This is an Accepted Manuscript of an article published by Taylor & Francis in the *International Journal of Geographical Information Science* on 27 Aug 2024, available at:
https://doi.org/10.1080/13658816.2024.2391969.
Cite as: Ito K, Quintana M, Han X, Zimmermann R, Biljecki F (2024): Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. *International Journal of Geographical Information Science*.

ARTICLE

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

Koichi Ito[a], Matias Quintana[b], Xianjing Han[c], Roger Zimmermann[c], and Filip Biljecki[a,d]

[a]Department of Architecture, National University of Singapore, Singapore; [b]Singapore-ETH Centre, Future Cities Lab Global Programme, Singapore; [c]School of Computing, National University of Singapore, Singapore; [d]Department of Real Estate, National University of Singapore, Singapore

ARTICLE HISTORY
Compiled August 29, 2024

ABSTRACT
Street view imagery (SVI), an emerging geospatial dataset, is useful for evaluating active transportation infrastructure, but it faces potential biases from its vehicle-based capture method, diverging from pedestrians' and cyclists' perspectives. Existing literature lacks both an examination of these biases and a solution. This study identifies and quantifies these biases by comparing conventional SVI with views from the road shoulder/sidewalk. To mitigate such perspective biases, we introduce a novel framework with generative adversarial network (GAN)-based image generation models (Pix2Pix and CycleGAN), an image regression model (ResNet-50), and a tabular model (LightGBM). Experiments assessed model effectiveness in translating car-centric views to those from pedestrian and cyclist perspectives. Results show significant differences in semantic indicators (e.g., green view index) between road center and road shoulder/sidewalk SVI, with low Pearson's correlation coefficients r (0.35-0.55 for road shoulders and 0.45-0.47 for sidewalks) indicating bias. The framework succeeded in creating realistic images and aligning pixel ratios between perspectives, achieving strong correlation coefficients (0.81 for road shoulders and 0.83 for sidewalks), thus reducing bias. This work contributes by providing a scalable and model-agnostic approach to produce accurate SVIs for urban planning and sustainability, setting a foundation for improving bikeability and walkability assessments and promoting active transportation.

KEYWORDS
Generative Adversarial Networks; Walkability; Bikeability; Active mobility; Spatial Data Infrastructures

1. Introduction

Active transportation (e.g., walking and cycling) plays an important role in improving sustainability and people's health in cities (Neves and Brand 2019, Cao and Shen 2019, Yap et al. 2023). Assessment of walkability and bikeability at a high resolution is beneficial for promoting such sustainable modes of transport for people in cities, and many studies have proposed various methods to assess them, such as on-site field observations/surveys (Clifton et al. 2007, Hoedl et al. 2010, Wahlgren et al. 2010,

CONTACT Filip Biljecki. Email: filip@nus.edu.sg

Horacek et al. 2012, Koh and Wong 2013), Geographic Information System-based approaches (Titze et al. 2012, Manton et al. 2016, Cain et al. 2018, Porter et al. 2020), and virtual audits (Gullón et al. 2015, Arellana et al. 2020). Recently, developments in machine learning have enabled researchers to propose scalable assessment methods by using street view imagery (SVI) and computer vision models (e.g., semantic segmentation), which allowed detailed evaluation of streets at a large scale without involving human labor (Ito and Biljecki 2021, Li et al. 2022a, Kang et al. 2023). However, these methods have limitations, such as variable perspectives. Google Street View, one of the primary data sources, provides images typically taken from a camera mounted on top of a car and, thus, rarely captures the perspectives of pedestrians or cyclists (Anguelov et al. 2010). Although some crowdsourced SVI platforms have images from cyclists and pedestrians, it is not viable to acquire them on a large scale due to the limited number of SVI from such perspectives (Biljecki et al. 2023, Hou et al. 2024). Studies on bikeability and walkability assessment using road center SVI have faced this issue and pointed it out as a potential major limitation that might create biases when analyzing data (Steinmetz-Wood et al. 2019, Ito and Biljecki 2021). For example, semantic segmentation is a popular technique adopted by such studies and used to calculate pixel ratios for different semantic classes (e.g., greenery, sky, and buildings), and the bias in this context refers to the difference in the pixel ratios between different perspectives. Although the bias can potentially cause inaccuracies in SVI-based assessments for active mobility, it is still unclear how large the bias is when using SVI with conventional perspectives (i.e., vehicular perspectives) due to a lack of research on this matter (Rui 2023).

Previous studies have also utilized SVI and machine learning models to predict and supplement urban features that are not widely available. For example, urban soundscape (Zhao et al. 2023, Zhuang et al. 2024), urban morphology (Zhang et al. 2021b), and building characteristics (Hu et al. 2020). Generative Adversarial Networks (GAN) have also demonstrated their capability to flexibly translate perspectives of images for cases where deterministic approaches cannot fully translate perspectives due to the absence of all the necessary information (i.e., camera parameters and visible areas in the image) in the input images to create views from other points of view. For example, GAN models have been used to translate aerial imagery to SVI (Toker et al. 2021, Regmi and Borji 2019, Wu et al. 2022a). However, no study has leveraged such machine/deep learning and GAN models to investigate the possibility of translating the perspective of SVI taken from highly positioned cameras on cars plying centers of motor lanes of roads to cyclists' views, which may be substantially different from existing works.

This study aims to fill this research gap by investigating the magnitude of biases when using road center SVI perspectives to assess active transportation modes' perspectives and examining whether GAN could be a suitable tool to translate such conventional perspectives into ones tailored to cyclists' and pedestrians' perspectives. To study these issues, we collected a training dataset of images taken from road shoulders and sidewalks. We aimed to quantify how large the biases are between road center SVI (i.e., conventional SVI perspectives) and road shoulder/sidewalk SVI (i.e., active mobility users' perspectives). Confirming the existence of biases, we created a new framework to overcome this bias by synthesizing Generative Adversarial Networks (GANs)-based image generation models (Pix2Pix and CycleGAN), an image regression model (ResNet-50), and a tabular model (LightGBM), thereby mitigating the aforementioned perspective bias.

This research contributes to the field by:

- understanding the bias introduced by road center SVI perspectives, particularly the lack of pedestrian and cyclist viewpoints in urban planning assessments, and raising awareness about it;
- developing a model-agnostic framework with generative/deep/machine learning models to mitigate the perspective bias in SVI, showcasing the potential of such an approach in urban planning and sustainability studies;
- contributing to urban planning and research by providing a framework to generate more accurate and reliable data for the assessment of bikeability and walkability, enhancing the sustainability and health aspects of city planning.

2. Literature review

2.1. *Applications of street view imagery*

Recent development of computer vision techniques and the proliferation of SVI data have enabled urban studies to map street-level features at large scales and high resolutions for various applications (Biljecki and Ito 2021, Wang et al. 2024a, Ito et al. 2024). For example, early studies used semantic segmentation to quantify urban greenery by analyzing the pixel ratio of greenery in SVI (Yang et al. 2009, Stubbings et al. 2019). Additionally, research has spanned multiple domains including spatial data infrastructure, urban health, urban perception, land use, building design, and transportation (Ogawa and Aizawa 2019, Keralis et al. 2020, Zhang et al. 2018, Cicchino et al. 2020, Wang et al. 2024b, Hu et al. 2023, Srivastava et al. 2018, Yao et al. 2019, Law et al. 2020, Fang et al. 2020, Qiao and Yuan 2021). SVI has also been used to evaluate active transportation infrastructures such as walkability and bikeability on urban scales. However, a significant limitation is the lack of perspectives from active transportation users due to the fact that the collection of SVI data is usually carried out using vehicle-mounted cameras, with limited data from road shoulders and sidewalks (Steinmetz-Wood et al. 2019, Ito and Biljecki 2021). This has resulted in biases that few studies have addressed or mitigated (Ki et al. 2023, Yin and Wang 2016, Li et al. 2018, Steinmetz-Wood et al. 2019, Ito and Biljecki 2021). Although some studies have collected their own images from the perspectives of active mobility users, it is still resource-intensive; therefore, there needs to be a more scalable solution (Chen et al. 2024).

2.2. *Synthetic data generation in urban science*

Numerous works have taken advantage of machine/deep learning models to mitigate the scarcity of data in urban science fields. Zhao et al. (2023), for example, used SVI, computer vision models, and tree-based machine learning models to predict and supplement soundscape characteristics on an urban scale, while Huang et al. (2024) estimated urban noise from SVI by combining a deep convolutional neural network model and machine learning models. Other studies have applied SVI and computer vision models for urban canyon classification (Hu et al. 2020) and building function and facade color classification (Zhang et al. 2021b).

Generative Adversarial Networks (GANs), a deep generative model, have also been developed to create synthetic data that mimic training data (Goodfellow et al. 2014). GANs have been used in fields such as medical science and geography to predict unobserved data and augment scarce data (Aggarwal et al. 2021, Zhao et al. 2021, Kang

et al. 2019, Abady et al. 2020, Isola et al. 2018, Li et al. 2020, Baier et al. 2022). In urban science, GANs have been applied to traffic volume and speed estimation (Xu et al. 2020, Zhang et al. 2019, Yu and Gu 2019, Lin et al. 2019), air/land traffic trajectory prediction (Wu et al. 2022b), enriching mobility data with socioeconomic attributes (Kim et al. 2022), predicting car accidents (Cai et al. 2020), master plan rendering from sketches (Ye et al. 2022, Choi et al. 2021), and creating synthetic population data (Garrido et al. 2020, Kim and Bansal 2023). In geographical information science, GANs have also been used for building footprint generation (Wu and Biljecki 2022), spatial interpolation (Zhu et al. 2020a), map image generation (Courtial et al. 2023), and creating privacy-preserving synthetic trajectory data (Rao et al. 2023).

A more relevant research area is the translation of views using GANs, such as the conversion of aerial images to SVI (Toker et al. 2021, Regmi and Borji 2019). Bajbaa et al. (2024) reviewed this topic, noting that GAN models are commonly used despite the availability of diffusion models (Ho et al. 2020, Dhariwal and Nichol 2021, Rombach et al. 2022). Diffusion models have been applied in remote sensing image fusion/generation (Sebaq and ElHelw 2023, Cao et al. 2023), floor plan generation (Ploennigs and Berger 2023), and interior design generation (Chen et al. 2023). While diffusion models produce photo-realistic images from text prompts and alter input image styles, GANs excel in generating images from input images, making them suitable for perspective-shifting tasks (Bajbaa et al. 2024). GANs have also been explored for image in-painting to remove occlusion in SVI (Zhang et al. 2021a), and a recent study proposed a pipeline to generate images with varying semantic features using a GAN model (Law et al. 2023).

As discussed so far, GAN models' flexibility to generate images from various points of view suggests its better suitability for perspective-shifting tasks than more deterministic approaches that reproject pixels in the input images based on a set of parameters. It is often the case that input images do not come with the parameters necessary for accurate reprojection of pixels, such as intrinsic (i.e., focal length, principal point, skew coefficient, and distortion coefficients) and extrinsic camera parameters (i.e., rotation matrix and translation vector). A naive solution for this issue is a simple projection of pixels onto a sphere around the camera position; however, this causes other issues, such as difficulty in determining the shift in X, Y, and Z coordinates, absence of pixels in some areas when shifting the camera position, and inaccurate distance to objects due to a lack of a precise depth map. We demonstrated these issues in Figure 1, where one can observe that this naive approach suffers from sensitivity to different parameters, voids, and distortion.

These existing studies show the usefulness of using image generation models to overcome issues in image-based analysis in various domains. However, no studies have explored the possibility of translating road center SVI taken by vehicles into the perspectives of active transportation users, despite the importance of overcoming the perspective gaps in achieving a more accurate and reliable assessment of the urban environment. This research seeks to bridge existing knowledge gaps through a detailed case study that examines the bias resulting from varying perspectives. Central to our approach is the exploration of a novel, model-agnostic framework that leverages the capabilities of Generative Adversarial Networks (GANs) to address and reduce this bias. Our focus on a model-agnostic strategy underscores the innovative aspect of our methodology, emphasizing its adaptability and potential applicability across a diverse range of models and contexts.

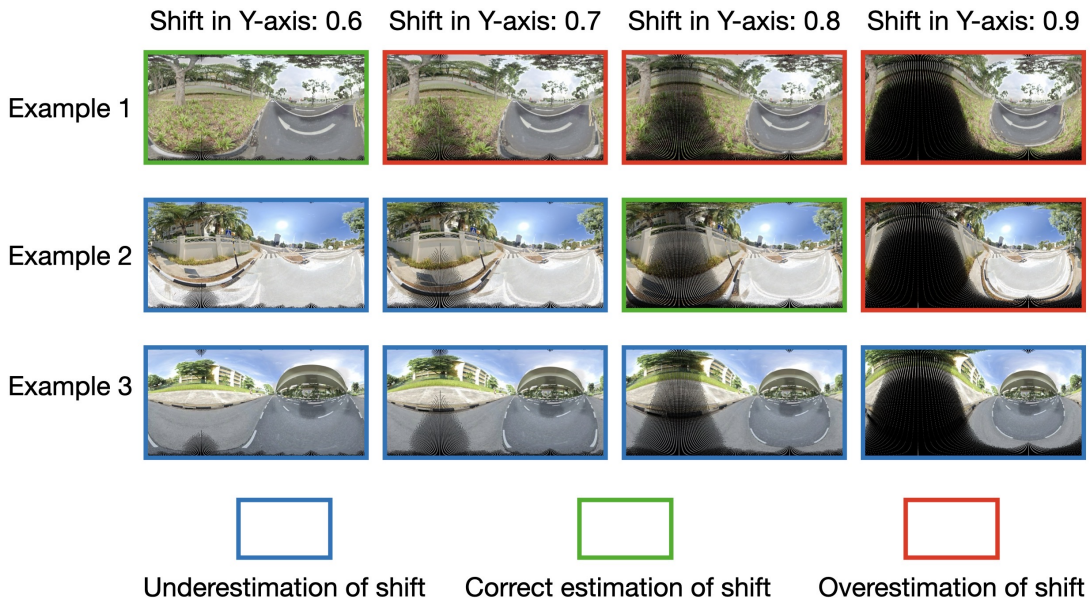


Figure 1.: Examples of different shifts in the Y-axis for three different street view images with varying road width.

3. Data and methods

Singapore was selected as a study site because it has sidewalks for most streets and is moderately safe to cycle, and three types of SVI have been collected: road center SVI from Google Street View (GSV), road shoulder SVI taken on road shoulders, and sidewalk SVI taken on sidewalks. These three types of SVI were used to train machine/deep learning and GAN models to translate pixel ratios of road center SVI data (i.e., building, vegetation, and sky) to those of the other two. Figure 2 shows the overall flow chart of the data and methodology used in this study, and Figure 3 and Figure 4 illustrate the models and approaches used in the experiments, respectively. Each approach uses a different combination of models, and we refer to the five approaches as A1, A2, A3, A4, and A5 hereafter as shown in Figure 4. We further elaborate on each component in the following subsections.

3.1. *Data collection*

The collection of images on road shoulders and sidewalks was conducted by cycling on public roads in Singapore with a smartphone camera installed onto a bicycle. The data collection took place during the daytime to ensure good lighting for the images, and the process was facilitated with the use of Mapillary to automatically capture SVI every five meters when cycling and record the locations of images (see Figure 5).

The total number of images collected on the road shoulder was 7,514, and 3,057 on the sidewalk. We downloaded them via Mapillary's Python SDK (Mapillary 2022).

3.2. *Preprocessing and semantic segmentation*

Once the image data were downloaded, images with too many occlusions were filtered out as part of preprocessing. This filtering procedure was performed with a semantic

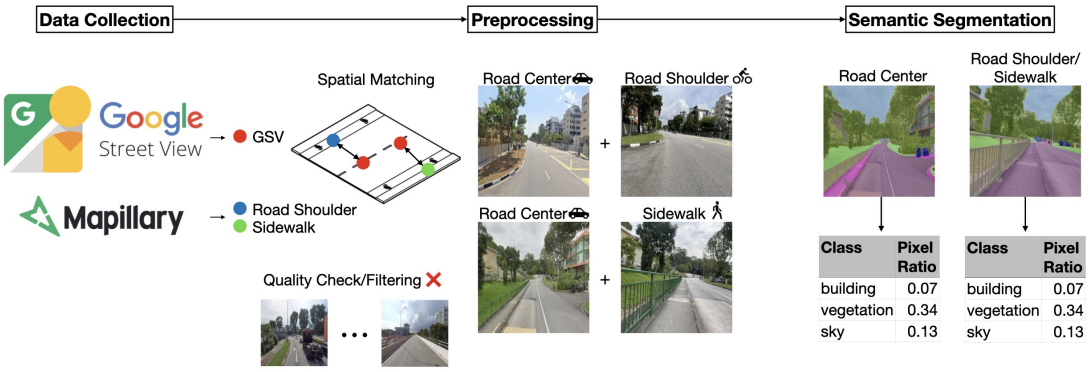


Figure 2.: The workflow of this study and the introduced methodology: data collection, processing and dataset construction, and semantic segmentation. We conducted data collection by downloading road center SVI and collecting road shoulder and sidewalk SVI and implemented processing and dataset construction by matching them and manually collecting images with spatial operations after filtering out images based on their quality. We ran semantic segmentation to quantify both datasets, which were used in model building as training data and approach evaluation as test data.

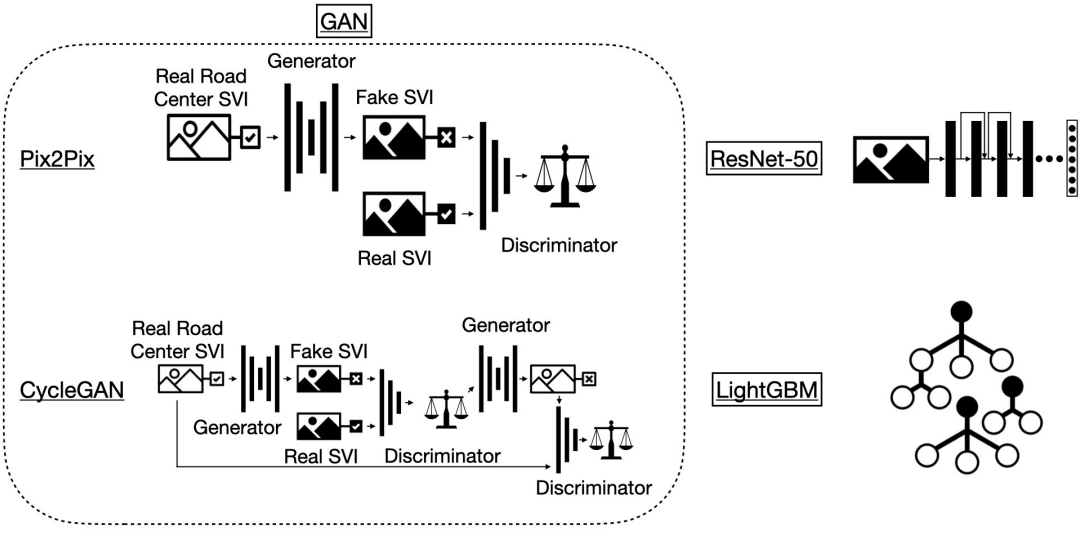


Figure 3.: The four machine/deep learning and GAN models used in this study are illustrated in this figure. GAN models are Pix2Pix and CycleGAN, consisting of generators and discriminators. As for machine/deep learning models, we used ResNet-50 and LightGBM.

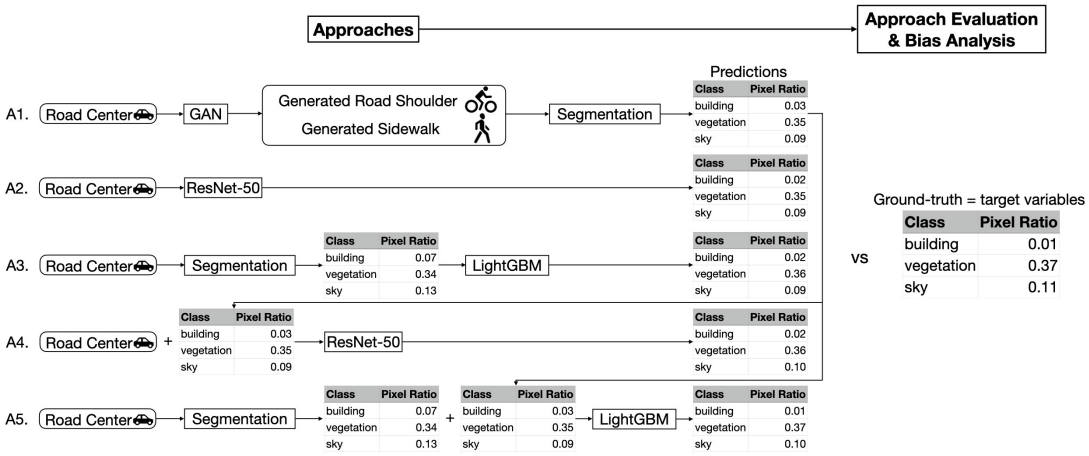


Figure 4.: This figure shows five approaches (i.e., A1-A5) used in this study and their evaluation. A1 generates road shoulder/sidewalk SVI using GAN models (i.e., Pix2Pix and CycleGAN) and obtains target pixel ratios through segmentation. A2 employs ResNet-50 to predict the target pixel ratios. A3 utilizes LightGBM with road center SVI pixel ratios as input to forecast the target pixel ratios. A4 inputs both road center SVI and pixel ratios from the GAN output into ResNet-50 to predict the target pixel ratios. A5 uses LightGBM with pixel ratios from both the road center SVI and the GAN output to predict the target pixel ratios.

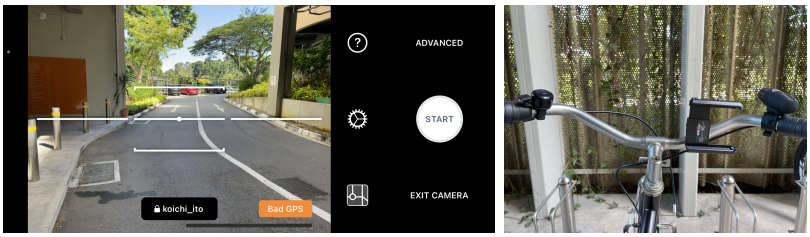


Figure 5.: The left image is the user interface of the Mapillary app for iOS devices, and the right image is how we installed a phone holder onto the bicycle.

segmentation model called Mask2Former pretrained on the CityScapes dataset with 84.5% mIoU and 19 categories (Cheng et al. 2022). We also used this segmentation model for all parts of this study that require semantic segmentation. This model was selected for its high accuracy. Subsequently, based on the locations of road shoulder and sidewalk SVI, road center SVI data were downloaded from GSV. To ensure high-quality alignment between different perspectives, road center SVI data were also filtered based on the spatiotemporal criteria and image classification model to remove any images taken in places that do not match the counterpart (e.g., images of highways). After filtering and matching, the resulting dataset consisted of 6,503 pairs of road shoulders images and 1,547 images of sidewalks. These were then split into train and test data sets by a 9:1 ratio. This high number of train data was chosen due to GAN's data-hungry training requirements to avoid overfitting (Karras et al. 2020). The pairs of road center SVI and road shoulder/sidewalk SVI were then quantified with a pre-trained Mask2Former model (Cheng et al. 2022), and the results were used for model training and evaluation. Among a few approaches to quantifying images (e.g., object detection), we decided to quantify semantic information of images with semantic segmentation because it has been widely used by previous works due to its capability to distinguish between different entities in the images, such as roads, sidewalks, vehicles, and pedestrians (Ito and Biljecki 2021).

3.3. *Models*

With the preprocessed dataset, we trained two GAN models — CycleGAN and Pix2Pix — to generate road shoulder/sidewalk SVI and two other models — ResNet-50 and LightGBM — as illustrated in Figure 3. As shown in Figure 4, we employed these different models and constructed five approaches shown below to predict the ground-truth pixel ratios of selected semantic classes from road shoulder/sidewalk SVI (i.e., building, vegetation, and sky) and only used road center SVI as input throughout the process. They are named as A1, A2, A3, A4, and A5.

- A1 Use GAN models (i.e., Pix2Pix and CycleGAN) to generate road shoulder/sidewalk SVI, from which we obtained the target pixel ratios with segmentation.
- A2 Use ResNet-50 to predict the target pixel ratios.
- A3 Use LightGBM with pixel ratios of the road center SVI as input to predict the target pixel ratios.
- A4 Use road center SVI and the pixel ratios from the GAN output (i.e., the output of A1 above) as input for ResNet-50 to predict the target pixel ratios.
- A5 Use LightGBM with pixel ratios of the road center SVI and pixel ratios from the GAN output (i.e., the output of A1 above) as input to predict the target pixel ratios.

We further provided details of each model used in this study in the following subsections.

3.3.1. GAN models

After the data cleaning, we utilized two types of GAN models to generate road shoulder and sidewalk views. The first model is the Pix2Pix model proposed by Isola et al. (2018), which achieved high flexibility of applications without engineering hyperparameters for different uses by adopting a U-Net architecture and a new discriminator

architecture called PatchGAN that judges whether N × N patch is real or fake. It replaces Gaussian noise z with dropout for diversified outputs. The Pix2Pix model's loss function is defined as:

$$\mathcal{L}cGAN(G, D) = \mathbb{E}x, y[\log D(x, y)] \tag{1}$$
$$+ \mathbb{E}x[\log(1 - D(x, G(x)))], \tag{2}$$

where x is the input image and y the target image. The generator G aims to minimize this loss, and the discriminator D to maximize it. The final loss function for generator G^* is:

$$G^* = \arg\min_G \max_D \mathcal{L}cGAN(G, D) + \lambda \mathcal{L}L1(G), \tag{3}$$

with L1 distance defined as $\mathcal{L}L1(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1]$. This architecture enabled the model to learn the overall structure of the images while producing relatively sharp output images, and such improvement allowed a wide range of studies to apply it in their fields (Wu and Biljecki 2022).

The second model is CycleGAN, developed by Zhu et al. (2020b). It introduces two domains, X and Y, and two generators, $G : X \to Y$ and $F : Y \to X$. CycleGAN uses a cycle-consistency loss for image fidelity. Its loss functions are:

$$\mathcal{L}\text{GAN}(G, D_Y, X, Y) = \mathbb{E}y \sim p_{\text{data}}(y)[\log D_Y(y)]$$
$$+ \mathbb{E}x \sim p\text{data}(x)[\log(1 - D_Y(G(x)))], \tag{4}$$

for generator G, and similarly for F. The cycle-consistency loss is:

$$\mathcal{L}\text{cyc}(G, F) = \mathbb{E}x \sim p_{\text{data}}(x)[|F(G(x)) - x|1]$$
$$+ \mathbb{E}y \sim p_{\text{data}}(y)[|G(F(y)) - y|1]. \tag{5}$$

The full objective combines these functions:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}\text{GAN}(G, D_Y, X, Y)$$
$$+ \mathcal{L}\text{GAN}(F, D_X, Y, X)$$
$$+ \lambda \mathcal{L}\text{cyc}(G, F), \tag{6}$$

with λ controlling the significance of the first two objectives. The optimal generators are found by:

$$G^*, F^* = \arg\min_{G,F} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y). \tag{7}$$

The models were implemented using the PyTorch implementation written by Isola et al. (2018) and Zhu et al. (2020b). These two models were selected because of their extensive use in different domains, customizability for various contexts, and relatively low cost of computational resources. The values and descriptions of the basic parameters used in this study are displayed in Table 4, and a list of parameters experimented in this study is shown in Table 1. We experimented with two GAN models (i.e., CycleGAN and Pix2Pix), two input data formats (i.e., panorama and perspective images), and two types of losses (i.e., default and default + mIoU). As for the calculation of

the mean intersection over union (mIoU), we used the aforementioned Mask2Former model. This metric provides a single performance figure that balances both the detection of the object (whether the pixels are identified as belonging to a particular class) and the delineation of the object (how accurately the segmentation outlines the object). A higher mIoU score indicates better segmentation performance, with a maximum value of 1 indicating perfect segmentation. It was added to test whether minimizing the discrepancy between semantically segmented ground-truth images and generated images at the pixel level can lead to a better overall composition of semantic information in the generated images.

Table 1.: Model parameters with their default values.

Parameter names	Tested values	Descriptions
GAN models	CycleGAN, Pix2Pix	Different types of GAN models
Input data format	Panorama (360 degrees), Perspective (90 degrees of front view)	Field of view for road center SVI used in the training.
Loss	Default, Default + mIoU from segmentation	mIoU was calculated between the target image and the generated images. λ for it was set at 10 for CycleGAN (i.e., the same value as its default λ_A and λ_B) and 100 for Pix2Pix (i.e., the same value as its default λ_L1).

3.3.2. LightGBM and ResNet-50

We also experimented using two types of deep learning to investigate which type of approach can yield the highest accuracy in predicting the ground-truth segmentation results, namely (1) image-to-table approach (i.e., A2 in Figure 2) and (2) table-to-table approach (i.e., A3 in Figure 2)

In the first approach, we conducted the image-to-table approach by inputting road center SVI and predicting the pixel ratios of the target semantic classes and implemented this approach with ResNet-50 (He et al. 2016), which uses 50 layers of a residual block. ResNet models have also been widely used by many urban studies for their simple architecture and readily available high-performing pre-trained models (Wei et al. 2022, Thackway et al. 2023, Quang et al. 2022, Chen et al. 2021). Specifically, ResNet-50 has been favored for its balance of depth and efficiency, making it suitable for complex tasks while maintaining manageable computational demands (He et al. 2016, Szegedy et al. 2016).

In the second approach, we input pixel ratios of different semantic classes calculated from segmented road center SVI in a tabular format and predict the pixel ratios of the target categories. We selected LightGBM for this approach due to its numerous advantages, including high accuracy, efficiency, and built-in regularization (Ke et al. 2017). LightGBM is recognized for its exceptional accuracy while maintaining a low

computational cost compared to similar gradient boosting algorithms, which are also known for their high accuracy (Florek and Zagdański 2023). Consequently, LightGBM has been widely adopted in urban studies for these reasons (Cui et al. 2021, Zhong et al. 2021, Xu et al. 2023). These two models — ResNet-50 and LightGBM — were chosen for this specific task for their balanced efficiency and accuracy, which is important as this study is the first study to explore this topic and it had been unclear how much computational complexity it requires to effectively and efficiently address the issue.

In addition to these two approaches, we also experimented with using GAN outputs to augment the accuracies of these two approaches to examine the potential supplementary value of GAN models. More specifically, we added the pixel ratios of the semantic classes calculated from segmenting the GAN-generated images and inputted them into the two models as additional features (i.e., A4 and 5 in Figure 2). For the ResNet-50 model, we concatenated these additional features to the main feature vector at the last fully connected layer.

3.4. *Approach evaluation and bias analysis*

We assessed the bias of road center SVI (i.e., differences in segmentation results between road center SVI and road shoulder/sidewalk SVI) as well as the values predicted by our models against ground truth views from pedestrians and cyclists using pixel ratios from semantic segmentation with these metrics:

- **Mean Squared Error (MSE)**: Reflects the average of the squared differences between estimated and actual values, providing a clear picture of the overall deviation from the ground truth.
- **Mean Absolute Error (MAE)**: Indicates the average absolute difference between estimated values and actual observations, providing an intuitive understanding of the average size of the errors.
- **R-squared**: Shows the proportion of variance in the dependent variable that is predictable from the independent variable(s), revealing how well the pixel ratios from road center SVI and predicted values align with the ground truth data in terms of variability.
- **Pearson's r**: Measures the strength and direction of the linear relationship between two variables, indicating the strength of the linear agreement.

For this study, we focused on three major semantic classes — building, sky, and vegetation — commonly used in assessing walkability and bikeability (Kang et al. 2023, Li et al. 2022b, Ito and Biljecki 2021, Ki et al. 2022) and predicting walking and cycling activities (Doiron et al. 2022, Koo et al. 2022, Huang et al. 2023). While other features like road conditions are also important, these three classes are critical in current literature. We measured the biases between road center SVI and road shoulder/sidewalk SVI in these classes using a consistent evaluation procedure across all models.

We assessed GAN model performance using three methods: train loss, Fréchet Inception Distance (FID), and qualitative evaluation. Evaluating GAN models solely on train loss is challenging, hence the need for these methods. For Pix2Pix, four losses are considered at training end: generator's GAN model loss, L_1 loss, discriminator's loss on real images, and discriminator's loss on fake images. For CycleGAN, the losses include discriminator's loss, generator's loss, cycle-consistency loss, and identity loss.

FID measures the similarity between the distributions of real and generated images using activations from an intermediate layer of the Inception network. Qualitative evaluation focused on overall structure, image quality/sharpness, and street feature

details. Additionally, we performed semantic segmentation to evaluate the models' ability to generate images with less bias compared to road center SVI. This segmentation analysis was conducted on a specific road as a case study to identify attributes causing larger errors.

4. Results

4.1. *Quantitative and qualitative assessments of GAN results*

Quantitative Assessments: Both Pix2Pix and CycleGAN models were trained using default parameters set by Isola et al. (2018) and Zhu et al. (2020b), shown in Table 4 as mentioned in section 3. Training losses for both models are depicted in Figure 6. The left plots show that Pix2Pix losses did not converge visibly. The right plots indicate that CycleGAN losses converged, with both cycle-consistency and identity losses decreasing over epochs. However, mIoU losses struggled to converge, possibly due to challenges in accurately predicting target image semantic classes at the pixel level.

Table 3 presents FID scores, where lower scores indicate better performance. Road shoulder models generally outperformed sidewalk models, probably because views from road shoulders are closer to road centers, allowing easier learning of pixel value distributions. CycleGAN models often achieved twice as good FID scores as Pix2Pix models, contradicting Saxena and Teli (2022), which found Pix2Pix to perform better with well-paired datasets. In this study, CycleGAN excelled due to its cycle consistency loss aiding in consistent translation output despite scalability and data quality challenges.

Perspective image input consistently yielded better FID scores, suggesting that limiting model input to relevant information improves output. Adding mIoU loss provided some improvement over default parameters but was less effective than perspective image input.

Qualitative Assessments: To further evaluate the GAN models' performance, we qualitatively analyzed the output images. Figure 7 shows real and generated images from different models. The first three columns display real images: road center panorama SVI, road center perspective SVI, and road shoulder/sidewalk SVI, with the first two used as input and the third as the target image.

The remaining columns show generated images from CycleGAN and Pix2Pix models with different parameter configurations: default, mIoU loss, and perspective input images. The first two rows display road shoulder images, while the last two rows show sidewalk images.

For road shoulder images, CycleGAN mIoU generated slightly sharper images than CycleGAN default with minimal distortion, and CycleGAN perspective produced the sharpest images closest to the real cyclist view. CycleGAN models preserved minor details like lane marker color. Pix2Pix models, however, produced grainier images with lower contrast and less defined object boundaries. Among Pix2Pix models, image quality improved in the order of perspective, mIoU, and default. Pix2Pix maintained a consistent structural composition but missed elements like lane markers and overhead roads.

For sidewalk images, GAN models performed worse. CycleGAN models showed distortion and overfitting, with CycleGAN default and mIoU displaying duplicated footpath lines and CycleGAN perspective resembling input images with slightly more

greenery. Pix2Pix models produced grainier, incomplete images with many white pixels, struggling with perspective translation due to the greater distance between input and target images and fewer training images.

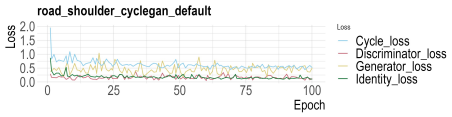
Overall, the qualitative evaluation suggests CycleGAN models, especially with perspective input images, might better produce understandable images while preserving real image details.

4.2. *Approach evaluation and bias analysis*

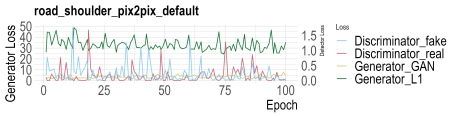
This study also utilized semantic segmentation on the test data set (i.e., 653 images for road shoulder and 155 images for sidewalk) to further analyze the bias of road center SVI compared to the road shoulder and sidewalk SVI and evaluate model performance. Only the major elements — building, sky, and vegetation — were examined because it is more difficult to accurately segment minor elements, such as traffic lights and bike lanes. Table 2 presents the performance for the vegetation's pixel ratio from the road shoulder view: mean squared error (MSE), mean absolute error (MAE), R-squared, and Pearson's r. The rest of the performance for building and sky from road shoulder and for building, sky, and vegetation from sidewalk can be found in the appendix (see Table 5, Table 6, Table 7, Table 8, and Table 9). The MSE and MAE quantify the discrepancies between predicted values and ground truth, with lower values being preferable. Conversely, R-squared represents the proportion of the variance in the ground truth values that is predictable from the predicted values, while Pearson's r measures the correlation between them. For both, higher values are desirable.

The findings from the road center SVI panorama and perspective indicate that the road center SVI, on its own, is not a dependable tool for evaluating greenery views from a cyclist's viewpoint. This is evidenced by the substantial errors reflected in its MSE values (0.0136 - 0.018) and MAE values (0.0935 - 0.107). Its R-squared values, ranging from -0.847 to -0.393, suggest a poor ability to account for variance — in fact, they perform worse than random predictions. Additionally, the correlation with ground truth, which lies between 0.355 and 0.553, remains relatively low. This result answers our first research question about the existence of bias between different views and confirms the concerns raised by previous papers (Steinmetz-Wood et al. 2019, Ito and Biljecki 2021). Moreover, we showed the metrics of SVI shifted with a naive method for road shoulder, which show even worse biases due to the issues discussed in Section 2.

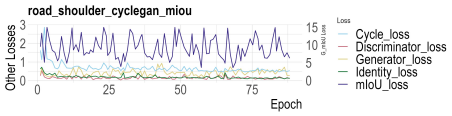
GAN-based models alone demonstrated their capabilities to mitigate the bias by improving the metrics. All the metrics for CycleGAN panorama (A1) and Pix2Pix panorama (A1) outperformed their road center SVI counterpart, and MSE, MAE, and R-squared were also improved by CycleGAN perspective and Pix2Pix perspective over their road center SVI counterpart. ResNet performed better than GAN-based models in all the metrics, and LightGBM performed even better than ResNet models. However, it is also important to note that we cannot simply compare them directly as their purposes are different (i.e., GAN models can produce images, but LightGBM can only take tabular data and produce numeric values). The best-performing model was the LightGBM model trained on segmentation results of perspective images together with segmented images generated by CycleGAN, which scored the best in all the metrics (MSE = 0.00334, MAE = 0.0427, R-squared = 0.658, and Pearson's $r = 0.812$). Thus, our findings suggest that GAN models can generate images from different perspectives and also provide useful features that can improve the predictions in some cases. The



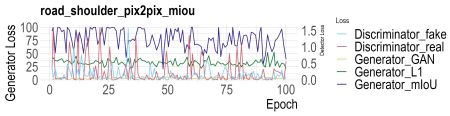
Road shoulder CycleGAN with default parameters.



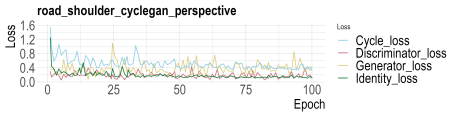
Road shoulder Pix2Pix with default parameters.



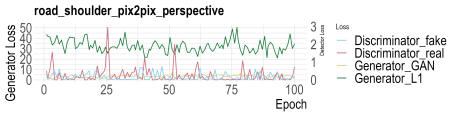
Road shoulder CycleGAN with mIoU loss.



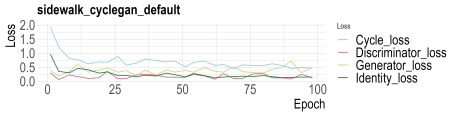
Road shoulder Pix2Pix with mIoU loss.



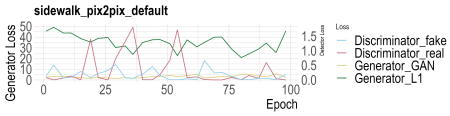
Road shoulder CycleGAN with perspective input images.



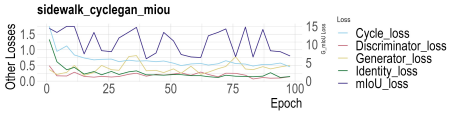
Road shoulder Pix2Pix with perspective input images.



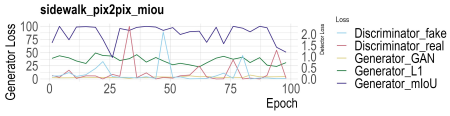
Sidewalk CycleGAN with default parameters.



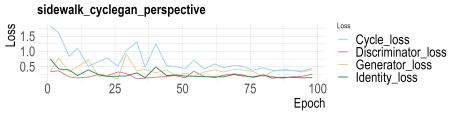
Sidewalk Pix2Pix with default parameters.



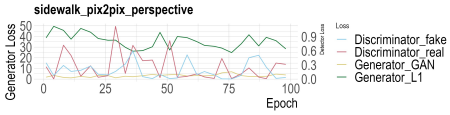
Sidewalk CycleGAN with mIoU loss.



Sidewalk Pix2Pix with mIoU loss.



Sidewalk CycleGAN with perspective input images.



Sidewalk Pix2Pix with perspective input images.

Figure 6.: Train losses for CycleGAN and Pix2Pix models. The left column shows losses of CycleGAN — including cycle consistency loss, discriminator loss, generator loss, and identity loss — and losses of Pix2Pix — including discriminator loss for fake images, discriminator loss for real images, generator's GAN loss, and generator's L1 loss.

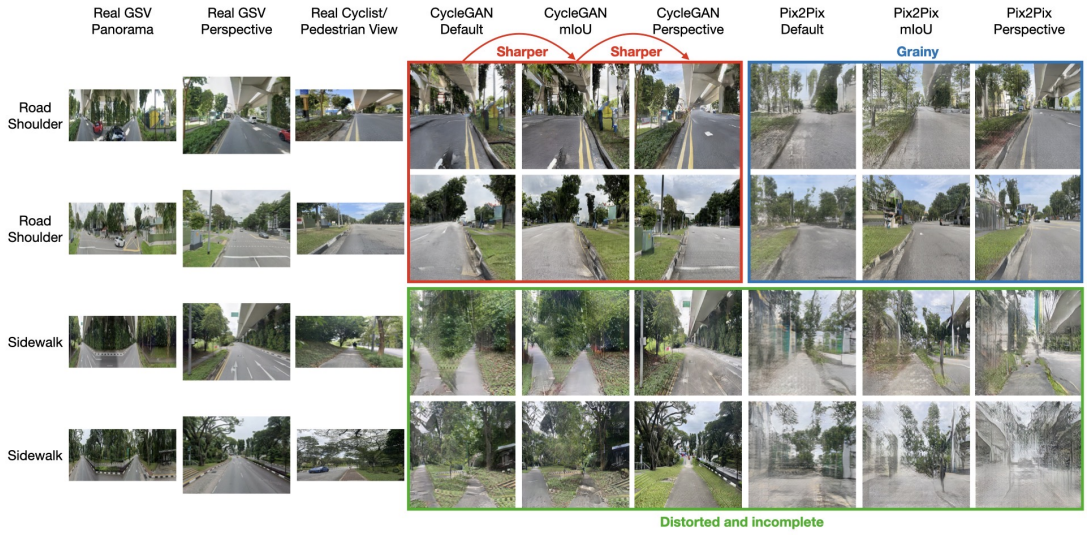


Figure 7.: This diagram shows an array of real images and generated images both for road shoulder and sidewalk. From left to right, it displays real road center SVI panorama, real road center SVI perspective, real cyclist and pedestrian view, CycleGAN with default parameters, CycleGAN with mIoU loss, CycleGAN with perspective input images, Pix2Pix with default parameters, Pix2Pix with mIoU loss, and Pix2Pix with perspective images. From top to bottom, it depicts two examples of road shoulder views and two examples of sidewalk views.

Table 2.: Performance of models trained on vegetation's pixel ratios from road shoulder perspectives. For MSE and MAE, the lower the number is, the more accurate the model is. And for R-squared and Pearson's r, the higher the number is, the more accurate the model is.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.018	0.107	-0.847	0.355
SVI panorama shifted with a naive method	0.020	0.114	-1.104	0.352
CycleGAN panorama (A1)	0.00979	0.0749	-0.00568	0.419
Pix2Pix panorama (A1)	0.00801	0.0668	0.177	0.494
LightGBM panorama without GAN (A3)	0.00365	0.0434	0.625	0.798
LightGBM panorama CycleGAN (A5)	0.00341	0.0433	0.65	0.806
LightGBM panorama Pix2Pix (A5)	0.00405	0.0478	0.584	0.767
ResNet-50 panorama without GAN (A2)	0.00439	0.0508	0.549	0.744
ResNet-50 panorama CycleGAN (A4)	0.00552	0.0578	0.433	0.664
ResNet-50 panorama Pix2Pix (A4)	0.00645	0.0623	0.337	0.596
Road center SVI perspective	0.0136	0.0935	-0.393	0.553
SVI perspective shifted with a naive method	0.064	0.232	-5.572	0.108
CycleGAN perspective (A1)	0.0107	0.0795	-0.0936	0.468
Pix2Pix perspective (A1)	0.00883	0.067	0.097	0.46
LightGBM perspective without GAN (A3)	0.0034	0.043	0.653	0.811
LightGBM perspective CycleGAN (A5)	**0.00334**	**0.0427**	**0.658**	**0.812**
LightGBM perspective Pix2Pix (A5)	0.00405	0.0473	0.586	0.768
ResNet-50 perspective without GAN (A2)	0.00382	0.0461	0.61	0.782
ResNet-50 perspective CycleGAN (A4)	0.00542	0.0568	0.446	0.683
ResNet-50 perspective Pix2Pix (A4)	0.00624	0.0617	0.362	0.624

findings above are consistent in other semantic classes on both road shoulder and sidewalk views.

4.3. *Case study*

Lastly, a case study was conducted on Pasir Panjang Road, a bidirectional secondary road with two lanes and sidewalks in Singapore, by running an inference with the best-performing models and performing semantic segmentation on the output. More specifically, we selected the LightGBM model trained on perspective images with CycleGAN output for the road shoulder and the LightGBM model trained on panorama images with CycleGAN output for the sidewalk based on their performance on the major semantic classes.

To analyze the variance in predicted pixel ratios from the ground truth, we visualized the discrepancy distributions for buildings, sky, and vegetation. These distributions are detailed in Figure 8, illustrating how our predictions diverge from actual measurements. The upper plot represents the road shoulder, while the lower one denotes the sidewalk. Distributions closer to 0 indicate a greater similarity to the ground truth. The density plots colored in red showcase the distribution of discrepancies among Google Street View images, while those in blue highlight the pixel ratios predicted by the CycleGAN-aided LightGBM models.

For the vegetation on the road shoulder, the CycleGAN-aided LightGBM model exhibits a higher density of around 0. In contrast, road center SVI's distribution centers more to the right of 0 and exhibits more extended tails on both ends. Regarding the sky on the road shoulder, both the combined LightGBM and CycleGAN model and road center SVI show denser distributions around a difference of 0. However, they are left-skewed, implying that the sky is frequently underrepresented in the road center

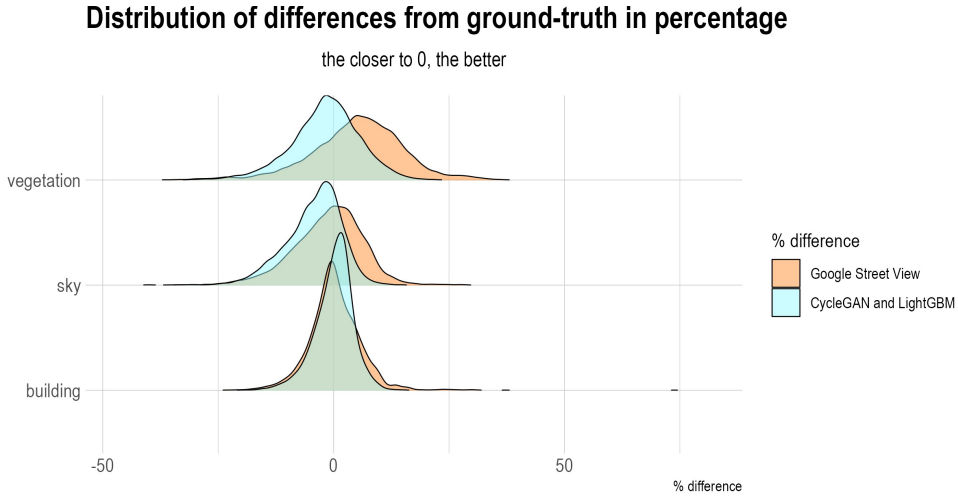
SVI and often underestimated by the model. This could be due to overrepresentation of the road in the generated images and blurs in the sky, which could cause inaccuracies in segmentation. For the building on the road shoulder, both the road center SVI and the CycleGAN-aided LightGBM model demonstrate dense distributions around a difference of 0. Yet, the latter's density is even more pronounced, highlighting its advantage over road center SVI.

As for the sidewalk, it is clearer that our CycleGAN-aided LightGBM model can predict pixel ratios of the three major classes much more accurately than road center SVI. Road center SVI's difference density plot for the vegetation is much longer-tailed, more left-skewed, and most importantly less dense at the difference of 0 than our model. Density plots for the sky also exhibit a similar pattern to vegetation, and those for the building also display longer tails of road center SVI and a highly dense distribution of our model's prediction around the difference of 0. Overall, our model's predictions are closely distributed around the difference of 0, and they are mostly within $\pm 25\%$ differences while road center SVI shows shifted centers and longer tails.

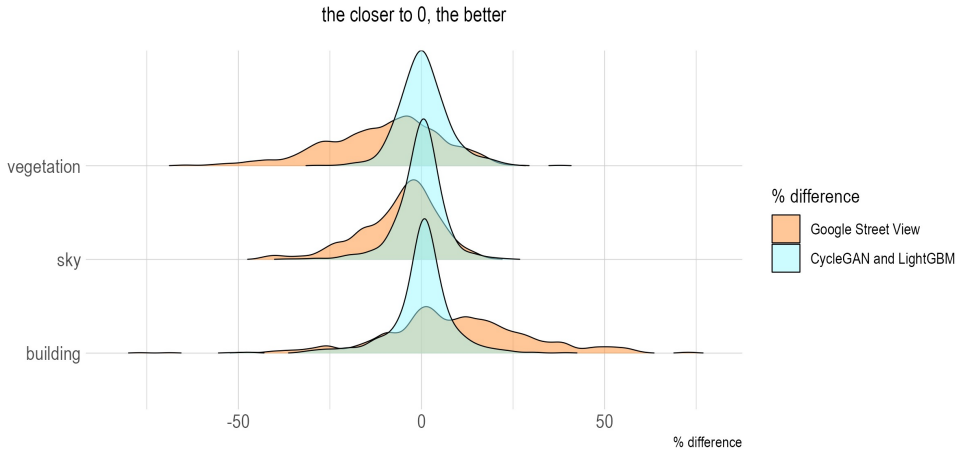
To assess the geographic spread of our models' enhancements, we created maps depicting the improvements for buildings, sky, and vegetation along streets in the study area. These enhancements are showcased in Figure 9, providing a visual representation of model performance across different road segments. These maps show blue color to represent positive improvements and red color to indicate negative changes. They also display images of input road center SVI, and both the most and least improved real and generated output from the CycleGAN model across the three categories. For buildings on the road shoulder, the map reveals improvements that are fairly uniformly distributed around and above 0 across the road. This suggests our model introduced consistent, albeit moderate, enhancements. In contrast, the map for buildings on the sidewalk predominantly features darker blue hues, pointing to the model's success in rectifying road center SVI-induced biases. Specific improvements for the road shoulder ranged between a high of 70% and a low of -10%. For the sidewalk, these values were 64% and -30%, respectively. Notably, despite achieving significant improvements, there's a discernible discrepancy between real and generated images in the plots. The map for the sky reveals some deterioration for the road shoulder, but improvements for the sidewalk span the entire road. For the road shoulder, improvements fluctuated between 22% and -22%, and between 37% and -19% for the sidewalk. Interestingly, generated images for the road shoulder closely mirror the real ones, but those for the sidewalk markedly differ. Regarding vegetation, both the road shoulder and sidewalk maps display consistent improvements by our model over road center SVI. Improvements on the road shoulder ranged from 29% to -20%, and from 61% to -18% on the sidewalk. Similar to the sky, while generated images for the road shoulder are nearly indistinguishable from real ones, those for the sidewalk deviate noticeably.

5. Discussion

This pioneering study introduces a novel, model-agnostic framework to explore the use of Generative Adversarial Networks (GANs) and deep/machine learning models for translating SVI from road centers to road shoulders and sidewalks. Despite its novelty, we also faced some limitations.

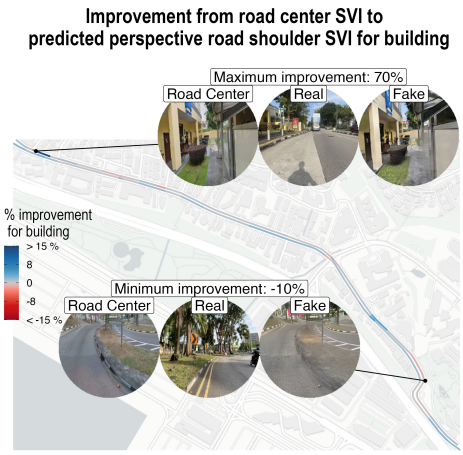


Density plots for road shoulder.

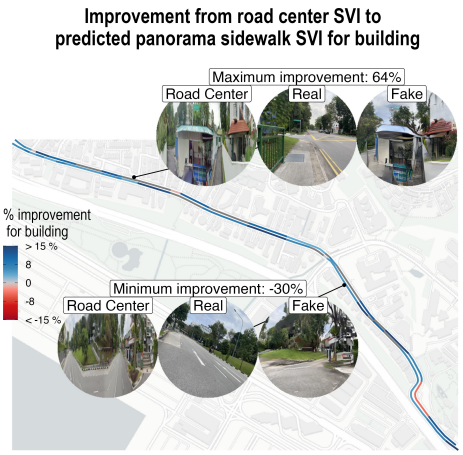


Density plots for sidewalk.

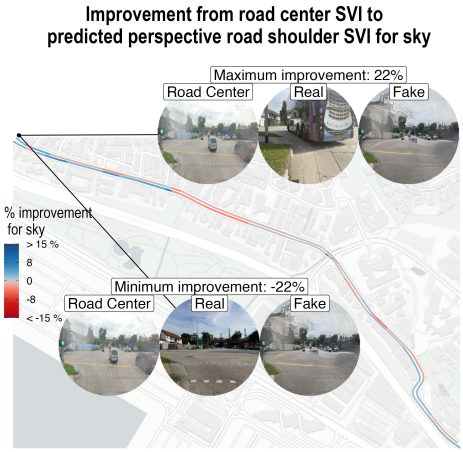
Figure 8.: Density plots of semantic segmentation's pixel ratios from Google Street View imagery (i.e., raw images taken from vehicular point of view) in red and from predictions by CycleGAN and LightGBM models for road shoulder and sidewalk in blue. The selected semantic classes are vegetation, sky, and building from top to bottom in the plots.



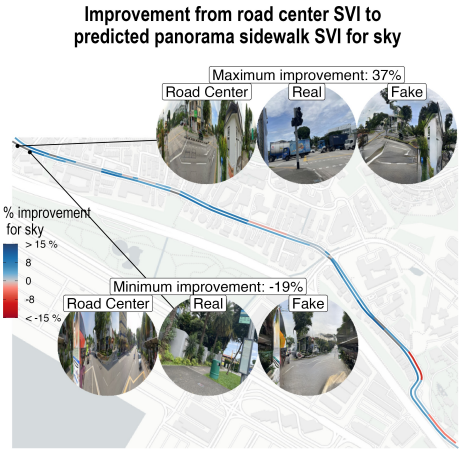
Building on road shoulder



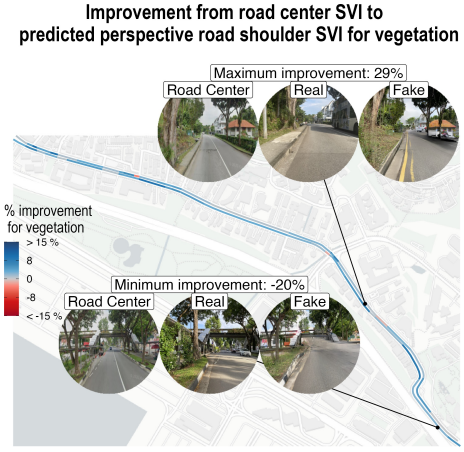
Building on sidewalk



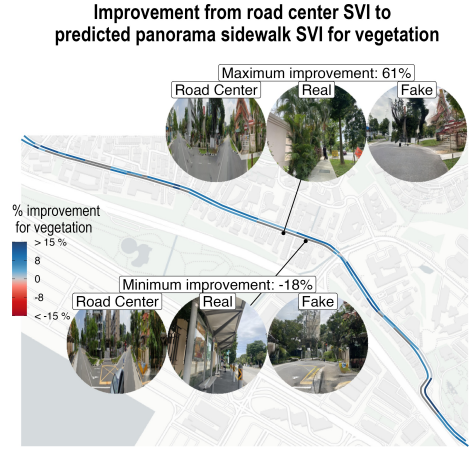
Sky on road shoulder



Sky on sidewalk



Vegetation on road shoulder



Vegetation on sidewalk

Figure 9.: These maps depict the enhancements made by the best models over Google Street View images. Street segments in blue signify improvements, while those in red highlight deteriorations. The top-right triplets show the largest improvements in percentage, and the bottom-left pairs show the smallest improvements in percentage. The left-most images in the triplets are Google Street View images used as input images for the model, the middle images are the real road shoulder/sidewalk view images, and the right-most images are the fake images generated by the model. Basemap credit: (c) OpenStreetMap contributors; CARTO light theme.

Category	Method	Parameters	FID
road_shoulder	cyclegan	perspective	60
road_shoulder	cyclegan	default	81
road_shoulder	cyclegan	miou	90
road_shoulder	pix2pix	perspective	131
road_shoulder	pix2pix	miou	138
road_shoulder	pix2pix	default	158
sidewalk	cyclegan	perspective	115
sidewalk	cyclegan	miou	138
sidewalk	cyclegan	default	144
sidewalk	pix2pix	miou	179
sidewalk	pix2pix	perspective	184
sidewalk	pix2pix	default	188

Table 3.: This table displays FID scores for different models. The category column shows the target perspectives, the method column shows the names of the GAN models, the parameter column shows different parameter settings, and the FID column shows FID scores.

5.1. *Data quality*

The first challenge was data quality. This study used road center SVI and road shoulder/sidewalk SVI as input data by matching their locations. However, it was difficult to standardize their locational relationships, despite our extensive effort in collecting SVI for a few months in a standardized manner (e.g., time, camera positions on road shoulders and sidewalks) and pre-processing the data (e.g., filtering spatiotemporally irrelevant SVI and SVI with occlusions). More specifically, to ensure their proximity, we set a threshold of five meters, within which both input SVI should be located; nonetheless, the angles of the shift from road center SVI to road shoulder/sidewalk SVI in relation to the streets were different for each pair. Thus, the unstandardized locational relationships might have made it difficult for GAN models to accurately learn the patterns and predict road shoulder/sidewalk views. Locational accuracies of SVI were difficult to judge simply from their distances as well. Such an example is shown on the first row of Figure 10. In some cases, there were overhead highways above the roads where road shoulder/sidewalk SVI was collected: when we used the coordinates of the road shoulder/sidewalk SVI to retrieve GSV (i.e., road center SVI), its API returned images of highways, not the road under it (see the second image of Figure 10. Despite the effort to remove such road center SVI with image classification, we could not manage to remove all of them which caused the GAN models to suffer from low-quality data. Occlusions in images were also problematic. Not only did road shoulder/sidewalk SVI have many occlusions like vehicles, but also road center SVI had occlusions, which created blurs and voids in the output images (see the third image in Figure 10).

Another important aspect of SVI data and artificial intelligence technologies is privacy and data security. The collection and processing of such imagery can inadvertently capture private moments and personal information of individuals without their consent, leading to potential privacy violations. Furthermore, the storage and analysis

of large datasets necessitate robust data security measures to prevent unauthorized access and misuse of sensitive information. For these reasons, we chose Mapillary as a platform to store our data as they take strict measures to protect privacy and security by, for example, blurring people's faces and vehicles' license plates.

5.2. *Scalability and generalizability*

Another challenge is the scalability of the study. The GAN models trained in this study are highly focused on the context of the study area (i.e. Singapore); thus, simply applying this study's models in different contexts may be difficult. Moreover, as this paper's methodology involves field data collection, it is resource-intensive to develop GAN models that are applicable in many different contexts. In sum, the generalizability and applicability of our model and approach to other cities would depend much on the similarity of the target context to our study area in Singapore, and it remains to be evaluated in future studies.

Lastly, the size/dimension of the input image might have been a bottleneck for the GAN models. Road center SVI is a 360-degree panorama, whereas the road shoulder/sidewalk SVI only has about 90 degrees of field of view, so their width-height ratios are quite different. Input images are usually fit into set sizes by either resizing or cropping them, which creates a trade-off in this study's context (i.e. distortion and completeness of the data). This study chose to resize them to retain as much information in the input images as possible, but one can observe that output images are affected by the distortion consequently.

5.3. *Future opportunities*

To overcome these challenges above, future studies can explore the possibility of utilizing more controlled settings such as 3D city models. In doing so, the data quality-related issues can be easily avoided. Moreover, by modifying the characteristics of urban environments, different contexts (e.g., cities outside of this study's study area) can be simulated with lower costs, and the input image size can be standardized by using the same image size (e.g., 360 degrees) for both the road center SVI perspective and the road shoulder/sidewalk SVI.

Our approach is just one method of mitigating perspective bias, and there may be others that can address this issue more effectively. The significance of GAN-based models cannot be overstated. While neither CycleGAN nor Pix2Pix's outputs substantially offset the aforementioned perspective bias, several important utilities of GANs have emerged. Firstly, they serve as invaluable tools for feature extraction and representation learning. The rich feature representations learned by GANs, whether from the generator or discriminator, can be leveraged as inputs for other models, enhancing their performance. Moreover, as demonstrated, GAN outputs can support other models and provide insightful visual aids for urban planners and policymakers. In terms of visualization and understanding of the environment, GANs can generate images from varied perspectives, offering a profound means to comprehend and map urban landscapes more effectively. Future studies should aim to refine and build upon our method to improve its effectiveness and applicability.

Considering future research and improvements, it is essential to note that GANs remain at the forefront of contemporary research. The likelihood that with deeper research and optimization, GANs might outperform other methods is compelling.

Adopting GANs can be viewed as a forward-thinking strategy, anticipating imminent advancements in GAN technology. Given the swift advancements in generative models' performances recently, there is an optimistic prospect that GANs will yield even more impressive outcomes in the near future.

In the selection of methodologies for our study, we opted for Generative Adversarial Networks (GANs) over diffusion models — a technique that has gained popularity in image generation tasks—for several reasons. Notably, diffusion models have been primarily employed in text-to-image generation Saharia et al. (2022b) and partial image-inpainting and image-style transformation applications Saharia et al. (2022a), not image-to-image translation, which was our focus. In contrast, GANs, with established variants such as Pix2Pix and CycleGAN, offer mature frameworks with a strong focus on image-to-image translation, facilitating their immediate application to our study's objectives. These GAN architectures are specifically optimized for image translation tasks, aligning with our goal of bias mitigation in SVI. Moreover, GANs are generally more computationally efficient than diffusion models — a crucial advantage for the extensive and data-intensive nature of our SVI analysis. GANs are particularly effective at capturing and translating the nuanced differences in road shoulder/sidewalk SVI and road center SVI, which was vital for addressing our research questions. While diffusion models are indeed an exciting and promising area of research for future SVI analysis, the established frameworks, practicality, and ready accessibility of GANs made them the more suitable choice for this study. We acknowledge the rapid development in the field of diffusion models and suggest their potential integration in future work as they become more tailored for complex image-to-image translation tasks in domains such as SVI. Our framework can still be leveraged when using different models.

Lastly, the implications of this study are particularly transformative for the scalable assessment of bikeability and walkability. By identifying and quantifying the bias inherent in road center SVI through Google Street View, our research has laid the groundwork for a more accurate representation of active mobility users' perspectives. The deployment of GAN and other machine learning models to mitigate these biases represents a substantial leap forward. Our approach corrects the skew in data interpretation that has previously gone unaddressed, enabling urban planners and researchers to harness the convenience and reach of SVI with newfound confidence in its accuracy. This methodological advancement allows for the broad application of SVI in urban studies, ensuring that assessments of walkability and bikeability can be conducted more reliably at scale. The corrected SVI, reflective of actual pedestrian and cyclist experiences, can be instrumental in shaping cities that are better designed for sustainable transportation, ultimately encouraging walking and biking through informed, data-driven urban design.

6. Conclusion

Street view imagery has been widely applied in geographical information science (Yao et al. 2021, Dai et al. 2024, Hu et al. 2023, Ito et al. 2024), and used to assess walkability and bikeability, typically with data taken from cars with elevated cameras. Given its increasing popularity, addressing potential bias due to mismatching perspectives is important.

This research is the first to investigate the extent of this bias and explore the potential of GAN and other machine/deep learning models to overcome it by predicting the

Unstable locational relationship due to unusual angle of the road shoulder view image.

A highway image paired with a non-highway image.

Occlusions by vehicles.

Figure 10.: Examples of the issues that we identified in the dataset, namely, unstable locational relationship, inclusion of highway images, and occlusions in images.

pixel ratios of semantic classes based on views from the road shoulder and sidewalk. The results indicate that road center SVI bias in assessing active transportation exists, and our framework with the Pix2Pix and CycleGAN models alleviates this by predicting major visual feature categories (i.e., building, sky, and vegetation) more accurately. Additionally, predictions improve when combining LightGBM with the segmentation results from CycleGAN-generated images.

Our findings are useful in understanding the previous research on active mobility and SVI with a perspective bias, justifying crowdsourced and human-oriented SVI collection efforts, and obtaining more accurate walkability and bikeability assessments by calibrating SVI perspectives.

Further research can leverage our model-agnostic approach using different models for more accurate and scalable results, utilizing controlled settings such as 3D models and realistic renderings. The technical contributions and findings of this study will be valuable for future studies that assess bike and walkability with SVI.

Acknowledgements

We are grateful for the valuable advice given by Wenmiao Hu (National University of Singapore). We thank the members of the NUS Urban Analytics Lab for the discussions and the editor and reviewers for their comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data and codes availability statement

Data use agreements prohibit the redistribution of street view imagery used in the research. The shared codes include codes to download the images, image processing, and image segmentation models to reproduce findings reported in the research at this figshare link. The link also includes a few sample images to show how the codes work. Due to commercial restrictions, supporting data is not available.

Funding

This research was supported by the Singapore International Graduate Research Award scholarship. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133.

Notes on contributor(s)

Koichi Ito is a PhD student at the National University of Singapore. His research interests include human mobility with emerging spatial data sources, such as street-view imagery, and machine/deep learning techniques.

Matias Quintana is a Postdoctoral Researcher and Module Coordinator at the Singapore-ETH Centre (SEC) at the Future Cities Lab (FCL) Global project. His research interests include remote sensing, urban data science, GeoAI, human-building interaction, and machine learning with emerging urban datasets such as street-level imagery and dynamic/sensor data.

Xianjing Han is a research fellow of the School of Computing, National University of Singapore. Her research interests include multimedia analysis and computer vision.

Roger Zimmermann is a Professor of Computer Science at the National University of Singapore. His research interests include streaming media architectures, media networking, applications of machine/deep learning, and spatio-temporal data management and urban computing, in combination with image/video data.

Filip Biljecki is an Assistant Professor at the National University of Singapore and the Principal Investigator of the NUS Urban Analytics Lab. He holds a Ph.D. in 3D GIS from the Delft University of Technology. His research interests include 3D city modelling, GeoAI, and geographic data science.

References

Abady, L., Barni, M., Garzelli, A., Tondi, B., 2020. GAN generation of synthetic multispectral satellite images, in: Image and Signal Processing for Remote Sensing XXVI, SPIE. pp. 122–133. doi:.

Aggarwal, A., Mittal, M., Battineni, G., 2021. Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights 1, 100004. doi:.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J., 2010. Google Street View: Capturing the World at Street Level. Computer 43, 32–38. doi:.

Arellana, J., Saltarín, M., Larrañaga, A.M., González, V.I., Henao, C.A., 2020. Developing an urban bikeability index for different types of cyclists as a tool to prioritise bicycle infrastructure investments. Transportation Research Part A: Policy and Practice 139, 310–334. doi:.

Baier, G., Deschemps, A., Schmitt, M., Yokoya, N., 2022. Synthesizing Optical and SAR Imagery From Land Cover Maps and Auxiliary Raster Data. IEEE Transactions on Geoscience and Remote Sensing 60, 1–12. doi:.

Bajbaa, K., Usman, M., Anwar, S., Radwan, I., Bais, A., 2024. Bird's-Eye View to Street-View: A Survey. doi:, arXiv:2405.08961.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning 215, 104217. doi:.

Biljecki, F., Zhao, T., Liang, X., Hou, Y., 2023. Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. International Journal of Applied Earth Observation and Geoinformation 122, 103385. doi:.

Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. Transportation Research Part C: Emerging Technologies 117, 102697. doi:.

Cain, K.L., Geremia, C.M., Conway, T.L., Frank, L.D., Chapman, J.E., Fox, E.H., Timperio, A., Veitch, J., Van Dyck, D., Verhoeven, H., Reis, R., Augusto, A., Cerin, E., Mellecker, R.R., Queralt, A., Molina-García, J., Sallis, J.F., 2018. Development and reliability of a streetscape observation instrument for international use: MAPS-global. The International Journal of Behavioral Nutrition and Physical Activity 15, 19. doi:.

Cao, Y., Shen, D., 2019. Contribution of shared bikes to carbon dioxide emission reduction

and the economy in Beijing. Sustainable Cities and Society 51, 101749. doi:.

Cao, Z., Cao, S., Wu, X., Hou, J., Ran, R., Deng, L.J., 2023. DDRF: Denoising Diffusion Model for Remote Sensing Image Fusion. arXiv:2304.04774.

Chen, J., Shao, Z., Hu, B., 2023. Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. Buildings 13, 1861. doi:.

Chen, W., Wu, A.N., Biljecki, F., 2021. Classification of urban morphology with deep learning: Application on urban vitality. Computers, Environment and Urban Systems 90, 101706. doi:.

Chen, Y., Huang, X., White, M., 2024. A study on street walkability for older adults with different mobility abilities combining street view image recognition and deep learning - The case of Chengxianjie Community in Nanjing (China). Computers, Environment and Urban Systems 112, 102151. doi:.

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-attention Mask Transformer for Universal Image Segmentation. doi:, arXiv:2112.01527.

Choi, S., Kim, J., Yeo, H., 2021. TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. Transportation Research Part C: Emerging Technologies 128, 103091. doi:.

Cicchino, J.B., McCarthy, M.L., Newgard, C.D., Wall, S.P., DiMaggio, C.J., Kulie, P.E., Arnold, B.N., Zuby, D.S., 2020. Not all protected bike lanes are the same: Infrastructure and risk of cyclist collisions and falls leading to emergency department visits in three U.S. cities. Accident Analysis & Prevention 141, 105490. doi:.

Clifton, K.J., Livi Smith, A.D., Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. Landscape and Urban Planning 80, 95–110. doi:.

Courtial, A., Touya, G., Zhang, X., 2023. Deriving map images of generalised mountain roads with generative adversarial networks. International Journal of Geographical Information Science 37, 499–528. doi:.

Cui, Z., Qing, X., Chai, H., Yang, S., Zhu, Y., Wang, F., 2021. Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. Journal of Hydrology 603, 127124. doi:.

Dai, S., Li, Y., Stein, A., Yang, S., Jia, P., 2024. Street view imagery-based built environment auditing tools: A systematic review. International Journal of Geographical Information Science 38, 1136–1157. doi:.

Dhariwal, P., Nichol, A., 2021. Diffusion Models Beat GANs on Image Synthesis. doi:, arXiv:2105.05233.

Doiron, D., Setton, E.M., Brook, J.R., Kestens, Y., McCormack, G.R., Winters, M., Shooshtari, M., Azami, S., Fuller, D., 2022. Predicting walking-to-work using street-level imagery and deep learning in seven Canadian cities. Scientific Reports 12, 18380. doi:.

Fang, F., Yu, Y., Li, S., Zuo, Z., Liu, Y., Wan, B., Luo, Z., 2020. Synthesizing location semantics from street view images to improve urban land-use classification. International Journal of Geographical Information Science , 1–24doi:.

Florek, P., Zagdański, A., 2023. Benchmarking state-of-the-art gradient boosting algorithms for classification. doi:, arXiv:2305.17094.

Garrido, S., Borysov, S.S., Pereira, F.C., Rich, J., 2020. Prediction of rare feature combinations in population synthesis: Application of deep generative modelling. Transportation Research Part C: Emerging Technologies 120, 102787. doi:.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Networks. doi:, arXiv:1406.2661.

Gullón, P., Badland, H.M., Alfayate, S., Bilal, U., Escobar, F., Cebrecos, A., Diez, J., Franco, M., 2015. Assessing Walking and Cycling Environments in the Streets of Madrid: Comparing On-Field and Virtual Audits. Journal of Urban Health: Bulletin of the New York Academy of Medicine 92, 923–939. doi:.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising Diffusion Probabilistic Models, in: Advances in

Neural Information Processing Systems, Curran Associates, Inc.. pp. 6840–6851.
Hoedl, S., Titze, S., Oja, P., 2010. The bikeability and walkability evaluation table reliability and application. American Journal of Preventive Medicine 39, 457–459. doi:.
Horacek, T.M., White, A.A., Greene, G.W., Reznar, M.M., Quick, V.M., Morrell, J.S., Colby, S.M., Kattelmann, K.K., Herrick, M.S., Shelnutt, K.P., Mathews, A., Phillips, B.W., Byrd-Bredbenner, C., 2012. Sneakers and spokes: An assessment of the walkability and bikeability of U.S. postsecondary institutions. Journal of Environmental Health 74, 8–15.
Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., Wang, Z., Zhao, T., Biljecki, F., 2024. Global Streetscapes — A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. ISPRS Journal of Photogrammetry and Remote Sensing 215, 216–238. doi:.
Hu, C.B., Zhang, F., Gong, F.Y., Ratti, C., Li, X., 2020. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. Building and Environment 167, 106424. doi:.
Hu, S., Xing, H., Luo, W., Wu, L., Xu, Y., Huang, W., Liu, W., Li, T., 2023. Uncovering the association between traffic crashes and street-level built-environment features using street view images. International Journal of Geographical Information Science 37, 2367–2391. doi:.
Huang, G., Yu, Y., Lyu, M., Sun, D., Zeng, Q., Bart, D., 2023. Using google street view panoramas to investigate the influence of urban coastal street environment on visual walkability. Environmental Research Communications 5, 065017. doi:.
Huang, J., Fei, T., Kang, Y., Li, J., Liu, Z., Wu, G., 2024. Estimating urban noise along road network from street view imagery. International Journal of Geographical Information Science 38, 128–155. doi:.
Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2018. Image-to-Image Translation with Conditional Adversarial Networks. doi:, arXiv:1611.07004.
Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transportation Research Part C: Emerging Technologies 132, 103371. doi:.
Ito, K., Kang, Y., Zhang, Y., Zhang, F., Biljecki, F., 2024. Understanding urban perception with visual data: A systematic review. Cities 152, 105169. doi:.
Kang, Y., Gao, S., Roth, R.E., 2019. Transferring multiscale map styles using generative adversarial networks. International Journal of Cartography 5, 115–141. doi:.
Kang, Y., Kim, J., Park, J., Lee, J., 2023. Assessment of Perceived and Physical Walkability Using Street View Images and Deep Learning Technology. ISPRS International Journal of Geo-Information 12, 186. doi:.
Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T., 2020. Training Generative Adversarial Networks with Limited Data. doi:, arXiv:2006.06676.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems, Curran Associates, Inc.
Keralis, J.M., Javanmardi, M., Khanna, S., Dwivedi, P., Huang, D., Tasdizen, T., Nguyen, Q.C., 2020. Health and the built environment in United States cities: Measuring associations using Google Street View-derived indicators of the built environment. BMC Public Health 20. doi:.
Ki, D., Lieu, S., Chen, Z., Lee, S., 2022. A Novel Walkability Index Using Google Street View and Deep Learning. doi:.
Ki, D., Park, K., Chen, Z., 2023. Bridging the gap between pedestrian and street views for human-centric environment measurement: A GIS-based 3D virtual environment. Landscape and Urban Planning 240, 104873. doi:.
Kim, E.J., Bansal, P., 2023. A deep generative model for feasible and diverse population synthesis. Transportation Research Part C: Emerging Technologies 148, 104053. doi:.
Kim, E.J., Kim, D.K., Sohn, K., 2022. Imputing qualitative attributes for trip chains extracted from smart card data using a conditional generative adversarial network. Transportation Research Part C: Emerging Technologies 137, 103616. doi:.
Koh, P.P., Wong, Y., 2013. Influence of infrastructural compatibility factors on walking and

cycling route choices. Journal of Environmental Psychology 36, 202–213. doi:.
Koo, B.W., Guhathakurta, S., Botchwey, N., 2022. How are Neighborhood and Street-Level Walkability Factors Associated with Walking Behaviors? A Big Data Approach Using Street View Images. Environment and Behavior 54, 211–241. doi:.
Law, S., Hasegawa, R., Paige, B., Russell, C., Elliott, A., 2023. Explaining holistic image regressors and classifiers in urban analytics with plausible counterfactuals. International Journal of Geographical Information Science 37, 2575–2596. doi:.
Law, S., Seresinhe, C.I., Shen, Y., Gutierrez-Roig, M., 2020. Street-Frontage-Net: Urban image classification using deep convolutional neural networks. International Journal of Geographical Information Science 34, 681–707. doi:.
Li, J., Chen, Z., Zhao, X., Shao, L., 2020. MapGAN: An Intelligent Generation Model for Network Tile Maps. Sensors 20, 3119. doi:.
Li, X., Santi, P., Courtney, T.K., Verma, S.K., Ratti, C., 2018. Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data. Transactions in GIS 22, 1029–1044. doi:.
Li, Y., Yabuki, N., Fukuda, T., 2022a. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. Sustainable Cities and Society 86, 104140. doi:.
Li, Y., Yabuki, N., Fukuda, T., 2022b. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. Sustainable Cities and Society 86, 104140. doi:.
Lin, Y., Dai, X., Li, L., Wang, F.Y., 2019. Pattern Sensitive Prediction of Traffic Flow Based on Generative Adversarial Framework. IEEE Transactions on Intelligent Transportation Systems 20, 2395–2400. doi:.
Manton, R., Rau, H., Fahy, F., Sheahan, J., Clifford, E., 2016. Using mental mapping to unpack perceived cycling risk. Accident Analysis & Prevention 88, 138–149. doi:.
Mapillary, 2022. Mapillary Python SDK. mapillary.
Neves, A., Brand, C., 2019. Assessing the potential for carbon emissions savings from replacing short car trips with walking and cycling using a mixed GPS-travel diary approach. Transportation Research Part A: Policy and Practice 123, 130–146. doi:.
Ogawa, M., Aizawa, K., 2019. Identification Of Buildings In Street Images Using Map Information, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 984–988. doi:.
Ploennigs, J., Berger, M., 2023. Diffusion Models for Computational Design at the Example of Floor Plans. arXiv:2307.02511.
Porter, A.K., Kohl, H.W., Pérez, A., Reininger, B., Pettee Gabriel, K., Salvo, D., 2020. Bikeability: Assessing the Objectively Measured Environment in Relation to Recreation and Transportation Bicycling. Environment and Behavior 52, 861–894. doi:.
Qiao, Z., Yuan, X., 2021. Urban land-use analysis using proximate sensing imagery: A survey. International Journal of Geographical Information Science 35, 2129–2148. doi:.
Quang, D.L., Sy, K.V., Viet, H.L., Bao, S.P., Quang, H.B., 2022. Signboards Detection From Street-view Image Using Convolutional Neural Network: A Case Study in Vietnam, in: 2022 RIVF International Conference on Computing and Communication Technologies (RIVF), pp. 394–397. doi:.
Rao, J., Gao, S., Zhu, S., 2023. CATS: Conditional Adversarial Trajectory Synthesis for privacy-preserving trajectory data publication using deep learning approaches. International Journal of Geographical Information Science 37, 2538–2574. doi:.
Regmi, K., Borji, A., 2019. Cross-view image synthesis using geometry-guided conditional GANs. Computer Vision and Image Understanding 187, 102788. doi:.
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis with Latent Diffusion Models. doi:, arXiv:2112.10752.
Rui, J., 2023. Measuring streetscape perceptions from driveways and sidewalks to inform pedestrian-oriented street renewal in Düsseldorf. Cities 141, 104472. doi:.
Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.,

2022a. Palette: Image-to-Image Diffusion Models. arXiv:2111.05826.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S.K.S., Gontijo-Lopes, R., Ayan, B.K., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M., 2022b. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, in: Advances in Neural Information Processing Systems.

Saxena, S., Teli, M.N., 2022. Comparison and Analysis of Image-to-Image Generative Adversarial Networks: A Survey. doi:, arXiv:2112.12625.

Sebaq, A., ElHelw, M., 2023. RSDiff: Remote Sensing Image Generation from Text Using Diffusion Model. arXiv:2309.02455.

Srivastava, S., Muñoz, J.E.V., Lobry, S., Tuia, D., 2018. Fine-grained landuse characterization using ground-based pictures: A deep learning solution based on globally available data. International Journal of Geographical Information Science 34, 1117–1136. doi:.

Steinmetz-Wood, M., Velauthapillai, K., O'Brien, G., Ross, N.A., 2019. Assessing the microscale environment using Google Street View: The Virtual Systematic Tool for Evaluating Pedestrian Streetscapes (Virtual-STEPS). BMC Public Health 19. doi:.

Stubbings, P., Peskett, J., Rowe, F., Arribas-Bel, D., 2019. A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning. Remote Sensing 11, 1395. doi:.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. doi:, arXiv:1602.07261.

Thackway, W., Ng, M., Lee, C.L., Pettit, C., 2023. Implementing a deep-learning model using Google street view to combine social and physical indicators of gentrification. Computers, Environment and Urban Systems 102, 101970. doi:.

Titze, S., Krenn, P., Oja, P., 2012. Developing a bikeability index to score the biking-friendliness of urban environments. Journal of Science and Medicine in Sport 15, S29–S30. doi:.

Toker, A., Zhou, Q., Maximov, M., Leal-Taixe, L., 2021. Coming Down to Earth: Satellite-to-Street View Synthesis for Geo-Localization. doi:, arXiv:2103.06818.

Wahlgren, L., Stigell, E., Schantz, P., 2010. The active commuting route environment scale (ACRES): Development and evaluation. International Journal of Behavioral Nutrition and Physical Activity 7, 58. doi:.

Wang, S., Huang, X., Liu, P., Zhang, M., Biljecki, F., Hu, T., Fu, X., Liu, L., Liu, X., Wang, R., Huang, Y., Yan, J., Jiang, J., Chukwu, M., Reza Naghedi, S., Hemmati, M., Shao, Y., Jia, N., Xiao, Z., Tian, T., Hu, Y., Yu, L., Yap, W., Macatulad, E., Chen, Z., Cui, Y., Ito, K., Ye, M., Fan, Z., Lei, B., Bao, S., 2024a. Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review. International Journal of Applied Earth Observation and Geoinformation 128, 103734. doi:.

Wang, Z., Ito, K., Biljecki, F., 2024b. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. Cities 145, 104704. doi:.

Wei, J., Yue, W., Li, M., Gao, J., 2022. Mapping human perception of urban landscape from street-view images: A deep-learning approach. International Journal of Applied Earth Observation and Geoinformation 112, 102886. doi:.

Wu, A.N., Biljecki, F., 2022. GANmapper: Geographical data translation. International Journal of Geographical Information Science 36, 1394–1422. doi:.

Wu, A.N., Stouffs, R., Biljecki, F., 2022a. Generative adversarial networks in the built environment: A comprehensive review of the application of gans across data types and scales. Building and Environment 223, 109477. doi:.

Wu, X., Yang, H., Chen, H., Hu, Q., Hu, H., 2022b. Long-term 4D trajectory prediction using generative adversarial networks. Transportation Research Part C: Emerging Technologies 136, 103554. doi:.

Xu, D., Wei, C., Peng, P., Xuan, Q., Guo, H., 2020. GE-GAN: A novel deep learning framework for road traffic state estimation. Transportation Research Part C: Emerging Technologies 117, 102635. doi:.

Xu, K., Han, Z., Xu, H., Bin, L., 2023. Rapid Prediction Model for Urban Floods Based on

a Light Gradient Boosting Machine Approach and Hydrological–Hydraulic Model. International Journal of Disaster Risk Science 14, 79–97. doi:.

Yang, J., Zhao, L., Mcbride, J., Gong, P., 2009. Can you see green? Assessing the visibility of urban forests in cities. Landscape and Urban Planning 91, 97–104. doi:.

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., Guan, Q., 2019. A human-machine adversarial scoring framework for urban perception assessment using street-view images. International Journal of Geographical Information Science 33, 2363–2384. doi:.

Yao, Y., Zhang, J., Qian, C., Wang, Y., Ren, S., Yuan, Z., Guan, Q., 2021. Delineating urban job-housing patterns at a parcel scale with street view imagery. International Journal of Geographical Information Science 35, 1927–1950. doi:.

Yap, W., Chang, J.H., Biljecki, F., 2023. Incorporating networks in semantic understanding of streetscapes: Contextualising active mobility decisions. Environment and Planning B: Urban Analytics and City Science 50, 1416–1437. doi:.

Ye, X., Du, J., Ye, Y., 2022. MasterplanGAN: Facilitating the smart rendering of urban master plans via generative adversarial networks. Environment and Planning B: Urban Analytics and City Science 49, 794–814. doi:.

Yin, L., Wang, Z., 2016. Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. Applied Geography 76, 147–153. doi:.

Yu, J.J.Q., Gu, J., 2019. Real-Time Traffic Speed Estimation With Graph Convolutional Generative Autoencoder. IEEE Transactions on Intelligent Transportation Systems 20, 3940–3951. doi:.

Zhang, F., Zhang, D., Liu, Y., Lin, H., 2018. Representing place locales using scene elements. Computers, Environment and Urban Systems 71, 153–164. doi:.

Zhang, J., Fukuda, T., Yabuki, N., 2021a. Automatic Object Removal With Obstructed Façades Completion Using Semantic Segmentation and Generative Adversarial Inpainting. IEEE Access 9, 117486–117495. doi:.

Zhang, J., Fukuda, T., Yabuki, N., 2021b. Development of a City-Scale Approach for Façade Color Measurement with Building Functional Classification Using Deep Learning and Street View Images. ISPRS International Journal of Geo-Information 10, 551. doi:.

Zhang, K., Jia, N., Zheng, L., Liu, Z., 2019. A novel generative adversarial network for estimation of trip travel time distribution with trajectory data. Transportation Research Part C: Emerging Technologies 108, 223–244. doi:.

Zhao, B., Zhang, S., Xu, C., Sun, Y., Deng, C., 2021. Deep fake geography? When geospatial data encounter Artificial Intelligence. Cartography and Geographic Information Science 48, 338–352. doi:.

Zhao, T., Liang, X., Tu, W., Huang, Z., Biljecki, F., 2023. Sensing urban soundscapes from street view imagery. Computers, Environment and Urban Systems 99, 101915. doi:.

Zhong, J., Zhang, X., Gui, K., Wang, Y., Che, H., Shen, X., Zhang, L., Zhang, Y., Sun, J., Zhang, W., 2021. Robust prediction of hourly PM2.5 from meteorological data using LightGBM. National Science Review 8, nwaa307. doi:.

Zhu, D., Cheng, X., Zhang, F., Yao, X., Gao, Y., Liu, Y., 2020a. Spatial interpolation using conditional generative adversarial neural networks. International Journal of Geographical Information Science 34, 735–758. doi:.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2020b. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. doi:, arXiv:1703.10593.

Zhuang, Y., Kang, Y., Fei, T., Bian, M., Du, Y., 2024. From hearing to seeing: Linking auditory and visual place perceptions with soundscape-to-image generative artificial intelligence. Computers, Environment and Urban Systems 110, 102122. doi:.

Appendix

Table 4.: Model parameters used in this study

Parameters	Values	Descriptions
input_nc	3	# of input image channels
output_nc	3	# of output image channels
ngf	64	# of gen filters in the last conv layer
ndf	64	# of discrim filters in the first conv layer
netD	basic	specify discriminator architecture
netG	resnet_9blocks	specify generator architecture
n_layers_D	3	only used if netD==n_layers
norm	instance	instance normalization or batch normalization
init_type	normal	network initialization
init_gain	0.02	scaling factor for initialization
no_dropout	true	no dropout for the generator
serial_batches	true	if true, takes images in order to make batches
num_threads	4	# threads for loading data
batch_size	1	input batch size
load_size	286	scale images to this size
crop_size	256	then crop to this size
max_dataset_size	inf	Maximum number of samples allowed per dataset
preprocess	resize_and_crop	scaling and cropping of images at load time
no_flip	false	if specified, do not flip the images
n_epochs	50	number of epochs with the initial learning rate
n_epochs_decay	50	number of epochs to linearly decay learning rate to zero
beta1	0.5	momentum term of adam
lr	0.0002	initial learning rate for adam
gan_mode	lsgan	type of GAN objective
pool_size	50	size of image buffer that stores previously generated images
lr_policy	linear	learning rate policy
lr_decay_iters	50	multiply by a gamma every lr_decay_iters iterations

Table 5.: Performance of models trained on building's pixel ratios from road shoulder perspectives.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.0698	0.211	-9.19	0.405
SVI panorama shifted with a naive method	0.0943	0.239	-12.757	0.210
CycleGAN panorama (A1)	0.00641	0.0569	0.065	0.473
Pix2Pix panorama (A1)	0.006	0.057	0.124	0.407
LightGBM panorama without GAN (A3)	0.00217	0.0308	0.683	0.829
LightGBM panorama CycleGAN (A5)	0.00222	0.0324	0.676	0.822
LightGBM panorama Pix2Pix (A5)	0.00264	0.0366	0.615	0.785
ResNet-50 panorama without GAN (A2)	0.00241	0.0339	0.648	0.806
ResNet-50 panorama CycleGAN (A4)	0.00246	0.0351	0.642	0.805
ResNet-50 panorama Pix2Pix (A4)	0.00245	0.0346	0.642	0.804
Road center SVI perspective	0.00997	0.0647	-0.448	0.533
SVI perspective shifted with a naive method	0.100	0.194	-13.704	0.103
CycleGAN perspective (A1)	0.0066	0.0544	0.0423	0.543
Pix2Pix perspective (A1)	0.00562	0.0495	0.184	0.509
LightGBM perspective without GAN (A3)	0.00195	0.0296	0.717	0.847
LightGBM perspective CycleGAN (A5)	0.00204	0.0302	0.704	0.839
LightGBM perspective Pix2Pix (A5)	0.00231	0.0327	0.665	0.824
ResNet-50 perspective without GAN (A2)	0.00204	0.0304	0.703	0.837
ResNet-50 perspective CycleGAN (A4)	0.00221	0.032	0.676	0.827
ResNet-50 perspective Pix2Pix (A4)	0.00222	0.0318	0.674	0.827

Table 6.: Performance of models trained on sky's pixel ratios from road shoulder perspectives.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.00589	0.0541	-0.49	0.371
SVI panorama shifted with a naive method	0.00748	0.0676	-0.887	0.294
CycleGAN panorama (A1)	0.00294	0.0392	0.257	0.614
Pix2Pix panorama (A1)	0.0014	0.0261	0.647	0.837
LightGBM panorama without GAN (A3)	0.000902	0.0205	0.772	0.88
LightGBM panorama CycleGAN (A5)	0.000874	0.0202	0.779	0.883
LightGBM panorama Pix2Pix (A5)	0.000915	0.0215	0.769	0.878
ResNet-50 panorama without GAN (A2)	0.000919	0.0224	0.768	0.877
ResNet-50 panorama CycleGAN (A4)	0.0015	0.0294	0.621	0.804
ResNet-50 panorama Pix2Pix (A4)	0.00294	0.043	0.256	0.604
Road center SVI perspective	0.00643	0.0513	-0.628	0.61
SVI perspective shifted with a naive method	0.00898	0.0734	-1.264	0.172
CycleGAN perspective (A1)	0.00311	0.04	0.213	0.627
Pix2Pix perspective (A1)	0.00139	0.0245	0.648	0.835
LightGBM perspective without GAN (A3)	0.000782	0.0193	0.802	0.897
LightGBM perspective CycleGAN (A5)	0.000789	0.0193	0.8	0.897
LightGBM perspective Pix2Pix (A5)	0.000917	0.0215	0.768	0.878
ResNet-50 perspective without GAN (A2)	0.000849	0.0215	0.785	0.887
ResNet-50 perspective CycleGAN (A4)	0.00136	0.0297	0.655	0.813
ResNet-50 perspective Pix2Pix (A4)	0.00304	0.0436	0.231	0.627

Table 7.: Performance of models trained on building's pixel ratios from sidewalk perspectives.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.065	0.192	-2.03	0.26
CycleGAN panorama (A1)	0.0237	0.103	-0.104	0.273
Pix2Pix panorama (A1)	0.0309	0.108	-0.438	0.153
LightGBM panorama without GAN (A3)	0.0125	0.0774	0.418	0.648
LightGBM panorama CycleGAN (A5)	0.0129	0.0777	0.399	0.635
LightGBM panorama Pix2Pix (A5)	0.0141	0.0845	0.341	0.586
ResNet-50 panorama without GAN (A2)	0.0156	0.0834	0.271	0.544
ResNet-50 panorama CycleGAN (A4)	0.0137	0.0861	0.361	0.615
ResNet-50 panorama Pix2Pix (A4)	0.0152	0.0857	0.29	0.556
Road center SVI perspective	0.0246	0.105	-0.131	0.262
CycleGAN perspective (A1)	0.0252	0.108	-0.16	0.251
Pix2Pix perspective (A1)	0.0253	0.103	-0.165	0.241
LightGBM perspective without GAN (A3)	0.0153	0.0809	0.295	0.552
LightGBM perspective CycleGAN (A5)	0.0147	0.0807	0.323	0.574
LightGBM perspective Pix2Pix (A5)	0.0177	0.0899	0.185	0.464
ResNet-50 perspective without GAN (A2)	0.0155	0.0871	0.289	0.545
ResNet-50 perspective CycleGAN (A4)	0.0165	0.0892	0.24	0.513
ResNet-50 perspective Pix2Pix (A4)	0.0147	0.0813	0.325	0.593

Table 8.: Performance of models trained on sky's pixel ratios from sidewalk perspectives.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.0151	0.086	-0.414	0.318
CycleGAN panorama (A1)	0.00857	0.0673	0.197	0.55
Pix2Pix panorama (A1)	0.00523	0.0471	0.509	0.773
LightGBM panorama without GAN (A3)	0.00465	0.0477	0.564	0.754
LightGBM panorama CycleGAN (A5)	0.0042	0.0447	0.606	0.78
LightGBM panorama Pix2Pix (A5)	0.00423	0.0452	0.603	0.782
ResNet-50 panorama without GAN (A2)	0.00505	0.0477	0.526	0.753
ResNet-50 panorama CycleGAN (A4)	0.0058	0.0557	0.456	0.682
ResNet-50 panorama Pix2Pix (A4)	0.007	0.0631	0.344	0.658
Road center SVI perspective	0.00948	0.0707	0.11	0.535
CycleGAN perspective (A1)	0.00922	0.0673	0.135	0.546
Pix2Pix perspective (A1)	0.00608	0.0508	0.429	0.74
LightGBM perspective without GAN (A3)	0.00445	0.0475	0.582	0.766
LightGBM perspective CycleGAN (A5)	0.00414	0.0445	0.612	0.785
LightGBM perspective Pix2Pix (A5)	0.00479	0.0461	0.551	0.758
ResNet-50 perspective without GAN (A2)	0.0052	0.05	0.512	0.76
ResNet-50 perspective CycleGAN (A4)	0.0067	0.0625	0.371	0.617
ResNet-50 perspective Pix2Pix (A4)	0.00798	0.0667	0.251	0.703

Table 9.: Performance of models trained on vegetation's pixel ratios from sidewalk perspectives.

Model	MSE	MAE	R-squared	Pearson's r
Road center SVI panorama	0.0381	0.149	-0.0943	0.47
CycleGAN panorama (A1)	0.0355	0.144	-0.0187	0.476
Pix2Pix panorama (A1)	0.0441	0.159	-0.265	0.577
LightGBM panorama without GAN (A3)	0.0112	0.0726	0.679	0.828
LightGBM panorama CycleGAN (A5)	0.0107	0.0748	0.694	0.835
LightGBM panorama Pix2Pix (A5)	0.0155	0.0914	0.554	0.753
ResNet-50 panorama without GAN (A2)	0.014	0.0878	0.599	0.776
ResNet-50 panorama CycleGAN (A4)	0.0166	0.102	0.524	0.74
ResNet-50 panorama Pix2Pix (A4)	0.0196	0.107	0.436	0.667
Road center SVI perspective	0.0313	0.131	0.106	0.451
CycleGAN perspective (A1)	0.0315	0.135	0.102	0.447
Pix2Pix perspective (A1)	0.0483	0.169	-0.378	0.52
LightGBM perspective without GAN (A3)	0.0131	0.0755	0.626	0.796
LightGBM perspective CycleGAN (A5)	0.0116	0.0771	0.669	0.819
LightGBM perspective Pix2Pix (A5)	0.0153	0.0898	0.562	0.757
ResNet-50 perspective without GAN (A2)	0.0113	0.0795	0.677	0.824
ResNet-50 perspective CycleGAN (A4)	0.015	0.092	0.572	0.757
ResNet-50 perspective Pix2Pix (A4)	0.0175	0.104	0.5	0.715