Sidewalk the Talk:

Translating street view imagery to correct perspectives to enhance bikeability and walkability studies

Koichi Ito, Matias Quintana, Filip Biljecki

Active transport, bikeability, Generative adversarial networks, street view imagery, walkability

URBAN SOLUTIONS AND SUSTAINABILITY R&D CONGRESS 2023
BUILDING SUSTAINABLE, RESILIENT, AND LIVEABLE CITIES OF TOMORROW
4TH - 5TH OCTOBER 2023

NUS National University of Singapore (SEC) SINGAPORE-ETH CENTRE (FCL) FUTURE CITIES LABORATORY GLOBAL urban analytics lab

BACKGROUND

- Active transport plays an important role in improving urban sustainability, and accurate and timely assessment of roads are important.

RESEARCH GAP

- Street view imagery has been used for walkability and bikeability assessment, but potential perspective bias have not been quantified or attempted to overcome.

AIM

- Quantify the bias among platforms.
- Try different models to minimize the bias.

METHODOLOGY

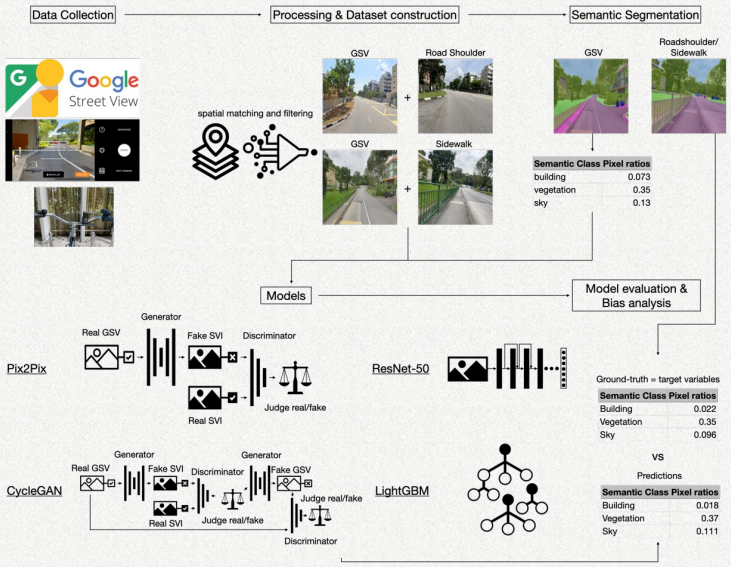

Figure 1: Framework overview. The Street-View Images (SVIs) are pre-processed and then loaded into two Generative Adversarial Networks (GAN) models, CycleGAN (bidirectional translations) [1] and Pix2Pix (unidirectional translation) [2] for training and inference

CONCLUSIONS

- This study revealed the presence of perspective bias when using GSV for active mobility analysis.
- CycleGAN [1] was able to mitigate the bias.

Future applications

- Future studies need to work on further improvement in the data collection/cleaning process
- More accurate assessment/intervention of the cycling/walking environment can nudge citizens towards greener, low-carbon transit options.

FINDINGS

- Google Street View (GSV), road shoulder view, and sidewalk views are weakly correlated, indicating the presence of bias
- CycleGAN with perspective input was found to be the best model to mitigate the bias with clear output images

Road Shoulder

Model	MSE	MAE	R-squared	Pearson's r
Gsv Panorama	0.018	0.107	-0.847	0.355
Cyclegan Panorama	0.00979	0.0749	-0.00568	0.419
Pix2pix Panorama	0.00801	0.0668	0.177	0.494
Lightgbm Panorama Without gan	0.00365	0.0434	0.625	0.798
Lightgbm Panorama Cyclegan	0.00341	0.0433	0.65	0.806
Lightgbm Panorama Pix2pix	0.00405	0.0478	0.584	0.767
Resnet Panorama Without gan	0.00439	0.0508	0.549	0.744
Resnet Panorama Cyclegan	0.00552	0.0578	0.433	0.664
Resnet Panorama Pix2pix	0.00645	0.0623	0.337	0.596
Gsv Perspective	0.0136	0.0935	-0.393	0.553
Cyclegan Perspective	0.0107	0.0795	-0.0936	0.468
Pix2pix Perspective	0.00883	0.067	0.097	0.46
Lightgbm Perspective Without gan	0.0034	0.043	0.653	0.811
Lightgbm Perspective Cyclegan	0.00334	0.0427	0.658	0.812
Lightgbm Perspective Pix2pix	0.00405	0.0473	0.586	0.768
Resnet Perspective Without gan	0.00382	0.0461	0.61	0.782
Resnet Perspective Cyclegan	0.00542	0.0568	0.446	0.683
Resnet Perspective Pix2pix	0.00624	0.0617	0.362	0.624

Sidewalk

Model	MSE	MAE	R-squared	Pearson's r
Gsv Panorama	0.0381	0.149	-0.0943	0.47
Cyclegan Panorama	0.0355	0.144	-0.0187	0.476
Pix2pix Panorama	0.0441	0.159	-0.265	0.577
Lightgbm Panorama Without gan	0.0112	0.0726	0.679	0.828
Lightgbm Panorama Cyclegan	0.0107	0.0748	0.694	0.835
Lightgbm Panorama Pix2pix	0.0155	0.0914	0.554	0.753
Resnet Panorama Without gan	0.014	0.0878	0.599	0.776
Resnet Panorama Cyclegan	0.0166	0.102	0.524	0.74
Resnet Panorama Pix2pix	0.0196	0.107	0.436	0.667
Gsv Perspective	0.0313	0.131	0.106	0.451
Cyclegan Perspective	0.0315	0.135	0.102	0.447
Pix2pix Perspective	0.0483	0.169	-0.378	0.52
Lightgbm Perspective Without gan	0.0131	0.0755	0.626	0.796
Lightgbm Perspective Cyclegan	0.0116	0.0771	0.669	0.819
Lightgbm Perspective Pix2pix	0.0153	0.0898	0.562	0.757
Resnet Perspective Without gan	0.0113	0.0795	0.677	0.824
Resnet Perspective Cyclegan	0.015	0.092	0.572	0.757
Resnet Perspective Pix2pix	0.0175	0.104	0.5	0.715

Figure 2: Tables of different models' performance metrics. LightGBM + CycleGAN + Perspective images had the lowest/best MAE and MSE and highest/best R-squared and Pearson's r for the road shoulder, and LightGBM + CycleGAN + Panorama had the best metrics for the sidewalk.

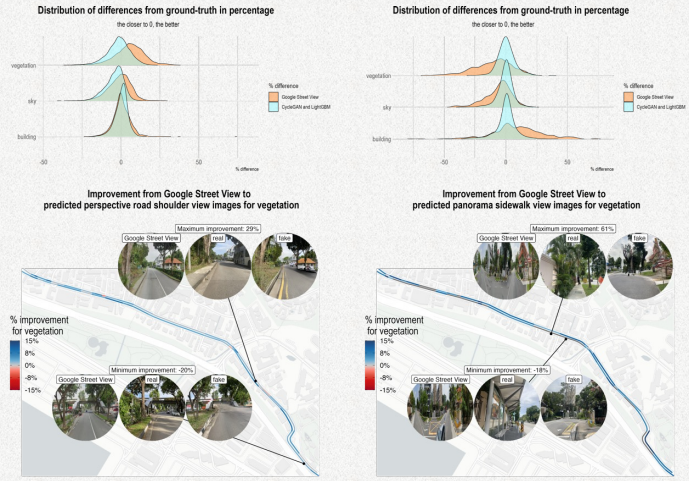

Figure 3: Example of available vegetation comparison in a real Street-View Images (SVI) and a generated one.

- Generative models can enhance the reliability of scalable SVI-based assessment.
- GAN models performed roughly 2 times better (average of original vs generated image similarity) on road shoulder images compared to sidewalk images.
- There are still challenges of 1. Data quality and 2. Scalability.
- Future studies can leverage simulated 3D virtual environment to overcome the challenges

[1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.

[2] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Contact
Koichi Ito, Ph.D. Student, National University of Singapore, Department of Architecture, Singapore, koichi.ito@u.nus.edu

Acknowledgements
This project was funded by the Singapore International Graduate Award (SINGA) Scholarship. This research was conducted at the Future Cities Lab Global at Singapore-ETH Centre and ETH Zurich. Future Cities Lab Global is supported and funded by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme and ETH Zurich (ETHZ), with additional contributions from the National University of Singapore (NUS), Nanyang Technological University (NTU), Singapore and the Singapore University of Technology and Design (SUTD).