International Cartographic Association
Association Cartographique Internationale

Developing deep learning model for predicting perceived walkability using street view images

Youngok Kang*, Jiyeon Kim

Ewha Womans University, ykang@ewha.ac.kr, pbmejy@gmail.com

* Corresponding author

Keywords: walkability, image analysis, GeoAI, street view image, crowd-sourced data, Siamese network, ranking loss

Abstract:

The visual appearance of a city is an important factor in shaping people's perception of the neighborhood environment. With the recent development of computer vision, research has attempted to analyze perceptions of urban built environments such as 'beauty', 'lively', 'bored', 'safety', and 'walking' in cities. However, if the built environment in the city is not diverse or the characteristics are not distinctive, there is a limit to increasing the accuracy of a model. In particular, unlike large cities, small and medium-sized cities are such cases. To mitigate this problem, we propose a model that learns both regional features on images by dividing each image into patches and global features of images. To evaluate the effectiveness of our model, we collected Kakao street view images of Jeonju city, and implemented the model to predict the walkability score using the result of the pairwise comparison as training data and measured accuracy. It is demonstrated that the proposed model records the highest accuracy compared to the models of the existing studies. Our model effectively learns the visual characteristics of street view images regardless of the size of the city in learning people's perception of the urban visual environment.

The study consists of the following four stages: 1) collecting street view images, 2) generating dataset via pairwise comparison, 3) training a model, and 4) predicting walkability score (Figure 1). For image collection, a total of 49,156 sample points were extracted at 30 m intervals along the road network in Jeonju. We obtained a total of 196,624 images by collecting four-way (0, 90, 180, 270 degrees) images for each sample point using the Kakao View Roadmap API. Through a website, we showed volunteers two street view images and asked them to decide which one is better for walking. Through this, we constructed 134,069 paired comparison datasets consisting of two street view images and its label. If a volunteer selected the left image, the label has a value of 0 and 1 in the opposite case. Siamese network was used as baseline of the model, which has a structure that learns similarities between input images. As a loss function, RankNet, which learns the ranking of pairwise data, is utilized. The baseline of the model is RSS-CNN (Dubey et al., 2015). Additionally the model consists of a global branch that learns global features of images, a patch branch that learns regional features, and a score branch that predicts scores (Figure 2).

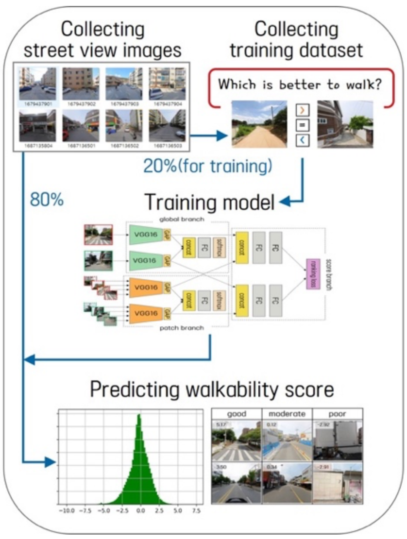

Figure 1. Framework of research

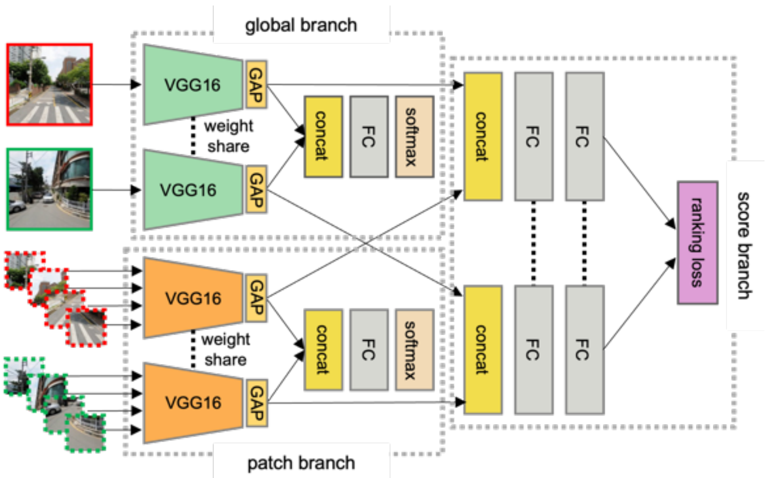

Figure 2. Proposed model architecture

https://doi.org/10.5194/ica-abs-6-118-2023

We confirmed the performance in the case of learning both global and regional features and compared the accuracy of our model with the models proposed in previous studies (Table 1). In Table 1, the base model (Dubey et al., 2016) and the semantic model (Xu et al., 2019) are ones proposed in previous studies. The patch model is a model that reflects only regional characteristics and is a structure excluding global branches in Figure 2. Looking at the accuracy, the global-patch model proposed in this study shows the highest accuracy.

Table 1. result of accuracy comparison

Model	Accuracy
Base model	73.87%
Semantic model	73.64%
Patch model	74.62%
Global-patch model	75.01%

Figure 3 shows the distribution of the perceived walkability score for collected street view images predicted by the global-patch model. Figure 4 shows an example of the perceived walkability score. The score at one point was calculated as the average value of the predicted score in the street view image of four directions (0 degrees, 90 degrees, 180 degrees, and 270 degrees). The mean is -0.21, the minimum is –10.21, and the maximum is 7.52.

In this work, we propose a model that can predict the perception of the visual properties of the urban street by using pairwise comparison datasets. This model has the characteristic of learning both the global and regional characteristics of street view images even in areas where urban built environments are not diverse. The accuracy comparison results show higher accuracy than previous studies.

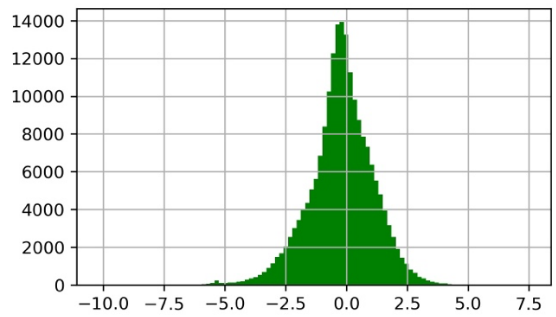

Figure 3. Distribution of perceived walkability score

Figure 4. example of perceived walkability score

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143782)

References

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A., 2016. Deep learning the city: Quantifying urban perception at a global scale. *In European conference on computer vision*, pp. 196-212.

Xu, Y., Yang, Q., Cui, C., Shi, C., Song, G., Han, X., & Yin, Y., 2019. Visual urban perception with deep semantic-aware network. *In International Conference on Multimedia Modeling*, pp. 28-40

Abstracts of the International Cartographic Association, 6, 118, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-118-2023 |