Gas Prices of America: The Machine-Augmented Crowd-Sourcing Era

Kevin Dick, François Charih, Jimmy Woo, James R. Green
Systems & Computer Engineering, Institute of Data Science
Carleton University, Ottawa, Canada
Email: {kevin.dick, james.green}@carleton.ca

***Abstract*—Google Street View (GSV) comprises the largest collection of vehicle-based imagery of the natural environment. With high spatial resolution, GSV has been widely adopted to study the natural environment despite its relatively low temporal resolution (*i.e.* limited time-series imagery available at a given location). However, vehicular-based imagery is poised to grow dramatically with the prophesied circulation of fleets of highly instrumented autonomous vehicles (AVs), producing high spatio-temporal resolution imagery of urban environments. As with GSV, leveraging these data presents the opportunity to extract information about the lived environment, while their high temporal resolution enables the study and annotation of time-varying phenomena. For example, circulating AVs will often capture location-coded images of gas stations. With a suitable CV system, one could extract the advertised numerical gas prices and automatically update crowd-sourced applications, such as GasBuddy. To this end, we assemble and release the Gas Prices of America (GPA) dataset, a large-scale, benchmark dataset of advertised gas prices from GSV imagery across the 49 mainland United States of America. Comprising 2,048 high quality annotated images, the GPA dataset enables the development and evaluation of CV models for gas price extraction from complex urban scenes. More generally, this dataset provides a challenging benchmark against which CV models can be evaluated for multi-number, multi-digit recognition tasks in the wild. For the digit-level classification task, the YOLO digit detection model trained on the Street View House Numbers dataset performed comparably to a random classifier, highlighting the difficulty of this task. Conversely, for the full-sign segmentation task, transfer learning of a DeepLabV3 ResNet101 model achieved a test F1 performance of 0.7125, following 100 epochs. Highly accurate models, when integrated with AV platforms, will represent the first opportunity to automatically update the traditionally human crowd-sourced GasBuddy dataset, heralding an era of machine-augmented crowd-sourcing. The dataset is available online at cu-bic.ca/gpa and at doi.org/10.5683/SP2/KQ6VNG. Accompanying code can be found at github.com/GreenCUBIC/Gas-Prices-of-America.**

***Keywords*-dataset preparation; remote sensing; character recognition**

I. Introduction

The last decade has seen the widespread adoption of Google Street View (GSV) imagery in diverse applications studying the natural environment. Recent examples include its use in health research [1], neighbourhood auditing [2], mapping air pollution [3], and assessing urban greenery [4]. With high spatial resolution and global coverage, GSV comprises the most comprehensive source of street-level imagery (*i.e. streetscapes*) in the world, lending to its utility to investigate and understand the natural environment.

Unfortunately, GSV is limited in its temporal resolution; a given spatial location is typically only photographed once every couple of years thereby reducing its utility for meaningful longitudinal analyses. Such analyses would be tremendously valuable for capturing time-varying events observable from the street-level perspective. At present, satellite photography offers both high spatial and temporal resolution imagery; however it is limited to the bird's eye view, limiting it's utility in observing phenomena more readily visible to the first-person perspective. To increase the temporal resolution of GSV-like imagery, the regular circulation of camera-laden vehicles is required.

A. Automating Crowd-Sourcing using Computer Vision

With the prophesied deployment of fleets of autonomous vehicles (AVs), we can expect a tremendous influx of such vehicle-based imagery. While these visual data have primary use for vehicular navigation, they have additional potential secondary use for investigating streetscape environments. More specifically, the collection, aggregation, and analysis of these data hold considerable potential to extract specific information in (near) real-time from the natural environment in an automated way. To date, the current most reliable source of (near) real-time street-level events relies on crowd-sourced information wherein human annotators will update a distributed repository with the most current information. Popular examples include the Waze app [5], for traffic-related events, and the GasBuddy app, for gas price data. Such human crowd-sourcing initiatives depends on the generosity of individual data contributors and is further limited by restrictions on the use of mobile devices whilst driving.

Machine-augmented crowd-sourcing of environmental data will first require the development of accurate machine learning (ML) and computer vision (CV) models, suitable for real-time deployment. Generating large-scale and annotated datasets comprising streetscape imagery capturing these events would be valuable for the training and evaluation of such algorithms (*e.g.* Convolutional Neural Networks; CNNs). Resultant models capable of extract-

978-1-7281-9891-0/20/$31.00 ©2020 IEEE
DOI 10.1109/CRV50864.2020.00029

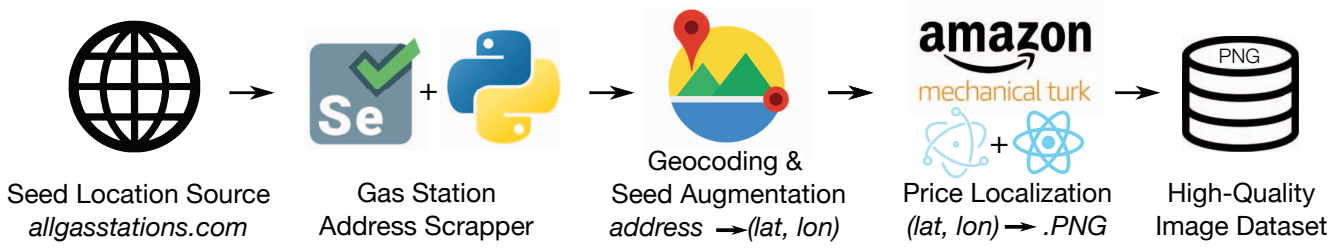

Figure 1. Overview of the Image Acquisition Pipeline. The published street addresses of known gas stations in the USA were scraped from AGS using a Selenium-based crawler and were geocoded to (lat,lon) pairs. The Google Places API was used to refine these coarse-grained seed addresses by searching for known stations within a 2km radius. For each refined seed, AMT Workers localized, segmented, and transcribed the advertised gas prices.

ing information from the natural environment with high performance will rival their human annotating counterpart, signaling the beginning of a machine-augmented crowd-sourcing era. Appositely, what reason has a human annotator to update a crowd-sourced app if their vehicle-integrated CV model does so automatically with equivalent (or possibly superior) accuracy?

B. Benchmark Dataset for 'Prices in the Wild'

In this work, we focus on the task of automating the extraction of advertised gas prices of fueling stations throughout the United States of America. To this end, a large-scale dataset of streetscape imagery of advertised gas prices is required. Beyond its utility in developing automated gas price detection models, the collection, annotation, and publication of such a dataset promises to be a boon to the CV research community as a resource for developing and benchmarking methods for the extraction of multi-digit, and multi-number information from natural images. Numerous large-scale & high-quality image datasets are made freely available to the ML community as benchmarks to evaluate the effectiveness of CV models. One class of these datasets comprise discrete-type character/digit representations amenable to classification tasks. Notable examples include the 1998 MNIST dataset of hand-written digits (one digit per image) [6] and the 2011 Street View House Numbers (SVHN) dataset (several digits per image) [7]. However, to the authors' knowledge, no benchmark of multi-number, multi-digit values in the natural environment exists.

Using CV to read textual information from images has been the focus of extensive research in the last decade. Under highly constrained conditions, character recognition is effectively a solved problem (*e.g.* classification of hand-written digits as with the MNIST dataset, optical character recognition of printed documents); however, the recognition of textual characters in unconstrained natural scenes is more challenging. However, images of text "in the wild" introduce a gamut of challenges such as illumination, orientation, blurring, obfuscation, distortion, and scaling, thereby making the task commensurately more complex. Moreover, combinations of characters, variations in font faces, and differing styles each exacerbate this problem. Previous datasets, such as the Street View House Number (SVHN) dataset [7], have helped address some of these issues by serving as a benchmark for development and evaluation of CV models. Whereas the MNIST dataset comprises a collection of 28×28 grey-scaled bitmaps, each depicting a single handwritten digit, the SVHN dataset represents multiple characters per image which, when combined, form a single, sequential, multi-digit house number sourced from Google Street View imagery. While the latter is useful to extract multi-digit values "in the wild", it does not address the challenge of distinguishing characters among several multi-digit numbers. The availability of such a dataset would function as a benchmark to evaluate models on their ability to accurately distinguish between multiple numerical values within natural environments. In this work we introduce the Gas Prices of America (GPA) dataset, a large-scale collection of advertised gas prices from GSV imagery throughout the United States of America. The dataset is available online at **cu-bic.ca/gpa** and in this **DataVerse Repository**.

II. Methods

A conceptual overview of the data acquisition pipeline is depicted in Figure 1. The collection of GSV imagery depicting advertised gas prices first required a set of "seed" coordinates: latitude and longitude (lat,lon) pairs corresponding to locations sufficiently proximal to known gas stations to facilitate the localization of the advertised price. To restrict our search to locations with a higher density of gas stations, we identified a database with listed addresses as opposed to performing a systematic grid search, exemplified in [8]. While a systematic grid search would constitute a geographically unbiased search strategy, the parameterization of such a search with a dense grid (*e.g.* very short distance between search points) would result in an exorbitant number of sample points, the majority of which are unlikely to be proximal to a gas station resulting in a high *external complexity* (first introduced in [8]) and wasteful APIs calls.

We refined our search strategy by considering a number of publicly available websites that list addresses for known gas stations throughout the United States. Two notable sources are All Gas Stations (AGS; allgasstations.com) and

the popular gas price crowd-sourcing platform, GasBuddy (GB; gasbuddy.com). Neither of these sources supports a machine-friendly interface to access location data. Fortunately, the AGS website implemented a navigable hierarchical structure, where $i \in \{1, 2, \ldots, 50\} \mid \langle$state-code-$i\rangle \in \{$'AL','AK',...,'WY'$\}$:

```
allgasstations.com/
└─ <state-code-i>
   ├─ <city-name-1>/
   │  ├─ <gas-station-1>
   │  ├─ ⋮
   │  └─ <gas-station-x>
   ├─ ⋮
   └─ <city-name-n>
      ├─ <gas-station-1>
      ├─ ⋮
      └─ <gas-station-y>
```

A custom Selenium-based Python scraper was developed to crawl the domain and extract all listed street-based addresses for each of the cities within each of the 50 states, resulting in 50,043 addresses. Unique addresses were geocoded to (lat,lon) pairs, providing 34,732 initial coarse-grained seed locations. To further augment this dataset, we leverage the *Google Places* API to obtain (lat,lon) pairs for all gas stations within a 2km radius of each seed location, resulting in an approximately ten-fold increase in refined seed locations. A further advantage of using this API was the improved *quality* of the returned seed locations which, conveniently, almost always localised in near proximity to a gas station.

A. Gas Price Localization & the GSV Image Collection App

The resultant dataset of $>300,000$ seed locations was then used to obtain a GSV image of the advertised gas price in the environs of each seed. Each seed represents a geographical location with an available GSV panorama in relative proximity to a gas station. The GSV interface was used to manually navigate the view to identify the advertised gas price(s). To perform this at scale, the Amazon Mechanical Turk (AMT) service was used in combination with a cross-platform GSV Image Collection application built upon the *Electron* and *React* frameworks. The application interface is depicted in Figure 2. Workers were instructed to navigate the interactive GSV interface to center and maximally zoom-in the view onto the advertised gas price before capturing the image as a 640×640 pixel scene. Consequently, a given seed location is "transformed" into an image encapsulating the desired gas price information. This data collection did not require research ethics board approval and AMT workers were compensated for each annotation task in accordance to the AMT best practices for pricing.

Figure 2. Screen-capture of the Image Collection Application interface. (1) indicates the input file of seed URLs; (2) indicates the directory to which images are saved; (3) indicates the current link number among all seeds; (4) a progress bar to visually compliment (3); (5) the interactive GSV interface to navigate the environment and frame the resultant image; (6) interactive map to teleport the view; (7) a link to the current seed URL; (8) previously completed seed URLs; (9) button to navigate to the previous set of seed URLs; (10) button to capture the currently framed view and save to file; (11) button to indicate an invalid seed link and skip to the next; (12) button to skip to the next set of seed URLs.

B. Ground Truth Annotation of Gas Prices

From the resultant dataset of images, the ground truth annotation of the contained gas prices is required. The concept of *ground truth* in this case is somewhat complex, given that several types and levels of annotations are possible. The segmentation of an image associates specific pixels to a given class. For example, a bounding box encapsulating the pixel depicting the digit "3" is assigned the class label "3". However, the detection of prices requires that individual digits be intelligently combined into the price they represent. Thus, for multi-digit numbers, such as gas prices, a bounding box might additionally encapsulate the pixels depicting the price "3.14" which, analogously, receives a class label of "3.14". Determining the grade of fuel is also necessary to relate the correct price to the correct fuel. Segmented regions of the image that encapsulate the pixels depicting the grade of a given fuel are also needed. To eliminate distracting elements within the image or to ignore unrelated digits within the frame, a bounding box over the gas price sign would additionally support focusing the attention of a learning model towards pixels relevant to the task of extracting prices (Figure 3).

On the other hand, an image can also be assigned a specific class; the prototypical image of a cat receives the class "cat". In this case, a learning algorithm learns to

focus on the salient features throughout the entire image to learn "cat-ness" facets. For our purposes, the class label of an image may be assigned the represented price(s). In it's simplest form, the ground truth label of an image comprises the most represented element: the cash price of regular, unleaded gasoline. However, to support the extraction of the variable number of all prices defined in the image, we devised the following ground truth labelling convention, supporting a flexible number of advertised prices, multiple grades of fuel, missing digits, and fractional digits:

$$\langle \text{gas-price-1} \rangle; \langle \text{gas-price-2} \rangle; \ldots; \langle \text{gas-price-}n \rangle \quad (1)$$

where n is the number of gas prices displayed in the image, and the prices are ordered by appearance when processing the image row-by-row and left-to-right from the top-left to bottom-right. Each given gas price $\langle \text{gas-price-}i \rangle$ in the image is assigned a unique *price id* and its ground truth label has the string form:

$$`x.xx(x/xx)g' \quad (2)$$

where $x \in \{0, 1, \ldots, 9\}$ and $g \in \{`r', `m', `p', `d'\}$ as the *grade* of fuel, representing 'Regular', 'Mid-Grade', 'Premium', and 'Diesel', respectively. The character '-' is used whenever a given digit or grade value is missing or deemed uninterpretable.

AMT was used to transcribe the identifiable prices within each image using the above convention. In a second round of AMT annotation, the bounding box of the sign within each image was determined, generating masks useful for transfer learning a semantic segmentation model. For digit classification tasks, the GPA currently contains 3-digit annotations for the cash price of regular, unleaded fuel. For segmentation tasks, the sign-level masks are currently available for a subset of images. Collection of digit-, price-, and label-segmented masks, as depicted in Figure 3, is a continuing effort. Progressively, AMT will be leveraged to obtain complete ground truth labels for all gas prices depicted in each GPA image.

C. Baseline Performance of the GPA

To gauge the difficulty of the GPA, we generate baseline models for both the digit classification and the sign segmentation tasks.

For digit classification, we implemented a YOLO ("You Only Look Once") digit detector trained on the SVHN dataset. Inferences were made for each image and performance was reported using the average precision and average recall of the model across all images. Varying the detection threshold parameter of the YOLO model produced a cluster of such average precision and average recall points.

Recall and precision are defined as:

$$\text{recall} = \frac{|\text{correctly predicted digits}|}{|\text{predicted digits}|} \quad (3)$$

$$\text{precision} = \frac{|\text{correctly predicted digits}|}{|\text{ground truth digits}|} \quad (4)$$

To better understand the expected baseline performance for models predicting a variable number of digits for images that may contain a variable number of true digits, we ran several Monte Carlo simulations using the algorithm depicted in Algorithm 1. In keeping the number of "target" values (*i.e.* y) fixed and varying the number of "predicted" values (*i.e* x), we repeatedly simulate random predictors, resulting in a curve in the average precision-recall space. Repeating the same for different target values produces a

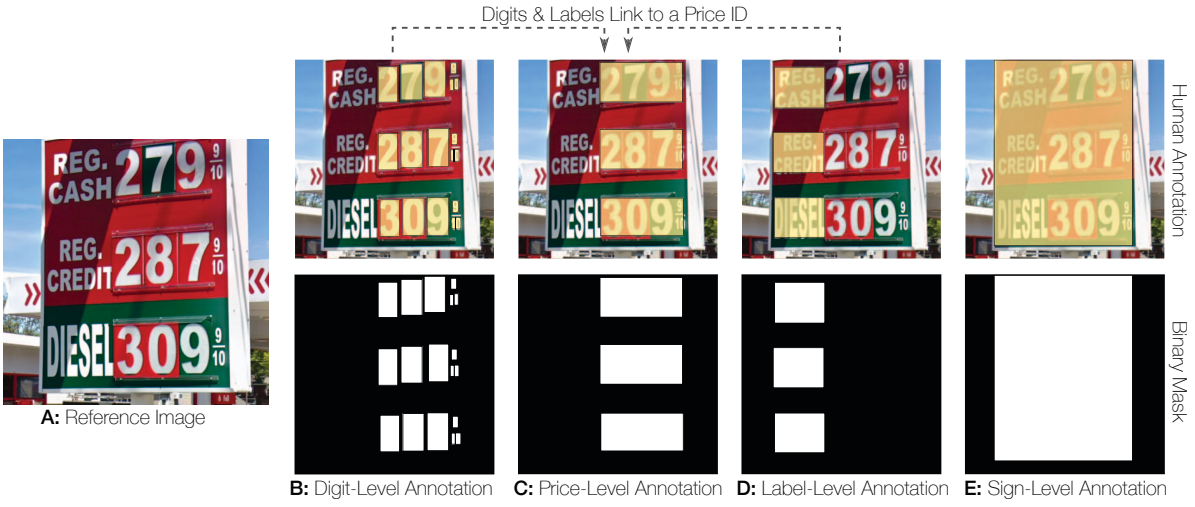

Figure 3. Example segmentation of a reference image (A) at the digit-level (B), price-level (C), label-level (D), and sign-level (E). A mapping based on a unique price ID is used to relate digit- and label-based information to each multi-digit price.

family of distinct curves in this space. We simulated values for $x \in [1, 2, \ldots, 100]$ and $y \in [1, 2, \ldots, 10, 20, \ldots, 100]$, using $i = 1000$ iterations for each simulation.

Finally, to establish the baseline performance for a segmentation-type task, we used the available sign-level masks to perform transfer learning using Google's DeepLabV3 ResNet101 semantic segmentation model implemented in PyTorch [9]. Accompanying code can be found at github.com/GreenCUBIC/Gas-Prices-of-America.

III. Results & Discussion

The preparation of a large-scale image dataset of advertised gas prices in America makes contributions to the broader scientific community in three ways:

1) Enables the development of CV models to automatically detect and report gas prices as part of AV crowdsourcing infrastructure.
2) Provides a benchmark to evaluate the performance of non-integer, multi-number, multi-digit values in natural environments.
3) Enables the development of various CV applications based on the extraction and analysis of various *metrics in the wild* (discussed below).

The dataset of images is made publicly available at the following locations cu-bic.ca/gpa/ and the GPA DataVerse Repository. The data are prepared such that minimal data preprocessing or formatting is required. As with the SVHN dataset, these images are sourced from the natural environment; however, given the variability in the number of prices, the GPA dataset presents a considerably more challenging, unsolved, real-world problem: the recognition of multi-digits numbers in natural scene images.

A. Characteristics of the GPA Dataset

At present, the GPA dataset comprised 2,048 640 $\times$ 640 pixel images, each averaging 540MB for 1.0GB in total storage. For each image, the single-price, ground truth label is available and sign-level masks are available for a subset of 1,056 images. These images were obtained by applying the localization procedure to approximately 3,000 randomly sampled, refined seed locations. As this represents only 1% of all available seed links, ongoing image collection and annotation work is expected to dramatically increase the size of the GPA dataset.

Algorithm 1 Monte Carlo Baseline Simulation

Input: num. of "target" digits, y, num. of "predicted" digits, x, num. of iterations, i
Output: avg. recall, avg. precision, 95% CIs

```
P ← []
R ← []
for i iterations do
    Y ← getRandomDigits(y)
    X ← getRandomDigits(x)
    R.append(getRecall(X, Y))
    P.append(getPrecision(X, Y))
end for
return avg(R), avg(P), getCIs(R), getCIs(P)
```

To ensure that the GPA imagery were not geographically biased we examined the number of coarse-grained seed locations obtained from AGS for each state v.s. the 2019 population census [10]. As expected, we observe a strong positive relation (Spearman 0.876) between the number of gas stations and state population with a noted exception of Idaho. Analogously, a moderate positive relationship (Spearman 0.477) between number of gas stations and population density, with the noted exception of Washington, DC.

In addition to depictions of these relationships, the GPA website additionally illustrates a set of six exemplar images representing relatively "easy" cases. Of these, images A-B each sport a single price comprising large digits and few environmental distractions. Images C-D depict images for which the gas price dominates the view, while exhibiting multiple prices, either varying in size, or consistently represented. Finally, images E and F depict high contrast and well defined digital displays. Even among these relatively easy images, the diversity of the dataset is clear, with multiple font styles, colours, sizes, etc.

Conversely, we also selected a set of images to highlight the challenging facets of accurately extracting visible gas prices within the images. Images A-F depict environmental, spatial, or temporal factors that render the digit recognition task more difficult. Blurring artifacts are often due to the GSV Licence Plate Anonymization process which apparently has mistaken advertised prices as vehicular identification, perhaps due to the aspect ratio of the price. Generally speaking, price and digit obfuscation were avoided at the price localization step by navigating the street-level view to a non-obstructing position; however, such a vantage point is not always readily available nor should it be expected within all AV-sourced imagery. Variations in lighting, price scale, backdrop, image resolution, and angle vary considerably throughout the dataset.

Images G-L depict contextual challenges to price extraction. Notably, the presence of multiple prices based on multiple grades of fuel (*e.g.* regular, mid-grade, premium, diesel) and payment method (*e.g.* cash vs. credit) commensurately increase the difficulty in combining individual digits into the extracted price and relating that price to the correct fuel grade. This challenge is exacerbated when multiple pricing signs appear together (Image I). Inter-price variations in fonts styles are prevalent throughout the dataset; however, rare instances of intra-price variation in font style do exist (*e.g.* Image J). Finally, partial digits or uniterpretable digits result from partially rendered digital displays, broken displays, missing values, or distant displays (Images K-L).

As previously described, the GPA dataset supports multiple levels of ground truth annotations. Contrary to other CV benchmark digit-specific datasets, such as MNIST and SVHN, the GPA can be leveraged in numerous evaluations of extracting price information, each presenting an increased level of difficulty.

B. Multiple Levels of Benchmark Difficulty

A distinct advantage of the GPA dataset over other benchmark datasets is the opportunity to extract various information, with differing levels of complexity, from the images. The GPA dataset is annotated such that it supports multiple levels of difficulty when evaluating a given CV model. The baseline task is the accurate extraction of the multiple gas prices, which can optionally include the recognition of the fractional portion of the price (*e.g.* $^9/_{10}$) and determine the grade of fuel attributed to a given price. The following prediction tasks are listed in order of difficulty:

1) **Sign-Level Segmentation:** Localize the sign which displays the advertised prices.
2) **Price-Level Segmentation:** Localize individual prices.
3) **Digit-Level Segmentation:** Localize individual digits and group together those contributing to the same price.
4) **Label-Level Segmentation:** Perform digit-level segmentation and grouping label-specific segmentations.
5) **Single-Price Classification:** Accurately extract *one* of the listed prices (no fraction) in the image (*e.g.* the regular, unleaded, cash price).
6) **Multi-Price Classification 1:** Accurately extract *all* of the listed prices (no fractions) in the image.
7) **Multi-Price Classification 2:** Accurately extract *all* of the listed prices, including fractions when present, in the image.
8) **Multi-Price Classification 3:** Accurately extract *all* of the listed prices, including fractions when present, as well as the *grade* of fuel (*e.g.* 'Regular', 'Mid-Grade', 'Premium', 'Diesel').

In alignment with the state-of-the-art section of paperswithcode.com, maintaining a scoreboard of methods and model architectures succeeding in one, multiple, or all categories will help establish the state-of-the-art approaches in the identification of numbers in the wild. A CV model, aptly denoted "GasBotty", capable of achieving a consistently high performance on all of these tasks, would be suitable to deploy as part of AV infrastructure to automate the process of updating the GasBuddy database with the most up-to-date gas prices.

C. Baseline Performance of State-of-the-Art Methods

To understand the difficulty of each benchmark, we leveraged various state-of-the-art CV models designed to address related tasks. Since only the annotated labels for the regular, unleaded, cash price and sign-level masks are available for each image, we establish this baseline performance for the benchmark five (Single-Price Classification) and benchmark one (Sign-Level Segmentation).

For the classification of a single gas price, we implemented a vanilla YOLO digit detection model trained on the SVHN dataset and made inferences for each of the 2,048 images and reported the average precision and average recall. We varied the YOLO threshold parameter and repeated this procedure producing a cluster of points (Figure 4A,C). The performance clusters for three random digit predictors (predicting $x = [1, 3, 10]$ random digits, respectively) are plotted in comparison along with the MC simulation for varying x and fixed $y = 3$. Interestingly, the YOLO model performance does not outperform random and may be explained by the discrepancy in training and evaluation image sizes and complexity (Figure 5).

The MC simulations for a varying number of target digits reveals that even a random model can appear to perform well with an increasingly large number of target digits (Figure 4B,D). Moreover, the non-uniform distribution of gas price digits (Figure 5A) suggest that a model biased to predicting twos, nines, and threes will see improved precision over one that samples digits at uniform random (Figure 4C). These results demonstrate the non-trivial nature of detecting multiple digits among multiple numbers within imagery sourced from the natural environment.

The baseline performance for the segmentation-type task, however, demonstrated that transfer learning using Google's state-of-the-art DeepLabV3 ResNet101 could rapidly produce a near-optimal segmentation model. Following 100 training epochs, we obtained test F1 performance of 0.7125 (Figure 6A). Interestingly, a quasi-periodic drop in performance was observed which we posit to reflect a subset of challenging images. Sample input and model output for a

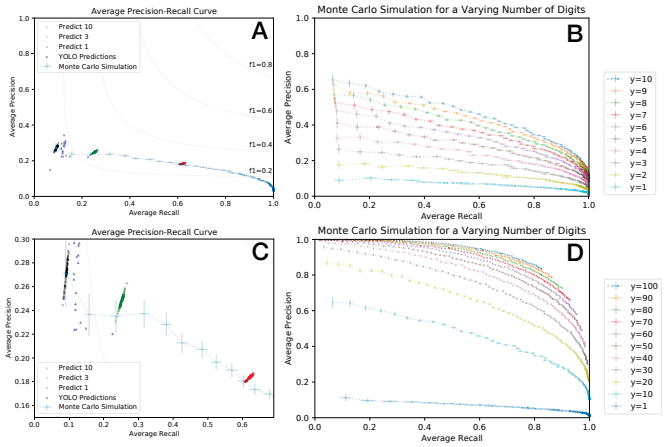

Figure 4. Digit Classification Baseline Performance. Panels A & C depict the $y = 3$ MC simulated baseline performance, the YOLO model predictions, and the performance of three random schemas ($x = \{1, 3, 10\}$) evaluated for the GPA dataset. Panels B & D depict each of the MC simulated baselines for a variable number of target digits.

given image is depicted in Figure 6B-D).

The ongoing collection of digit-, price-, and label-level annotated masks will enable determination of baseline performance for the remaining segmentation-type benchmarks. The combined use of these models is expected to contribute to the development of a robust method capable of accurately identifying metrics in the wild.

D. GasBotty: Machine-Augmented Crowd-Sourcing

Effective CV models that successfully localize and extract gas prices represent the first opportunity for leveraging AV infrastructure to remotely monitor time-varying events in the natural environment. When incorporated in the edge/fog layer of a fleet of AVs, the onus of updating GasBuddy's database will no longer fall to the human, relieving them of the responsibility to communicate gas price updates.

GasBotty represents the first of what is anticipated to be many applications wherein AV-based CV models contribute crowd-sourced information. Conceivably, the crowd-contributed traffic updates of the Waze application might eventually be automated from a similar AV-based CV model. Arguably, the task of identifying dynamic traffic events such as stalled vehicles, hidden police traps, and traffic accidents are far more complicated. Inter-vehicle communication systems have been proposed for alerting vehicles within the vicinity of a possible hazard [11]. Our proposed vision is complementary in that the AV-sourced detections would also be aggregated within a centralized repository and distributed to all interested parties.

Further development of AV systems to accommodate this streetscape crowd-sourced information would foreshadow

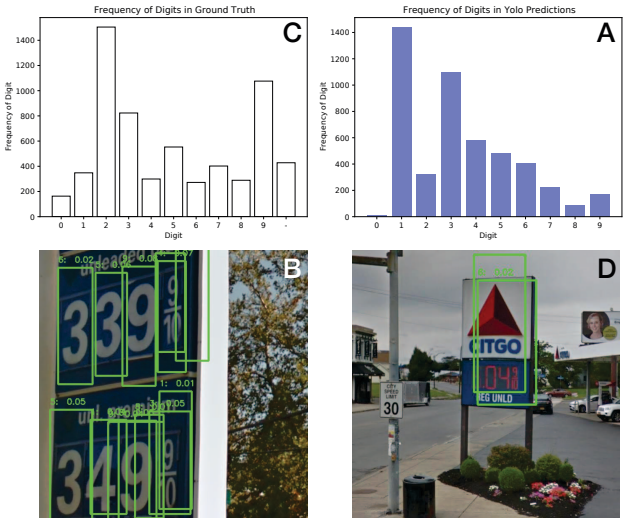

Figure 5. Digit Frequency and YOLO Predictions. Panels A & B depict the GPA digit frequency as compared to those predicted by the YOLO digit detector. Panels C & D depict two sample YOLO predictions and emphasize the discrepancy in predicted digit size and actual digit size.

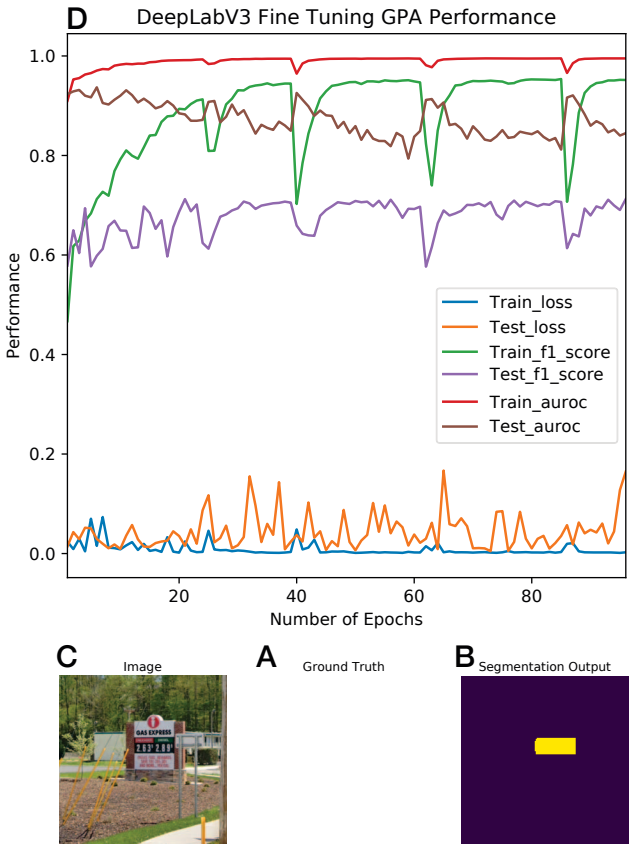

Figure 6. Sign Segmentation Baseline Performance using Transfer Learning with the DeepLabV3 ResNet101 Model.

the impending era of machine-augmented crowd-sourced data. It is expected that machine-augmented crowd-sourced data would increase the quality and consistency of crowd-sourced data. Furthermore, users of a given platform are more likely to contribute data if the detection and update process is made effortless (*i.e.* through automation). Promisingly, the data will likely have greater temporal resolution and, therefore, greater utility.

E. Extracting Metrics in the Wild

Beyond its utility as a benchmark for the evaluation of CV models and development of sophisticated AV-destined CV models, the GPA dataset promises further application in the generalized extraction of *metrics in the wild*. That is, analogous to the anticipated GasBotty model, capable of extracting gas prices from the surround environment with high fidelity, a myriad of similar models able to extract various metrics from the wild are made possible. As visual creatures ourselves, urban environments are densely packed with advertisement and media in subtle pursuit of our attention. Quickly perusing a sample of images from the GPA dataset reveals widespread pricing information for commodities other than gas, such as cigarettes, soft drinks,

alcohol, and fast food.

Businesses for which the advertised commodity is completely interchangeable (*e.g.* regular gas from Shell $\approx$ regular gas from Exxon $\approx$ regular gas from Valero $\approx \cdots$) must uphold their end of an implicit *social contract* with the consumer by advertising their price. A wise consumer is likely to peruse the competition in search of the lowest gas price at a given point in time. In their proverbial "race to the bottom", these businesses are reluctant to advertise their prices online at the expense of their profits. The demand for omniscient knowledge of the "current best deal" gave rise to the GasBuddy platform.

Related examples exist; the detection of any time-varying, visually advertised value from a streetscape can benefit from the GPA dataset. From the perspective of healthcare, the road-side advertisement of a clinic's wait time can be extracted and updated. Entirely new applications may result from the increased temporal resolution and accuracy of such machine-augmented crowd-sourcing. For example, prior to and following a natural disaster, crowd-sourced information about local resources is often provided by fellow citizens through platforms such as Twitter. Moving into the machine-augmented crowd-sourcing era, a fleet of AVs capable of the high-temporal vision-based detection of the (un)availability of a specific resources would revolutionize crisis management for the betterment of humanity.

IV. Conclusion

Leveraging the high spatio-temporal resolution imagery from circulating fleets of AVs with state-of-the-art CV methods promises to revolutionize our ability to crowd-source time- and space-varying events within the natural environment. The GPA dataset described in this work represents a first step towards the development of highly accurate CV models with utility as part of AV infrastructure. The dataset further serves as a benchmark against which the CV community might evaluate their models for variable numbered multi-digit, multi-number classification tasks. The GSV Image Annotation app described, herein, also contributes a useful tool for the rapid localization of desirable GSV imagery from a set of seed locations. In summary, the secondary use of these AV-sourced data heralds a new era of machine-augmented crowd-sourcing when fused with state-of-the-art machine learning methods. The GPA dataset is available online at cu-bic.ca/gpa and at doi.org/10.5683/SP2/KQ6VNG. Accompanying code can be found at github.com/GreenCUBIC/Gas-Prices-of-America.

References

[1] A. Rzotkiewicz, A. L. Pearson, B. V. Dougherty, A. Shortridge, and N. Wilson, "Systematic review of the use of google street view in health research: major themes, strengths, weaknesses and possibilities for future research," *Health & place*, vol. 52, pp. 240–246, 2018.

[2] A. G. Rundle, M. D. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler, "Using google street view to audit neighborhood environments," *American journal of preventive medicine*, vol. 40, no. 1, pp. 94–100, 2011.

[3] J. S. Apte, K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. Vermeulen, and S. P. Hamburg, "High-resolution air pollution mapping with google street view cars: exploiting big data," *Environmental science & technology*, vol. 51, no. 12, pp. 6999–7008, 2017.

[4] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using google street view and a modified green view index," *Urban Forestry & Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015.

[5] C. Guo, T. H. Kim, A. Susarla, and V. Sambamurthy, "Understanding content contribution behavior in a geo-segmented mobile virtual community: The context of waze," *Available at SSRN 3065303*, 2019.

[6] Y. LeCun, C. Cortes, and C. J. Burges, "The mnist database of handwritten digits, 1998," *URL http://yann. lecun. com/exdb/mnist*, vol. 10, p. 34, 1998.

[7] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.

[8] K. Dick, F. Charih, Y. S. Dosso, L. Russell, and J. R. Green, "Systematic street view sampling: High quality annotation of power infrastructure in rural ontario," in *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE, 2018, pp. 134–141.

[9] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[10] United States Census Bureau, "National population totals and components of change: 2010-2019," https://www.census.gov/data/datasets/time-series/demo/popest/2010s-national-total.html.

[11] C. Brenzel, C. Passmann, and R. Meschenmoser, "Wireless inter-vehicle communication for hazard warning," *IFAC Proceedings Volumes*, vol. 34, no. 9, pp. 275–277, 2001.