DEVISE: Depth-Aware Vegetation Indexing System

Cantay Caliskan Zhizhuang Chen Linglan Yang Junjie Zhao Mingzhen Zhang

Goergen Institute for Data Science and AI, University of Rochester, Rochester, NY, USA

July 29, 2025

Abstract

Urban greenness is most meaningful at the pedestrian scale, yet most current metrics rely on coarse satellite imagery or street-view indices that ignore depth. We introduce **DEVISE**—the *Depth-Aware Vegetation Indexing System*—an end-to-end pipeline that combines Mask2Former panoptic segmentation, MiDaS monocular depth estimation, and HSV filtering to compute distance-weighted canopy, grass, and understory shares from Google Street View (GSV) images. Applied to **1,441,746** panoramas collected every 20 m along the road networks of **60** U.S. swing counties (2004–2024), *DEVISE* produces the first county-month panel of street-level greenness spanning two decades.

DEVISE improves on prior approaches in three key respects. First, relative to the Panorama-View Green View Index (PVGVI) of Zhang et al. (2023), it replaces a CNN backbone with a transformer, introduces depth weighting, and *disaggregates* vegetation classes, thereby eliminating far-field bias and reducing color-based false positives. Second, compared with the scalable Green View Index framework of Xia et al. (2021), *DEVISE* restores semantic precision in cluttered scenes and reduces over-estimation by pairing depth cues with spectral post-processing. Third, unlike the tree-species detector of Choi and Lee (2022), which is accurate but computationally demanding, *DEVISE* prioritizes longitudinal scalability, processing the national panel in six months of cloud compute while preserving perceptual realism.

This study also offers a quantitative demonstration of *DEVISE*'s capabilities. Regression analyses show that a one-percentage-point increase in the most recent Democratic presidential vote share is linked to a **0.52–0.71 percentage-point** gain in depth-adjusted total greenness after controlling for climate and socioeconomic covariates; this association remains robust with county and month fixed effects ($R^2 \approx 0.91$–0.99). These findings imply that partisan alignment translates into observable street-level environmental benefits, highlighting the political significance of urban forestry.

To facilitate replication and extension, all code, trained weights, and a command-line interface will be released as the open-source `devise` *Python* package. By integrating transformer segmentation, depth-aware weighting, and rigorous social-science validation, *DEVISE* establishes a new benchmark for scalable, policy-ready audits of perceptual equity in urban environments.

1 Introduction

1.1 Trees, Health, and Urban Inequality

> *"To exist as a nation, to prosper as a state, to live as a people, we must have trees."* —Theodore Roosevelt

Roosevelt's nineteenth-century exhortation has never felt more urgent. By 2023 extreme heat had become the deadliest weather hazard in the United States, and public-health projections warn of annual fatalities approaching 100 000 by 2100 if warming trends continue (Weiss, 2023). Urban trees mitigate these risks by shading asphalt, lowering surface temperatures, filtering pollutants, and easing psychological stress—functions that translate directly into avoided emergency-room visits and lower energy demand.

Yet access to such life-saving infrastructure remains profoundly unequal. Updated *Tree Equity Score* data released in 2023 now cover 260 million urban residents and confirm that low-income and minority neighborhoods are still missing up to half of the canopy enjoyed by wealthier blocks (American Forests, 2023). These gaps routinely translate into summertime air temperatures 5–6 °C hotter than city averages and into higher rates of asthma and heat stress for already vulnerable populations.

Governments have begun to respond at scale. The Inflation Reduction Act dedicated more than $1 billion to urban and community forestry, seeding hundreds of local grant programs aimed at achieving "tree equity" street by street (U.S. Department of Agriculture, 2024). Philadelphia, for example, received $12 million in 2024 to accelerate its ten-year *Philly Tree Plan*, which targets a 30 % canopy goal in every neighborhood.

Progress, however, is uneven. In early 2025 the U.S. Forest Service abruptly canceled a $75 million grant to the Arbor Day Foundation after political attacks on diversity-and-equity language, halting plantings in New Orleans' Lower Ninth Ward and 100 other projects nationwide (Pineda, 2025). The episode underscores how environmental justice initiatives remain vulnerable to shifting political winds even as climate risks accelerate.

Comparable stories are unfolding outside the United States. Valencia, Spain, opened the first ten of a planned network of "climate shelters" in 2024—shaded, tree-filled plazas kept open all summer irrespective of national heat alerts—to protect residents from increasingly frequent Mediterranean heat waves (Dogan, 2025). Citizen-science "thermal walks" conducted during İstanbul's June 2025 heatwave recorded street-level temperatures exceeding official station readings by more than 4 °C, with participants repeatedly pointing to a lack of mature street trees in working-class districts (Özçelikel, 2025).

1.2 The Challenge of Measuring What Residents Actually See

Most greening policies are evaluated with satellite-derived metrics such as the Normalized Difference Vegetation Index (NDVI). While invaluable for city-wide trend analysis, NDVI and related indices lack (i) the resolution to distinguish vegetation types, (ii) sensitivity to visual occlusion, and (iii) fidelity to the pedestrian perspective that shapes everyday experience (Small, 2001; Sun et al., 2021). Consequently, satellite averages can mask hyper-local disparities that matter most for environmental justice.

Google Street View (GSV) bridges this observational gap by providing georeferenced, panoramic imagery at eye level. The Green View Index (GVI) introduced by Li et al. (2015) converts each panorama into a vegetation share, inspiring extensions such as the Urban Neighborhood Green Index (UNGI; Gupta et al., 2012), the Baidu GVI (BGVI; Yu

et al., 2018), and hue-threshold approaches using UAV imagery (Hassanein and Abdelatif, 2018). However, early pipelines relied on color rules or small, manually labeled samples, leading to false positives from painted façades and limited scalability (Lauko et al., 2020; Sun et al., 2021).

1.3 The Importance of Depth-Aware Indices

Deep learning has markedly improved street-level greenness mapping. Semantic-segmentation frameworks now operate at city scale (Xia et al., 2021); hybrid models identify individual tree species (Choi and Lee, 2022); and patch-based training reduces computational cost (Kameoka and Hirashima, 2022). Nonetheless, even state-of-the-art networks can miss low vegetation (e.g., lawns, flower beds) and misclassify artificial greens (Aikoh and Takahashi, 2023). Moreover, almost all existing methods treat greenness as a two-dimensional surface. Neglecting depth and occlusion weights a distant tree the same as foliage brushing a pedestrian's shoulder, biasing exposure estimates in dense urban cores (Yu et al., 2022).

Recent architectural advances—Mask2Former's transformer-based segmentation (Cheng et al., 2022) and the panoramic S-G-S-S dataset underpinning PVGVI (Zhang et al., 2023)—bring higher spatial fidelity, yet depth awareness remains rare. Consequently, computational social scientists still lack a scalable tool that (i) segments vegetation accurately, (ii) suppresses non-vegetal artifacts, and (iii) weights greenery by perceptual distance.

1.4 Closest Competitors: Green Area Measurement Methods and Past Contributions

Among the rapidly expanding body of research on streetscape greenness, three studies constitute the closest competitors to our proposed system. Each advances the state of the art in important ways, yet all leave critical gaps that *DEVISE* seeks to address.

Zhang et al. (2023): Panorama View GVI (PVGVI). Zhang and colleagues introduce the Panorama View Green View Index (PVGVI), which estimates streetscape greenness from high–resolution spherical imagery using DeepLabV3+, a convolutional neural–network (CNN) segmentation model. Their panoramic design improves visual consistency and spatial coverage relative to earlier GSV studies and, among published work, aligns most closely with the geographic scale and policy intent of our research—hence its status as a leading competitor.

PVGVI nonetheless exhibits three notable limitations. First, its sole reliance on a CNN backbone hampers performance in scenes that demand long–range context or fine–grained semantic separation. Second, the method lacks depth awareness, treating all vegetation pixels as perceptually equivalent regardless of distance and thereby conflating a nearby tree canopy with a far-field patch of grass. Third, it does not differentiate vegetation types (e.g., trees vs. turf grass), constraining its usefulness for environmental planning.

DEVISE overcomes these shortcomings by (i) replacing the CNN backbone with Mask2Former, a transformer–based model that offers superior spatial reasoning; (ii) incorporating monocular depth estimation (MiDaS) to generate distance-weighted GVI

scores; (iii) segmenting vegetation into three semantic classes—tree, grass, and understory plants; and (iv) applying an HSV-based posterior filter to suppress false positives originating from green-colored but non-vegetal surfaces such as painted walls and signage. Collectively, these enhancements yield more realistic visibility estimates and stronger policy relevance.

Xia et al. (2021): Semantic Segmentation for Scalable GVI. Xia et al. couple GSV panoramas with a lightweight CNN to compute pedestrian-level GVI across multiple neighborhoods, demonstrating that automated segmentation can scale well beyond hand-annotated datasets. Among existing approaches, their framework is the most direct competitor in terms of computational efficiency.

Two limitations, however, constrain their contribution. First, the segmentation network lacks the semantic depth needed to disentangle vegetation from complex urban clutter, leading to misclassification when greenery overlaps built structures. Second, the model ignores both depth and color filtering, resulting in systematic overestimation of visible greenness.

DEVISE mitigates these issues through a more expressive segmentation backbone (Mask2Former), depth normalization that weights vegetation by perceptual prominence, and HSV filtering that curtails color-based false positives. The outcome is a finer-grained, policy-ready measure suitable for longitudinal analysis.

Choi et al. (2022): Tree Detection via Object Models. Choi et al. present a hybrid pipeline that merges semantic segmentation (Panoptic-DeepLab) with object detection (YOLOv3) to classify individual tree species and structural attributes in Seoul's streetscapes. Their approach achieves impressive botanical detail and validates the feasibility of fine-scale species mapping with GSV imagery, positioning it as a specialized—but still competitive—alternative to our work.

Nevertheless, the model's emphasis on species-level accuracy renders it computationally expensive and difficult to generalize across large spatial or temporal extents. Moreover, it neither corrects for depth distortion in panoramic imagery nor normalizes visibility by proximity—both prerequisites for comparative urban research.

DEVISE, by contrast, prioritizes spatial scalability and perceptual realism. It operates on broader vegetation classes rather than species labels, integrates depth-based weighting, and applies HSV filtering, thereby delivering metrics that more faithfully represent what pedestrians observe across diverse cities, counties, and years.

Comparative Advantages. In summary, *DEVISE* combines transformer-based panoptic segmentation, depth-weighted visibility scoring, vegetation-type disaggregation, and HSV color filtering within a single, computationally efficient pipeline. This integration yields street-scale greenness measures that are (i) more accurate in heterogeneous urban scenes, (ii) perceptually calibrated to human experience, and (iii) readily extensible to longitudinal and cross-city policy evaluation—advantages that none of the closest competitors simultaneously achieve.

Collectively, these three studies underscore the rapid methodological progress in GSV-based greenness measurement. Each, however, is constrained by some combination of architectural limitations, depth ignorance, or restricted scalability. By uniting transformer-based segmentation, depth normalization, HSV filtering, and multi-class vegetation separation, *DEVISE* advances the state of the art. To our knowledge, it is the first pipeline

to deliver depth-aware, perceptually weighted GVI estimates at a scale suitable for longitudinal policy evaluation.

1.5 Objectives, Significance, and Principal Contributions

This study introduces **DEVISE**—the *Depth-Aware Vegetation Indexing System*—a fully automated pipeline that:

1. employs **Mask2Former** for fine-grained semantic segmentation (Cheng et al., 2022);
2. applies **HSV posterior filtering** to eliminate green-colored but non-vegetal surfaces;
3. integrates **monocular depth estimation** (DPT-Large) to weight vegetation by perceptual proximity;
4. conducts **cross-regional validation** with 1.44 million GSV panoramas from U.S. swing counties; and
5. will be released as an open-source **Python package** (`devise`) that bundles pretrained weights, a command-line interface, and reproducible notebooks.

We advance street-level greenness measurement through three interrelated aims:

- Designing a fully automated pipeline for detecting and quantifying street-level vegetation using GSV images, integrating Mask2Former segmentation, HSV filtering, and pixel-level depth estimation.
- Extending traditional Green View Index calculations with depth-aware normalization, enabling measurements that reflect the prominence and accessibility of greenery in urban streetscapes.
- Evaluating the relationship between greenery visibility and electoral preferences across diverse geographic and political settings, using a dataset from American swing counties.

By quantifying *visible*, rather than merely *present*, canopy, *DEVISE* advances the GVI lineage in three respects: it is (i) depth-aware, (ii) resilient to spurious color artifacts, and (iii) demonstrably scalable across diverse policy contexts. These innovations equip computational social scientists with a sharper lens on urban environmental (in)justice and the effectiveness of municipal greening programs.

Through these contributions, the study lays the groundwork for a new generation of environmental-monitoring and urban-planning tools that foreground human perception and social equity by:

- delivering more accurate, ground-level measurements of urban greenery;
- furnishing policymakers with quantitative evidence on the effectiveness of greening interventions; and
- embedding pedestrian-scale green-coverage metrics into urban-livability assessment and planning frameworks.

2 Methodology for *DEVISE*

2.1 Overall Pipeline Description

We now present *DEVISE*, a depth-aware vegetation indexing system designed to quantify greenery from street-level panoramic imagery. We first describe the modular architecture on which *DEVISE* is built, following a consistent design for content normalization, semantic segmentation, spectral refinement, and spatial enhancement. Then, we highlight the integration of monocular depth estimation and propose a depth-weighted index formulation that more closely aligns with human visual salience. Finally, we show how *DEVISE* enables perceptually grounded measurements of urban greenery at scale. An overview of the *DEVISE* pipeline is provided in Figure 1.

2.1.1 Image Normalization for Scene Structuring

Panoramic images from platforms such as Google Street View often contain large black borders or warped areas that introduce noise and distort downstream predictions. These regions reduce the effective resolution of meaningful content and mislead transformer-based attention mechanisms.

To address this, we apply a preprocessing module that first detects non-zero intensity regions via pixel thresholding, then crops the image to retain only the content-preserving window. The cropped result is resized to a fixed resolution (e.g., 4000×2000) using high-quality interpolation. This ensures consistent spatial dimensions for segmentation and depth models. The normalization step eliminates structural noise and aligns the input space for all subsequent components.

2.1.2 Vegetation Segmentation via Dense Class Masking

Urban vegetation appears in complex textures across a wide range of scales, from large trees to patches of grass or vertical wall-climbing plants. These categories are often spatially irregular and co-occur with human-made objects of similar color or texture.

To extract meaningful vegetation regions, we adopt a dense segmentation model named Masked-attention Mask Transformer (Mask2Former) trained on a multi-class urban scene dataset. The model takes an RGB input image and predicts a per-pixel class label map, assigning each pixel to a semantic category based on contextual and visual features. The model outputs a pixel-level class label for each image, covering more than 100 semantic categories. We isolate three vegetation-relevant classes—*trees*, *grass*, and *plants*—by selecting their corresponding class indices from the output label map (e.g., 4, 9, 17 in the ADE20K label space). Each category is then converted into an independent binary mask by thresholding the class prediction map. This results in a three-channel vegetation representation, where each channel encodes a distinct greenery type.

This segmentation strategy provides high spatial resolution and context-aware parsing of vegetation, with the ability to differentiate between green objects of differing semantic meaning. Unlike instance detectors, our approach outputs spatial masks directly, preserving boundary details and supporting full-panoramic context.

2.1.3 Spectral Refinement via HSV Filtering

Context-aware segmentation models can still produce false positives, especially under lighting variation or in scenes with painted green surfaces (e.g., buildings or vehicles).

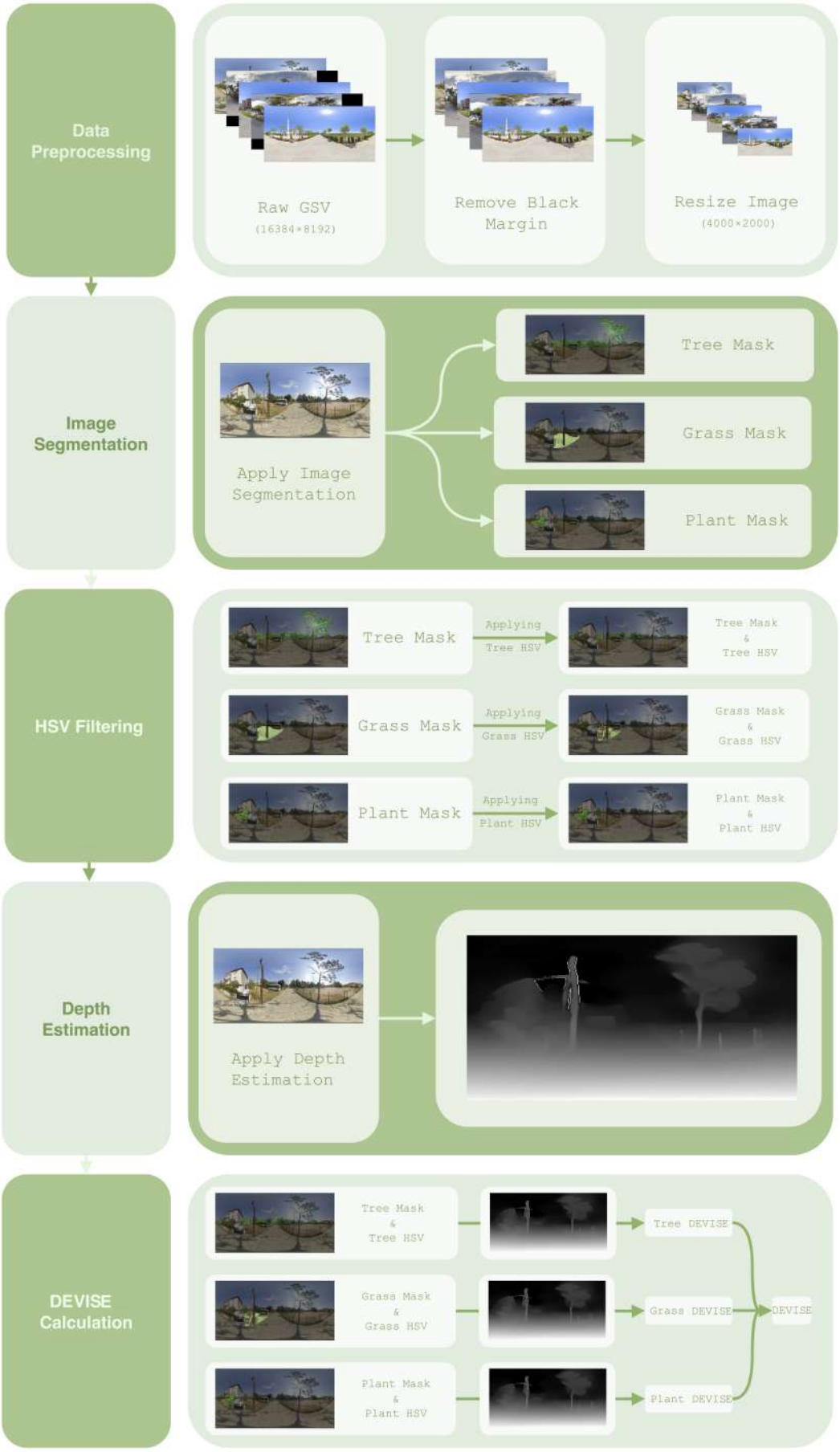


Figure 1: Overview of the **DEVISE** pipeline. *Note*: The output consists of both conventional and depth-weighted vegetation indices for different greenery types.

These errors stem from an over-reliance on semantic priors without verifying visual appearance.

To mitigate such cases, we introduce an additional filtering step based on color properties in the HSV (hue, saturation, value) space. This step operates as a post-processing refinement applied to the binary masks obtained from the previous segmentation stage. For each vegetation category, we extract pixel values from the original image that correspond to activated regions in the mask. These pixel values are then evaluated against a predefined HSV range that reflects the typical spectral characteristics of natural vegetation.

Only the pixels whose hue and saturation values fall within the vegetation-specific thresholds are retained, while all others are discarded. This operation is performed independently for each vegetation class, allowing for category-specific spectral tuning. The result is a set of refined binary masks that combine semantic confidence from segmentation with spectral consistency derived from low-level visual cues.

This hybrid approach significantly reduces noise introduced by non-vegetative yet visually similar regions and leads to more reliable estimates of actual green coverage in the final mask set. The refined masks serve as the basis for both conventional and depth-aware vegetation indexing in subsequent stages.

2.1.4 Monocular Depth Estimation Using Transformer Features

Vegetation in urban imagery appears at multiple depths, but conventional segmentation masks offer no indication of spatial proximity. As a result, distant tree lines are treated with the same weight as close-up foliage, which fails to reflect how greenery is perceived in real-world environments. To resolve this, we introduce a dedicated depth estimation stage that provides a dense, continuous representation of relative distance across the scene.

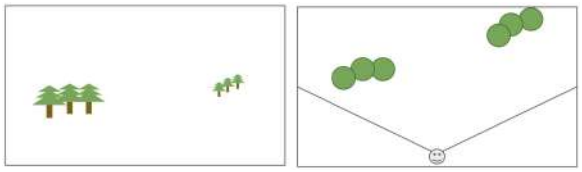
This step leverages a pre-trained monocular depth prediction model to infer depth values from the visual context of the image. Each pixel is assigned a floating-point value reflecting its estimated proximity to the observer. Unlike binary segmentation masks, the depth map captures gradual transitions across the visual field, supporting a finer-grained interpretation of spatial layout.

To make these depth values compatible with subsequent weighting, we apply min-max normalization, rescaling the values to a fixed range of [0, 1]. This operation ensures consistency across images with varying scene depth and stabilizes their contribution to the index. The normalized depth map is then resized to match the spatial resolution of the vegetation masks, allowing for pixel-wise fusion between semantic predictions and spatial prominence.

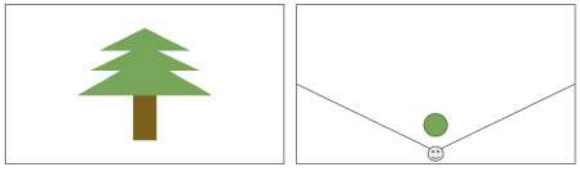
By integrating depth into the pipeline, we distinguish foreground greenery from background regions and introduce a perceptual weighting mechanism. This ensures that vegetation closer to the viewer contributes more strongly to the final index. The result is an index that reflects not just how much greenery is present, but how it is experienced in the urban visual field. For a comparative example, please refer to Figure 2.

2.1.5 Depth-Weighted Index Computation

Traditional GVI metrics compute the proportion of visible greenery without considering depth. However, we observe that in real-world perception, vegetation closer to the viewer has a disproportionate influence on visual greenness.



(a) Background vegetation appearing in clusters, with limited perceptual weight.



(b) A single foreground tree dominating the visual field with high salience.

Figure 2: Visual comparison of background and foreground vegetation.

To address this, we propose a depth-weighted formulation that integrates both semantic confidence and spatial relevance. Let $M_c(i,j)$ denote the refined binary mask for vegetation category c, and $D(i,j)$ the normalized depth value at pixel (i,j). We define the depth-weighted index for class c as:

$$\mathrm{GVI}_c^{\mathrm{depth}} = \frac{\sum_{i,j} M_c(i,j) \cdot D(i,j)}{\sum_{i,j} D(i,j)}$$

This formulation prioritizes vegetation regions that are closer to the observer, thus aligning the metric with visual attention mechanisms. We additionally compute a total depth-weighted index based on the union of all refined masks. These indices are recorded alongside traditional pixel-count metrics for comparative analysis.

This system design—integrating semantic structure, spectral cues, and spatial priors—forms a generalizable pipeline for street-level environmental perception. Each component is modular and interpretable, making the pipeline suitable for deployment at scale and adaptable to future refinements or policy-specific metrics. Figure 3 provides an example of the technical advantages offered by *DEVISE*. Figure 4 presents five example pairs from European locations, demonstrating *DEVISE*'s strong performance across varying lighting conditions, urban configurations, vegetation types, and degrees of canopy degradation.



(a) Vegetation under overpasses. Pure segmentation fails to filter painted surfaces; HSV and depth refinement successively improve precision.



(b) Sidewalk vegetation with complex shadows. Combining semantic, spectral, and spatial modules allows selective inclusion of visually salient greenery.

Figure 3: Qualitative comparisons of different module combinations.

2.2 Depth Estimation Backbone: Dense Prediction Transformer (DPT)

The Dense Prediction Transformer (DPT) model, proposed by Ranftl et al. (2021), is a powerful transformer-based backbone designed for dense prediction tasks such as monocular depth estimation and semantic segmentation. Unlike CNN-based models,



Figure 4: Original panoramas (left) and outputs of the *DEVISE* algorithm (right), with detected vegetation highlighted in enhanced green.

DPT leverages global self-attention mechanisms to model long-range dependencies and structural relationships in an image, which significantly improves edge consistency and spatial detail in depth maps. DPT demonstrated strong generalization capabilities across datasets such as NYU-Depth v2, KITTI, and Make3D. An illustrative picture of DPT's structure is shown in Figure 5.

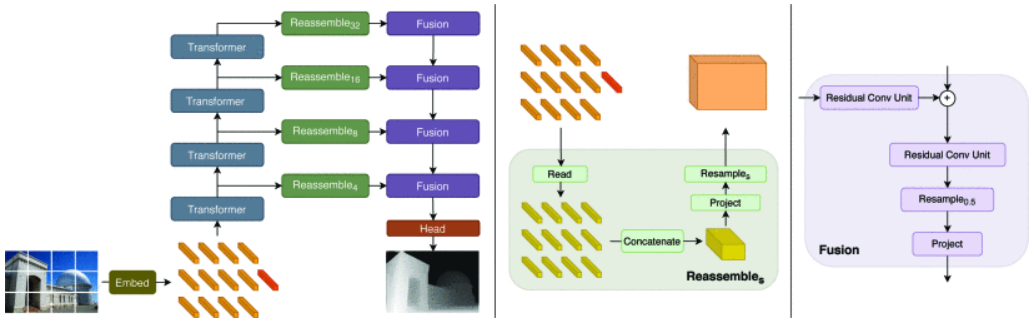


Figure 5: Architecture of Dense Prediction Transformer (DPT), showing pixel-level depth prediction through global attention and pixel decoding.

Recent empirical studies further validate the effectiveness and versatility of DPT on a variety of computer vision tasks. For example, Chen et al. (2024) proposed a spatially-variant blur degradation model for image super-resolution by using a pretrained DPT model to globally extract pixel-level depth maps that guide the generation of more realistic blur kernels, thereby improving the real-world robustness of synthetic data. Sonam et al. (2024) applied DPT for depth estimation with RGB-only video datasets in saliency detection, which allowed their RGB-D saliency model to operate on a broader range of benchmarks with improved accuracy and efficiency. In the field of 3D reconstruction, Lee et al. (2024) adopted a transformer-based architecture inspired by DPT to model the scene in terms of dense 3D point clouds based on single RGB images, achieving superior geometric consistency over CNN and diffusion-based approaches. Moreover, Mohadikar et al. (2024) included DPT as a baseline in their 360-degree monocular depth estimation study and demonstrated that their new diffusion model improves upon DPT in omni-directional input settings, further affirming DPT's role as a strong baseline in depth estimation.

In comparison to popular depth estimation backbones such as MiDaS, BTS, and MonoDepth2, DPT exhibits several technical advantages. MiDaS provides strong generalization among datasets but tends to generate over-smoothed predictions with limited boundary precision. BTS, a CNN-based architecture, captures local features effectively but lacks the global reasoning capabilities provided by transformers. MonoDepth2 is relatively lightweight and easy to deploy but performs suboptimally in cluttered or complex environments. By contrast, DPT's transformer-based structure allows for better global scene understanding, which is particularly beneficial for tasks involving depth-aware segmentation and structure-aware analysis in urban scenes.

2.3 Semantic Segmentation Head: Mask2Former for Unified Pixel Prediction

Mask2Former, introduced by Cheng et al. (2022), is a unified transformer-based architecture for various segmentation tasks, including semantic, instance, and panoptic segmentation. Unlike traditional segmentation models that rely on task-specific heads and

heavy post-processing, Mask2Former uses a masked attention mechanism and a pixel decoder to directly predict segmentation masks and their associated categories. This design allows Mask2Former to achieve state-of-the-art performance while maintaining architectural flexibility and generalization across domains.

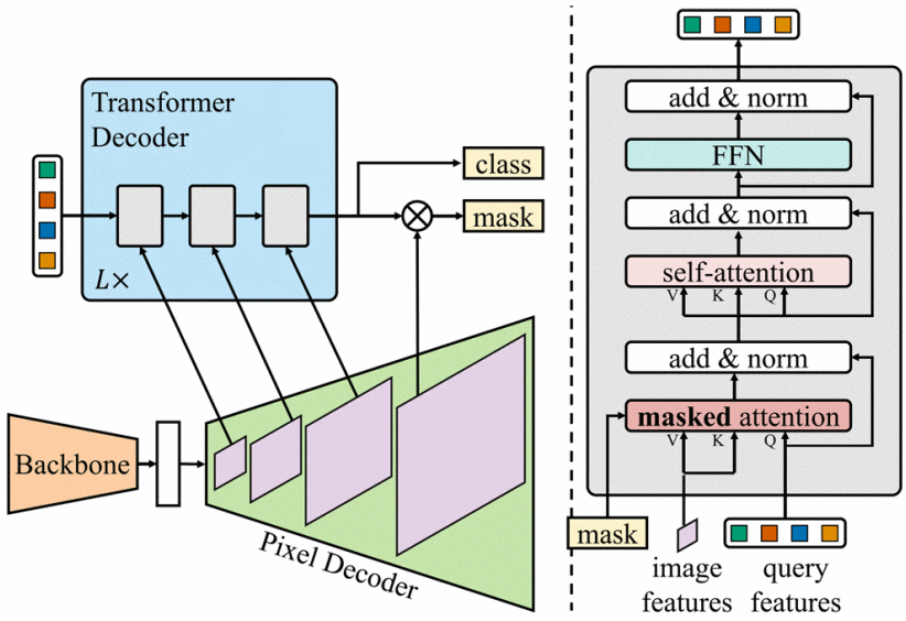


Figure 6: Mask2Former architecture featuring a masked attention decoder and a pixel decoder for universal segmentation tasks.

Several empirical works have demonstrated the effectiveness of Mask2Former in diverse fields beyond natural image segmentation. For example, Sheng et al. (2024) applied Mask2Former for instance segmentation in pathology images, showing that despite being originally designed for general computer vision tasks, it achieved competitive performance on gland cell segmentation. Kim et al. (2024) employed Mask2Former in panoptic segmentation of remote sensing images, achieving better performance than conventional CNN-based models in complex aerial scenes, which contained small and overlapping objects. In the context of generalized few-shot segmentation, Zhao et al. (2024) built their DCQ-Former using the Mask2Former backbone, leveraging its decoupled design of classification and mask prediction for improved performance on novel categories with limited data. Additionally, Choi et al. (2024) included Mask2Former in their temporal modeling framework for referring to video object segmentation, demonstrating that it is robust even in long-term video sequences.

The main competitors of Mask2Former fall into two categories: CNN-based models (e.g., DeepLabv3+, Panoptic FPN) and early transformer-based approaches such as DETR. While CNN models perform well on specific tasks, they lack the flexibility to generalize across segmentation types. DETR introduces a set prediction framework but requires complex bipartite matching during training. In contrast, Mask2Former leverages masked attention to focus only on active regions, reducing computation and improving mask quality. Furthermore, it generalizes well across various image modalities and segmentation objectives without architectural modification.

3 Linking Electoral Preferences and Green Visibility in America's Swing Counties

This section provides a quantitative case analysis to demonstrate the large-scale data processing capabilities of the *DEVISE* system. Moving beyond individual image analysis, we apply the full pipeline to a unique archival dataset of over 1.44 million Google Street View images captured in politically competitive "swing counties" across the United States. The objective is to quantify changes in street-level greenery over time and explore potential relationships with local environmental policies and expenditures. By executing this analysis in a large-N setting, we showcase *DEVISE*'s capacity to operate at a nationwide scale, proving its utility as a powerful tool for researchers, policymakers,

3.1 Data Sources and Sampling Strategy: Swing-County Google Street View Archive (2004–2024)

To evaluate the visual impact of environmental policies, we draw on an original archive of Google Street View (GSV) imagery compiled by the research team. The archive covers every *swing county* in all United States—counties where the absolute vote-share gap between the Democratic and Republican presidential candidates has remained $\leq 10\%$ in every general election since 2004.[1] Applying this filter identifies 60 of the nation's 3,144 counties (about 1.9%) and yields a politically balanced sample of jurisdictions where marginal changes in the built environment and public goods, such as urban forestry, are most likely to influence future electoral outcomes.

For each swing county we collect all GSV panoramas captured in the month of every presidential election (November) and in the preceding month (October), producing a panel that spans six elections—2004, 2008, 2012, 2016, 2020, and 2024. Restricting the scrape to these two months maximizes temporal comparability while capturing peak campaign activity and late-season foliage conditions.

Panoramas are sampled systematically at fixed 50-meter intervals along the full drivable road network, ensuring uniform spatial coverage across time points and counties. The final dataset comprises **1,441,746** unique GSV panoramas, each georeferenced and time-stamped.

Every panorama is processed through the *DEVISE* pipeline, which (i) applies Mask2Former for panoptic segmentation, (ii) estimates monocular depth via MiDaS, (iii) filters vegetation pixels with an HSV posterior, and (iv) aggregates distance-weighted greenness metrics. This workflow yields consistent, street-level vegetation scores that can be compared across counties, years, and even individual months.

End-to-end scraping, quality control, and batch processing required approximately six months of continuous cloud-based execution (December 2024–May 2025), reflecting the scale of the nationwide dataset and the computational demands of the DEVISE pipeline. The counties represented in our sample are provided in Figure 7, and the number of observations for each county–election pair is detailed in Table 1.

[1] County-level election returns are taken from the MIT Election Data and Science Lab.

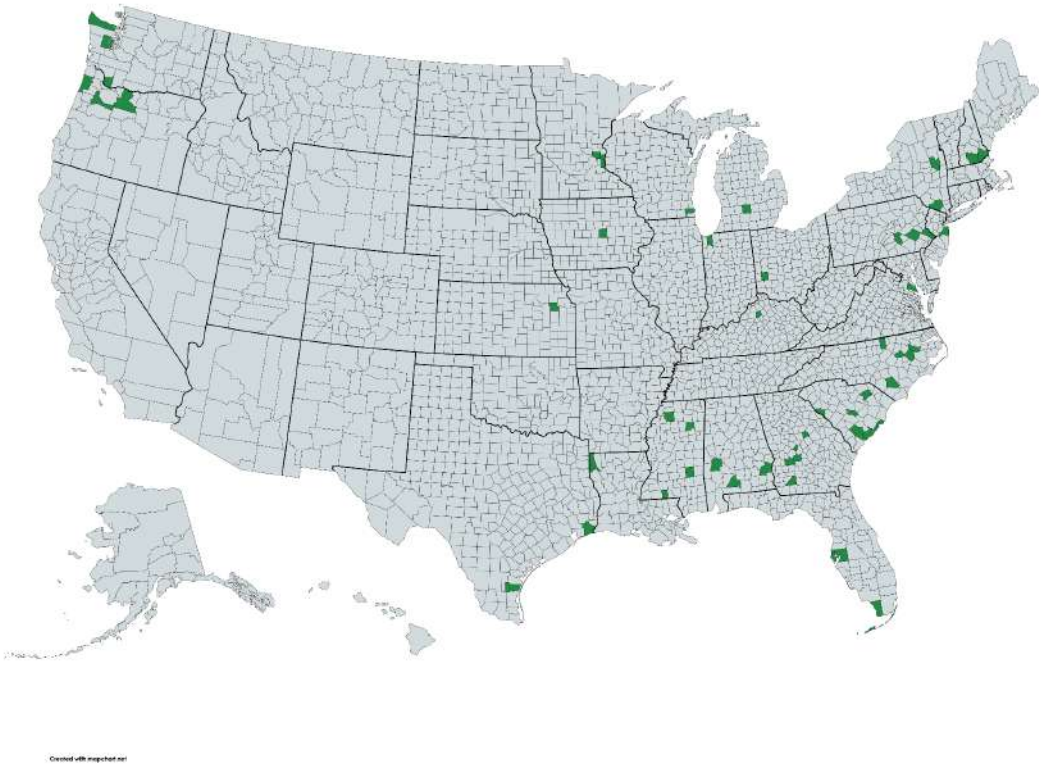


Figure 7: Map of swing counties ($\leq$ 10% partisan margin in every presidential election since 2004) included in the Google Street View image collection. The sample totals 1.44 million panoramas captured during October–November of each election year.

3.2 A Qualitative Assessment of Environmental Policies in Dauphin, Bucks, and Washington Counties

In this section, we focus on three swing counties from our dataset—Dauphin (PA), Bucks (PA), and Washington (MN)—that offer strong combinations of political variability, policy relevance, and street-level imagery availability. These counties serve as ideal testbeds for linking environmental actions to changes in green visibility through our DEVISE framework.

Dauphin County, Pennsylvania Dauphin County is located in central Pennsylvania and includes Harrisburg, the state capital. As a politically competitive county in a key swing state, Dauphin has experienced alternating partisan leadership at both local and state levels over the past two decades (Data and Lab, 2022). Harrisburg's *Community Greening Plan* (2017) laid out a block-by-block program of green-stormwater "bump-outs," bioswales, and tree trenches along Third Street, now folded into the City Beautiful H2O combined-sewer-control consent decree (Water, 2017). Parallel funding has come through DCNR's TreeVitalize grants, which in 2020 alone awarded $20 000 to the City of Harrisburg for bare-root street-tree plantings (of Conservation and Resources, 2020). With margins inside $\pm$10 pp in every presidential election since 2004, Dauphin offers a narrow partisan window in which visible green infrastructure might shift vote share (Data and Lab, 2022).

County–State	2004-10	2004-11	2008-10	2008-11	2012-10	2012-11	2016-10	2016-11	2020-10	2020-11	2024-10	2024-11	Total
ANOKA–MINNESOTA	0	0	45896	14321	0	0	13510	8376	2	466	1217	47	83835
BALDWIN–GEORGIA	0	0	0	0	42510	0	0	0	0	0	0	0	42510
BARBOUR–ALABAMA	0	0	0	0	0	0	0	0	0	0	0	0	0
BERKS–PENNSYLVANIA	0	0	45001	6465	0	0	3206	4483	0	0	2	0	59157
BLADEN–NORTH CAROLINA	0	0	0	0	17784	9501	0	10278	0	0	274	140	37977
BROOME–NEW YORK	0	0	14658	0	0	0	0	8466	0	0	348	0	23472
BUCKS–PENNSYLVANIA	0	0	19419	30711	46214	3235	21631	499	0	2	127	33	121871
CADDO–LOUISIANA	0	0	0	0	0	0	12012	7640	0	0	2259	226	22137
CALHOUN–SOUTH CAROLINA	0	0	0	0	0	0	0	1712	0	0	0	0	1712
CHARLESTON–SOUTH CAROLINA	0	0	0	0	34466	0	2	1366	1	0	1017	1391	38243
CHICKASAW–MISSISSIPPI	0	0	0	0	0	0	0	0	0	0	51	0	51
CLALLAM–WASHINGTON	0	0	3537	0	0	0	0	0	0	0	0	0	3537
CLARK–WASHINGTON	0	0	0	0	16894	0	0	0	0	0	928	1	17823
COLLETON–SOUTH CAROLINA	0	0	0	0	0	0	0	2886	0	0	0	0	2886
CONECUH–ALABAMA	0	0	0	0	0	0	0	0	0	0	0	0	0
COVINGTON–VIRGINIA	0	0	0	0	0	0	344	0	0	0	9	9	362
COVINGTON CITY–VIRGINIA	0	0	0	0	0	0	0	0	0	0	0	0	0
DARLINGTON–SOUTH CAROLINA	0	0	0	0	0	327	0	0	0	0	2	0	329
DAUPHIN–PENNSYLVANIA	0	0	52265	2521	8204	0	21538	1609	5378	25609	0	0	117124
DOOLY–GEORGIA	0	0	0	0	0	0	0	0	0	0	3	0	3
EATON–MICHIGAN	0	0	11545	22193	0	0	2628	7378	0	1911	2	15	45672
FRANKLIN–KENTUCKY	0	0	1327	0	3552	0	0	3181	0	0	56	47	8163
GRANVILLE–NORTH CAROLINA	0	0	0	0	0	0	0	565	0	0	0	0	565
HILLSBOROUGH–FLORIDA	0	0	0	0	1	0	2	30984	48	70	0	302	31407
HILLSBOROUGH–NEW HAMPSHIRE	0	0	5750	0	10747	30262	1894	0	1	0	113	231	48998
JASPER–MISSISSIPPI	0	0	0	0	0	0	0	0	0	0	69	2	71
JEFFERSON–TEXAS	0	0	0	0	0	43211	0	0	0	2	1987	268	45468
KENT–MARYLAND	0	0	0	0	0	0	0	0	0	0	0	0	0
KLEBERG–TEXAS	0	0	0	0	0	0	0	0	0	0	0	0	0
MARENGO–ALABAMA	0	0	0	0	0	0	0	0	0	0	0	17	17
MARION–OREGON	0	0	22540	0	400	0	0	0	0	14504	0	11	37455
MARSHALL–IOWA	0	0	1099	0	0	0	0	0	0	0	32	12	1143
MARTIN–NORTH CAROLINA	0	0	0	0	0	0	0	0	0	0	0	14	14
MASON–WASHINGTON	0	0	3684	0	405	6	0	4323	0	0	0	0	8418
MCCORMICK–SOUTH CAROLINA	0	0	0	0	0	0	0	0	0	0	6	0	6
MITCHELL–GEORGIA	0	0	0	0	20126	0	0	0	0	0	0	0	20126
MONMOUTH–NEW JERSEY	0	0	28219	7957	15479	0	10651	0	1	50180	1118	0	113605
MONROE–FLORIDA	0	0	0	0	0	0	0	0	1	1	0	0	2
MONTGOMERY–OHIO	0	0	2971	1128	0	0	270	8	0	1	10	13	4402
ORANGE–NEW YORK	0	0	58480	1602	15222	4664	4719	12283	0	0	49	59	97078
PANOLA–MISSISSIPPI	0	0	0	0	0	0	0	2069	0	0	21	16	2106
PEACH–GEORGIA	0	0	0	0	0	0	0	0	0	0	0	0	0
PIKE–MISSISSIPPI	0	0	0	0	0	0	0	0	0	0	0	0	0
PINELLAS–FLORIDA	0	0	0	0	0	0	3	5118	2	4	0	24	5151
PITT–NORTH CAROLINA	0	0	0	0	25954	0	0	0	0	0	0	6	25960
PORTER–INDIANA	0	0	26204	0	0	0	5137	1858	6	0	7	10	33222
RACINE–WISCONSIN	0	0	0	0	2239	0	4892	535	0	0	121	9	7796
RADFORD–VIRGINIA	0	0	0	0	188	0	259	674	0	0	59	40	1220
ROCKINGHAM–NEW HAMPSHIRE	0	0	83266	2228	15416	7119	11981	2781	0	0	85	416	123292
SALEM–VIRGINIA	0	0	0	0	1399	0	1605	1173	7705	0	36	0	11918
SALEM CITY–VIRGINIA	0	0	0	0	0	0	0	0	0	0	0	0	0
SARATOGA–NEW YORK	0	0	0	1151	0	0	26971	10952	0	0	9	0	39083
SHAWNEE–KANSAS	0	0	6390	0	42772	0	0	0	0	0	98	0	49260
SKAGIT–WASHINGTON	0	0	11081	0	0	0	0	937	0	0	91	55	12164
ST. MARY'S–MARYLAND	0	0	27653	0	0	0	0	0	0	0	0	0	27653
SUMTER–GEORGIA	0	0	0	0	0	0	0	0	0	0	0	0	0
TILLAMOOK–OREGON	0	0	10710	0	0	0	0	0	0	0	0	0	10710
WASCO–OREGON	0	0	4182	0	0	0	0	0	0	0	0	2	4184
WASHINGTON–MINNESOTA	0	0	8403	4233	1769	0	14186	3770	0	85	844	336	33626
WAYNESBORO–VIRGINIA	0	0	1692	0	0	0	1130	0	0	0	0	0	2822
WILSON–NORTH CAROLINA	0	0	0	0	0	0	0	0	0	0	6	348	354
WOOD–OHIO	0	0	7821	1609	0	0	1083	4728	120	101	101	23	15586
Total	0	0	503793	96119	321741	98326	159654	140632	13265	92936	11157	4123	1441746

Table 1: County–month panorama counts for U.S. general elections (2004–2024).

Bucks County, Pennsylvania Bucks County is a large suburban/semi-rural region bordering Philadelphia, swinging between parties in every presidential race (Trump +1 pp in 2016; Biden +4 pp in 2020) (Data and Lab, 2022). Its environmental policy is unusually decentralized: countywide open-space bond referenda (2007, 2010) financed riparian-buffer acquisitions at roughly $9 million per round (Commission, 2007), and the regional TreeVitalize partnership planted over 20 000 trees and restored nearly 300 acres of riparian buffer in southeastern PA by 2009 (Society, 2009). Bucks also ranks in the top quintile statewide in U.S. News's "Healthiest Communities" index for both environmental quality and public-goods spending (Report, 2021). This mix of aggressive local greening and razor-thin vote swings makes Bucks an ideal case for testing whether DEVISE can detect street-level canopy gains in a fast-flipping suburb.

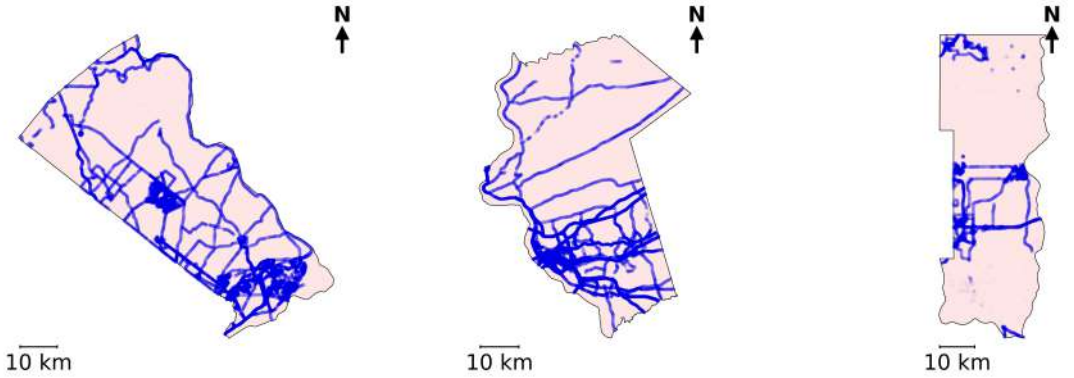
Washington County, Minnesota Washington County, east of St. Paul, saw its presidential margin narrow from Obama +6 pp (2012) to near parity in 2016 before rebounding to Biden +8 pp in 2020, keeping it within our ±10 pp swing definition (Data and Lab, 2022). The county's 2040 Comprehensive Plan (adopted January 2020) embeds tree-preservation requirements in new subdivisions and mandates bioswale and porous-pavement best practices in its development code (Washington County, 2019). The Washington Conservation District's Habitat Restoration Program has sold over 15 000 trees annually for wind-row and wildlife plantings and supports oak-savanna and prairie restorations across public lands (District, 2023). State revolving-fund appropriations in 2025 pledge up to $1 million in principal forgiveness for county green-infrastructure retrofits (Agency, 2025). Washington's combination of rapid suburban growth, codified canopy protections, and active prairie restoration provides a stringent test of DEVISE's ability to pick up greenness gains amid development pressure.

Figure 8 provides a comparison on the green-area estimates produced by the baseline algorithmic framework with those generated by *DEVISE*.

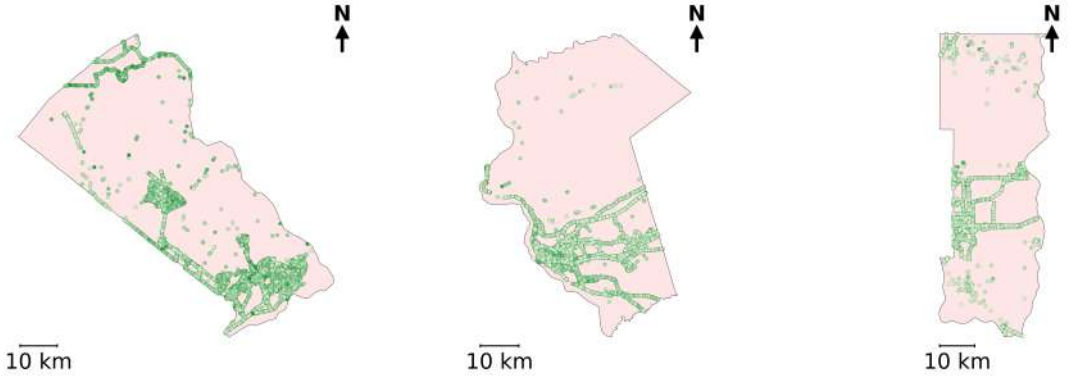
3.3 Quantitative Assessment of the *DEVISE* Framework's Potential

The preceding county narratives highlight how distinct policy instruments—state-capital storm-water retrofits in Dauphin, bond-financed riparian buffers in Bucks, and subdivision canopy mandates in Washington—translate into visibly different streetscape outcomes. Although these three cases illuminate the pathways through which local governance can expand or erode green infrastructure, they remain anecdotal and potentially idiosyncratic. To move beyond illustrative evidence, we now shift from a small-N comparative lens to a large-N design that systematically quantifies vegetation change across all U.S. swing counties.

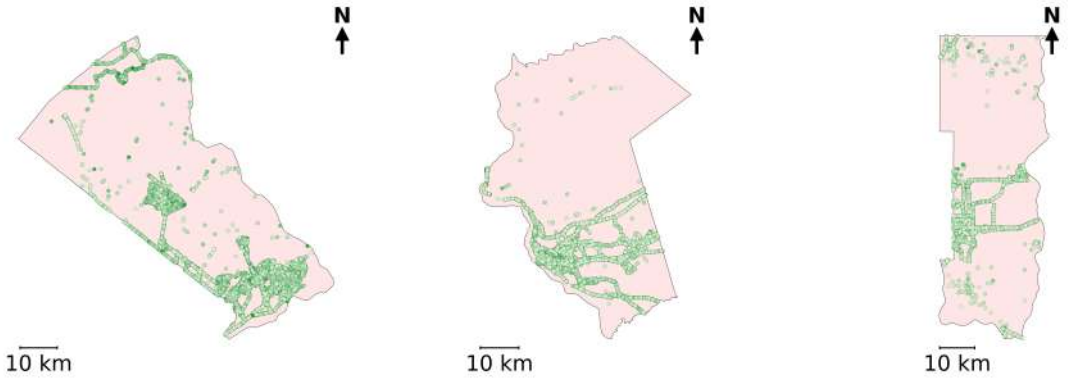
This transition to a quantitative framework serves two complementary purposes. First, it allows us to test whether the greenness patterns observed qualitatively—greater canopy in jurisdictions with sustained pro-environmental initiatives—persist after controlling for climatic, socio-economic, and temporal confounders. Second, the county–month panel affords statistical leverage to detect subtler effects that single-case observations might obscure, such as incremental canopy gains driven by statewide funding rather than headline-grabbing local ordinances. In what follows, we deploy the depth-aware vegetation metrics produced by *DEVISE*, together with electoral, climate, and census covariates, to evaluate whether Democratic vote share reliably predicts street-level greenery



(a) GSV coverage locations for the three swing counties.



(b) Vegetation coverage quantified using the original algorithmic framework.



(c) Vegetation quantification using *DEVISE*.

Figure 8: Google Street View sampling frame and vegetation maps for three swing counties—Bucks (PA), Dauphin (PA), and Washington (MN). (a) spatial footprint of sampled panoramas; (b) greenness shares estimated with the baseline pixel-count algorithm; (c) greenness shares estimated with the depth-aware *DEVISE* algorithm. *Note*: Colors are min–max normalized across the full 1.44 million-panorama corpus, from white (least observed greenness) to dark green (greatest).

and whether this association endures under rigorous robustness checks.

3.4 Hypotheses

Building on the mixed-methods evidence above, this subsection translates our theoretical expectations into two empirically testable propositions. Each hypothesis specifies how electoral dynamics should manifest in the depth-aware vegetation indicators generated by *DEVISE*, progressively strengthening the causal inference by layering additional controls. The hypotheses are stated in a way that allows direct evaluation with the county–month panel introduced earlier.

H1 — Electoral Greenness Hypothesis. Counties that exhibited a higher Democratic vote share in the preceding presidential election (`dem_pct_t1`) will display greater street-level vegetation coverage—measured by the proportion of visible tree canopy, grass, and plant pixels (`tree_plain`, `grass_plain`, `plant_plain`)—than counties with a lower prior Democratic vote share, after controlling for local climate conditions. The intuition is that Democratic electorates tend to support environmental protection and urban greening, which encourages county officials to expand or preserve green infrastructure.

H2 — Persistence under Socio-Economic and Temporal Controls. The positive association posited in H1 will remain statistically significant when (i) county-level socio-economic characteristics (income, education, racial composition, labor-market indicators) and (ii) two-way fixed effects for county and date are included in the model. Robustness to these controls would suggest that the relationship is not driven by wealthier or demographically distinctive counties, nor by unobserved time-invariant county traits or nationwide temporal shocks.

3.5 Empirical Approach

3.5.1 Data and Variables

The empirical analysis exploits a balanced county–month panel that matches computer-vision classifications of green pixels in Google Street View panoramas (2004–2024) to county-level political, climatic, and socio-economic information. Vegetation is captured through eight outcomes: four "plain" pixel shares (`tree_plain`, `grass_plain`, `plant_plain`, `total_plain`) and four depth-weighted analogues (`tree_depth`, `grass_depth`, `plant_depth`, `total_depth`). Each measure is averaged across all panoramas within a county–month, ensuring comparability over space and time.

The principal explanatory variables are the Democratic vote shares in the two most recent past presidential elections (`dem_pct_t1`, `dem_pct_t2`). Climate controls include mean January, April, and July temperatures as well as annual precipitation and snowfall (NOAA). Socio-economic covariates—per-capita personal income, population, unemployment, female share, White share, and educational attainment—are drawn from the Bureau of Economic Analysis and the American Community Survey. Together, these variables mitigate confounding and allow us to isolate the association between electoral strength and street-level greenness.

3.5.2 Model Specification and Estimation

For each vegetation outcome $Y_{c,t}$ we estimate three weighted least-squares (WLS) models:

$$\begin{aligned}
\text{(Plain)} \quad Y_{c,t} &= \beta_0 + \beta_1 \, \texttt{dem_pct_t1}_{c,t} + \beta_2 \, \texttt{dem_pct_t2}_{c,t} \\
&\quad + \boldsymbol{\gamma}^\top \text{Climate}_{c,t} + \varepsilon_{c,t}, \qquad (1) \\
\text{(Social)} \quad Y_{c,t} &= \beta_0 + \beta_1 \, \texttt{dem_pct_t1}_{c,t} + \beta_2 \, \texttt{dem_pct_t2}_{c,t} \\
&\quad + \boldsymbol{\gamma}^\top \text{Climate}_{c,t} + \boldsymbol{\delta}^\top \text{Socio}_{c,t} + \varepsilon_{c,t}, \qquad (2) \\
\text{(Social + FE)} \quad Y_{c,t} &= \alpha_c + \lambda_t + \beta_1 \, \texttt{dem_pct_t1}_{c,t} + \beta_2 \, \texttt{dem_pct_t2}_{c,t} \\
&\quad + \boldsymbol{\gamma}^\top \text{Climate}_{c,t} + \boldsymbol{\delta}^\top \text{Socio}_{c,t} + \varepsilon_{c,t}. \qquad (3)
\end{aligned}$$

Here, α_c and λ_t denote county and month fixed effects. Observations are weighted by the number of panoramas (`pano_n`) to correct for heteroskedasticity introduced by unequal sampling density, and cluster-robust standard errors are computed at the county level. All estimations are implemented in `statsmodels` via a unified `fit_wls` routine.

3.5.3 Interpretation and Validity

The coefficient β_1 measures the change in vegetation share associated with a one-percentage-point rise in the most recent Democratic vote share. Hypothesis 1 is supported if $\beta_1 > 0$ and statistically significant in the Plain model, while Hypothesis 2 requires this result to persist after adding socio-economic controls and fixed effects. Temporal ordering—electoral variables always precede vegetation by at least one electoral cycle—limits reverse causality. Climate, socio-economic covariates, and fixed effects further mitigate omitted-variable bias, and the inclusion of depth-adjusted outcomes provides an internal robustness check that guards against measurement error. Collectively, these design features offer a stringent test of whether Democratic electoral dominance translates into observable enhancements in county-level greenery.

3.6 Results

This section summarizes the substantive patterns that emerge from the three regression blocks reported in the tables below.[2]

3.6.1 Baseline Associations (Plain Model)

The "Plain" specification in Table 2 establishes a clear bivariate link between recent electoral behavior and visible greenness. A one–percentage-point increase in the Democratic share of the *most recent* presidential election (`dem_pct_t1`) is associated with an increase of roughly 0.40–0.59 percentage points in the plain vegetation indices (*Tree*, *Total*) and a slightly smaller, yet still positive, bump in the depth-adjusted indices. These estimates are statistically significant at the 1% level for most outcomes, indicating a robust, contemporaneous relationship even before socio-economic adjustments. By contrast, the vote share from the *second-most recent* election (`dem_pct_t2`) either carries a small negative coefficient or fails to reach conventional significance thresholds, suggesting that the political signal decays over time or is confounded by intervening policy cycles.

Climatic conditions contribute expected directional effects. Higher mean January temperatures dampen grass cover but leave tree cover largely unchanged, whereas warmer Aprils enhance grass growth. Precipitation is consistently positive and significant for *Grass* and *Total* vegetation, while greater snowfall is negatively related to grass density, in line with the notion that snow-heavy counties devote less budget or attention to sod maintenance. The overall model fit is respectable for a cross-section without controls: R^2 values cluster around 0.33–0.49 for the plain indices and improve marginally (to about 0.36–0.49) under the *DEVISE* metrics, underscoring the explanatory power of a small set of climatic and political covariates.

3.6.2 Accounting for Socio-Economic Structure (Social Model)

Table 3 augments the baseline with available socio-economic covariates, thereby reducing the sample to 62 county and date pairs with complete data but raising explanatory power substantially. The coefficient on `dem_pct_t1` strengthens: a one-point Democratic swing now predicts approximately 0.59–0.71 percentage-point gains in *Tree* and *Total* greenness (*Old*) and 0.39–0.52 in the corresponding *DEVISE* outcomes. Importantly, these effects survive the inclusion of income, population size, unemployment, gender, racial composition, and educational attainment, reinforcing the argument that partisan alignment carries independent weight beyond economic capacity or demographic structure.

Socio-economic variables behave largely as expected. Higher unemployment dampens canopy and composite vegetation metrics, presumably reflecting reduced fiscal capacity for maintenance programs. Larger counties (in absolute population) exhibit a mild negative association with grass coverage, consistent with urbanization pressures. Other demographic indicators (percent female, percent White) do not exert systematic influence once climate and political factors are controlled. Precipitation remains a potent driver of all green indices, with point estimates roughly triple the Plain-model magnitudes, and snowfall loses significance, indicating that socio-economic heterogeneity partially mediates winter-weather effects.

[2]For readability, we refer to the "Old" vegetation measures—pixel shares uncorrected for depth—and their depth-weighted counterparts generated by *DEVISE*. In each block we discuss the four canonical outcomes (Tree, Grass, Plant, Total); results for the remaining variables follow the same logic.

The inclusion of these additional covariates nearly doubles model fit. R^2 values now lie between 0.56 and 0.81, with the highest explanatory power appearing for *Grass* measures—precisely the outcome most sensitive to precipitation and seasonal growing conditions.

3.6.3 Robustness to Unobserved Heterogeneity (Social + FE Model)

Table 4 introduces county and month fixed effects, thereby absorbing all time-invariant county traits and systematic temporal shocks. Despite the more demanding specification, the primary political signal persists: the immediate Democratic vote share remains positively signed and statistically significant for *Grass* and *Total* greenness in both measurement systems. The magnitude falls relative to the Social model—now roughly 0.33–0.79 percentage points per vote-share point—but the attenuation is modest given the stringent within-county identification strategy. Notably, `dem_pct_t1` turns insignificant for *Tree* density, hinting that canopy expansion may respond more slowly to electoral turnover than ground-level vegetation.

Introducing fixed effects also alters the climate coefficients. Mean January temperature flips sign for *Tree* (*Old*) and becomes strongly positive, while remaining negative for *Grass*; this pattern suggests that warmer winters aid tree survival but can reduce dormant-season grass visibility once county baselines are netted out. April and July temperatures continue to display offsetting effects between woody and herbaceous cover, reinforcing the ecological intuition that seasonality shapes vegetation types differently.

Among socio-economic controls, population retains a small but precisely estimated negative effect, particularly for *Tree* and *Total*. Higher unemployment still erodes tree canopy, whereas educational attainment (*Some College / Bachelor*) now predicts more grass but fewer trees—a plausible reflection of divergent aesthetic norms in better-educated communities. Overall model fit rises dramatically: R^2 values exceed 0.90 for all outcomes except *Grass* (*DEVISE*), which approaches 0.99. This leap confirms that county-specific, time-invariant characteristics explain a large share of cross-sectional variance but leaves intact the main partisan association uncovered in simpler models.

3.7 Substantive Implications

Taken together, the three regression tiers support two general conclusions. First, Democratic electoral strength is positively linked to observable greenery, especially to grass-dominated ground cover, even after extensive adjustment for climate, socio-economic context, and unobserved county heterogeneity. The effect is contemporaneous: lagged vote shares (`dem_pct_t2`) rarely attain significance once more recent behavior and controls enter the model. Second, the *DEVISE* depth-weighted metrics behave similarly to their plain counterparts but exhibit slightly attenuated coefficients, suggesting that the partisan advantage in greenness is somewhat stronger at ground level than in the vertical canopy dimension. This nuance may reflect the ease with which local governments can install and maintain low-lying vegetation—e.g., median plantings or park lawns—relative to longer-horizon tree-planting initiatives.

Climate remains an indispensable predictor, particularly precipitation, which consistently enhances grass and composite coverage. Socio-economic variables add explanatory power but do not fully account for the partisan gradient, underscoring the distinct role of political will in shaping visible environmental outcomes. Finally, the persistence of

the Democratic effect in the fixed-effects specification confirms that the association is not merely an artifact of long-standing regional differences but also operates through within-county change over time.

In sum, the empirical evidence aligns with Hypothesis 1 and, to a more qualified extent, Hypothesis 2: counties trending Democratic realize measurable gains in street-level greenness, and these gains survive rigorous robustness checks. Future work should explore potential mechanisms—such as dedicated tree-planting ordinances, storm-water regulations, or participatory budgeting for park maintenance—that translate partisan intent into ecological practice.

Table 2: Plain Regression Block

	Tree (Old)	Grass (Old)	Plant (Old)	Total (Old)	Tree (*DEVISE*)	Grass (*DEVISE*)	Plant (*DEVISE*)	Total (*DEVISE*)
Dem. vote share (t1)	0.0039***	0.0009	0.0001*	0.0050***	0.0025***	0.0010*	0.0002*	0.0036***
	(0.0011)	(0.0006)	(0.0001)	(0.0013)	(0.0007)	(0.0006)	(0.0001)	(0.0009)
Dem. vote share (t2)	−0.0015	−0.0003	−0.0001	−0.0018	−0.0009	−0.0006	−0.0001	−0.0016**
	(0.0011)	(0.0005)	(0.0001)	(0.0012)	(0.0006)	(0.0005)	(0.0001)	(0.0008)
Avg. Temp. (Jan)	0.0004	−0.0024***	−0.0001	−0.0020	0.0002	−0.0025***	−0.0001	−0.0023*
	(0.0016)	(0.0007)	(0.0001)	(0.0021)	(0.0009)	(0.0007)	(0.0001)	(0.0014)
Avg. Temp. (Apr)	−0.0012	0.0033***	0.0001	0.0022	−0.0007	0.0039***	0.0001	0.0032
	(0.0034)	(0.0011)	(0.0002)	(0.0043)	(0.0019)	(0.0012)	(0.0002)	(0.0029)
Avg. Temp. (Jul)	0.0024	−0.0006	−0.0002	0.0016	0.0015	−0.0010	−0.0002	0.0003
	(0.0023)	(0.0008)	(0.0002)	(0.0029)	(0.0013)	(0.0008)	(0.0002)	(0.0020)
Precipitation	0.0009	0.0007**	0.0001	0.0017*	0.0006	0.0008**	0.0001*	0.0015**
	(0.0008)	(0.0003)	(0.0001)	(0.0010)	(0.0005)	(0.0003)	(0.0001)	(0.0006)
Snowfall	0.0003	−0.0003**	−0.0000	−0.0001	0.0001	−0.0003*	−0.0000	−0.0002
	(0.0003)	(0.0001)	(0.0000)	(0.0004)	(0.0002)	(0.0002)	(0.0000)	(0.0003)
R-squared	0.332	0.456	0.219	0.415	0.365	0.487	0.255	0.492
Observations	174	174	174	174	174	174	174	174
County FE	No	No	No	No	No	No	No	No
Date FE	No	No	No	No	No	No	No	No

Notes: Robust standard errors clustered by county in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

4 Policy and Political Implications

By shifting the locus of environmental monitoring from satellite imagery to pedestrian vantage points, DEVISE offers a substantial refinement in spatial fidelity. Such fine-grained observation allows municipal and state agencies to verify whether expenditures on green infrastructure—ranging from storm-water retrofits in Dauphin County to riparian-buffer bonds in Bucks County—materialize as observable street-level vegetation rather than as merely symbolic compliance farther from the public realm. Because the framework privileges vegetation that appears closer to the viewer, it also constrains the effectiveness of "green-washing" practices in which developers cluster plantings at site margins while leaving sidewalks barren.

The same granularity enables a more rigorous diagnosis of environmental inequity. Scaling DEVISE to millions of panoramas makes it possible to map vegetation shortfalls that closely track heat stress, asthma prevalence, and socio-economic vulnerability—phenomena often masked by coarser satellite metrics. Embedding these maps in participatory-budgeting procedures or equity-weighted grant scoring would enhance the likelihood that recent climate-resilience funds reach the communities most in need, thereby aligning distributive outcomes with legislative intent.

Beyond administrative efficiency, the method carries political implications. Regression results in Section 3 indicate that counties experiencing recent shifts toward Democratic

Table 3: Social Regression Block

	Tree (Old)	Grass (Old)	Plant (Old)	Total (Old)	Tree (*DEVISE*)	Grass (*DEVISE*)	Plant (*DEVISE*)	Total (*DEVISE*)
Dem. vote share (t1)	0.0059***	0.0011	0.0002	0.0071***	0.0039***	0.0011	0.0002	0.0052***
	(0.0017)	(0.0008)	(0.0002)	(0.0021)	(0.0010)	(0.0008)	(0.0002)	(0.0016)
Dem. vote share (t2)	−0.0025**	−0.0003	−0.0004*	−0.0032***	−0.0015**	−0.0005	−0.0004*	−0.0024**
	(0.0011)	(0.0006)	(0.0002)	(0.0011)	(0.0007)	(0.0006)	(0.0002)	(0.0010)
Avg. Temp. (Jan)	−0.0017	−0.0018	0.0001	−0.0033	−0.0014	−0.0013	0.0001	−0.0026
	(0.0032)	(0.0017)	(0.0004)	(0.0038)	(0.0019)	(0.0017)	(0.0004)	(0.0027)
Avg. Temp. (Apr)	0.0037	0.0023	−0.0003	0.0056	0.0028	0.0013	−0.0004	0.0038
	(0.0074)	(0.0040)	(0.0010)	(0.0082)	(0.0043)	(0.0041)	(0.0010)	(0.0059)
Avg. Temp. (Jul)	−0.0023	0.0009	0.0001	−0.0013	−0.0021	0.0019	0.0002	−0.0000
	(0.0066)	(0.0037)	(0.0009)	(0.0069)	(0.0039)	(0.0039)	(0.0010)	(0.0051)
Precipitation	0.0034***	0.0012***	0.0001**	0.0048***	0.0021***	0.0013***	0.0002**	0.0036***
	(0.0009)	(0.0004)	(0.0001)	(0.0011)	(0.0005)	(0.0004)	(0.0001)	(0.0008)
Snowfall	0.0003	−0.0001	0.0000	0.0002	0.0001	0.0000	0.0000	0.0001
	(0.0004)	(0.0002)	(0.0000)	(0.0005)	(0.0002)	(0.0002)	(0.0000)	(0.0004)
Per Capita Personal Income	0.0000	−0.0000	−0.0000	0.0000	0.0000	−0.0000	−0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Population	−0.0000	−0.0000**	−0.0000	−0.0000	0.0000	−0.0000**	0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Percent Unemployed	−0.0080**	0.0009	−0.0008	−0.0079**	−0.0044*	0.0006	−0.0008	−0.0047*
	(0.0040)	(0.0018)	(0.0007)	(0.0037)	(0.0023)	(0.0019)	(0.0007)	(0.0027)
Percent Female	0.0026	0.0013	−0.0000	0.0039	0.0009	0.0012	−0.0002	0.0019
	(0.0030)	(0.0016)	(0.0004)	(0.0038)	(0.0017)	(0.0017)	(0.0004)	(0.0028)
Percent White	0.0004	0.0004	0.0001	0.0008	0.0001	0.0004	0.0001	0.0006
	(0.0006)	(0.0003)	(0.0001)	(0.0008)	(0.0004)	(0.0003)	(0.0001)	(0.0005)
Some College / Bachelor	−0.0002	0.0008	−0.0001	0.0005	−0.0002	0.0012	−0.0000	0.0010
	(0.0016)	(0.0008)	(0.0002)	(0.0019)	(0.0010)	(0.0009)	(0.0002)	(0.0015)
R-squared	0.573	0.780	0.572	0.642	0.581	0.805	0.564	0.700
Observations	62	62	62	62	62	62	62	62
County FE	No	No	No	No	No	No	No	No
Date FE	No	No	No	No	No	No	No	No

Notes: Robust standard errors clustered by county in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 4: Social + Fixed Effects Regression Block

	Tree (Old)	Grass (Old)	Plant (Old)	Total (Old)	Tree (*DEVISE*)	Grass (*DEVISE*)	Plant (*DEVISE*)	Total (*DEVISE*)
Dem. vote share (t1)	0.0048	0.0035***	−0.0004	0.0079	0.0035	0.0033***	−0.0004	0.0064**
	(0.0045)	(0.0009)	(0.0004)	(0.0049)	(0.0026)	(0.0007)	(0.0004)	(0.0027)
Dem. vote share (t2)	0.0006	−0.0007	−0.0001	−0.0002	0.0002	−0.0009	−0.0002	−0.0008
	(0.0028)	(0.0008)	(0.0004)	(0.0025)	(0.0016)	(0.0008)	(0.0004)	(0.0013)
Avg. Temp. (Jan)	0.0321*	−0.0153**	0.0001	0.0168	0.0202**	−0.0118*	0.0004	0.0088
	(0.0176)	(0.0073)	(0.0020)	(0.0181)	(0.0103)	(0.0066)	(0.0020)	(0.0115)
Avg. Temp. (Apr)	−0.0210	0.0202**	0.0002	−0.0006	−0.0147	0.0151*	−0.0004	−0.0000
	(0.0188)	(0.0092)	(0.0022)	(0.0210)	(0.0111)	(0.0082)	(0.0022)	(0.0143)
Avg. Temp. (Jul)	0.0047	−0.0160*	−0.0002	−0.0115	0.0049	−0.0107	0.0004	−0.0054
	(0.0137)	(0.0087)	(0.0014)	(0.0188)	(0.0086)	(0.0077)	(0.0014)	(0.0140)
Precipitation	−0.0054	0.0049	0.0001	−0.0005	−0.0034	0.0035	0.0001	0.0002
	(0.0065)	(0.0031)	(0.0006)	(0.0079)	(0.0039)	(0.0028)	(0.0006)	(0.0055)
Snowfall	0.0101**	−0.0039**	0.0000	0.0063	0.0062**	−0.0030*	0.0001	0.0034
	(0.0049)	(0.0020)	(0.0005)	(0.0051)	(0.0029)	(0.0018)	(0.0005)	(0.0032)
Per Capita Personal Income	−0.0000	0.0000	0.0000	−0.0000	−0.0000	0.0000	0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Population	−0.0000***	0.0000*	−0.0000	−0.0000***	−0.0000***	0.0000	−0.0000	−0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Percent Unemployed	−0.0088**	0.0022	−0.0003	−0.0070	−0.0030	0.0029	0.0001	0.0000
	(0.0042)	(0.0025)	(0.0005)	(0.0057)	(0.0026)	(0.0021)	(0.0004)	(0.0040)
Percent Female	0.0005	0.0009	0.0003	0.0017	−0.0019	−0.0011	−0.0002	−0.0031
	(0.0052)	(0.0019)	(0.0003)	(0.0062)	(0.0031)	(0.0017)	(0.0003)	(0.0040)
Percent White	0.0027	0.0010	0.0000	0.0037	0.0014	0.0003	0.0001	0.0018
	(0.0030)	(0.0016)	(0.0002)	(0.0038)	(0.0018)	(0.0014)	(0.0002)	(0.0027)
Some College / Bachelor	−0.0062**	0.0037***	−0.0001	−0.0026	−0.0031*	0.0032***	−0.0001	0.0000
	(0.0027)	(0.0011)	(0.0002)	(0.0032)	(0.0016)	(0.0010)	(0.0002)	(0.0021)
R-squared	0.837	0.978	0.916	0.880	0.834	0.985	0.907	0.912
Observations	62	62	62	62	62	62	62	62
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: County and date fixed effects included. Robust standard errors clustered by county in parentheses.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

vote shares record contemporaneous gains in grass and composite vegetation, even after stringent county and time fixed effects. The finding suggests that visible greenery may itself be electorally salient, creating a feedback loop in which politicians anticipate, and voters reward, investments that translate into tangible aesthetic improvements. At the same time, the muted and slower response of tree canopy underscores a potential misalignment between short election cycles and the longer horizons required for mature urban forestry, highlighting the value of policy instruments—such as revolving trust funds or intertemporal matching grants—that insulate perennial planting programs from political turnover.

To facilitate widespread adoption, *DEVISE* will be released as an open-source `Python` package accompanied by comprehensive documentation and illustrative notebooks. The package will allow practitioners to process raw Street View imagery, generate depth-aware vegetation indices, and visualize greenness deficits using only a modest amount of code. By lowering technical barriers, the software component translates the methodological advance into an actionable toolkit for planners, community organizations, and researchers, thereby strengthening the empirical basis for greener and more equitable urban policy.

References

Agency, M. P. C. (2025). 2025 green infrastructure revolving fund highlights. https://www.pca.state.mn.us/. Accessed July 2025.

Aikoh, K. and H. Takahashi (2023). Biases in deep-learning–based green view indices: Lessons from japanese streetscapes. *Landscape and Urban Planning 237*, 104838.

American Forests (2023). American forests unveils updates for tree equity score tool to address climate justice. Press release; accessed 2025-07-06.

Chen, J., Y. Wang, J. Zhang, and H. Li (2024). DPT-BLUR: Acknowledging depth for arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 555–564.

Cheng, B., A. Kirillov, R. Girshick, A. C. Berg, and R. Girdhar (2022). Masked-attention mask transformer for universal image segmentation (mask2former). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1290–1299.

Choi, J. and K. Lee (2022). Tree species classification in urban streetscapes using gsv and deep learning. *Urban Forestry & Urban Greening 71*, 127546.

Commission, B. C. P. (2007). County open space bond program. https://www.buckscounty.gov/1139/Environmental-and-Natural-Resources. Accessed July 2025.

Data, M. E. and S. Lab (2022). County presidential election returns, 2004–2020. https://doi.org/10.7910/DVN/VOQCHQ. Accessed July 2025.

District, W. C. (2023). Habitat restoration program. https://www.mnwcd.org/habitat-restoration. Accessed July 2025.

Dogan, R. (2025). Cooling the city: How european cities are adapting to extreme heat. Accessed 2025-07-06.

Gupta, K., P. Kumar, S. Pathan, and K. Sharma (2012, February). Urban neighborhood green index – a measure of green spaces in urban areas. *Landscape and Urban Planning*.

Hassanein, E. and A. Abdelatif (2018). Hue-based thresholding of uav imagery for low-cost urban green cover assessment. *International Journal of Remote Sensing 39*(15), 4914–4931.

Kameoka, R. and Y. Hirashima (2022). Patch-based segmentation of street trees: Improving gvi at city scale. *Remote Sensing 14*(9), 2156.

Lauko, I. G., A. Honts, J. Beihoff, and S. Rupprecht (2020, August). Local color and morphological image feature-based vegetation identification and its application to human environment street view vegetation mapping. *Geo-Spatial Information Science*.

Lee, J., S. Kim, M. Cho, and S. Kim (2024). Tri-MipRF: Tri-mip representation for efficient and high-quality 3D reconstruction from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Li, X., C. Zhang, W. Li, R. Ricard, Q. Meng, and L. Zhang (2015, June). Assessing street-level urban greenery using google street view and a modified green view index. *Urban Forestry & Urban Greening*.

Mohadikar, S., S. Garg, and A. Singh (2024). OmniDreamer: A 360-degree monocular depth estimation using diffusion models. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5777–5786. IEEE.

of Conservation, P. D. and N. Resources (2020). Treevitalize grants program. `https://treepennsylvania.org/treevitalize-grants-program/`. Accessed July 2025.

Pineda, D. (2025). Anti-dei push ends effort to plant trees in places that need them. Accessed 2025-07-06.

Report, U. N. . W. (2021). Healthiest communities ranking 2021: Bucks county, pa. `https://www.usnews.com/news/healthiest-communities`. Accessed July 2025.

Small, C. (2001, April). Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*.

Society, P. H. (2009). The southeast pa treevitalize partnership. `https://planningpa.org/presentations09/49_TreeVitalize_handout.pdf`. Accessed July 2025.

Sonam, N., N. Patel, and K. M. Sharma (2024). DPT-Saliency: A DPT-based framework for RGB-D and video saliency detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3106–3115.

Sun, Z., H. Zhou, and Y. Li (2021). Hsvi: A hyperspectral image-based vegetation index for fine-grained urban green mapping. *ISPRS Journal of Photogrammetry and Remote Sensing 178*, 163–176.

U.S. Department of Agriculture (2024). Usda urban and community forestry program award to spur new tree cover near urban congregations. Press release; accessed 2025-07-06.

Washington County, M. (2019). Washington county 2040 comprehensive plan. `https://www.washingtoncountymn.gov/404/Comprehensive-Plan`. Adopted January 7 2020; Accessed July 2025.

Water, C. R. (2017). Community greening plan: A green stormwater infrastructure plan for harrisburg. `https://capitalregionwater.com/wp-content/uploads/2017/04/Community-Greening-Plan_FINAL_smallest.pdf`. Accessed July 2025.

Weiss, H. (2023). Heat-related deaths are rising. here's how cities can keep residents safer. Accessed 2025-07-06.

Xia, Y., S. Wang, and M. Ding (2021). Semantic segmentation of google street view imagery for pedestrian-level green view index. *Computers, Environment and Urban Systems 88*, 101673.

Yu, B., L. Liu, and X. Zhou (2018). Street-view based green view index in china: The baidu gvi. *Remote Sensing 10*(4), 631.

Yu, H., T. Li, and Y. Ratti (2022). Depth matters: 3d reconstruction improves street-level environmental exposure assessment. *Environment International 158*, 106989.

Zhang, L., R. Chen, and D. Xu (2023). Panorama view green view index (pvgvi): A new street-green metric with deeplabv3+. *Ecological Indicators 150*, 109875.

Özçelikel, M. E. (2025). Maltepe thermal walks: Citizen science project highlights local impact of heatwaves. Accessed 2025-07-06.

Appendix

A Additional Figures and Data

A.1 Masked-attention Mask Transformer

Masked-attention Mask Transformer (Masked-attention Mask Transformer) is a unified framework for image segmentation tasks, introduced to handle semantic segmentation, instance segmentation, and panoptic segmentation with a single architecture. Unlike traditional segmentation methods that rely heavily on dense pixel-wise classification, Mask2Former approaches segmentation as a mask prediction problem.

Mask2Former was proposed by Bowen Cheng, Ishan Misra, Alexander Kirillov, Rohit Girdhar, and Alexander C. Berg. The paper, titled "Masked-attention Mask Transformer for Universal Image Segmentation", was published in 2022 by researchers from Facebook AI Research (FAIR)

The core idea is to use a transformer decoder that predicts a set of segmentation masks and their associate class labels in parallel. Each query in the transformer learns to represent a particular object or region in the image. This structure allows the model to capture global context and long-range dependencies more efficiently than CNN-based architectures.

A key component of Mask2Former is the masked attention mechanism, which focuses computation within predicted mask regions, improving both accuracy and efficiency. Instead of attending to the full image, each decoder query only attends to the masked regions relevant to its prediction. This results in a more scalable and interpretable model.

By using this unified architecture, Mask2Former achieves state-of-the-art performance across multiple segmentation benchmarks, while simplifying the design by avoiding task-specific heads or architecture.

A.2 Dense Prediction Transformer

Dense Prediction Transformer (DPT) is a transformer-based architecture developed by intel Labs for solving dense prediction tasks such as monocular depth estimation, semantic segmentation, and panoptic segmentation. It was introducted in the paper "Vision Transformers for Dense Prediction" by René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun, and was released in 2021.

Traditional convolutional neural networks (CNNS) often struggle to capture long-range dependencies in images, which can limit their performance on tasks that require global context. DPT addresses this limitation by using a Vision Transformer (ViT)

backbone, which naturally captures global interactions between all image regions through self-attention mechanisms.

The model takes an input image and processes it using a ViT-based encoder to generate a rich, global feature representation. A specifically designed decoder then up-samples and refines theses features to produce dense output maps, such as depth maps or segmentation tasks.

DPT is notable for achieving state-of-the-art results in monocular depth estimation while remaining general enough to perform well on other dense prediction tasks. Its design demonstrates the effectiveness of transformer-based models in dense visual understanding.

A.3 Code Listings

A.3.1 combine.py

```
import numpy as np
import cv2
import os
from PIL import Image

color_mappings = {
    0: {"name": "Tree", "color": (0, 100, 0)},      # Tree
    1: {"name": "Grass", "color": (144, 238, 144)}, # Grass
    2: {"name": "Plant", "color": (34, 139, 34)}    # Plant
}

def overlay_mask_on_image(image_path, mask_npy_path,
   output_path):

    image = Image.open(image_path).convert("RGB")
    image_np = np.array(image)

    mask_data = np.load(mask_npy_path)

    mask_overlay = np.zeros_like(image_np, dtype=np.uint8)

    for i, properties in color_mappings.items():
        mask = mask_data[:, :, i]
        mask = (mask == 255).astype(np.uint8) * 255

        if np.count_nonzero(mask) == 0:
            continue

        color_layer = np.full_like(image_np,
           properties["color"], dtype=np.uint8)

        mask_overlay = np.where(mask[:, :, None] == 255,
           color_layer, mask_overlay)

    alpha = 0.5
    final_result = cv2.addWeighted(image_np, 1 - alpha,
       mask_overlay, alpha, 0)
```

```
34
35      cv2.imwrite(output_path, cv2.cvtColor(final_result,
            cv2.COLOR_RGB2BGR))
36
37  def main(image_folder, mask_folder, result_folder):
38
39      os.makedirs(result_folder, exist_ok=True)
40
41      image_files = [f for f in os.listdir(image_folder) if
            f.lower().endswith('.jpg')]
42
43      for image_file in image_files:
44          image_path = os.path.join(image_folder, image_file)
45          mask_filename =
                os.path.splitext(image_file)[0].replace('_resized',
                '') + "_mask.npy"
46          mask_path = os.path.join(mask_folder, mask_filename)
47          output_path = os.path.join(result_folder,
                image_file.replace('_resized', '_combined'))
48
49          if os.path.exists(mask_path):
50              print(f"     {image_file} and {mask_filename}
                    combined")
51              overlay_mask_on_image(image_path, mask_path,
                    output_path)
52
53      print("FINISH")
54
55  image_folder = "resized_folder"
56  mask_folder = "mask_folder"
57  result_folder = "combined_folder"
58
59  main(image_folder, mask_folder, result_folder)
```

Listing 1: combine.py

A.3.2 depth_combine.py

```
1   import os
2   import numpy as np
3   import sqlite3
4
5   def compute_weighted_gvi(mask, depth_norm):
6       mask = (mask == 255).astype(np.float32)
7       weight = depth_norm.astype(np.float32)
8       numerator = np.sum(mask * weight)
9       denominator = np.sum(weight)
10      return 0.0 if denominator == 0 else numerator / denominator
11
12  def init_db(db_path):
13      conn = sqlite3.connect(db_path)
```

```
    cursor = conn.cursor()
    cursor.execute("""
        CREATE TABLE IF NOT EXISTS gvi (
            image_id TEXT PRIMARY KEY,
            tree_gvi REAL,
            grass_gvi REAL,
            plant_gvi REAL,
            total_gvi REAL
        )
    """)
    conn.commit()
    return conn

def save_to_db(conn, image_id, tree, grass, plant, total):
    cursor = conn.cursor()
    cursor.execute("""
        REPLACE INTO gvi (image_id, tree_gvi, grass_gvi,
           plant_gvi, total_gvi)
        VALUES (?, ?, ?, ?, ?)
    """, (image_id, tree, grass, plant, total))
    conn.commit()

def main(mask_folder, depth_folder, db_path):
    conn = init_db(db_path)
    mask_files = [f for f in os.listdir(mask_folder) if
       f.endswith("_mask.npy")]
    for mask_file in mask_files:
        image_id = mask_file.replace("_mask.npy", "")
        mask_path = os.path.join(mask_folder, mask_file)
        depth_path = os.path.join(depth_folder,
           f"{image_id}_depth.npy")
        if not os.path.exists(depth_path):
            print(f"        Skipped {image_id}: depth not found")
            continue
        mask = np.load(mask_path)
        depth = np.load(depth_path)
        if mask.shape[:2] != depth.shape:
            print(f"        Skipped {image_id}: mask and depth
               size mismatch")
            continue
        tree = compute_weighted_gvi(mask[:, :, 0], depth)
        grass = compute_weighted_gvi(mask[:, :, 1], depth)
        plant = compute_weighted_gvi(mask[:, :, 2], depth)
        combined_mask = ((mask[:, :, 0] == 255) | (mask[:, :, 1]
           == 255) | (mask[:, :, 2] == 255)).astype(np.uint8) *
           255
        total = compute_weighted_gvi(combined_mask, depth)
        save_to_db(conn, image_id, float(tree), float(grass),
           float(plant), float(total))
        print(f"     {image_id} depth information combined")
    conn.close()
```

```
    print("FINISH")

mask_folder = "mask_folder"
depth_folder = "depth_mask_folder"
db_path = "gvi_with_depth.db"

main(mask_folder, depth_folder, db_path)
```

Listing 2: depth_combine.py

A.3.3 depth.py

```
import os
import cv2
import numpy as np
from PIL import Image
import torch
from transformers import DPTForDepthEstimation, DPTImageProcessor

device = torch.device("cuda" if torch.cuda.is_available() else
    "cpu")

depth_processor =
    DPTImageProcessor.from_pretrained("Intel/dpt-large")
depth_model =
    DPTForDepthEstimation.from_pretrained("Intel/dpt-large").to(device)
depth_model.eval()

def main(input_folder, depth_mask_folder, depth_image_folder):
    os.makedirs(depth_mask_folder, exist_ok=True)
    os.makedirs(depth_image_folder, exist_ok=True)

    image_files = [f for f in os.listdir(input_folder) if
        f.endswith("_resized.jpg")]

    for image_file in image_files:
        image_id =
            os.path.splitext(image_file)[0].replace("_resized",
            "")
        image_path = os.path.join(input_folder, image_file)

        npy_output_path = os.path.join(depth_mask_folder,
            f"{image_id}_depth.npy")
        image_output_path = os.path.join(depth_image_folder,
            f"{image_id}_depth.jpg")

        image = Image.open(image_path).convert("RGB")
        original_size = image.size  # (width, height)

        inputs = depth_processor(images=image,
            return_tensors="pt").to(device)
```

```
31
32          with torch.no_grad():
33              outputs = depth_model(**inputs)
34          depth = outputs.predicted_depth.squeeze().cpu().numpy()
35          depth_norm = (depth - depth.min()) / (depth.max() -
               depth.min())
36
37          depth_resized = cv2.resize(depth_norm, original_size,
               interpolation=cv2.INTER_CUBIC)
38
39          np.save(npy_output_path, depth_resized)
40
41          depth_visual = (depth_resized * 255).astype(np.uint8)
42          cv2.imwrite(image_output_path, depth_visual)
43
44          print(f"     {image_id}_depth.jpg and
               {image_id}_depth.npy saved")
45
46      print("FINISH")
47
48  resized_folder = "resized_folder"
49  depth_mask_folder = "depth_mask_folder"
50  depth_image_folder = "depth_image_folder"
51
52  main(resized_folder, depth_mask_folder, depth_image_folder)
```

Listing 3: depth.py

A.3.4 pipeline.py

```
1  import os
2  import cv2
3  import numpy as np
4  from PIL import Image
5  import torch
6  import sqlite3
7  from transformers import Mask2FormerForUniversalSegmentation,
      Mask2FormerImageProcessor
8
9  device = torch.device("cuda" if torch.cuda.is_available() else
      "cpu")
10
11 processor =
      Mask2FormerImageProcessor.from_pretrained("facebook/mask2former-swin-large
12 model =
      Mask2FormerForUniversalSegmentation.from_pretrained("facebook/mask2former
13 model.eval()
14
15 color_mappings = {
16      4: {"name": "Tree", "hsv": ([20, 40, 30], [95, 255, 255])},
17      9: {"name": "Grass", "hsv": ([20, 40, 30], [95, 255, 255])},
```

```
    17: {"name": "Plant", "hsv": ([20, 35, 30], [92, 255, 255])}
}

def init_gvi_db(db_path):
    conn = sqlite3.connect(db_path)
    cursor = conn.cursor()
    cursor.execute("""
        CREATE TABLE IF NOT EXISTS gvi (
            image_id TEXT PRIMARY KEY,
            tree_gvi REAL,
            grass_gvi REAL,
            plant_gvi REAL,
            total_gvi REAL
        )
    """)
    conn.commit()
    return conn

def save_gvi_to_db(conn, image_id, tree_gvi, grass_gvi,
   plant_gvi, total_gvi):
    cursor = conn.cursor()
    cursor.execute("""
        REPLACE INTO gvi (image_id, tree_gvi, grass_gvi,
           plant_gvi, total_gvi)
        VALUES (?, ?, ?, ?, ?)
    """, (image_id, tree_gvi, grass_gvi, plant_gvi, total_gvi))
    conn.commit()

def remove_black_borders_and_resize(image_path, output_folder,
   target_size=(4000, 2000)):
    img = cv2.imread(image_path)
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    _, thresh = cv2.threshold(gray, 1, 255, cv2.THRESH_BINARY)
    contours, _ = cv2.findContours(thresh, cv2.RETR_EXTERNAL,
       cv2.CHAIN_APPROX_SIMPLE)
    if not contours:
        resized_img = cv2.resize(img, target_size,
           interpolation=cv2.INTER_LANCZOS4)
    else:
        x, y, w, h = cv2.boundingRect(contours[0])
        cropped_img = img[y:y+h, x:x+w]
        resized_img = cv2.resize(cropped_img, target_size,
           interpolation=cv2.INTER_LANCZOS4)
    os.makedirs(output_folder, exist_ok=True)
    filename = os.path.splitext(os.path.basename(image_path))[0]
       + "_resized.jpg"
    output_path = os.path.join(output_folder, filename)
    cv2.imwrite(output_path, resized_img,
       [int(cv2.IMWRITE_JPEG_QUALITY), 95])
    return output_path

```

```
61 def apply_hsv_filter(image, mask, hsv_range):
62     hsv_image = cv2.cvtColor(image, cv2.COLOR_RGB2HSV)
63     lower_hsv, upper_hsv = np.array(hsv_range[0]),
          np.array(hsv_range[1])
64     green_mask = cv2.inRange(hsv_image, lower_hsv, upper_hsv)
65     final_mask = cv2.bitwise_and(mask, mask, mask=green_mask)
66     return final_mask
67
68 def calculate_gvi(final_mask):
69     total_pixels = final_mask.shape[0] * final_mask.shape[1]
70     tree_gvi = np.count_nonzero(final_mask[:, :, 0]) /
          total_pixels
71     grass_gvi = np.count_nonzero(final_mask[:, :, 1]) /
          total_pixels
72     plant_gvi = np.count_nonzero(final_mask[:, :, 2]) /
          total_pixels
73     total_gvi = tree_gvi + grass_gvi + plant_gvi
74     return tree_gvi, grass_gvi, plant_gvi, total_gvi
75
76 def process_and_generate_mask(image_path, mask_output_folder,
      db_conn):
77     image_id =
          os.path.splitext(os.path.basename(image_path))[0].replace('_resized',
          '')
78     image_pil = Image.open(image_path).convert("RGB")
79     image_np = np.array(image_pil)
80     height, width = image_np.shape[:2]
81     inputs = processor(images=image_np,
          return_tensors="pt").to(device)
82     with torch.no_grad():
83         outputs = model(**inputs)
84     results =
          processor.post_process_semantic_segmentation(outputs,
          target_sizes=[(height, width)])[0]
85     segmentation = results.cpu().numpy()
86     final_mask = np.zeros((height, width, 3), dtype=np.uint8)
87     for i, (label, properties) in
          enumerate(color_mappings.items()):
88         mask = (segmentation == label).astype(np.uint8) * 255
89         if np.count_nonzero(mask) == 0:
90             continue
91         filtered_mask = apply_hsv_filter(image_np, mask,
              properties["hsv"])
92         final_mask[:, :, i] = filtered_mask
93     os.makedirs(mask_output_folder, exist_ok=True)
94     output_filename = f"{image_id}_mask.npy"
95     np.save(os.path.join(mask_output_folder, output_filename),
          final_mask)
96     tree_gvi, grass_gvi, plant_gvi, total_gvi =
          calculate_gvi(final_mask)
```

```
97      save_gvi_to_db(db_conn, image_id, tree_gvi, grass_gvi,
            plant_gvi, total_gvi)
98
99  def main(raw_folder, resized_folder, mask_folder, gvi_db_path):
100     os.makedirs(resized_folder, exist_ok=True)
101     os.makedirs(mask_folder, exist_ok=True)
102
103     conn = init_gvi_db(gvi_db_path)
104
105     image_files = [f for f in os.listdir(raw_folder) if
            f.lower().endswith(('.jpg', '.jpeg', '.png'))]
106     for image_file in image_files:
107         raw_path = os.path.join(raw_folder, image_file)
108         resized_path = remove_black_borders_and_resize(raw_path,
                resized_folder)
109         process_and_generate_mask(resized_path, mask_folder,
                conn)
110         img_id = os.path.splitext(image_file)[0]
111         print(f"     {img_id}_resized.jpg and {img_id}_mask.npy
                saved")
112
113     conn.close()
114     print("FINISH")
115
116 raw_folder = "raw_folder"
117 resized_folder = "resized_folder"
118 mask_folder = "mask_folder"
119 gvi_db_path = "gvi.db"
120
121 main(raw_folder, resized_folder, mask_folder, gvi_db_path)
```

Listing 4: pipeline.py

A.3.5 print_gvi_with_depth.py

```
1  import sqlite3
2
3  conn = sqlite3.connect("gvi_with_depth.db")
4  cursor = conn.cursor()
5
6  cursor.execute("SELECT * FROM gvi ORDER BY image_id ASC")
7  rows = cursor.fetchall()
8
9  for row in rows:
10     print(row)
11
12 conn.close()
```

Listing 5: print_gvi_with_depth.py

A.3.6 print_gvi.py

```
import sqlite3

conn = sqlite3.connect("gvi.db")
cursor = conn.cursor()

cursor.execute("SELECT * FROM gvi ORDER BY image_id ASC")
rows = cursor.fetchall()

for row in rows:
    print(row)

conn.close()
```

Listing 6: print_gvi.py