RAMPNET: A Two-Stage Pipeline for Bootstrapping Curb Ramp Detection in Streetscape Images from Open Government Metadata

John S. O'Meara[1,2] **Jared Hwang**[2] **Zeyu Wang**[2]
Michael Saugstad[2] **Jon E. Froehlich**[2]

[1]Issaquah High School [2]University of Washington

https://github.com/ProjectSidewalk/RampNet

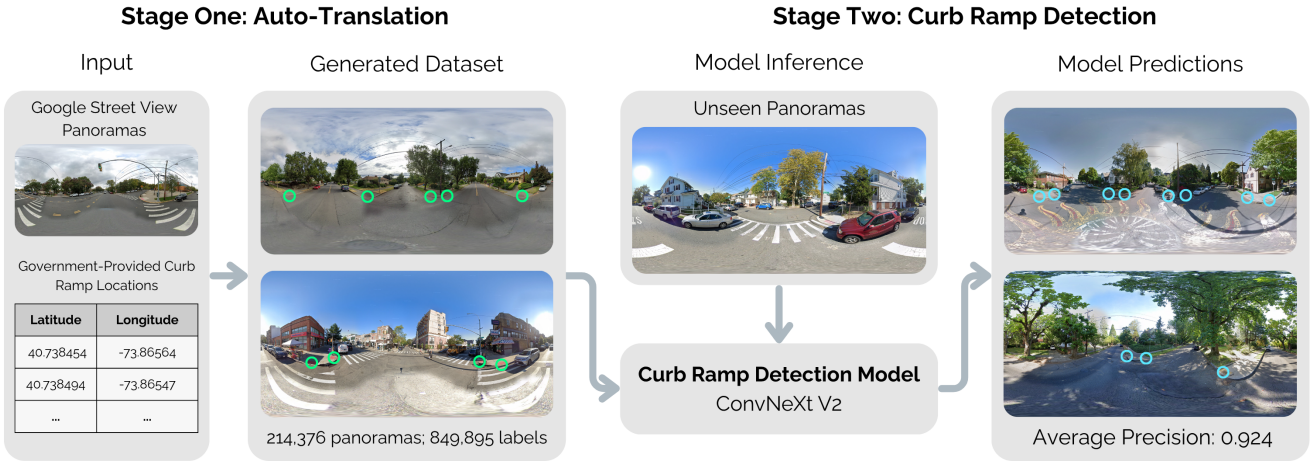


Figure 1. We introduce RAMPNET, a custom two-stage pipeline for bootstrapping curb ramp detection models in streetscape images from open government data. In Stage 1, we auto-generate a labeled streetview curb ramp dataset by translating government-provided curb ramp location data (*i.e.*, <lat, long> lists) into pixel labels on Google Street View panoramas. Stage 2 then uses this generated dataset to train a detection model that predicts curb ramp points in unseen panoramas.

Abstract

Curb ramps are critical for urban accessibility, but robustly detecting them in images remains an open problem due to the lack of large-scale, high-quality datasets. While prior work has attempted to improve data availability with crowdsourced or manually labeled data, these efforts often fall short in either quality or scale. In this paper, we introduce and evaluate a two-stage pipeline to scale curb ramp detection datasets and improve model performance. In Stage 1, we generate a dataset of more than 210,000 annotated Google Street View (GSV) panoramas by auto-translating government-provided curb ramp location data to pixel coordinates in panoramic images. In Stage 2, we train a curb ramp detection model (modified ConvNeXt V2) from the generated dataset, achieving state-of-the-art performance. To evaluate both stages of our pipeline, we compare to manually labeled panoramas. Our generated dataset achieves 94.0% precision and 92.5% recall, and our detection model reaches 0.9236 AP—far exceeding prior work. Our work contributes the first large-scale, high-quality curb ramp detection dataset, benchmark, and model.

1. Introduction

Curb ramps are critical to urban accessibility [24, 33], impacting both independence [26] and wellbeing [39] for people with mobility disabilities. However, information regarding curb ramp location and condition is limited, making it difficult to plan and maintain accessible infrastruc-

ture [10, 14]. To obtain this information, municipalities currently perform "*boots-on-the-ground*" sidewalk inspections [40], which can be prohibitively expensive and time-consuming. To address these challenges, computer vision-based techniques using streetscape images have been attempted [1, 12, 17, 45], but their results are limited by a lack of high-quality datasets [45]. While great strides have been made in the object detection space, curb ramp detection remains difficult due to data scarcity.

Existing curb ramp detection datasets are unsuitable for training deep learning models. *Project Sidewalk* [35] is a crowdsourcing initiative that allows users to identify and assess curb ramps and other accessibility features through a Google Street View (GSV) web interface. While this approach is scalable, its design does not require users to comprehensively label each panorama (pano), rendering the data suboptimal for deep learning methods. Indeed, Weld *et al*. discuss this "under-labeling" as a limitation of using Project Sidewalk data to train object detection models [45]. Despite its limitations, the Project Sidewalk dataset [38] is currently one of the only publicly available data sources for image-based curb ramp detection. One alternative is the *Mapillary Vistas* dataset [25], which has a class for "curb cuts"; however, we found that their categorization was overly broad and included driveways labeled as curb cuts.

In this paper, we introduce and evaluate a two-stage pipeline (Fig. 1), called RAMPNET, that leverages open government-provided curb ramp location datasets to auto-label curb ramps in streetscape images. While many local governments lack open data on curb ramp locations [10], the datasets that are available contain only metadata (*e.g*., text lists of <lat, long>) without corresponding images, precluding their usage with computer vision techniques. Thus, in Stage 1 of our pipeline, we introduce an automated method to translate these city-collected location coordinates to image pixel coordinates in GSV panoramas, allowing us to create a dataset comprising over 840,000 auto-generated curb ramp labels. During the translation process, we use a *ConvNeXt V2* [46] deep learning model to isolate the exact point of a curb ramp given a crop in the direction of the object's location (derived from the <lat, long>). In Stage 2, we use our generated dataset to train a separate deep learning model (again using ConvNeXt V2) that detects curb ramps in GSV panoramas, demonstrating real world applicability. By introducing a highly scalable, automatic, image-based curb ramp labeling pipeline, along with a new open dataset and initial benchmarks, our overarching goal is to help standardize and advance curb ramp detection research—similar to how *WIDER FACE* [48] and *VGGFace2* [7] advanced face detection research.

We evaluate both pipeline stages by comparing to manually labeled panoramas (ground truth). We randomly selected and manually labeled 1,000 panoramas, yielding 3,919 manual curb ramp labels. In comparing our Stage 1 output to ground truth, we find a high level of agreement with the manual annotations, correctly identifying 92.5% of curb ramps with 94.0% precision. For Stage 2, we find that our model achieves state-of-the-art performance (0.924 AP) for curb ramp detection, greatly surpassing a previous method [45] that relied solely on crowdsourced data.

In sum, our research contributes: (1) a new technique for translating government-provided curb ramp location data (<lat, long> lists) to pixel locations in streetscape panoramas; (2) the first large-scale, high-quality curb ramp detection dataset; (3) a comprehensive benchmark for measuring curb ramp detection performance; and (4) a state-of-the-art curb ramp detection model. Our code and datasets are also open source[1], allowing others to build off our work and establishing key benchmarks for the community. Importantly, while Stage 1 relies on pre-existing government metadata, this is only for bootstrapping. The Stage 2 model benefits *all* cities with GSV availability.

2. Related Work

We describe work in curb ramp auditing and automated streetscape analysis.

2.1. Curb Ramp Audit Methods

In the US, sidewalks are required to have curb ramps (or "curb cuts") to support mobility for all, including people in wheelchairs, caregivers pushing strollers, or even travelers pulling luggage [10, 43]. Traditionally, cities audit curb ramps through *"boots-on-the-ground"* sidewalk inspections [40]. However, these manual audits are expensive and time-consuming [34]. Indeed, in a recent study of 178 US cities, Deitz *et al*. found that only 10% published data on curb ramps [10]. Recent work has attempted to accelerate this auditing process with handheld LiDAR devices to reduce the amount of measurements performed manually [3, 9, 32, 42]. However, these tools still require workers to conduct on-site data collection, limiting scalability. Moreover, this approach focuses primarily on assessing the quality of curb ramps at known locations, rather than detecting *where* curb ramps are, as we do in this work. In general, in-person data collection still faces challenges with regards to cost and scalability despite attempts to improve efficiency. Some municipalities offer community-oriented tools [29] that allow residents to report accessibility issues, but these systems again rely on *in situ* observation and active citizen participation, limiting scalability.

Prior work has attempted to leverage sensor data (*e.g*., accelerometer, gyroscope, and magnetometer data) to aid in urban accessibility assessment. *Briometrix* [6], a company that specializes in sidewalk mobility audits, uses in-

[1] https://github.com/ProjectSidewalk/RampNet

strumented wheelchairs to collect sensing data. Similarly, *SideSeeing* [8], a multimodal dataset for sidewalk assessment, uses video and sensing data captured with chest-mounted mobile devices. Unlike traditional in-person sidewalk audits, this approach directly involves pedestrians in the auditing process. Still, the need for physical, on-site data collection and inspection hinders scalability.

Virtual crowdsourcing initiatives are a promising alternative [16]. For example, Project Sidewalk [35] allows users to identify and assess accessibility features like curb ramps through a GSV web interface. Compared to field audits, this technique scales rapidly and at a lower cost. Importantly, users need not be physically present in the city to contribute. Previous work has demonstrated a high level of agreement between these crowdsourced labels and government field data [4]. While Project Sidewalk's crowdsourcing approach scales, it is still limited due to its requirement for human labor. In addition, Project Sidewalk does not require users to comprehensively label panoramas, complicating efforts to use its data in deep learning applications [45].

Others have explored computer vision techniques to detect pedestrian features in aerial imagery [19], including for sidewalks [18, 27], crosswalks [2], and roads [36]. However, detecting curb ramps [13] remains challenging due to occlusions from shadows and trees, their relatively small size, and their tendency to visually blend in with sidewalks. It is also difficult to reliably assess curb ramp quality factors (*e.g.*, steepness, presence of tactile warnings) due to the low resolution and inherent limitations of aerial imagery. While our work focuses on detection rather than quality assessment, our usage of high-resolution streetscape imagery instead of aerial imagery enables future work for this task.

2.2. Automated Streetscape Analysis

With the rise of computer vision and deep learning technologies, streetscape imagery has increasingly become an important tool for large-scale urban analysis [5, 20]. GSV, the largest repository of such imagery, primarily collects panoramas along public roadways, offering a plethora of data about the built environment. This data has been used in multiple domains, including accessibility assessment [1, 12, 17, 45], demographic studies [15], neighborhood quality assessment [44], and real estate valuation [23]. When used in conjunction with computer vision techniques, streetscape imagery provides a low-cost and scalable method for understanding urban environments.

Prior work has attempted to leverage streetscape imagery and computer vision models to detect curb ramps and other accessibility features in images. For example, Hara *et al*. developed *Tohme* [17], a system that combines crowdsourcing and computer vision to semi-automatically detect curb ramps in GSV imagery. While far from fully automated curb ramp detection, their system results in a 13% reduction in time cost compared to manual labeling alone. Building on this work, Weld *et al*. introduced a customized *ResNet* model trained on crowdsourced data to automatically detect accessibility features (*e.g.*, curb ramps, missing curb ramps, obstructions, surface problems) in GSV panoramas [45].

Despite these efforts, no existing curb ramp detection model comes close to human-level labeling performance. Tohme achieves a recall of 67% and a precision of 26% when benchmarked against manual labels [17]. Weld *et al*. improve upon this with a recall of 78.7% and a precision of 33.7% [45], but these results are still infeasible for fully automatic urban accessibility assessment. While a variety of factors play into this poor performance, two key factors are: (1) the lack of large-scale, high-quality and open curb ramp image datasets; (2) the lack of standardized benchmarks. Our research attempts to address both, generating a dataset that is both larger and cleaner than previous efforts, and enabling a state-of-the-art curb ramp detection model that greatly outperforms prior work.

3. Stage 1: Dataset Generation

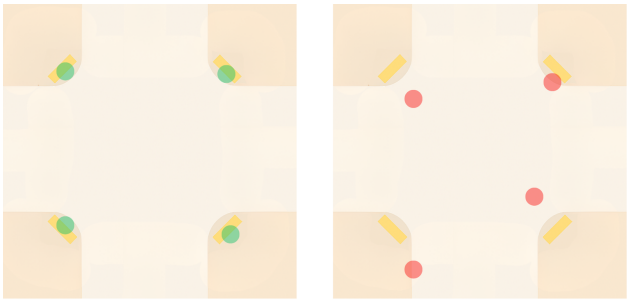
We introduce and evaluate a two-stage pipeline (Fig. 1), called RAMPNET, for bootstrapping curb ramp detection in streetscape images from open government metadata. We first describe Stage 1, which auto-generates a labeled image-based curb ramp dataset by translating government-provided curb ramp location data to image pixel coordinates in GSV panoramas (Fig. 3). Rather than traditional object detection datasets, which contain bounding boxes, we use single points to represent curb ramps, mirroring prior work in curb ramp detection [35, 45]. At a high level, our dataset generation process involves identifying relevant panoramas, extracting directional image crops, localizing curb ramps within these crops, and aggregating the results back onto the full panorama.

3.1. Dataset Selection

We describe our dataset selection process both for the open government data as well as corresponding GSV panoramas.

Selecting government datasets. Stage 1 is dependent on high quality curb ramp location data published by local governments. However, as noted previously, this data is rare: of the 178 US cities studied in [10], 90% published open street data but only 34% had sidewalk data and far fewer (10%) included curb ramps. Our goal is to leverage those cities that *do* publish curb ramp locations, translate those locations to pixels in GSV panos, and use this new dataset to train computer vision models.

Even when government curb ramp data is available, we observed location imprecision, which impacts our pipeline (see Fig. 2). To evaluate location quality, we manually examined curb ramp data from eight different local government datasets by overlaying curb ramp locations on top of



(a) High-Quality Location Data (b) Low-Quality Location Data

Figure 2. A visual comparison of location data quality. **(a)** High-quality coordinates precisely align with curb ramps' physical location. **(b)** In contrast, low-quality coordinates are often misplaced, rendering them unsuitable for our method.

City	# of Curb Ramps	Location Precision
Austin, TX	48,995	OK
Bend, OR	**13,611**	**Good**
Los Angeles, CA	91,759	OK
Nashville, TN	18,285	OK
New York City, NY	**217,680**	**Good**
Portland, OR	**45,324**	**Good**
Seattle, WA	45,653	Poor
Washington D.C.	34,859	OK

Table 1. Initial sample of government-provided curb ramp datasets in the US and their location precision compared to aerial imagery.

an aerial imagery base map. Of the eight cities, one city (Seattle) had *poor* location precision, four were *OK*, and three were *good*—see Tab. 1. For our purposes, we use all data from the *good* category: New York City, NY [30]; Portland, OR [31]; and Bend, OR [28]—all which offer precise and diverse curb ramp styles.

While formats vary, each government-provided dataset includes at least a unique identifier and a <lat, long> tuple (*e.g.,* see Figs. 1 and 3). For Stage 1, we need to convert these location tuples to corresponding pixel locations in a GSV panorama. For this, we need to determine which GSV panoramas best correspond to the curb ramp location data.

Selecting GSV panoramas. To identify and select corresponding GSV panoramas, we define a 10-meter radius around each government-defined curb ramp location and download all available panoramas within this boundary as an equirectangular image at 4096×2048 pixel resolution—a resolution we selected because both prior work [17] and our own testing conclude that further increasing the resolution results in minimal performance gain while significantly increasing computational complexity.

Selecting label candidates. For each panorama within the 10-meter radius, we then must determine which curb ramps to label within it. Using the government data, we select all curb ramps within a 35-meter radius of the panorama's location. This radius is intentionally larger than the 10-meter radius used for panorama selection, as it ensures that all curb ramps reasonably visible within the panorama are captured and labeled. Therefore, while a pano is selected based on a single nearby curb ramp that is within 10 meters, its final set of labels includes *all* curb ramps within a larger 35-meter radius. We also require the curb ramp installation date (defined in the government data) to be earlier than the panorama capture date. For each panorama, all curb ramp locations that meet these constraints will be marked as label candidates and saved for the auto-translation step, where we convert their locations to image pixel coordinates.

To reflect real-world scenarios, we also infuse our dataset with 20% null images (*i.e.*, a panorama that contains zero curb ramps). To download a null panorama, we first randomly select one of the three cities: New York City, Portland, or Bend. We then choose a random point on that city's street network and locate nearby panoramas. If none of the nearby panos are positioned at least 60 meters away from any known curb ramp location, we repeat the process, continuing until we find a panorama that is at least 60 meters away from all curb ramps.

In total, the above process yields 219,170 panoramas and 959,442 label candidates. Of these panoramas, 54.9% are from New York City, 8.5% are from Bend, and 36.6% are from Portland (left columns in Tab. 2). The discrepancy between cities can be attributed to differences in city size, GSV availability, and curb ramp quantity.

3.2. Auto-Translating Geo to Image Coordinates

With the government-based curb ramp location data and corresponding GSV panos downloaded, we now need to translate the curb ramp <lat, long> coordinates to image pixel coordinates on the GSV panos. We do this in two parts: first, for each potential curb ramp in a pano, we extract a directional crop around the curb ramp in the pano and then, second, we use a trained ConvNeXtV2 [46] model to isolate curb ramp points within these crops (Fig. 3).

Auto-cropping along curb ramp heading. To generate the crops around each curb ramp, we calculate the angle between the panorama capture location and the curb ramp location. Using this angle, we generate a square perspective crop in the direction of the curb ramp (1024×1024 pixels), akin to what a user would observe in GSV if they were looking toward it. The crop has a 90-degree field of view and a 30-degree downward pitch to better capture the ground level. To focus our analysis on relevant context, we retain only the middle third of the crop, producing a final image that is 341×1024 pixels.

Identifying curb ramp pixels. To isolate the exact pixel

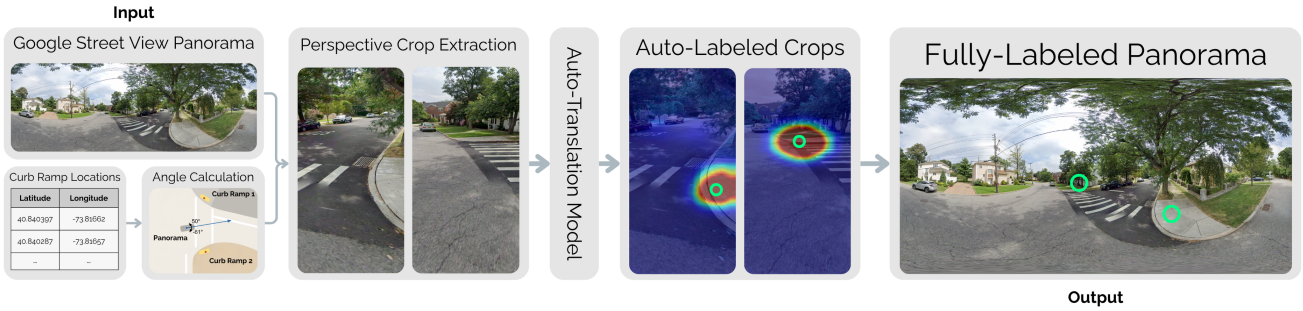


Figure 3. The auto-translation method used to generate our dataset. For each label candidate (a government-provided curb ramp location), we compute its angle with respect to the panorama location. With this angle, we extract a directional perspective crop from the full panorama. We then pass this crop to our auto-translation model, which outputs a heatmap whose peaks represent possible curb ramp points. These points are projected back onto the original equirectangular image, yielding a fully-labeled panorama.

coordinates of curb ramps within each crop, we use a modified ConvNeXt V2 [46] model (specifically, the *base* variant) that outputs a heatmap of probable curb ramp points. Our model is initially pretrained on crops extracted from 20,698 annotated Project Sidewalk [35] panoramas spanning 12 US cities, including Chicago, Seattle, Pittsburgh, and St. Louis. However, using Project Sidewalk data alone results in poor performance due to its many missing labels[2]. Indeed, a crop model that is solely trained on Project Sidewalk data alone achieves a recall of 76.7%, and a precision of 77.2%. To rectify this issue, we add a second round of training on crops extracted from a smaller, manually-labeled dataset of 312 panoramas from Bend, NYC, and Portland. Here, every curb ramp occurrence in each pano was labeled using Label Studio [41]. By combining both data sources, our final crop model achieves a recall of 89.0%, and a precision of 87.0%. This crop-level model is not to be confused with our Stage 2 curb ramp detection model, which identifies curb ramps in whole panoramas, rather than just crops.

After the auto-translation process, our dataset contains 214,376 fully labeled panoramas and 849,895 curb ramp labels, larger than any previous efforts (see Tab. 3).

3.3. Study Method

We now describe our evaluation approach. Though our immediate focus here is on evaluating Stage 1 performance, we describe methodological details shared between Stage 1 and 2. We first divide our auto-labeled Stage 1 dataset of 214,376 panoramas into 70% training, 20% validation, and 10% test splits. Because an individual curb ramp may be present in more than one panorama at differing angles or viewpoints, we must be wary of possible data leakage.

[2]As noted, users in Project Sidewalk are asked to label sidewalk features and obstacles in GSV imagery, including curb ramps; however, they may do so from multiple panoramas rather than comprehensively labeling a single panorama—which makes the data suboptimal for model training

City	Initial		After Auto-Trans.	
	Curb Ramp	Panoramas	Curb Ramp	Panoramas
Bend	20,451	18,545	19,082	18,205
New York	655,084	120,430	566,832	117,323
Portland	283,907	80,195	263,981	78,848
Total	**959,442**	**219,170**	**849,895**	**214,376**

Table 2. Table showing the number of initial curb ramp candidates and panoramas, and the final counts of successfully generated labels and panoramas after auto-translation in Stage 1.

To avoid this, we use a semi-random splitting strategy that groups panoramas located within 60 meters of each other into the same split. Our final experimental dataset contains 150,063 panoramas in the training split, 42,875 in validation, and 21,438 in test. While we do not use the validation split in this study, we provide it for other researchers for tasks such as hyperparameter tuning or model selection. We now describe our ground truth approach and correctness metrics; both are used to evaluate Stage 1 and 2.

Ground truth. To create a ground truth, we manually labeled curb ramps across 1,000 panoramas randomly sampled from the test split (independent of the 312 aforementioned manually labeled panoramas). We used Label Studio [41] for this process. In total, we labeled the center points of 3,919 curb ramps (3.9 ramps/pano).

Correctness metrics. To evaluate the accuracy of predicted curb ramp labels against our manually-created ground truth, we use a proximity-based comparison method. For each predicted label, we check if there are any ground-truth points within an 88-pixel radius in the 4096×2048 pixel panorama. If not, we mark the label as a false positive. Conversely, if it does match one or more ground-truth points, those within the radius will be counted as a true positives. If more than one predicted label matches the same ground-truth point, we pick the one with highest

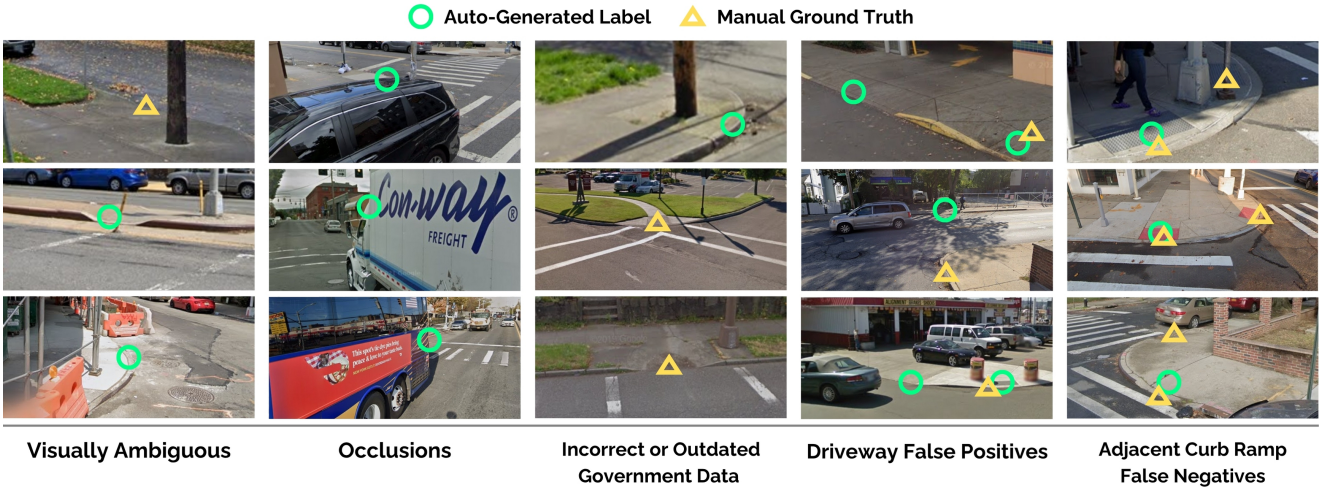


Figure 4. We performed a qualitative analysis of 100 randomly sampled Stage 1 errors and inductively categorized them into five groups: visually ambiguous (43% of errors), occlusion (27%), disagreements with government data (17%), driveways mistaken for curb ramps (8%), and unlabeled adjacent curb ramps (5%).

confidence and ignore the others entirely in our calculation. This is to ensure that no ground-truth points are counted more than once. If any of the ground-truth points are not detected, we count these as false negatives.

3.4. Results

In comparing the 3,919 ground truth labels to Stage 1 output across 1,000 panoramas, our findings demonstrate a high level of agreement, achieving 92.5% recall and 94.0% precision for an F-score of 0.932. Tab. 3 compares our dataset to Project Sidewalk's [35], demonstrating significant improvements in both scale and comprehensiveness.

To advance understanding of performance, we randomly sampled and analyzed 100 errors and inductively coded them thematically. The most common error was disagreements on visually ambiguous curb ramps (43%). These cases involved distant or low-resolution objects where a definitive identification is inherently subjective. Other common errors were the following: occlusions, where a car, bus, or other object obscures the curb ramp itself but our model still makes a speculative prediction (27%); disagreements in government data either due to underlying errors in the open datasets (*e.g.,* a missing curb ramp labeled as a curb ramp) or temporal differences between capture dates in the government metadata *vs.* the GSV pano (17%); driveways mistaken for curb ramps (*e.g.*, an adjacent driveway leaks into a crop) (8%); and unlabeled adjacent curb ramps (5%).

4. Stage Two: Curb Ramp Detection

While Stage 1 produces a large-scale dataset of auto-generated curb ramp labels on GSV panos, in Stage 2, we show how this generated dataset can be used to train a state-of-the-art curb ramp detection model using pixels alone–far exceeding prior work [17, 45]. Crucially, unlike Stage 1, which requires government data, the Stage 2 model works on any city with streetscape imagery available.

	Project Sidewalk	RAMPNET Stage 1
Comprehensiveness	Partially-Labeled	Fully-Labeled
Data Source	Crowdsourced	Auto-Generated
# of Ramp Labels	427,435	849,895
# of Panoramas	149,574	214,376
Labels / Panorama*	2.86	4.96

*Excludes panoramas with zero labels.

Table 3. Comparison of Project Sidewalk and RAMPNET datasets.

4.1. Model Architecture

For the computer vision model, we again use the ConvNeXt V2 [46] architecture, specifically the *base* model variant. To speed up convergence and improve performance, we load weights from a publicly available ConvNeXt V2 model that is pretrained on the ImageNet-1k dataset [11].

To enable point-based object detection, we modify the architecture's output layers to perform heatmap regression. We replace the classification head with a module consisting of a 3×3 convolution, a ReLU activation, and a bilinear upsampling layer, followed by a final 1×1 convolution that produces a single-channel heatmap, as shown in Fig. 5. Our model accepts a 4096×2048px panorama and outputs a scaled-down 1024×512px heatmap of probable curb ramp points. We represent each label in the heatmap by centering a 2D gaussian kernel ($\sigma = 10.0$) at its corresponding

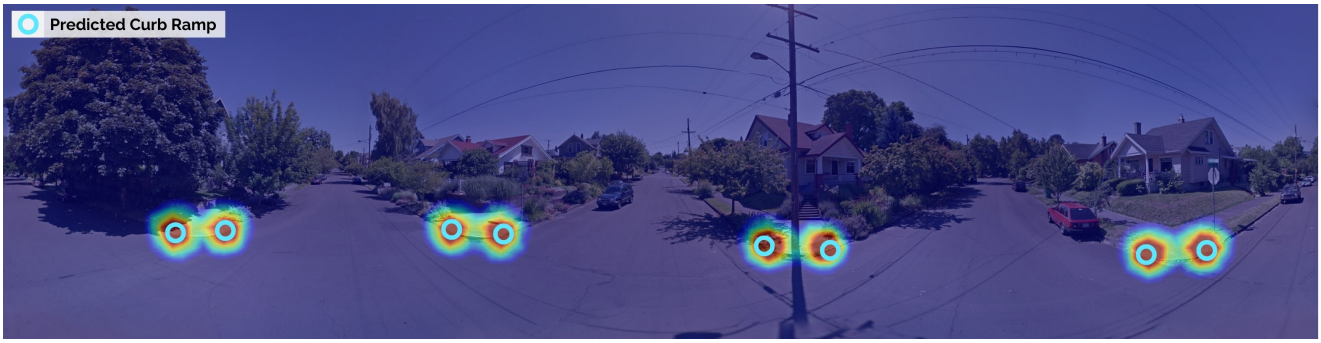


Figure 5. A heatmap generated by our Stage 2 curb ramp detection model. Peaks of the heatmap, which represent predicted curb ramp points, are circled in blue.

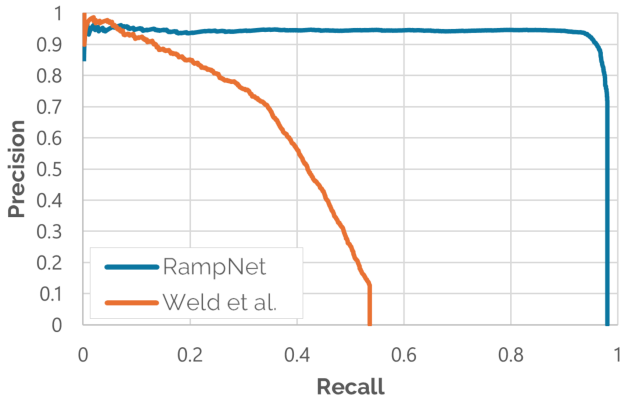


Figure 6. Precision-recall curves of RAMPNET's curb ramp detection model and the previous state-of-the-art [45]. Both are benchmarked against our manually-labeled dataset.

location in the image. This heatmap regression approach is commonly used in human pose estimation [47], but here we use it for point-based object detection. To extract the predicted points, we identify peaks in the heatmap whose maximum value is above a certain threshold (currently set to 0.55). This threshold can be tweaked depending on the user's desired balance of precision and recall.

4.2. Training

We use the same 70% training (150,063 panos), 20% validation (42,875 panos), and 10% test split (21,438 panos) described in Sec. 3. During the training process, we augment our data by randomly applying a horizontal flip, as both our own testing and prior work [37] has demonstrated that this increases performance. Our loss function is pixel-wise mean squared error. We use the Adam optimizer [22] with a learning rate of 1.0×10^{-5}. We trained the model for a single epoch on 16 NVIDIA L40s GPUs (distributed across four nodes). Our batch size was limited to one, as larger batch sizes were infeasible due to VRAM limitations.

	Weld *et al.* [45]	RAMPNET Stage 2
Input Type	Depth+Image+Geo Data	Image Data
Dataset Used	Project Sidewalk [35]	Auto-Generated
Avg. Precision	0.3803	0.9236

Table 4. Performance comparison of our curb ramp detection model against the previous state-of-the-art. Average precision is calculated by benchmarking against our manually-labeled subset.

4.3. Results

We then tested our trained model against the 1,000-panorama ground truth dataset (with 3,919 manually labeled curb ramps). We report an average precision (AP) of 0.924. To contextualize these results, we compared to the previous state-of-the-art curb ramp detection system released by Weld *et al.* [45]. Using their released model checkpoints and code [38], we evaluated their curb ramp detection system on our manually labeled dataset, finding that they achieve 0.380 AP, far lower than our 0.924 AP (see Tab. 4 and Fig. 6).

Further, we also tested our trained model against the test split of our auto-generated dataset (containing 21,438 panos and 82,055 labels). We report an AP of 0.873. However, we consider this benchmark to be less reliable than our above comparison with the manually-labeled ground truth, due to the higher likelihood of there being errors in the auto-generated dataset.

Naturally, because we are eliminating one input (the government curb ramp data), we might expect lower performance in Stage 2 than in Stage 1. However, we find this performance gap to be surprisingly small, with our Stage 2 model essentially matching the performance of our Stage 1 dataset. When benchmarking against the 1000-panorama ground truth dataset, our Stage 1 generated dataset had a recall of 92.5% and a precision of 94.0% versus our Stage 2 model's precision of 93.9% at the same recall threshold.

5. Discussion

In this paper, we introduced RAMPNET, a custom two-stage pipeline for bootstrapping curb ramp detection in streetscape images using open government metadata and deep learning. Below, we discuss limitations, suggest directions for future work, and describe the impact and potential applications of automatic curb ramp detection.

5.1. Limitations

While our dataset includes more than 840,000 curb ramp labels, many of the labels are duplicates of the same physical curb ramp captured from different panorama viewpoints. Indeed, an individual curb ramp appears in 4.5 different panoramas on average. While still useful for training, these repeated views may offer less value than entirely unique curb ramps. During the dataset splitting process, researchers must take care to avoid data leakage by ensuring that the same curb ramp is not distributed across multiple splits, a precaution taken in our work.

As with any deep learning system, our dataset and model contain biases that have the potential to limit generalizability. For example, our dataset was generated from just three U.S. cities. While we strategically selected these locations for their diverse curb ramp styles, the limited coverage area likely introduces regional bias. While we believe our dataset is sufficiently diverse to support the detection of curb ramps in the United States, more work is needed to create datasets for other regions, where curb ramp style differs significantly.

Our system is also limited by its dependence on streetscape imagery, which can sometimes be unavailable or outdated. While up-to-date streetscape imagery is generally available in major U.S. cities, coverage in rural areas, small towns, or rapidly changing neighborhoods can be sparse or outdated, reducing our system's accuracy and timeliness. Indeed, a recent study by Kim and Jang, which analyzed 45 small- and medium-sized cities, found that 44% of commute routes lacked adequate GSV imagery and that there was significant temporal variability in the coverage [21].

5.2. Future Work

Our long-term goal is to enable fully automatic urban accessibility assessment. We have addressed a key component of that goal by developing a method for automatic curb ramp detection. However, other unsolved problems remain.

To assess curb ramp accessibility, cities must (1) know *where* curb ramps are and (2) be able to evaluate key quality factors (*e.g.*, presence of tactile warnings, steepness). While our work addresses the first requirement by enabling the automatic detection of curb ramps, further work is needed to enable automatic quality assessment. We posit that our auto-translation technique could be extended for this task, as quality information is often included in government-provided curb ramp metadata. Although RAMPNET, like Project Sidewalk [35], outputs single-point labels, future work should explore generating bounding boxes to provide richer spatial information crucial for quality assessment (*e.g.*, ramp width and alignment).

While the focus of our work is auto-translating location coordinates to image pixel coordinates, the reverse process is equally important. Current city workflows are reliant on location data rather than the image pixel coordinates that our model outputs. The introduction of a system to translate back to location coordinates would greatly enhance the usability of our curb ramp detection model.

Lastly, curb ramps are not the only urban accessibility feature that could benefit from automatic detection and analysis. Other features (*e.g.*, pedestrian signals, missing curb ramps, path obstacles, surface problems) play a similarly vital role in ensuring urban accessibility. We encourage future work to explore ways of automatically detecting and assessing these additional features to provide a more comprehensive understanding of accessibility conditions.

5.3. Impact on Urban Accessibility

Our research aids in urban accessibility planning by enabling a low-cost, scalable method for curb ramp detection. Our dataset achieves 92.5% recall and 94.0% precision, and our curb ramp detection model reaches 0.924 AP, significantly outperforming prior work and, for the first time, achieving near-human-level quality.

We envision a future where whole cities can be audited in a single day instead of over the course of multiple months. By reducing time and cost requirements, we allow more cities to conduct crucial accessibility audits. We also give accessibility advocates a powerful tool to hold cities accountable to the legal requirements defined by the *Americans with Disabilities Act* [43].

6. Conclusion

In conclusion, RAMPNET advances research in automatic curb ramp detection by (1) offering a novel technique for auto-translating open government datasets into GSV pano labels, by (2) producing the largest and most comprehensive curb ramp detection dataset, and by (3) establishing key performance benchmarks for automatic curb ramp detection (0.924 AP). We believe these contributions provide a foundation for fully automatic urban accessibility assessment and future research in the area.

7. Acknowledgments

This work was supported by NSF SCC-IRG #2125087 and NSF OAC #2411222.

References

[1] Marc A. Adams, Christine B. Phillips, Akshar Patel, and Ariane Middel. Training Computers to See the Built Environment Related to Physical Activity: Detection of Microscale Walkability Features Using Computer Vision. *International Journal of Environmental Research and Public Health*, 19 (8):4548, 2022. 2, 3

[2] Dragan Ahmetovic, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. Mind Your Crossings: Mining GIS Imagery for Crosswalk Localization. *ACM Trans. Access. Comput.*, 9(4):11:1–11:25, 2017. 3

[3] Chengbo Ai, Qing Hou, and University of Massachusetts at Amherst. Improving Pedestrian Infrastructure Inventory in Massachusetts Using Mobile LiDAR. Technical Report 19-007, 2019. 2

[4] Sajad Askari, Devon Snyder, Chu Li, Michael Saugstad, Jon E. Froehlich, and Yochai Eisenberg. Validating Pedestrian Infrastructure Data: How Well Do Street-View Imagery Audits Compare to Government Field Data? *Urban Science*, 9(4):130, 2025. 3

[5] Filip Biljecki and Koichi Ito. Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215:104217, 2021. 3

[6] Briometrix. Wheelchair Pilots. https://briometrix.com/wheelchair-pilot/. 2

[7] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74, 2018. 2

[8] R. J. P. Damaceno, L. Ferreira, F. Miranda, M. Hosseini, and R. M. Cesar Jr. SideSeeing: A multimodal dataset and collection of tools for sidewalk assessment, 2024. 3

[9] DeepWalk. DeepWalk ADA Solutions. https://www.deepwalk.com/. 2

[10] Shiloh Deitz, Amy Lobben, and Arielle Alferez. Squeaky wheels: Missing data, disability, and power in the smart city. *Big Data & Society*, 8(2):20539517211047735, 2021. 2, 3

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 6

[12] Michael Duan, Shosuke Kiami, Logan Milandin, Johnson Kuang, Michael Saugstad, Maryam Hosseini, and Jon E. Froehlich. Scaling Crowd+AI Sidewalk Accessibility Assessments: Initial Experiments Examining Label Quality and Cross-city Training on Performance. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–5, New York, NY, USA, 2022. Association for Computing Machinery. 2, 3

[13] Esri. Using GeoAI to Inventory ADA Curb Ramps. https://www.esri.com/en-us/lg/industry/public-works/stories/county-innovates-using-geoai-to-inventory-ada-curb-ramps-saving-significant-time-money. 3

[14] Jon E. Froehlich, Anke M. Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. Grand challenges in accessible maps. *interactions*, 26(2):78–81, 2019. 2

[15] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113, 2017. 3

[16] Kotaro Hara, Vicki Le, and Jon Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 631–640, New York, NY, USA, 2013. Association for Computing Machinery. 3

[17] Kotaro Hara, Jin Sun, Robert Moore, David Jacobs, and Jon Froehlich. Tohme: Detecting curb ramps in google street view using crowdsourcing, computer vision, and machine learning. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, pages 189–204, New York, NY, USA, 2014. Association for Computing Machinery. 2, 3, 4, 6

[18] Maryam Hosseini, Mikey Saugstad, Fabio Miranda, Andres Sevtsuk, Claudio T. Silva, and Jon E. Froehlich. Towards Global-Scale Crowd+AI Techniques to Map and Assess Sidewalks for People with Disabilities, 2022. 3

[19] Maryam Hosseini, Andres Sevtsuk, Fabio Miranda, Roberto M. Cesar, and Claudio T. Silva. Mapping the walk: A scalable computer vision approach for generating sidewalk network datasets from aerial imagery. *Computers, Environment and Urban Systems*, 101:101950, 2023. 3

[20] Yujun Hou, Matias Quintana, Maxim Khomiakov, Winston Yap, Jiani Ouyang, Koichi Ito, Zeyu Wang, Tianhong Zhao, and Filip Biljecki. Global streetscapes—a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215:216–238, 2024. 3

[21] Junghwan Kim and Kee Moon Jang. An examination of the spatial coverage and temporal variability of Google Street View (GSV) images in small- and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems*, 102:101956, 2023. 8

[22] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 7

[23] Stephen Law, Brooks Paige, and Chris Russell. Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. *ACM Trans. Intell. Syst. Technol.*, 10(5):54:1–54:19, 2019. 3

[24] Iva Mrak, Tiziana Campisi, Giovanni Tesoriere, Antonino Canale, and Margareta Cindrić. The role of urban and social factors in the accessibility of urban areas for people with motor and visual disabilities. *AIP Conference Proceedings*, 2186(1):160008, 2019. 1

[25] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5000–5009, Venice, 2017. IEEE. 2

[26] Rita Newton, Marcus Ormerod, Elizabeth Burton, Lynne Mitchell, and Catharine Ward-Thompson. Increasing Independence for Older People through Good Street Design. *Journal of Integrated Care*, 18(3):24–29, 2010. 1

[27] Huan Ning, Xinyue Ye, Zhihui Chen, Tao Liu, and Tianzhi Cao. Sidewalk extraction using aerial and street view images. *Environment and Planning B: Urban Analytics and City Science*, 49(1):7–22, 2022. 3

[28] City of Bend. Curb Ramps. https://bend-data-portal-bendoregon.hub.arcgis.com/datasets/bendoregon::curb-ramps/about. 4

[29] City of New York. Pedestrian Ramp Complaint · NYC311. https://portal.311.nyc.gov/article/?kanumber=KA-02276, . 2

[30] City of New York. Pedestrian Ramp Locations | NYC Open Data. https://data.cityofnewyork.us/Transportation/Pedestrian-Ramp-Locations/ufzp-rrqu/about_data, . 4

[31] City of Portland. Curb Ramps. https://gis-pdx.opendata.arcgis.com/datasets/PDX::curb-ramps/about. 4

[32] Michael Olsen, Ezra Che, John Caya, Michael Caya, Gene Roe, James Schneider, Rebecca Embacher, and Development United States. Department of Transportation. Federal Highway Administration. Office of Research, and Technology. Pocket Lidar Curb Ramp Assessments [TechNote]. Technical Report FHWA-HRT-25-014, 2024. 2

[33] Dori E. Rosenberg, Deborah L. Huang, Shannon D. Simonovich, and Basia Belza. Outdoor Built Environment Barriers and Facilitators to Activity among Midlife and Older Adults with Mobility Disabilities. *The Gerontologist*, 53(2): 268–279, 2013. 1

[34] Andrew G. Rundle, Michael D.M. Bader, Catherine A. Richards, Kathryn M. Neckerman, and Julien O. Teitler. Using Google Street View to Audit Neighborhood Environments. *American journal of preventive medicine*, 40(1):94–100, 2011. 2

[35] Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon Froehlich. Project Sidewalk: A Web-based Crowdsourcing Tool for Collecting Sidewalk Accessibility Data At Scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, New York, NY, USA, 2019. Association for Computing Machinery. 2, 3, 5, 6, 7, 8

[36] Pourya Shamsolmoali, Masoumeh Zareapoor, Huiyu Zhou, Ruili Wang, and Jie Yang. Road Segmentation for Remote Sensing Images Using Adversarial Spatial Pyramid Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):4673–4688, 2021. 3

[37] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):60, 2019. 7

[38] Project Sidewalk. sidewalk-cv-assets19. https://github.com/ProjectSidewalk/sidewalk-cv-assets19, 2025. 2, 7

[39] Melody Smith, Octavia Calder-Dawe, Penelope Carroll, Nicola Kayes, Robin Kearns, En-Yi (Judy) Lin, and Karen Witten. Mobility barriers and enablers and their implications for the wellbeing of disabled children and young people in Aotearoa New Zealand: A cross-sectional qualitative study. *Wellbeing, Space and Society*, 2:100028, 2021. 1

[40] Edward R. Stollof and Janet M. Barlow. Pedestrian Mobility and Safety Audit Guide. 2008. 2

[41] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. Label Studio: Data labeling software, 2020-2025. Open source software available from https://github.com/HumanSignal/label-studio. 5

[42] Yelda Turkan, Erzhuo Che, Michael J. Olsen, Yang Zhou, Oregon State University, and Pacific Northwest Transportation Consortium (PacTrans) (UTC). Automated Localization and Functional Condition Assessment of ADA Curb Ramps With Mobile Lidar Point Clouds. Technical Report 2020-S-OSU-1, 2022. 2

[43] U.S. Congress. Americans with disabilities act of 1990. 42 U.S.C. § 12101, 1991. 2, 8

[44] Zeyu Wang, Koichi Ito, and Filip Biljecki. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. *Cities*, 145:104704, 2024. 3

[45] Galen Weld, Esther Jang, Anthony Li, Aileen Zeng, Kurtis Heimerl, and Jon E. Froehlich. Deep Learning for Automatically Detecting Sidewalk Accessibility Problems Using Streetscape Imagery. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 196–209, New York, NY, USA, 2019. Association for Computing Machinery. 2, 3, 6, 7

[46] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders, 2023. 2, 4, 5, 6

[47] Bin Xiao, Haiping Wu, and Yichen Wei. Simple Baselines for Human Pose Estimation and Tracking. In *Computer Vision – ECCV 2018*, pages 472–487, Cham, 2018. Springer International Publishing. 7

[48] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5525–5533, 2016. 2