

Designing for Perception: Weather-Aware Streetscapes via Generative Modeling and Global Datasets

Haoyu Zhuang¹, Chenbo Zhao¹, Yoshiki Ogawa¹, Yoshihide Sekimoto¹

¹ Center for Spatial Information Science, The University of Tokyo, Japan - (hy.z, zhao cb, ogawa, sekimoto)@csis.u-tokyo.ac.jp

Keywords: Generative Modeling, Global Dataset, Street View Image, Subjective Perception.

Abstract

Urban streetscape evaluation has traditionally relied on computer vision methods to quantify semantic elements such as greenery, signage, and architectural features. However, these objective approaches often overlook subtle factors like lighting and atmospheric conditions that deeply affect human perceptions of beauty, safety, and ambiance. Moreover, existing expert-driven evaluations may not fully represent diverse public perspectives. To address these limitations, this study proposes a perception-oriented framework integrating generative AI and large-scale subjective feedback to investigate how weather conditions shape urban streetscape perception. Using TSIT-based image translation, we generated sunny, cloudy, and rainy versions of 88,000 street view images while preserving scene structure and semantics. Over 38,525 participants provided subjective ratings across 22 perceptual dimensions, such as beauty, cleanliness, safety, and warmth. Results revealed that sunny scenes scored higher than cloudy and higher than rainy scenes on the most of 22 dimensions, with notable improvements in brightness, neatness, and cleanliness. Rainy conditions elicited the highest negative ratings in dimensions like boredom and depression, while dimensions such as greenery and interestingness remained relatively stable across weather types.

1. Introduction

Recently, the evaluation of urban streetscapes has evolved significantly, with a growing emphasis on subjective perception-based methodologies that reflect how individuals experience and respond to their environments (Ogawa et al., 2024). This shift parallels broader trends in urban studies and environmental psychology, which increasingly recognize that the design and quality of public spaces have profound impacts on mental health, social interaction, and overall quality of life. The streetscape is not merely a physical corridor for movement but a complex sensory and emotional environment that influences how people feel, behave, and connect with their surroundings.

Traditionally, streetscape analysis has focused on quantifiable semantic elements—such as greenery, signage, street furniture, and architectural features—using computer vision techniques to identify and classify these components within urban imagery (Xu et al., 2024, Yap et al., 2023). Such approaches have provided urban planners and designers with efficient, objective metrics to evaluate physical characteristics of streets and to support data-driven decisions in city planning. For example, measuring the amount of visible greenery or the density of pedestrian infrastructure has been shown to correlate with walkability indices and perceived environmental quality.

However, while these semantic analysis methods have advanced our ability to assess physical attributes objectively, they often fail to account for the more subtle and complex factors that contribute to human perceptions of beauty, safety, comfort, and ambiance. Specifically, they tend to overlook the crucial influence of lighting conditions, weather variations, seasonal changes, and atmospheric effects on visual experience. These factors are often non-semantic yet deeply impactful at the perceptual level, shaping how spaces are emotionally and psychologically interpreted by users. For instance, the same street can appear inviting and lively under bright sunlight but dull or even threatening under gloomy skies.

Furthermore, existing approaches to evaluating visual quality and aesthetic merit in urban settings are often expert-driven, relying on architectural principles, design heuristics, or planning guidelines defined by professionals (Lee et al., 2024). While expert assessments offer valuable insights into formal aesthetics, structural coherence, and design functionality, they may not fully represent the diverse perspectives of ordinary citizens, who interact with urban spaces in more dynamic, spontaneous, and emotionally nuanced ways. People's everyday perceptions are shaped not only by compositional balance and spatial proportions but also by cultural background, personal experience, and contextual cues such as time of day and weather.

This discrepancy between expert judgment and public perception highlights the need for alternative frameworks that incorporate a broader range of subjective experiences and cultural contexts. Moreover, with the rise of smart city initiatives and human-centered urban design paradigms, there is a growing demand for methodologies that can systematically capture, quantify, and integrate user-centered perceptual data into planning and design processes. Without such integration, urban development risks remaining technocratic and detached from lived human experiences, potentially undermining the social and emotional sustainability of urban environments.

To address these specific research gaps, our study implements the following targeted measures:

1. Addressing the expert-centered paradigm in streetscape evaluation, we construct a more democratic and inclusive approach by collecting a large volume of real-world online survey data and training an effective subjective evaluation model. This model enables multi-dimensional assessment of streetscapes across 22 perceptual dimensions, integrating the perspectives of a broad range of everyday users rather than relying solely on expert judgments.
2. To overcome the limitations of conventional semantic analysis approaches that focus only on streetscape object features

while neglecting lighting and atmospheric ambiance, we leverage TSIT, a generative image-to-image translation model, to modify weather conditions in streetscape images without altering their semantic content. This allows us to systematically examine the effects of ambient light and weather conditions on people's subjective perceptions.

3. To enhance the generalizability and reliability of our analysis and minimize regional biases, we utilize a large-scale open-source global dataset that provides streetscape imagery across multiple countries and continents along with practical weather labels. This ensures that our findings reflect diverse urban contexts rather than being confined to a single geographical region.

2. Related Work

2.1 The Influence of Weather on SVIs

Weather conditions exert a profound influence on the visual characteristics and perceptual qualities of urban street view images (SVIs). Prior research has demonstrated that weather affects not only the brightness, contrast, and color composition of images but also shapes psychological impressions such as safety, attractiveness, and liveliness. Researchers analyzed the impact of weather on object detection accuracy in SVIs, revealing significant performance degradation under rainy and foggy conditions (Ibrahim et al., 2019). Similarly, Liu et al. (Liu et al., 2025) showed that sunny weather enhances perceived vibrancy and comfort, whereas cloudy and rainy conditions are often associated with dullness and negative affect. These findings align with environmental psychology studies, which indicate that ambient lighting and weather conditions modulate human mood and spatial cognition (He and Li, 2021). However, existing works primarily focus on the technical impacts of weather on image processing models or general psychological correlates, with limited exploration of systematic perceptual differences across multiple aesthetic and emotional dimensions in urban contexts.

2.2 Generation of Street View Image

Recently, diffusion-based generative models have emerged as state-of-the-art techniques for high-resolution and photorealistic image synthesis (Abdelraouf et al., 2022). Stable Diffusion (Rombach et al., 2022) in particular has demonstrated remarkable capabilities in generating diverse urban scenes conditioned on textual prompts, enabling flexible customization of architectural styles, vegetation layouts, and street configurations. Unlike GAN-based methods, diffusion models iteratively refine images through denoising processes (Zhang et al., 2023), achieving superior fidelity and fine-grained detail preservation. In the context of streetscape studies, Stable Diffusion and its variants have been employed to create hypothetical urban scenarios, visualize pedestrian-friendly designs, and generate synthetic training data for perception models (Li et al., 2024). However, current applications primarily focus on creating idealized or conceptual scenes rather than translating real street view images under varying conditions (Zhang and Yu, 2021). Integrating diffusion models into perceptual evaluation frameworks remains underexplored, yet holds promise for generating diverse, semantically coherent stimuli to investigate aesthetic, emotional, and behavioral responses to urban design elements at scale.

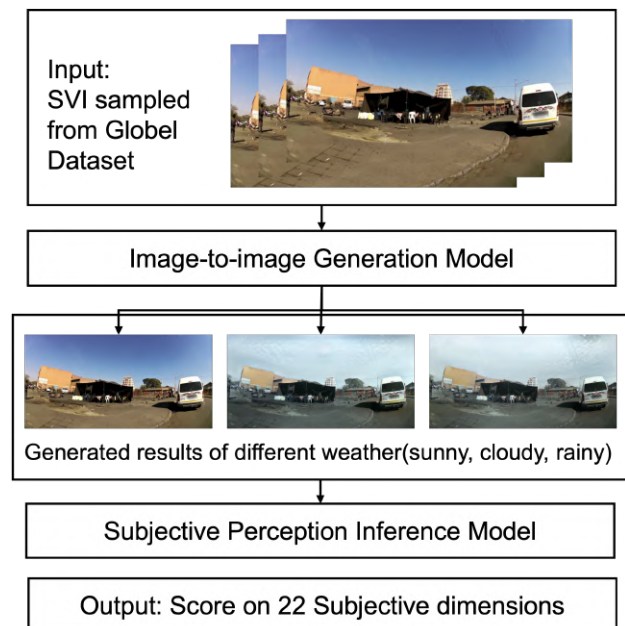


Figure 1. The Framework of research

The advancement of generative models has enabled controllable manipulation and synthesis of street view images under varying environmental conditions. Early approaches, such as CycleGAN (Chu et al., 2017), enabled unpaired image-to-image translation between weather domains (e.g. sunny-to-rainy). However, these methods often suffer from structural inconsistencies and low-fidelity texture details. Subsequent models, including TSIT (Two-Stage Image Translation) (Jiang et al., 2020), introduced a disentangled architecture to separately model structure and texture transformations, yielding improved realism and semantic preservation. In the urban computing field, generative models have been used to simulate street design interventions or predict future urban appearances, but rarely have they been applied to systematically examine weather-induced perceptual variations. The use of high-fidelity generative weather translation offers new opportunities for controlled experimentation on environmental influences, overcoming the limitations of natural dataset biases and enabling scalable perceptual evaluation under diverse conditions.

2.3 Subjective Perception of Street View Image

Subjective perception-based analysis has emerged as a critical complement to objective semantic evaluation in urban studies. Traditional computer vision research on SVIs focuses on extracting quantifiable features such as greenery ratio, façade types, or traffic elements (Wang et al., 2019, Yap et al., 2023, Hu et al., 2024). However, recent works emphasize the importance of incorporating human judgments to assess aesthetic, emotional, and psychological responses to streetscapes. For example, ?developed a crowdsourced dataset for measuring perceived safety from Google Street View images, while (Gu et al., 2025, Fan et al., 2025) constructed models to predict scenicness based on human ratings. Ogawa et al. (Ogawa et al., 2024) introduced a multi-dimensional subjective perception evaluation model covering 22 perceptual categories, highlighting diverse affective and aesthetic dimensions in urban imagery. Despite these advances, few studies (Zhao et al., 2025) integrate subjective perception models with generative transformations to systematically analyze how specific environmental factors such as weather shape multi-dimensional perceptual outcomes.

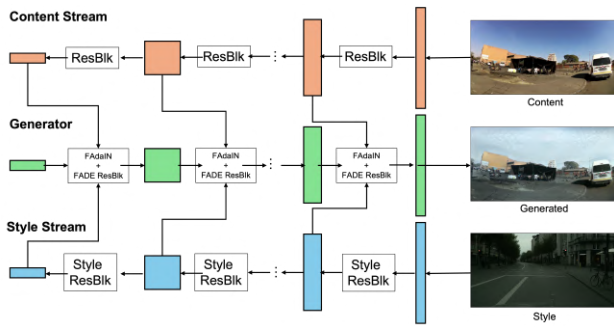


Figure 2. Generation Model Structure

Bridging this gap is essential for designing urban spaces that align with citizens' experiential and emotional needs.

3. Dataset

Our empirical study utilizes a dataset of 88,000 street view images, randomly sampled from two globally crowdsourced, open-access urban imagery platforms (Hou et al., 2024). This platform aggregates user-contributed street-level imagery from a wide range of geographic regions, climatic zones, cultural contexts, and imaging devices, ensuring that the dataset reflects the rich visual diversity of real-world urban environments.

These raw images served as the foundational input for both generative manipulation and perceptual evaluation. In the preprocessing stage, a subset of images was manually annotated with weather conditions, including sunny, rainy, and cloudy labels. Based on these manually labeled samples, this study employed a MaxVit multi-label classification model to automatically assign weather labels to the remaining large volume of images. The model's performance was evaluated using standard metrics such as Accuracy, Precision, Recall, and F1 Score (Hou et al., 2024).

As a result, in the image generation phase, we obtained a dataset containing nearly 88,000 globally sourced street-view images with reliable weather annotations. This large-scale labeled dataset not only facilitated condition-aware generative modeling but also enabled rigorous comparisons across diverse environmental contexts. Furthermore, it provided a strong foundation for analyzing how specific weather attributes influence human perceptions of streetscapes in different climatic, cultural, and urban settings worldwide.

For the training data of the subjective perception evaluation model, we used 610,017 street view images from Setagaya Ward, Tokyo, Japan. These images were standardized and served as labeled samples for subjective perception evaluation. During the annotation process, we conducted a large-scale questionnaire survey, with approximately 38,525 online participants contributing to the task. Each participant was randomly assigned a number of image pairs for subjective comparison, resulting in a total of nearly 8.8 million image-pair ratings (Ogawa et al., 2024). Furthermore, previous research has demonstrated that this dataset makes significant contributions to the optimization of urban streetscapes (Zhao et al., 2025). This large-scale, high-quality human feedback dataset provided a solid foundation for training the subjective perception evaluation model, ensuring its robustness and generalization across multiple subjective evaluation dimensions.

4. Methodology

To address these limitations, we propose a novel perception-oriented framework (Figure 1) that integrates generative artificial intelligence (AI) with large-scale subjective feedback across multiple perceptual dimensions. The core of our approach lies in applying a state-of-the-art image-to-image translation technique-TSIT (Two-Stage Image Translation) (Jiang et al., 2020)-to transform the weather(e.g., sunny, cloudy, rainy) conditions of street view images, while preserving the original scene's structure and semantics.

Following image generation, we employ a subjective rating model to systematically collect user-perceived scores across diverse emotional and aesthetic dimensions. These perceptual ratings reflect how environmental variations influence human judgment of visual scenes. Finally, we perform rigorous statistical analyses on the collected ratings to quantify the impact of weather conditions on subjective perception. This end-to-end pipeline—from controllable visual transformation via TSIT, to subjective perception evaluation model, and finally to statistical interpretation—provides a flexible and scalable methodology for investigating the perceptual effects of environmental ambiance in urban imagery.

4.1 Generative Model

To generate realistic environmental variations, we employ state-of-the-art diffusion-based generative models, trained on large-scale urban visual datasets. These models allow us to controllably manipulate weather attributes without altering the core layout or semantic content of each streetscape (Jiang et al., 2020).

By using a perception-oriented approach, we isolate the visual influence of ambient lighting, foliage coloration, atmospheric clarity, and other mood-affecting factors. This method leverages the strengths of generative AI to simulate highly plausible environmental scenarios, enabling a controlled and scalable framework for perceptual experimentation.

In this method (Figure 2), this research adopts the hinge loss as our adversarial loss. For the generator, the total loss consists of the hinge-based adversarial loss, a perceptual loss, and a feature matching loss. In the case of multi-scale discriminators, only the hinge-based adversarial loss is used to differentiate between real and generated images. The generator and discriminator are trained alternately in a minimax optimization framework.

The generator loss (Jiang et al., 2020) $\mathcal{L}_G$ and discriminator loss $\mathcal{L}_D$ are defined as:

$$\mathcal{L}_G = -\mathbb{E}[D(g)] + \lambda_P \mathcal{L}_P(g, x^c) + \lambda_{FM} \mathcal{L}_{FM}(g, x^s), \quad (1)$$

$$\mathcal{L}_D = -\mathbb{E}[\min(-1 + D(x^s), 0)] - \mathbb{E}[\min(-1 - D(g), 0)], \quad (2)$$

where $g = G(z_0, x^c, x^s)$ represents the output image generated from the input noise map z_0 , the content image x^c , and the style image x^s . The perceptual loss $\mathcal{L}_P$ measures the difference between feature representations extracted by the VGG-19 network. The feature matching loss $\mathcal{L}_{FM}$ aligns intermediate features at multiple layers of the discriminator. Here, λ_P and λ_{FM} denote the weights for the perceptual and feature matching losses (Jiang et al., 2020).

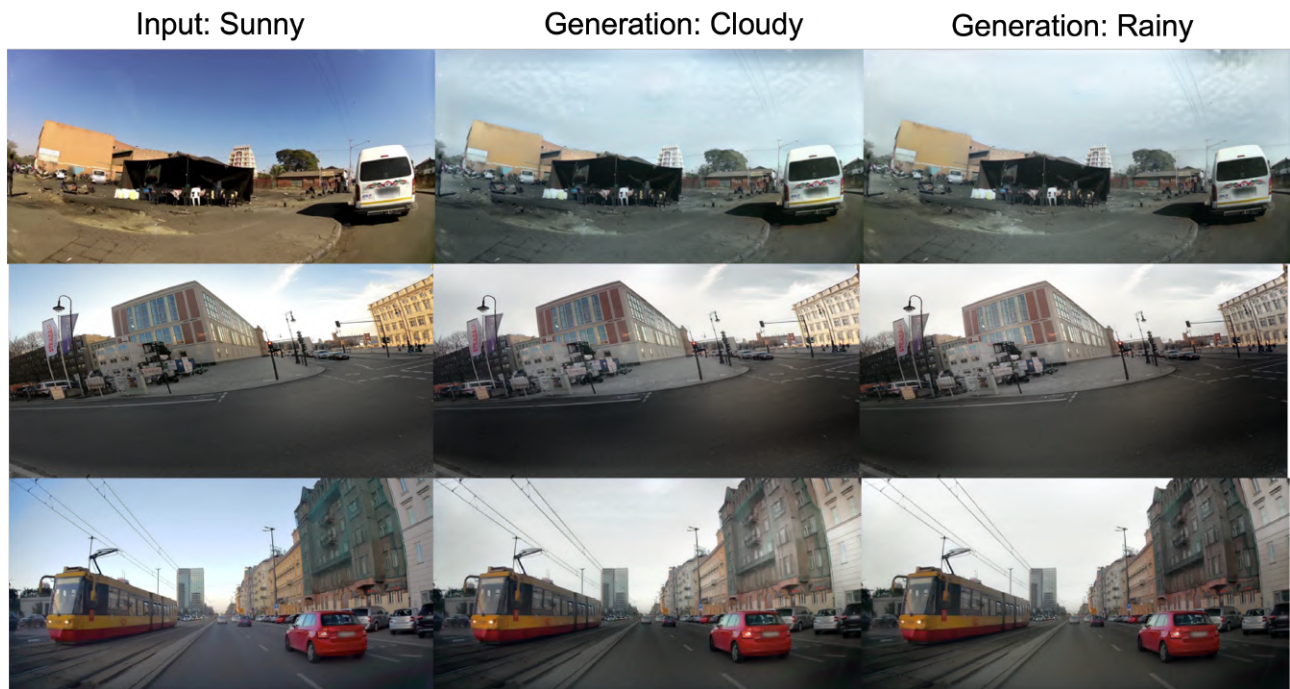


Figure 3. Generation Examples

This image-to-image translation model has been validated across multiple benchmark datasets, including BDD100K and ADE20K, demonstrating superior performance in image style transfer tasks such as day-to-night and sunny-to-rainy transformations. In particular, the model achieves significantly lower Fréchet Inception Distance (FID) scores compared to other baseline methods (Jiang et al., 2020), indicating its strong ability to generate photorealistic and perceptually consistent outputs.

Therefore, in our study, we deploy this state-of-the-art model to accomplish the critical task of transforming sunny street-view images into alternative weather conditions. By leveraging its proven generalization capability and robustness across diverse urban scenes, we ensure that environmental variations generated for perceptual experiments remain realistic and semantically faithful to the original streetscapes.

4.2 Subjective Perception Inference Model

We do not employ a predictive inference model in the traditional sense. Instead, we infer perceptual judgments based on large-scale human responses obtained through a comprehensive online questionnaire survey, in which over 38,525 participants provided subjective ratings on street view images (Ogawa et al., 2024).

This method formulates the subjective perception prediction task for street-view images as a multilabel classification problem, aiming to predict the differences across 22 subjective perception dimensions for each image pair simultaneously. The model employs a Siamese CNN architecture, where two images are input separately and processed through a ConvNeXt-B backbone pretrained on ImageNet with frozen weights to extract 1024-dimensional feature vectors. These vectors are then passed through a shared-weight layer to output 22-dimensional subjective perception scores for each image, and the difference between the two vectors is computed to obtain the final prediction results (Ogawa et al., 2024, Zhao et al., 2025).

The loss function is defined as the binary cross-entropy (BCE) between the predicted difference vector and the questionnaire ground truth. ConvNeXt-B was chosen as the backbone due to its balance between accuracy and computational efficiency (Ogawa et al., 2024). The Siamese structure enables direct comparison of subjective perception differences between two images, and upon training convergence, either branch can be used independently as an estimation model for subjective perception scores.

These ratings were collected across 22 carefully selected perceptual dimensions, such as perceived beauty, cleanliness, safety, vibrancy, warmth, and comfort (Ogawa et al., 2024). The dimensions were determined through extensive literature review and pre-study interviews to ensure both conceptual relevance and coverage of key emotional and aesthetic factors in urban perception.

The resulting subjective perception evaluation model, trained on this high-quality dataset, demonstrates strong performance, achieving over 85% accuracy across all dimensions in predicting user ratings (Ogawa et al., 2024, Zhao et al., 2025). This validates the reliability of the crowd-sourced data and the robustness of the inference process, providing a solid foundation for subsequent perceptual analysis.

5. Evaluation and Analysis

To ensure the validity of our findings, we evaluated the visual quality of the generated images using both automated metrics and manual inspection. The TSIT-based generation process produced high-fidelity, semantically consistent results with minimal artifacts (Figure 3). Specifically, the Fréchet Inception Distance (FID) scores were 6.6 for cloudy images and 7.1 for rainy images, indicating strong realism and supporting the reliability of our subsequent subjective perception analysis.

Using the annotated dataset, we performed advanced statistical analyses and regression modeling to uncover the relationships

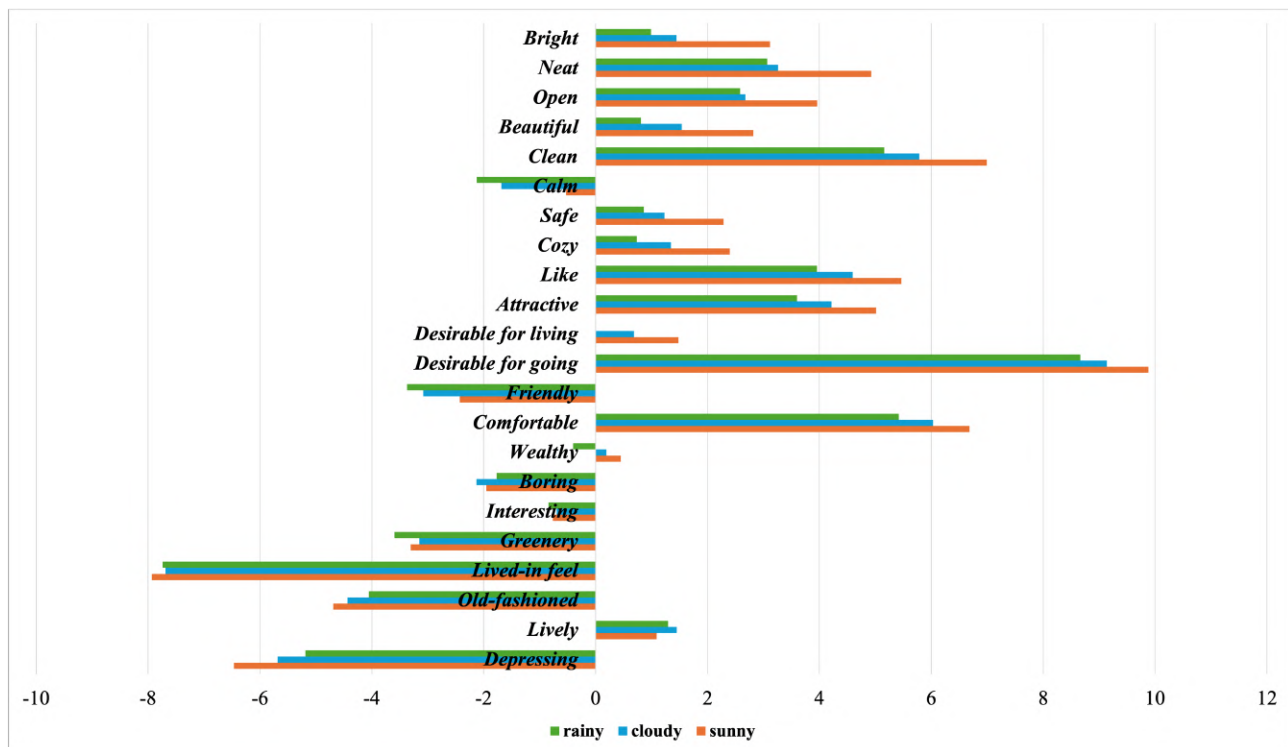


Figure 4. Subjective score of different weathers

between environmental variables and perceptual outcomes. Our results demonstrate that environmental ambiance—particularly factors such as light quality, shadow sharpness, foliage coloration, and atmospheric conditions—has a significant and measurable influence on aesthetic and emotional judgments of urban scenes.

At the overall level, we observe substantial differences in subjective scores across different weather conditions, which are in line with common intuition and prior environmental psychology findings. Images rendered under sunny conditions consistently receive significantly higher average subjective perception scores than those under cloudy or rainy settings. Specifically, across the sample of 88,000 images, the average score for sunny images is higher than that of cloudy images in most dimensions, while cloudy images still outperform rainy images (Figure 4). These findings clearly illustrate the dominant role of ambient brightness and visual clarity in shaping positive urban impressions and may reflect evolutionary psychological preferences for well-lit, visually accessible environments.

In addition to overall score comparisons, we conducted dimension-level analyses across the 22 perceptual categories to examine how specific environmental conditions influence distinct aspects of urban scene perception (Figure 5). Our findings reveal that sunny conditions lead to the most substantial rating increases over cloudy weather in the dimensions of bright, neat, open, beautiful, and clean, with all average improvements exceeding 110% (Figure 4). When comparing sunny to rainy conditions, the most pronounced increases are observed in bright, neat, cozy, beautiful, and clean, again indicating a strong positive perceptual shift under clear skies (Table 1). These results suggest that weather-induced brightness and visual comfort directly enhance evaluations of cleanliness, openness, and overall aesthetic appeal.

When comparing cloudy to rainy conditions, although the over-

all differences are less dramatic, cloudy scenes still achieve notably higher scores. In particular, the beautiful dimension shows an average increase of more than 240% (Table 1), while dimensions such as clean, like, and desirable for living also record improvements of over 200% (Table 1). This indicates that even moderate reductions in environmental gloominess can substantially improve people's affective and functional impressions of urban spaces.

Interestingly, in dimensions associated with negative emotional connotations, such as boring and depressing, rainy scenes consistently receive the highest ratings. This suggests that gloomy weather significantly enhances perceptions of dullness and emotional heaviness, potentially increasing feelings of discomfort or avoidance. Meanwhile, in certain dimensions—such as interesting and greenery—the scores remain relatively stable across sunny, cloudy, and rainy conditions, indicating that these perceptual aspects may be less sensitive to environmental ambiance. For example, greenery perception is likely influenced more by actual vegetative content than by atmospheric conditions, while perceived interestingness may depend on structural or semantic scene factors rather than transient weather variables.

Taken together, these results provide clear empirical evidence for the profound role of environmental ambiance in shaping subjective perceptions of urban streetscapes. They highlight the need for urban planners and designers to consider atmospheric and lighting factors in streetscape interventions aiming to promote aesthetic appreciation, psychological comfort, and spatial desirability. Moreover, the differential sensitivities across perceptual dimensions point to opportunities for targeted design strategies that address specific perceptual deficits under unfavorable weather conditions.

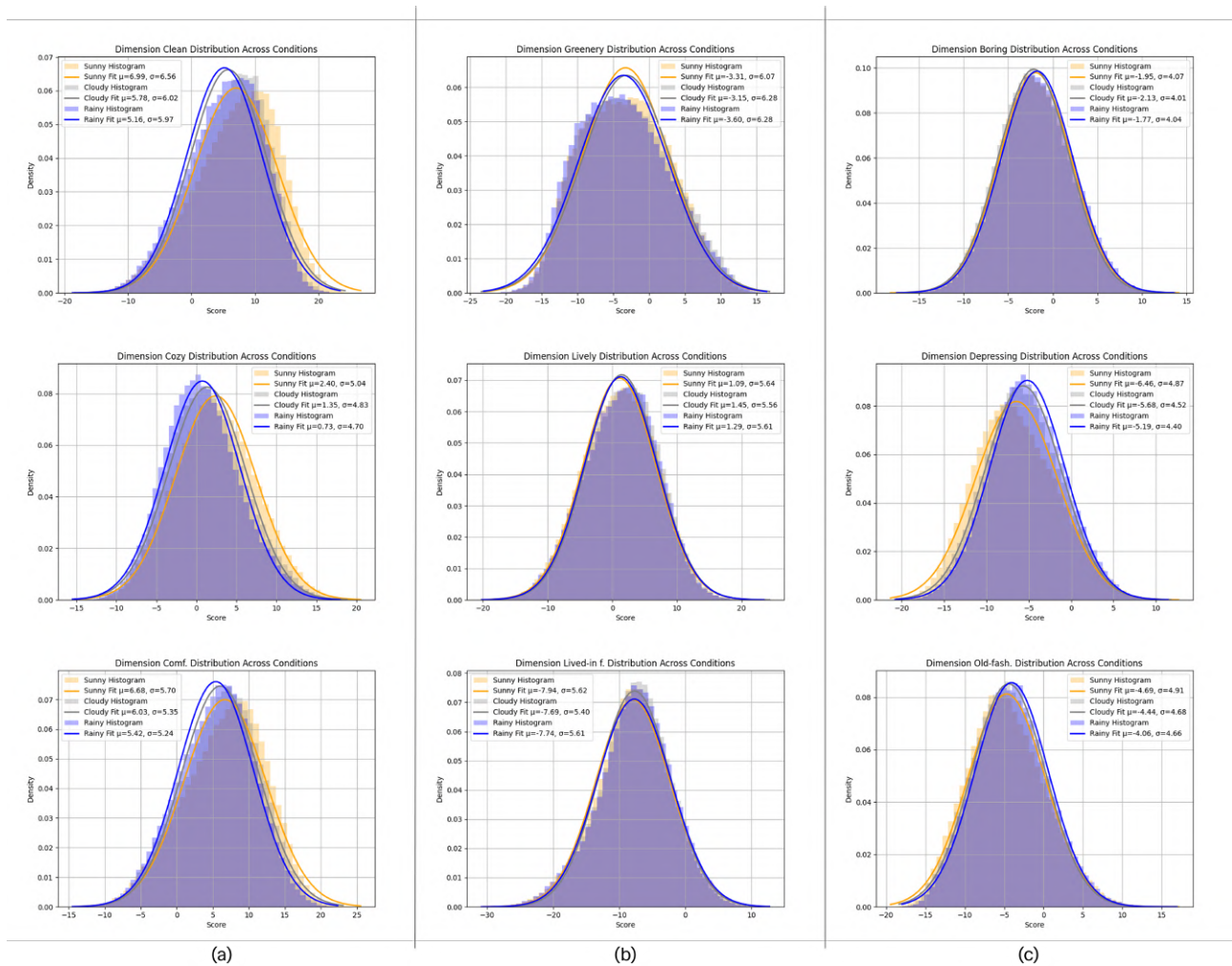


Figure 5. The distribution of 9 dimensions. (a) shows that the scores of sunny SVIs are higher than others. (b) shows in these dimensions, the scores of different weather are almostly same. (c) shows in these dimensions the scores of sunny SVIs are lower than others

6. Limitation

Despite the promising results demonstrated in this study, several limitations remain. Although the TSIT-based image generation framework generally produces visually realistic and semantically consistent weather translations, it exhibits notable instability under certain conditions. For example, when processing scenes involving complex lighting interactions or occluded weather-specific features, the model occasionally generates unrealistic artifacts or fails to achieve complete environmental transformation. These limitations highlight that achieving universally robust and precise weather translation across diverse street view images remains an open challenge for future generative model development.

Another important limitation concerns the cultural scope of the subjective perception evaluation model. The ratings used for training were collected exclusively through Japanese-language questionnaire surveys, reflecting the aesthetic judgments and perceptual tendencies of Japanese residents. As a result, the model's inferences may not generalize to populations with different cultural backgrounds, urban experiences, and architectural preferences. This issue is further compounded by the fact that the underlying street view imagery is sourced entirely from Setagaya Ward in Tokyo, a geographically and architecturally

specific area, potentially introducing location-based biases into perceptual evaluations.

Furthermore, the process of weather annotation in the dataset partially relied on manual labeling, which introduces inherent ambiguities due to the absence of standardized definitions distinguishing similar weather categories such as “sunny” and “partly cloudy” or “cloudy” and “rainy.” Such inconsistencies may affect both the conditional image generation process and the accuracy of perceptual comparison analyses. These observations underscore the necessity of developing rigorous, objective, and reproducible environmental labeling protocols to support more precise and generalizable computational perception studies in the future.

7. Conclusion

This study proposes a novel streetscape perception analysis framework by integrating a multi-dimensional subjective evaluation model with generative weather transformation techniques. Methodologically, we first constructed a large-scale subjective evaluation model covering 22 perceptual dimensions based on real-world online survey data. This enables streetscape evaluation to move beyond expert judgments and fully reflect the au-

Table 1. Subjective score of different weathers(* shows the highest score.)

	Depressing	Lively	Old-fashioned	Lived-in feel	Greenery	Interesting	Boring	Wealthy	Comfortable	Friendly	Desirable for going
Sunny	-6.46	1.09	-4.69	-7.94	-3.31	-0.76	-1.95	0.45*	6.68*	-2.43*	9.88*
Cloudy	-5.68	1.45*	-4.43	-7.69*	-3.15*	-0.69*	-2.13	0.19	6.03	-3.08	9.14
Rainy	-5.19*	1.29	-4.05*	-7.74	-3.59	-0.84	-1.77*	-0.41	5.42	-3.37	8.66

	Desirable for living	Attractive	Like	Cozy	Safe	Calm	Clean	Beautiful	Open	Neat	Bright
Sunny	1.48*	5.01*	5.46*	2.39*	2.29*	-0.53*	6.99*	2.81*	3.96*	4.92*	3.12*
Cloudy	0.69	4.21	4.59	1.35	1.23	-1.69	5.78	1.54	2.68	3.26	1.45
Rainy	0.01	3.61	3.96	0.73	0.86	-2.13	5.16	0.81	2.58	3.07	0.99

thetic experiences of everyday users, promoting a more democratic and inclusive approach to urban perception analysis.

Secondly, we leveraged TSIT, an image-to-image generative model, to transform weather conditions in streetscape images without altering their semantic content. This allows us to systematically and controllably analyze the effects of different weather conditions on streetscape perception, overcoming previous limitations due to the uncontrollability of natural weather for experimental validation.

Additionally, we utilized a global streetscape dataset spanning multiple countries and continents, combined with objective weather labels, to enhance the generalizability and external validity of our analyses. This approach avoids the regional biases associated with single-city or single-country datasets, ensuring that our findings are applicable not only to specific areas but also provide valuable references for streetscape design and urban management on a global scale.

Future research should first expand the subjective perception dataset to include multilingual and cross-cultural survey data, capturing a wider range of aesthetic and emotional responses from participants in different sociocultural contexts. Second, integrating more advanced and generalizable image-to-image translation models capable of finer weather gradations, such as diffusion-based or hybrid transformer-diffusion architectures, may improve visual realism and controllability. Third, incorporating objective weather metadata, including temperature, humidity, and precipitation intensity obtained from meteorological APIs, will enable precise alignment between visual conditions and physical environmental parameters. Finally, future studies could extend the framework beyond weather conditions to integrate other environmental variables such as time of day, seasonal changes, air pollution levels, and pedestrian density.

References

- Abdelraouf, A., Abdel-Aty, M., Wu, Y., 2022. Using vision transformers for spatial-context-aware rain and road surface condition detection on freeways. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 18546–18556.
- Chu, C., Zhmoginov, A., Sandler, M., 2017. Cyclegan, a master of steganography. *arXiv preprint arXiv:1712.02950*.
- Fan, Z., Feng, C.-C., Biljecki, F., 2025. Coverage and bias of street view imagery in mapping the urban environment. *Computers, Environment and Urban Systems*, 117, 102253.
- Gu, Y., Quintana, M., Liang, X., Ito, K., Yap, W., Biljecki, F., 2025. Designing effective image-based surveys for urban visual perception. *Landscape and Urban Planning*, 260, 105368.
- He, N., Li, G., 2021. Urban neighbourhood environment assessment based on street view image processing: A review of research trends. *Environmental Challenges*, 4, 100090.
- 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.
- Hu, Y., Qian, F., Yan, H., Middel, A., Wu, R., Zhu, M., Han, Q., Zhao, K., Wang, H., Shao, F. et al., 2024. Which street is hotter? Street morphology may hold clues-thermal environment mapping based on street view imagery. *Building and Environment*, 262, 111838.
- Ibrahim, M. R., Haworth, J., Cheng, T., 2019. WeatherNet: Recognising weather and visual conditions from street-level images using deep residual learning. *ISPRS International Journal of Geo-Information*, 8(12), 549.
- Jiang, L., Zhang, C., Huang, M., Liu, C., Shi, J., Loy, C. C., 2020. Tsit: A simple and versatile framework for image-to-image translation. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, Springer, 206–222.
- Lee, C., Park, A. H., Lee, H., Bratman, G. N., Hankey, S., Li, D., 2024. Measuring urban nature for pedestrian health: Systematic review and expert survey. *Landscape and Urban Planning*, 250, 105129.
- Li, W., He, J., Ye, J., Zhong, H., Zheng, Z., Huang, Z., Lin, D., He, C., 2024. Crossviewdiff: A cross-view diffusion model for satellite-to-street view synthesis. *arXiv preprint arXiv:2408.14765*.
- Liu, Y., Wang, Z., Ren, S., Chen, R., Shen, Y., Biljecki, F., 2025. Physical urban change and its socio-environmental impact: Insights from street view imagery. *Computers, Environment and Urban Systems*, 119, 102284.

Ogawa, Y., Oki, T., Zhao, C., Sekimoto, Y., Shimizu, C., 2024. Evaluating the subjective perceptions of streetscapes using street-view images. *Landscape and Urban Planning*, 247, 105073.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Wang, R., Liu, Y., Lu, Y., Zhang, J., Liu, P., Yao, Y., Grekousis, G., 2019. Perceptions of built environment and health outcomes for older Chinese in Beijing: A big data approach with street view images and deep learning technique. *Computers, Environment and Urban Systems*, 78, 101386.

Xu, Y., Tong, H., Liu, J., Su, Y., Li, M., 2024. An Assessment of the Urban Streetscape Using Multiscale Data and Semantic Segmentation in Jinan Old City, China. *Buildings*, 14(9), 2687.

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(6), 1416–1437.

Zhang, F., Yu, L., 2021. Street view imagery: Methods and applications based on artificial intelligence. *National Remote Sensing Bulletin*, 25(5), 1043–1054.

Zhang, L., Rao, A., Agrawala, M., 2023. Adding conditional control to text-to-image diffusion models. *Proceedings of the IEEE/CVF international conference on computer vision*, 3836–3847.

Zhao, C., Ogawa, Y., Chen, S., Oki, T., Sekimoto, Y., 2025. Street Space Quality Improvement: Fusion of Subjective Perception in Street View Image Generation. *Information Fusion*, 103467.