Green Sense

Noam Arian , Adi Inbar, Dor Halevy

[1] Affiliation 1; e-mail@e-mail.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Abstract

Monitoring the quality of urban green spaces is critical for urban resilience, resident wellbeing and efficient maintenance planning. Green Sense is a computer vision pipeline designed to assess green space quality automatically from street level imagery, distinguishing between three conditions of vegetation: Healthy, Dried and Contaminated. A labelled synthetic dataset has been constructed using a visual generative AI model (FIBO) configured to mimic typical street camera views. Methodologically, the work follows a staged design. First, a baseline classifier that replicates the approach of the study Green Space Quality Analysis Using Machine Learning Approaches is trained on the synthetic dataset to provide a comparable reference point. Second, a structured model improvement pipeline is applied, combining hyperparameter tuning, exploration of alternative architectures and refined data preparation, with all configurations evaluated on both the internal test set and additional generated images that simulate street camera conditions. The best performing configuration is designated as the Green Sense model. In the next stage, the selected model will be evaluated on authentic street camera images from a municipal authority, in order to compare synthetic and real world performance and to assess readiness for practical deployment in urban green space monitoring.

Keywords: Urban green spaces, street level imagery, synthetic dataset, generative AI, image classification, deep learning, transfer learning, hyperparameter tuning, environmental monitoring, municipal maintenance

1. Introduction

Urban green spaces are a critical component of modern cities, shaping residents' quality of life, social cohesion and public health. As these areas age and intensify in use, municipalities face growing pressure to systematically monitor their cleanliness, vegetation quality and overall maintenance level at scale. In parallel, advances in computer vision and spatial analysis have enabled new approaches to automatically classify and evaluate green spaces in civil environments, moving beyond traditional manual surveys. This literature review will first show the societal importance of clean and well maintained urban greenery and existing monitoring practices, and then examine the main methodological frameworks and models used to classify green spaces from visual and spatial data.

According to Bedimo Rung, Mowen and Cohen, urban green spaces and parks provide accessible settings for regular physical activity, which is associated with reduced risks of obesity, cardiovascular disease, diabetes and other chronic conditions. They also report psychological, social and environmental benefits of green spaces, including improved mood, lower stress and fewer symptoms of depression and anxiety, stronger neighborhood ties and social capital, and contributions to cleaner air and mitigation of urban heat [1].

Kaplan's evaluation of a "vest pocket" park in downtown Ann Arbor shows that even a very small green space in a dense business district functions as a valued refuge that offers quiet, rest and visual relief. Users reported that simply having this planted, shaded spot nearby improved their daily experience of the city, highlighting the strong restorative benefits of accessible greenery in urban cores [2].

According to Madureira et al., the most valued green space characteristics are cleanliness and good maintenance, richness in plant species, the presence of water bodies, sufficient benches and a generally tranquil atmosphere. Cleanliness and maintenance stand out as the single most important attribute across the three cities, more important than park size, parking or high visitor numbers [3].

Greenspaces are increasingly recognized as critical urban infrastructure rather than a "nice to have" amenity, because they support physical and mental health, reduce health inequalities and help cities adapt to challenges such as air pollution, heat and flooding. In response, municipalities and their partners now treat parks and wider green infrastructure as natural capital: they use spatial planning tools, inclusive design, community programs and formal valuation methods to demonstrate benefits and to prioritize investment. Funding typically comes from a mix of local authority budgets, national health and environment strategies, and dedicated mechanisms such as the Community Infrastructure Levy (a planning charge on new developments used to fund local infrastructure, including parks and green spaces) and Section 106 developer contributions (site specific legal agreements where developers fund or provide infrastructure to mitigate the impacts of their projects), as well as external grants from bodies like the National Lottery Heritage Fund and National Trust. While the review stresses that budgets are under pressure, it shows that targeted investment in high quality, accessible greenspace is seen as a relatively low cost way to deliver health, social and environmental priorities [4].

This paper [5] contrast conventional satellite based indicators of urban green cover, particularly the Normalized Difference Vegetation Index (NDVI), with a human scale measure of street greenery that better reflects residents' daily visual exposure. Using Google Street View imagery and a SegNet convolutional network, they segment vegetation at pixel level and compute a Green View Index (GVI) for each street segment, which is then classified into low, medium and high greenery based on expert judgements. This GVI based metric is integrated with space syntax measures of street accessibility to derive indicators of "daily accessed greenery," capturing where people are most likely to experience greenery during routine pedestrian and commuting trips. When compared with NDVI at planning area scale, the daily accessed greenery indicators show only partial correspondence, leading the authors to conclude that top down NDVI alone cannot represent the benefits actually enjoyed by city residents and should be complemented by street level, machine learning based assessments of visible greenery in planning practice.

In the study “Deep Green Diagnostics Urban Green Space Analysis Using Deep Learning and Drone Images,” [6] the authors present an operational framework for fine grained assessment of urban vegetation health and surface contamination using high resolution aerial imagery. The work addresses the limitation of satellite based indices such as NDVI that only capture overall vegetation quantity and cannot distinguish between healthy, dry or polluted green areas. To overcome this, the authors construct a dataset of geo referenced RGB images acquired by a DJI Phantom 4 drone flown over four distinct urban environments in Mexico. The raw images are tiled into patches of 200 by 200 pixels, from which 9 901 samples are manually labelled into eight classes that jointly encode vegetation condition healthy, dry, unhealthy or no vegetation and the presence or absence of visible solid waste contamination. On this dataset they train a convolutional neural network followed by a multilayer perceptron classifier that learns discriminative visual features directly from the image patches and assigns each tile to one of the eight semantic categories. The resulting system achieves around 72 percent accuracy on a held out test set, with particularly strong performance for healthy vegetation and no vegetation classes, and demonstrates reasonable transferability when retrained on small samples from a new area. Overall, this article shows that deep learning applied to drone imagery can provide spatially detailed, operational information on both vegetation condition and urban cleanliness, and offers an open source prototype that municipal or health authorities can adapt for monitoring and decision support.

The article “Green Space Quality Analysis Using Machine Learning Approaches” [7] develops a full pipeline for assessing the visual quality of urban green spaces using machine learning on eye level images. It defines quality through conditions such as cleanliness, maintenance and dryness, builds a labelled dataset with three classes healthy, dried and contaminated, and compares several transfer learning image classification models, selecting ResNet50 as the best performing model and deploying it as a practical web tool for green space monitoring. Methodologically, the study follows the CRISP DM lifecycle, collecting 944 smartphone photos of Kuala Lumpur green spaces in 2022, manually annotating them into the three quality classes, and then addressing class imbalance through data augmentation using flips, rotations, brightness and contrast changes and CLAHE to obtain a balanced dataset of 986 images. The augmented images are resized to 224 by 224 pixels and split into training, validation and test sets, and nine ImageNet pretrained convolutional networks ResNet50, ResNet101, DenseNet201, VGG16, VGG19, XCeption, MobileNet, InceptionResNetV2 and EfficientNetB7 are fine tuned using Adam optimisation, categorical cross entropy loss, early stopping and learning rate reduction, with performance evaluated via accuracy, precision, recall, F1 score, Cohen Kappa and ROC AUC. ResNet50 is chosen as the overall best model, reaching about 99 percent accuracy together with almost perfect precision, recall and ROC AUC, and therefore selected for deployment in a Streamlit web application that allows users to upload or capture a photo and receive an automatic quality label, while the authors conclude that such transfer learning pipelines on relatively small eye level datasets can already deliver operational tools for green space maintenance support and point to mobile deployment and segmentation based models as important future directions.

2. Materials and Methods

This chapter outlines the methodological pipeline of the Green Sense project, organized to ensure a logically consistent flow from synthetic dataset construction through model improvement and validation.

A labelled image dataset was constructed to simulate street-camera views of urban green spaces. Approximately 300 images were generated for each of three classes: Healthy, Dried, and Contaminated. Healthy denotes vital, well-maintained vegetation, Dried represents visible lack of irrigation or plant vitality, and Contaminated refers to vegetation affected by litter, waste, or similar pollution. The images were generated using the FIBO visual generative model, which is configured to mimic realistic street-level perspectives with variation in lighting, distance, camera angle, and surrounding urban environment. FIBO also allows fine-grained control over scene composition, object placement, color distribution, and visual layout, ensuring that the synthetic dataset closely mirrors the characteristics, visual style, and structural patterns of the images used in the original dataset described in the source paper.

Once generated, the synthetic dataset underwent a structured quality assessment. The quality of the synthetic image dataset was evaluated using four complementary criteria. First, visual fidelity and perceptual quality were assessed, both manually and automatically, to ensure that the images appear realistic and similar to those in the dataset used in the original paper. Second, dataset diversity was checked to confirm that all classes and variations are represented, preventing over-representation of certain conditions or omission of rare cases. Third, a machine-learning utility check was performed using TSTR (Train on Synthetic, Test on Real), to ensure that models trained on the synthetic data generalize well to the original dataset. Finally, basic statistical similarity was verified by comparing key features, such as class proportions and color distributions, to ensure that the synthetic dataset preserves essential properties of the original data. If the dataset failed to meet any of these criteria, an iterative refinement process was applied before proceeding to model training. Such refinements included adjusting generative constraints or prompts to better match the original dataset, targeted oversampling of underrepresented conditions or classes, latent-space editing to adjust texture, illumination, or vegetation appearance, diversity-aware sampling to prevent mode collapse, diffusion-based post-processing to reduce artifacts, and annotation refinements through multi-annotator consensus or automated consistency checks. The dataset was then re-evaluated, and this loop continued until all quality metrics were satisfied, ensuring strong alignment with the original dataset characteristics.

After obtaining a validated synthetic dataset, the original models and methodology from the source paper were fully reconstructed to ensure methodological alignment and reproduce previously reported results on the original dataset. These replicated models were also evaluated on the synthetic dataset to establish a baseline before any enhancements were introduced. Following baseline replication, the model improvement pipeline was implemented in three stages. First, hyperparameters including learning rate, batch size, number of epochs, and regularization strength were tuned on the validation set, and the optimal configuration was re-evaluated on both the internal test set and the synthetic street-camera dataset. Second, multiple alternative architectures, such as different convolutional backbones and efficient modern networks, were trained under the tuned hyperparameters and compared on the same evaluation sets. Third, the preprocessing pipeline was refined through data augmentations including random crop, flip, rotation, color jitter, alternative normalization schemes, and class balancing using class-weighted loss or oversampling where appropriate. The configuration achieving the highest overall performance across both evaluation sets was designated as the Green Sense mode.

Finally, the selected Green Sense model was evaluated on authentic street-camera images provided by a municipal authority. Frames containing public green spaces were labeled using the same protocol as for the synthetic dataset, enabling direct comparison between synthetic and real-world performance. This evaluation demonstrated the model’s readiness for practical deployment, confirming that the synthetic data, combined with iterative quality assurance and model refinement, produces robust and generalizable predictions in real urban environments.

References

P. A. J. M. P. D. A. C. M. Ariane L. Bedimo-Rung, "The Significance of Parks to Physical Activity and Public Health".

[2] R. Kaplan, "EVALUATION OF AN URBAN VEST-POCKET PARK °," USDA FOREST SERVICE.

[3] F. N. ,. J. V. O. a. T. M. Helena Madureira, "Preferences for Urban Green Space Characteristics: A Comparative Study in Three Portuguese Cities," enviornments.

[4] "Improving access to greenspace A new review for 2020," Public Health England.

[5] S. L. Donghwan Ki, "Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning," Landscape and Urban Planning.

[6] H. C. C. A. D. P. L.-J. I. A. V.-M. a. M. S. S.-C. Marco A. Moreno-Armendáriz, "Deep Green Diagnostics: Urban Green Space Analysis Using Deep Learning and Drone Images," sensors, 2019.

[7] Z. R. a. N. Z. Jaloliddin Rustamov, "Green Space Quality Analysis Using Machine Learning Approaches," sustainability.

[8] ∗. D. R. Y. L. X. S. Y. Z. W. Z. T. Z. Yu Yea, "Measuring daily accessed street greenery: A human-scale approach for informing better urban planning practices," Landscape and Urban Planning.