/time stamps). Once relevant posts have been collected, they can be processed by a natural language processing (NLP) algorithm, such as a 'text categorization' technique, to infer the nature of the poster's sentiment. Hot spot analyses can then be performed to assess the extent to which spatial dependency among posts along streetscapes may exist.

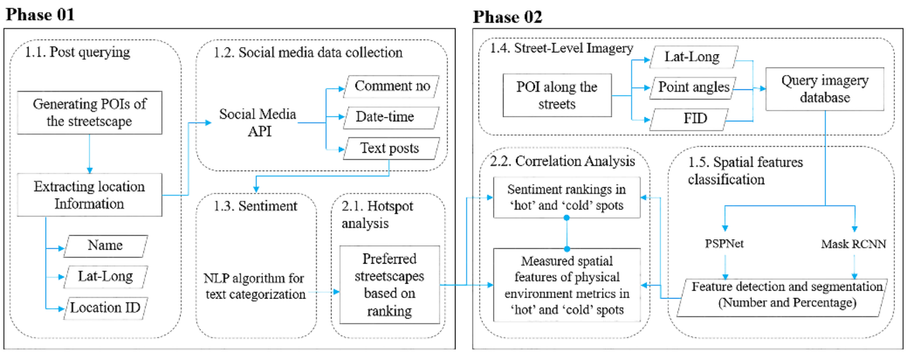

Figure 1. Overview of the methodology

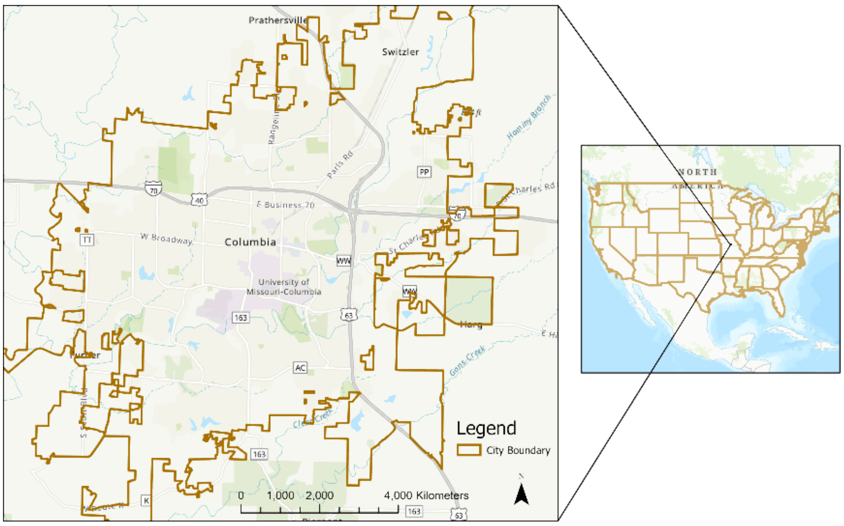

Figure 2. Study Region

In Phase 2, metrics of urban form such as enclosure, human dimension, and complexity in determining sentiment are derived for street segments. This can be accomplished via analysis of driver/pedestrian-perspective imagery (e.g., Google Street View (GSV)) for a set of points of interest (POI) along city streets. Spatial features, such as the sky, vegetation, roads, buildings, and sidewalks, can then be identified and quantified using computer vision methods such as PSPNet and Mask RCNN. Finally, level of enclosure, visibility, and complexity can be evaluated with respect to the sentiment metrics.

3. Application

To demonstrate the analysis framework (Figure 1), an application to the City of Columbia, MO, USA was examined. Columbia (Figure 2) is a medium sized city (population of 126,254) in the midwestern US which hosts several major universities/colleges (United States Census Bureau, 2020).

3.1. SOCIAL MEDIA POSTING LOCATIONS

In this application, public posts from the Instagram social media platform are used to represent expressions of sentiment for the city's streetscapes. Several major roads traversing the city were selected for analysis and 16 POIs along these roads were generated to assist in the search for proximate Instagram posting locations. The coordinates of the POIs were then used to search for Instagram posting sites using the instagraphi (https://github.com/adw0rd/instagrapi) Python package as shown in Figure 3. Querying the locations of the 16 POIs resulted in the identification of 135 unique Instagram posting sites in the study region.

```
#Set the coordinate system
columbia = arcpy.SpatialReference
            ("NAD 1983 UTM Zone 15N")
#Red featureclass stored in shp file format
os.path.join(workspace,'*Streets*.shp')

#Generate Points Along Lines
arcpy.management.GeneratePointsAlongLines
                (Input_Features="*Streets*.shp",
                Output_Feature_Class="*Points*.shp",
                x_field = "Lat"
                y_field = "Long"
                coordinate_system = columbia)

#Populate Lat-Long fields to the Output Feature
    arcpy.management.CalculateGeometryAttributes
                (in_features=Output_Feature_Class,
                geometry_property =
                [["Long", "POINT_X"],
                ["Lat", "POINT_Y"]])

#Scrap location using ./instagraphi
from instagraphi import Client
client = Client()

#Logging in the Instagram API
client.login("***", "***")

#Calling the name, lat-long & Location ID
client.location_search(Lat, Long)

#Importing the locations in GIS platform
arcpy.management.XYTableToPoint
                (in_table = "csv dataset",
                out_table = "Shape File",
                y_field = "Long"
                x_field = "Lat"
                coordinate_system = columbia)
```

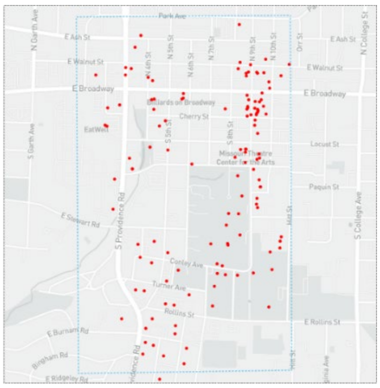

Figure 3. Code and mapping of Instagram posting locations

3.2. GEOTAGGED SOCIAL MEDIA DATA COLLECTION

Relevant social media posts were obtained from InstaLoadGram, a third-party data provider. A total of 111 locations (out of the initial 135 sites) featured usable public posts, yielding a total of 63,861 posts spanning a six-year period (1/1/2015 - 20/11/2021). A few example postings are shown in Table 1.

Location Name	Lat	Long	Caption	Date	Likes/ Comments
Francis Quadrangle	38.946111	-92.328889	Just some undergrads, nothing to see here. Keep scrolling	Nov 14th, 2021, 1:54:14 am	81/02
Francis Quadrangle	38.946111	-92.328889	No better way to introduce our new mellophone band family than with the classic pose!😎💛	Aug 16th, 2021, 7:27:08 am	48/00
Francis Quadrangle	38.946111	-92.328889	So excited to welcome our new members to the MU Spirit Executive family! FIGHT TIGER & MIZ!	May 6th, 2021, 4:30:08 pm	35/00
Francis Quadrangle	38.946111	-92.328889	Soaking up the sun	Apr 26th, 2021, 3:42:02 am	19/00

Table 1. Example Instagram posts acquired

3.3. SENTIMENT ANALYSIS

Individual sentiment can be inferred using a well-trained NLP model to score text-based posts. For example, 'safe', 'lively', 'bored', 'affluent', 'gloomy', and 'beautiful' are six commonly used sentiment markers. In order to reliably classify post by sentiment, the geotagged postings were processed using a normalized rating system (Zhang et al., 2018). Several deep learning (DL) models were investigated in this respect. Ultimately, the pre-trained Transformers model was selected given its ability to recognize the context giving meaning to each word in the sentence, allowing for more parallelization and shorter training and prediction durations (Colón-Ruiz and Cristóbal, 2021). Specifically, the Hugging Face API, a Bidirectional Encoder Representations from Transformers (BERT) based pre-trained Transformers model, was utilized to predict multilingual sentiment from the Instagram posts given the capacity of BERT to create contextually appropriate word embeddings (Devlin et al. 2019; Colón-Ruiz and Cristóbal, 2021). The model, trained using the BERT approach on one million human evaluations, automatically translates each post into the normalized rating ranging from 1 to 5 stars (Very Negative = 1 star; Negative = 2 stars; Neutral = 3 stars; Positive = 4 stars; Very Positive = 5 stars), an example of which is depicted in Figure 4 (Hugging Face, 2021).

```
ModelName = 'nlptown/bert-base-multilingual-uncased-sentiment'
TextClassifier = pipeline('sentiment-analysis', model = ModelName, Device = 0)

SentimentResults = TextClassifier(
                    ["Just some undergrads, nothing to see here. Keep scrolling",
                    "No better way to introduce our new mellophone band family than with the
                    classic pose!😎❤",
                    "So excited to welcome our new members to the MU Spirit Executive family!
                    FIGHT TIGER & MIZ!",
                    "Soaking up the sun"])

   o   {'label': '3 stars', 'score': 0.4023219347000122}
   o   {'label': '5 stars', 'score': 0.32480156421661377}
   o   {'label': '5 stars', 'score': 0.8477996587753296}
   o   {'label': '4 stars', 'score': 0.4071483314037323}
```

Figure 4. Example of sentiment scoring input/output

3.4. STREETVIEW IMAGERY

GSV images were retrieved for spatial feature classification. GSVs provide street-level and profile views of the urban environment and thus represent what individuals navigating the city may encounter.

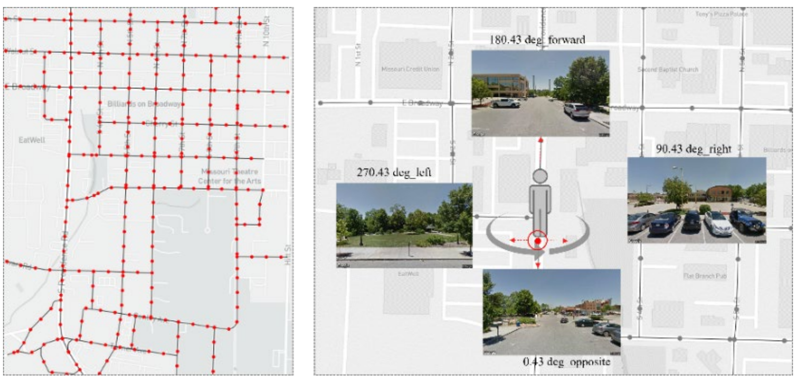

Figure 5. POIs along streets (left panel); GSV images per point (right panel)

Overall, 341 POIs along roads in the study region were generated to represent locations that would be traversed by individuals. Because the intention was to gather consecutive images of urban features, the POIs were collected at 30-meter intervals and then utilized to retrieve GSV images using Google API (Figure 4 - left panel). Four different viewing perspectives were considered for each POI i∈I given the sequence in which the POIs would be traversed along the roads: a) in the direction of the next POI i+1, b) 90.0 degrees from POI i+1, c) 180.0 degrees from POI i+1, and d) 270 .0 degrees from POI i+1 as illustrated in Figure 5 (right panel). Whereas the horizontal field of view was kept at 90.0 degrees, the "pitch" of the returned 800x400 pixel images was set to 0.0 degrees. In total, 1,364 images were retrieved for the sampled POIs.

3.5. SPATIAL FEATURE CLASSIFICATION

Enclosure, human scale, and complexity measurements are environment metrics thought to be strongly connected to the pedestrian experience but challenging to assess at urban scale (Miranda et al., 2021). Two prominent image detection and semantic segmentation models, PSPNet and Mask R-CNN, were used to classify features in each image. Enclosure and human scale measures were evaluated using PSPNet. The model was trained using the 150-category ADE20k dataset (Miranda et al., 2021). The proportion of an image classified as sky, walls, fences, trees, and buildings was used to represent the level of enclosure that persons were likely to experience. Human scale was estimated as the percentage of an image classified as sidewalk or road. The level of complexity present in the urban environment was characterized as the number of people, bicycles, motor bikes, cars, and streetlights present in each image as computed using the Mask R-CNN (Qiu et al., 2021).

4. Results

Computations for both the sentiment analyses and spatial feature classification were performed using the Google Colab Pro platform with a system configuration of Nvidia

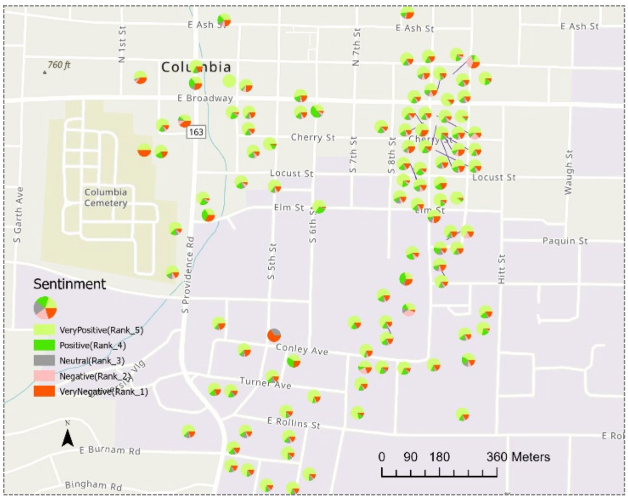

Figure 6. Summary output of sentiment analysis

Tesla P100-PCIE GPU and 16GB RAM. Assessing the sentiment associated with each of the 63,861 postings took 35,756 seconds of computational time. Figure 6 summarizes the proportion of each sentiment classification at each posting site. At many posting sites, a majority of the posts are classified as very positive. There is quite a bit of variation though in the spatial distribution of negative, very negative sentiment, with locations having higher proportions of such sentiment located in the Northwest, Northeast, and Central portions of the study region.

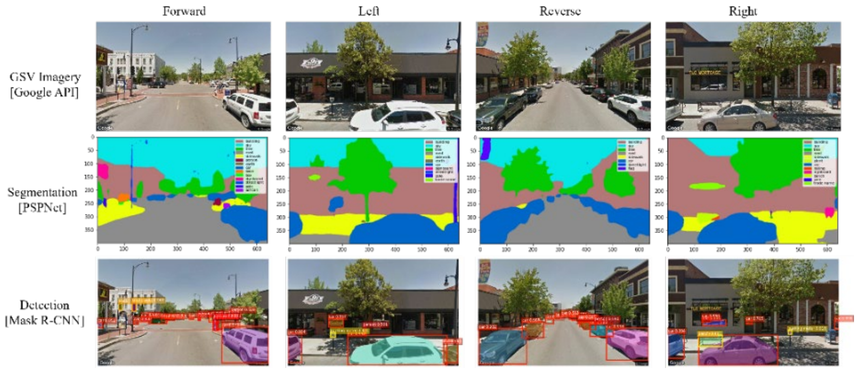

Figure 7. Spatial feature classification

The retrieval and classification of the 1,364 GSV images required 10,912 seconds of processing time. Figure 7 illustrates the output of the four images processed for one example POI. The first row depicts the images as retrieved. The second row depicts the percentage of pixels associated with different spatial features. The summary metrics associated with this are provided in the first and second sections of Table 2. The third row displays the number of complexity features identified in each image. The summary metrics for this are provided in the third section of Table 2.

	First section Enclosure					Second section Human Scale		Third section Complexity				
ID: 135	Sky	Wall	Fence	Tree	Building	Roads	Sidewalks	People	Bicycles	bikes	Cars	Streetlights
Forward	0.430	0.002	0.021	0.105	0.222	0.385	0.085	3	1	0	8	4
Left	0.222	0.000	0.002	0.158	0.238	0.010	0.012	1	0	2	4	1
Reverse	0.330	0.004	0.000	0.025	0.276	0.320	0.002	2	0	0	11	1
Right	0.150	0.010	0.005	0.071	0.484	0.103	0.183	1	0	0	2	0

Table 2. Spatial features calculation based on segmentation and detection

A heatmap was performed based on the proportion of negative sentiment rankings associated with each posting site (Figure 8). The heatmap (Figure 8-left panel) indicates several areas of denser proportion of negative sentiments (purple circle-marked). Improved can be found in areas to the Southeast (enclosed in a red circle). Similarly, an analysis of spatial autocorrelation, the Getis-Ord Gi*, of the proportion of positive postings at each location (Figure 8-right panel) reveals the presence of statistically significant spatial autocorrelation at several sites in the study region.

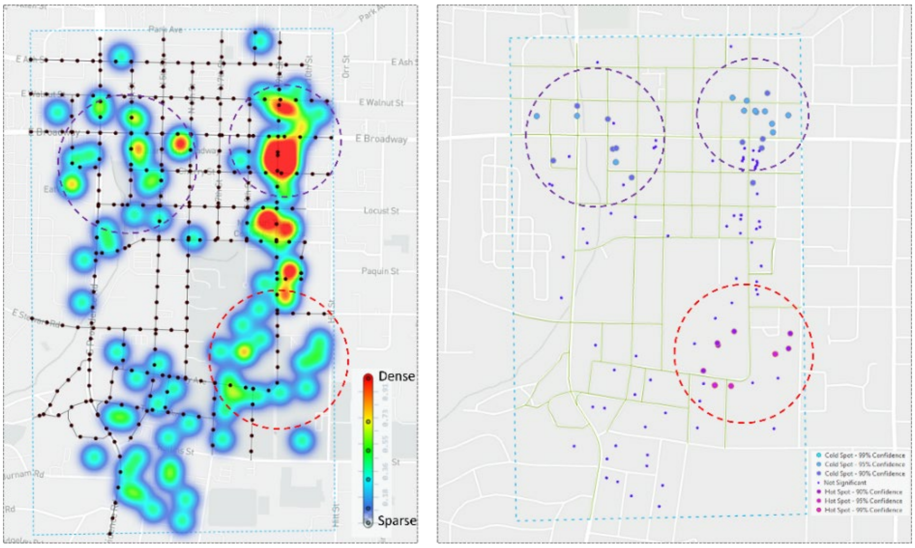

Figure 8. Heatmap based on negative posts (left panel); Hotspot analysis (right panel)

In particular, 19 sites in the South-eastern portion of the region were found to exhibit statistically significant positive spatial autocorrelation at the 0.90 confidence level or above (10 at the 0.95 confidence level or above, 4 at the 0.99 confidence level or above). Two areas to the North were found to exhibit statistically significant spatial autocorrelation of lower proportion of posts classified as positive at the 0.90 confidence level or above (11 at the 0.95 confidence level, 12 at the 0.99 confidence level or above).

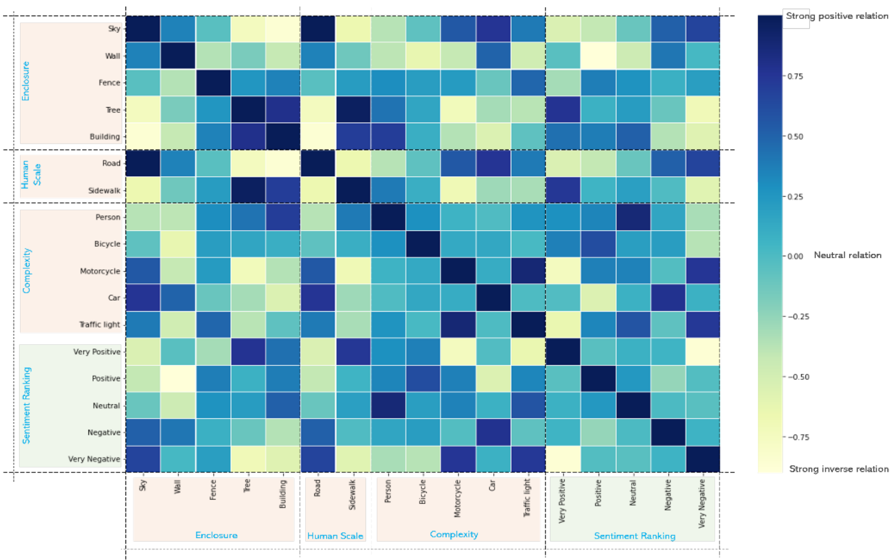

Figure 9. Correlation matrix of streetscape features and sentiment ratings

Finally, the characteristics of the streetscape were evaluated for posting locations associated with ratings of hot and cold spots to establish the extent to which streetscape features may be linked with individual sentiment. The results of the correlation analysis between streetscape features and sentiment ratings are summarized in Figure 9 and indicate that the streetscape features are strongly linked with individual sentiments.

In general, locations along streets of positive posting are strongly associated with the presence of vertical structures such as trees and buildings. Negative postings are more often associated with locations having fewer enclosure components and a greater amount of open sky. Both findings are consistent with those of Miranda et al (2021). Additionally, the findings show that streets with better pedestrian infrastructure, such as sidewalks, significantly contribute to positive sentiment, which is consistent with the human scale concept. Finally, areas with a higher concentration of human activity - people and cyclists - are directly associated with positive sentiments, whereas urban areas with heavy traffic - cars and motor vehicles - are inversely related to positive sentiments, as noted by Ye et al (2021).

5. Discussion and Future Works

Making urban human settlements more inclusive, safe, and resilient, as well as encouraging healthy lifestyles and well-being at all ages, are critical to long-term urban sustainability. The way people perceive and value places, infrastructures, and events in their daily lives has far-reaching implications for sustainability planning and implementation efforts. To this end, the aim of this research is to provide vital insights into street semantics to promote more informed community planning decisions (what SDG 11 aims to achieve by 2030). The paper details a framework for linking individual sentiment to characteristics of the streetscape. Open-source social media postings and are processed using deep learning methods to infer sentiment and characterize visibility at locations along a transportation system to make these associations possible. The framework is applied to a small case study to illustrate the procedures involved and demonstrate the applicability of the process.

There are various decision points, constraints, and extensions to this study that might be worthwhile directions for future research. Because of Meta's data policy at the time of the study, the investigators had limited access to social media. Furthermore, the sentiments expressed in the postings may not directly relate to the streetscape in all instances. Further, working with social media postings entails the complex task of interpreting the opinions of multiple individuals. In this study, a pre-trained model was utilized for this task. However, a more target-specific transformers model could be used instead, making use of a large amount of Instagram data, segmenting location-related data, and then training the transformers model for more accurate and relevant sentiment analysis. Given that social media postings are generally accompanied by a time/date stamp, it would also be intriguing to analyse sentiment in a more temporally disaggregated form if more frequent data on urban environment were available.

In the future, real-time social media datasets that incorporate both space and time into analysis, such as Emerging Hot Spot Analysis (EHSA) as a spatiotemporal application of the Getis-Ord Gi* statistical analysis, would be more robust and efficient in analysing urban streetscape. Finally, developing a website-based user-friendly interface for the city residents can be quite beneficial in obtaining real-time responses.

Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-21-2-0276. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. Finally, the authors would like to express their gratitude to Raj Roy for sharing his expertise in the Transformers DL model for NLP.

References

Batty, M., Hudson-Smith, A., Milton, R., & Crooks, A. (2010). Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS,* 16(1), 1–13. https://doi.org/10.1080/19475681003700831

Colón-Ruiz, C., & Segura-Bedmar, I. (2020). Comparing deep learning architectures for sentiment analysis on drug reviews. *Journal of Biomedical Informatics,* 110, 103539. https://doi.org/10.1016/j.jbi.2020.103539

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs].* http://arxiv.org/abs/1810.04805

Guerrero, P., Møller, M. S., Olafsson, A. S., & Snizek, B. (2016). Revealing Cultural Ecosystem Services through Instagram Images: The Potential of social media Volunteered Geographic Information for Urban Green Infrastructure Planning and Governance. *Urban Planning*, 1(2), 1–17. https://doi.org/10.17645/up.v1i2.609

Hugging face API. (n.d.). Hugging face API. Retrieved November 22, 2021, from https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

Ilieva, R. T., & McPhearson, T. (2018). Social-media data for urban sustainability. *Nature Sustainability*, 1(10), 553–565. https://doi.org/10.1038/s41893-018-0153-6

Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation, 3*(2–3), 185–204. https://doi.org/10.1080/13875868.2003.9683761

Qiu, W., Li, W., Liu, X., & Huang, X. (2021). Subjectively Measured Streetscape Perceptions to Inform Urban Design Strategies for Shanghai. *ISPRS International Journal of Geo-Information,* 10(8), 493. https://doi.org/10.3390/ijgi10080493

Salazar Miranda, A., Fan, Z., Duarte, F., & Ratti, C. (2021). Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment. *Computers, Environment and Urban Systems, 86,* 101563. https://doi.org/10.1016/j.compenvurbsys.2020.101563

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLOS ONE,* 8(7), e68400.

United States Census Bureau. (2020). *Census of Columbia [Data].* https://data.census.gov/cedsci/table?q=columbia,%20mo&tid=DECENNIALPL2020.P1

Ye, C., Zhang, F., Mu, L., Gao, Y., & Liu, Y. (2021). Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science,* 48(6), 1430–1444. https://doi.org/10.1177/2399808320935467

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning,* 180, 148–160. https://doi.org/10.1016/j.landurbplan.2018.08.020.