Visual Census Using Cars

Anonymous CVPR submission

Paper ID 1611

Figure 1. Look at the car on the *left*. It is a Prius 2012. From the car only, can you guess where the owner lives and drives the car? What his or her highest degree is? And what is the household income of the owner? A car can tell us a thousand things, as it turns out. In this paper, we combine results from a fine-grained car detection algorithm with a number of social census data provided by the U.S. government, and are able to infer that Prius is the most popular car in the city of San Francisco, its owner often holds a college degree and earns a household income of $30,000 (*right*)

1. Introduction

Abstract

Detecting a large number of BMWs in images informs us that those images may be of a wealthy area. Conversely, knowing that our images were obtained from a wealthy neighborhood increases the likelihood of detecting expensive cars. We explore this relationship between demographic factors and fine-grained classes by performing large scale detection of over 2600 car classes and conducting a social analysis of unprecedented scale in computer vision. Using 45 million images from 200 of the biggest cities in the United States, we predict demographic factors such as neighborhood wealth, education levels and show that our results correlate well with census data. To facilitate our work, we have collected the largest and most challenging fine-grained dataset reported to date consisting of over 2600 classes of cars comprised of images from google Street View and other web sources,classified by car experts to account for even the most subtle of visual differences.

The artifacts which we choose to surround ourselves with tell us not only about ourselves but also about the society in which we live. In the 21st century, some of the most relevant objects defining people and their lifestyles are houses, clothes and cars. The cars people own can provide significant personal information: by knowing that the woman in Fig. 1 drives a Prius we can guess that she is probably from San Francisco and earns an income greater than $30,000. Traditionally, the most prevalent method for gathering such personal and demographic information is through surveys such as census and American community survey (ACS) projects. However, the emergence of large and diverse sets of data generated by people has enabled computer scientists and computational sociologists to gain interesting insights by analyzing massive user texts and social networks [26, 7]. For example, recent work from [20] analyzed over one million books and presented results related to the evolution of the English language as well as various cultural phenomena.

On the computer vision side, a few pioneering works by Zhou et al, Ordonez et al, and Naik et al have recently started to apply visual scene analysis techniques to infer characteristics of neighborhoods and cities [28, 13, 21, 23]. In this work, we are also interested in using images to understand cities, neighborhoods and the demographic makeup of their inhabitants. However, instead of using global image statistics, we achieve this goal by detecting and classifying cars on the street. 95% of American households own cars [6], and as seen in fig. 1 cars give a lot of information about individuals as well as neighborhoods. By using cars as a lens into understanding society, we are able to gain insights ranging from the demographic makeup of cities to neighborhood pollution levels.

Our contributions are two-fold. First, we offer a fine-grained car dataset of unprecedented scale. It has 2657 car classes consisting of nearly all car types produced in the world after 1990: with a total of 700,000 images from websites such as edmunds.com, cars.com, craigslist.com and Google Street View (Fig. 2). We use our dataset to train a large scale fine-grained detection system, detecting cars in more than 45,000,000 Google Street View images collected from 200 of the largest American cities.

Figure 2. Examples of cars from our fine-grained dataset. Left: examples of cars from edmunds.com, cars.com and craigslist.com. Right: examples of cars from streetview images. Cars from streetview images tend to be lower resolution and are often occluded where as those from other sources are typically unoccluded and centered in the image.

Second, we present a suite of interesting social analysis of American people and cities using our car detections. In Section 5, we show that from a single source of data, Google Street View images, we are able to predict diverse sets of important societal information typically gathered by different entities. We not only show good correlation with census data related to demographics, but can also predict information not present in census data such as neighborhood pollution levels. Finally, Section 6 presents preliminary results in exploring the use of demographic information to improve fine-grained car classification.

2. Related Work

City analysis via image features. There has been recent interest in using images to characterize cities [24, 23, 9, 28, 21, 13]. Salesses et al [24] created scores for perceptions of wealth, safety and uniqueness by asking people to rate images from three cities on a scale of 1–10 and [21, 23] predicted these scores using various global image features such as GIST [22] and CNN [17]. In another line of work analyzing cities, [28] perform city identity recognition after representing each city with higher level attributes. Doersch et al [9] identify unique qualities of cities such as Paris and Prague and [13] shows that given an image of a particular city location, it is possible to predict the most likely direction for the location of a McDonalds. While our work also shares the motive of city analysis through imagery, it differs in that we use fine-grained object recognition to achieve this goal. As shown in section 5, this allows us to perform much more extensive social and demographic analysis through imagery and also easily extend our analysis to many other cities without the need for additional labeling.

Fine-grained object recognition. Fine-grained object recognition is a difficult problem due to the high visual similarity between classes. Nevertheless, recent works such as [27, 5] show impressive results where [27] augments state of the art object detection algorithms such as RCNN [11] and [5] uses a strongly supervised DPM model [2]. Although there are many fine-grained datasets such as [25, 14, 15, 19, 4], none of them match object classification datasets like imagenet [8] in the number of classes or images. Recent works such as [4] have introduced larger scale fine-grained datasets and [19] has introduced a 3D car dataset annotated with metadata such as location information. We introduce a geotagged car dataset with unprecedented scale in both the number of classes and images.

Using GPS data to improve classification. Although an increasing number of images that we interact with daily are associated with GPS tags, there are very few computer vision algorithms that take advantage of location based metadata. However, recent works such as [1] use location information to assist in detecting objects such as trash cans and street lamps, [4] learns a spatio-temporal prior to improve bird classification and [19] uses some location information such as elevation to assist in car classification. Following the work of [4], we explore the use of

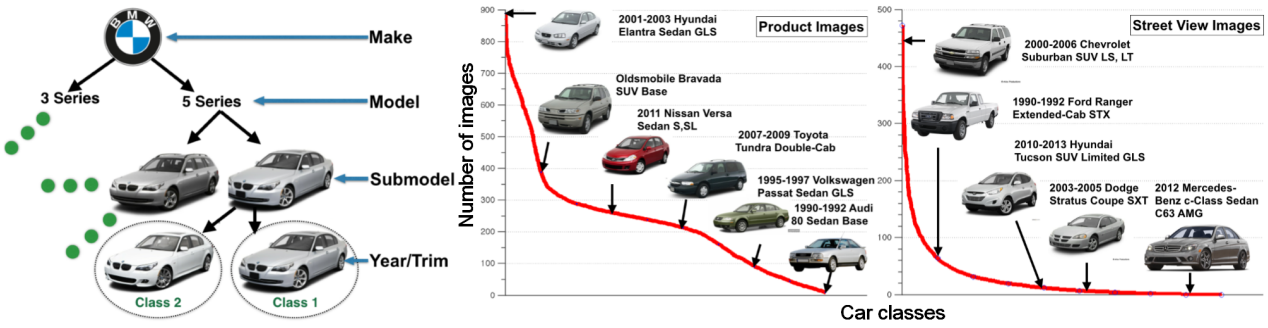

Figure 3. Left: A hierarchy of car classes in our dataset. Classes become more difficult to distinguish lower in the hierarchy, with differences extremely subtle at the year and trim level. Center: The number of images per class obtained from edmunds.com, cars.com, and craigslist.com. Right: The number of images per class from Street View.

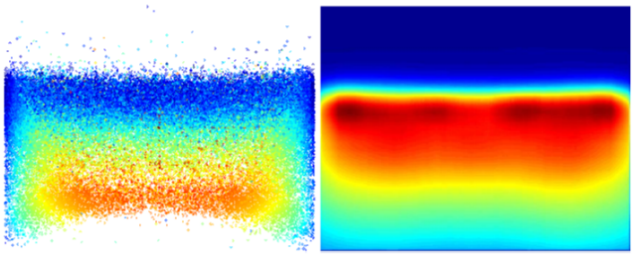

Figure 4. (a) Heatmap of bounding box scale. Each point represents a single bounding box from Street View. The color of each point color indicates the area, with red higher area. The magnitude at each point indicates the size of a bounding box centered at that location in a Street View image. (b) Heatmap of Street View bounding box coverage in an image. The magnitude at each point indicates the number of bounding boxes that include that point.

Attribute	Training	Validation	Test
Street View Images	199,666	39,933	159,732
Street View BBoxes	34,712	6,915	27,865
Product Shot Images	313,099	-	-
Product Shot BBoxes	313,099	-	-
Total Images	512,765	39,933	159,732

Table 1. Dataset statistics for our training, validation, and test splits. "BBox" is shorthand for Bounding Box. Product shot bounding boxes and images are from craigslist.com, cars.com and edmunds.com.

demographic data to improve fine-grained classification.

3. Cars and Cities Dataset

We collected 45 million images from 8 million GPS points by sampling the roads of 200 of the biggest US cities every 25m. For each GPS coordinate we gathered Street View images at 0, 60, 120, 180, 240, 300 degree rotations. In order to train a fine-grained car classifier to detect and classify the cars in our images, we created the largest ever reported fine-grained dataset of cars consisting of all the cars listed on a prominent car information website: edmunds.com (all cars manufactured after 1990). To assemble our car dataset we obtained images of all cars in edmunds.com (∼18k cars) and grouped the cars into visually indistinguishable classes using a series of amazon mechanical turk tasks. After creating a class list, we collected additional images of cars from cars.com as well as craigslist and used AMT to annotate them with bounding boxes. Finally, we hired car experts to label ∼ 70k bounding boxes of cars from the street view images with fine-grained lables. We also labeled cars from craigslist.com and cars.com by parsing the car posting titles.

Table **??** summarizes the statistics of our dataset, Fig. 2 shows example images and Fig. 4 visualizes the location and scale of car bounding boxes in the Street View images. As seen in Fig. 4 and Fig. 2, Street View images typically contain multiple cars per image where as images from sources such as craigslist.com contain one car per image. Bounding boxes from Street View images can also be located in many parts of the image, have a large range of sizes and are typically blurry and occluded. These aspects of our dataset differentiate our data from other fine-grained datasets such as [25] containing one centered and focused object per image. Fig. 3a, shows a hierarchy of classes in our dataset where classes become increasingly visually indistinguishable while traveling down the tree. It can be seen that the difference between the fine-grained classes in our dataset is extremely subtle making it difficult for non-expert humans to distinguish. Finally Figs. 3a and b show the distribution of Street View images and product shots (images from craigslist.com, cars.com and edmunds.com) for different classes. The image/class distribution for images from

Attribute	Accuracy
Make	66.38%
Model	51.83%
Submodel	77.74%
Price	61.61%
Domestic/Foreign	87.71%
Country	84.21%

Table 2. Classification accuracy on the test set for various car attributes.

the 3 websites is very different from that of the Street View images. By collecting images from different sources we minimize the amount of bias present in our dataset.

4. Detecting and classifying cars

Although RCNN based fine-grained detection algorithms have reported state of the art results, [11] [27], its computational and memory requirements make it impractical for use in a large scale detection setting such as ours. In addition to memory requirements, detecting cars in our images would take ∼20s per image using a machine with a single GPU. Thus our pipeline, instead, consists of using DPM [10] to detect cars and a CNN [16] to classify them. We present details in the sections below.

4.1. Car Detection

Inorder to evaluate accuracy/speed trade off, we trained DPMS with different numbers of components and parts and used the model with the best tradeoff for detection. Our final algorithm employs a single component 8 part DPM and achieves an AP of 64.2% while taking 5 secs per image. As a point of comparison, the highest AP (68.7%) was achieved with a 5 component 8 part DPM which takes ∼22 secs per image. Detailed plots and timing measurements of other DPMs we have trained are discussed in our supplementary material.

4.2. Car Classification

We use a standard CNN from [16] with [12] to classify the DPM detections into one of 2657 fine-grained classes. Although our test set consists of street view images, as mentioned in sec. 3 61% of our positive training images are composed of cars from other sources such as craigslist. Thus we add deformations such as blurring to these images during training to liken them to street view images. We give details of training the CNN in our supplementary material. At test time we take the top 10% scoring DPM bounding boxes and classify them. This results in a 10x increase in speed but only a ∼2% drop in AP as compared to using all the bounding boxes detected by DPM. We achieve an accuracy of 33.15% on the true positive DPM bounding boxes

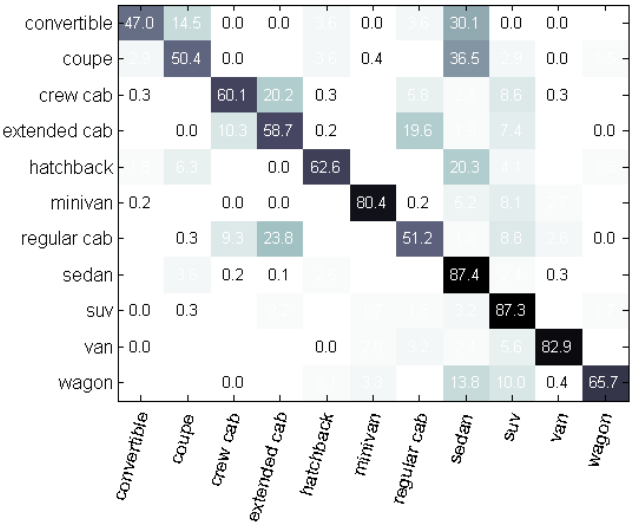

Figure 5. Confusion matrix for car body types. Most misclassifications are between similar types of submodels such as sedans and coupes or extended cabs and a crew cabs (different types of trucks).

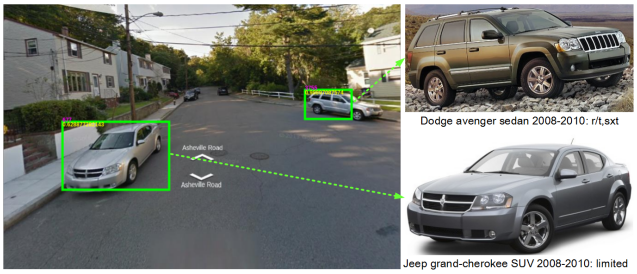

Figure 6. An example of our car detections and classifications. The two cars in the image are correctly detected and classified as the classes shown on the right.

and 31.27% on the ground truth bounding boxes. Fig. 6 shows an example of our Street View detections. The cars are detected and classified correctly even though blurriness and occlusions in street view images makes this a difficult task. Fig. 5 shows a confusion matrix for submodel level classifications. It can be seen that most of the errors are between highly similar submodels such as sedan and coupe.

4.3. Analyzing Hierarchical Classification Accuracy

Some types of classification mistakes are more costly than others for the task of social analysis. For example, an error misclassifying 2001 Honda Accord lx to 2001 Honda Accord dx is not significant. However, misclassifying a 2012 BMW 3-series to a 1996 Honda Accord,for example, would create large errors in an analysis measuring the realtionship between the average car price or age in a zip code

and median household income. In order to gain more insight into the types of errors our classifier makes, we measure the accuracy of classifying different car attributes. Table 2 lists accuracies by various car attributes such as those in fig. 3a and others like car price, year etc... We can see that the accuracy is much higher after aggregating by different attributes.

5. Visual Census

There are many government and non-government projects such as census and American community survey (ACS) that are dedicated to obtaining personal and demographic information and each of these projects studies a particular aspect of demographics and cities. For example, ACS collects data related to the demographic makeup of the US, the American Environmental Agency collects data pertaining to city pollution and private organizations such as car dealerships gather information regarding the relationship between cars and demographics. In this section we ask: how much information about individuals, neighborhoods and cities can we learn from a single source of data —Google Street View images? Surprisingly, we show that by detecting cars in 45,000,000 images across 200 cities, we can predict data traditionally obtained from multiple sources.

In cases with available ground truth data, we show good correlation with results from disparate sources such as census data, Massachusetts vehicle registration, San Francisco air quality data and other market research sources. In cases where we do not have ground truth data we show that our results follow conventional wisdom. Finally, we present spatial statistics based analysis that can only be performed due to the dense GPS sampling in our data.

We divide our analysis into three sections presenting individual person level, zip code level and city level results.

5.1. What cars on the street tell us about people

Sanity Check: Street View car detections correlate well with registered cars. Do detected cars in a zip code correlate with cars driven by its residents? To answer this question we obtained vehicle census data for Massachusetts, which is the only state to release extensive vehicle registration data, and found an extremely high Pearson correlation coefficient of 0.82 ($p \ll 0.01$) between the expected number of cars we detected per zip code and the number of registered cars. Fig. 8 plots Pearson correlation coefficients between the number of detected and registered cars for each make. The high correlation values verify that cars detected from Google Street View images contain useful information pertaining to the types of cars driven by people in a particular zip code.

Census Variable	Car Attribute	Pearson r
Household income	# 1990-1994 cars	-0.42
Household income	# 1995-1999 cars	-0.40
Household income	# 2000-2004 cars	0.21
Household income	# 2005-2009 cars	0.46
Household income	% of foreign cars	0.59
Household income	% of German cars	0.57
Household income	% of US cars	-0.59
Household income	Avg. price	0.49
Education: highschool	Avg. price	-0.21
Education: college	Avg. price	0.32
Education: graduate scl.	Avg. price	0.39

Table 3. Pearson correlation coefficient between various census variables and detected car attributes. All p values are $\ll 0.01$.

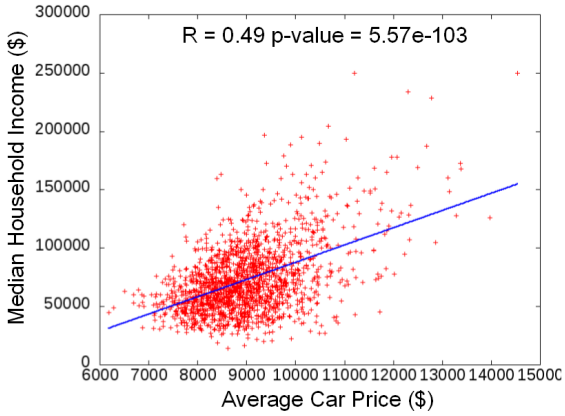

Figure 7. Scatter plot of average price of detected cars per zip code vs. median household income per zip code for all zip codes in our dataset.

The relationship between individual wealth and cars. Do wealthy people drive expensive cars? To answer this question we gathered zip code level as well as census tract level 2007-2012 American Community Survey data for the 200 cities in our dataset and analyzed how the census data relates to statistics from our detected cars. Table 3 shows correlation values between various attributes of the detected cars and census variables and Fig. 7 plots average car price per zip code vs. median household income. As expected, there is a high correlation between median household income and the average car price in a zip code (r=0.49) which indicates that wealthy people do drive expensive cars. The correlation values in Table **??** show results consistent with conventional wisdom as well as ones that may not be obvious. For example, a result consistent with conventional wisdom is that wealthy people drive foreign, especially German cars and that poor people drive old cars with low miles per gallon (MPG). What is perhaps surprising is that poor people also drive American cars.

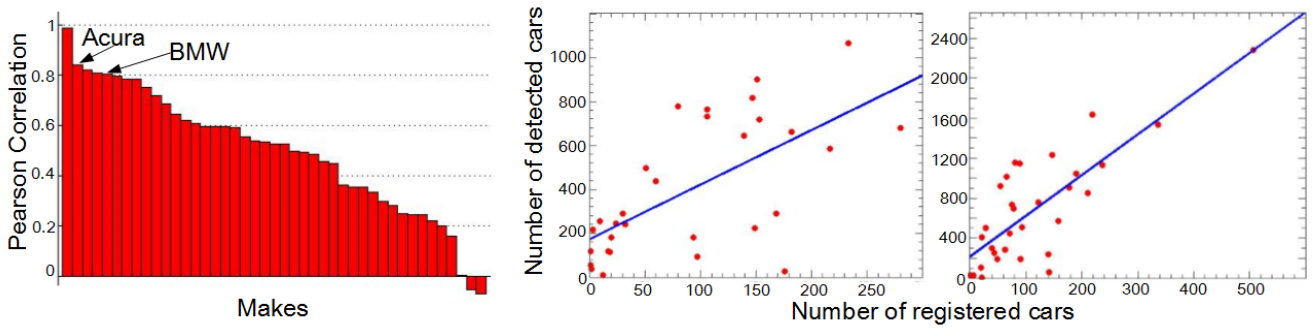

Figure 8. Left: Pearson correlation coefficient between the number of detected and registered cars in Boston for each car make. Center: Scatter plot of detected BMWs per zip code vs. registered BMWs per zip code in Boston. Right: number of detected Honda's per zip code vs. registered Honda's per zip code in Boston.

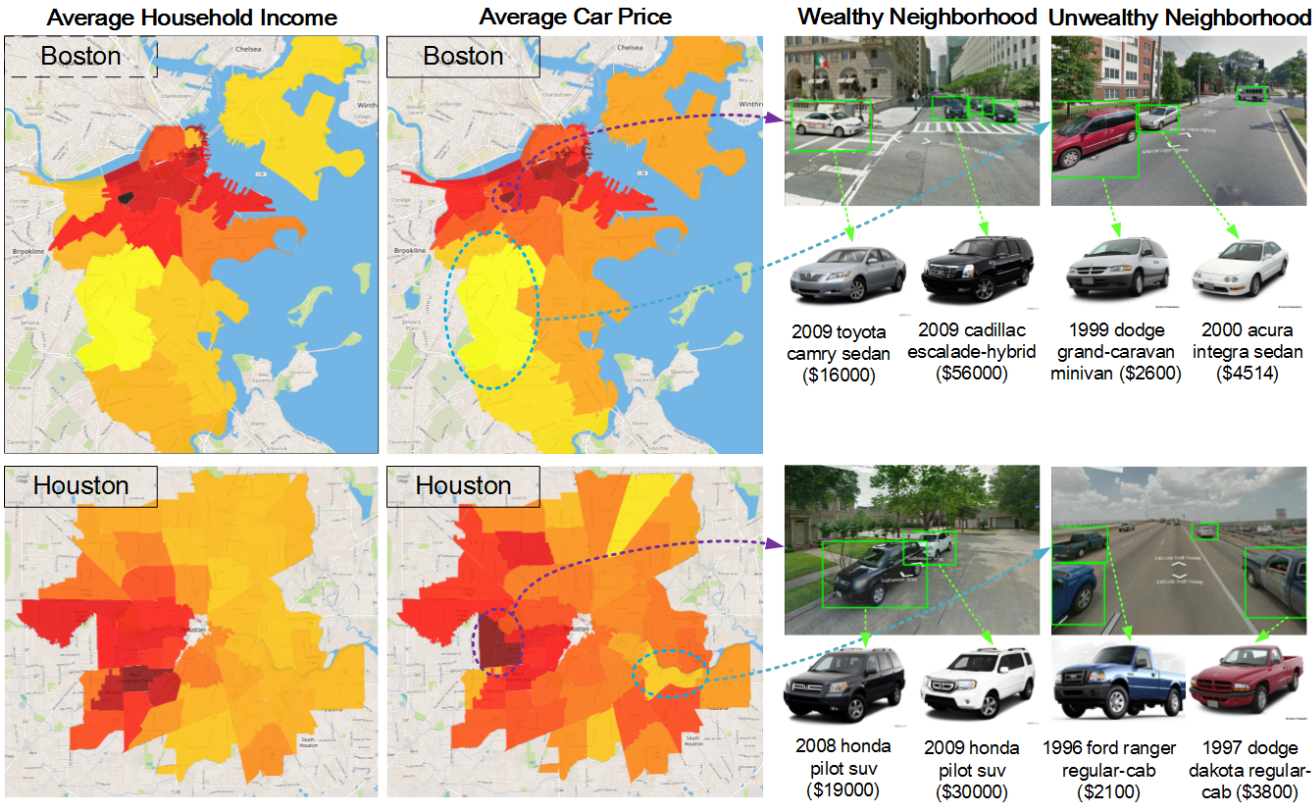

Figure 9. Left: Heat map of average household income per zip code obtained from census data. Center: Heat map of average car price per zip code detected from our data. Right: Example of Street View images and car detections corresponding to zip codes with high and low incomes respectively. We detect many cheap cars in low income areas and expensive cars in high income areas.

How does education relate to cars on the street? As shown in Table 3 there is a high negative correlation between the number of people with only a high school education and the average price of a car in a zip code. As expected, we also found a high correlation between the number of college educated people in a zip code and the average car price. What is perhaps surprising is that although there is a large increase in correlation coefficient from high school to college educated, the jump from college to graduate school is very low. Thus, there is a negligable difference in the price of cars driven by people who hold bachelors vs. graduate degrees. From the analysis conducted in this

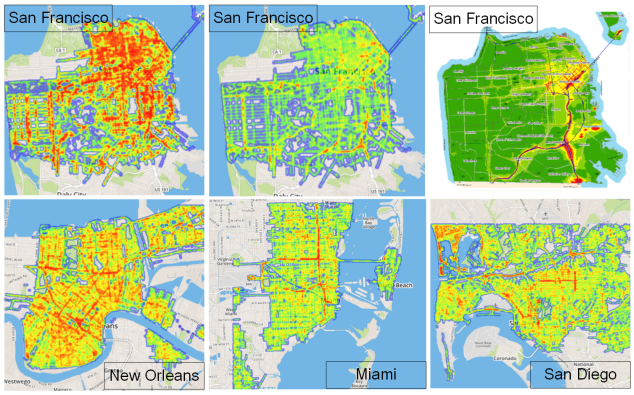

Figure 10. A. Density of cars in San Francisco inversely weighted by their expected miles per gallon (MPG). B. The inverse of weighted average of car MPG in San Francisco where the weights are the expected number of cars in a GPS point. C. Ground Truth for Air quality (measured in annual particulate matter) in San Francisco from [3]. D-F inverse weighted average of car MPG for three additional cities.

section as well as additional ones presented in our supplementary material, it is evident that many important demographic census variables can be predicted by zip code level car attributes obtained through fine-grained car detection.

5.2. What cars on the street tell us about neighborhoods

In this section we show that in addition to predicting individual level demographic information, we can use Google Street View car detections to learn about the characteristics of neighborhoods.

Which neighborhoods are wealthy/poor? Intuitively, seeing many expensive cars on the street indicates that we are in a rich neighborhood and vice versa. Figure 9 shows a heat map of the average price of detected cars within a zip code for Boston and Houston. We can see that in both cities, the expected car price is a good indicator of neighborhood wealth.

Which neighborhoods have high car pollution? We investigate the possibility of predicting neighborhoods with high pollution using our car detection results. To do this we plotted a heat map of the expected number of detected cars per sample inversely weighted by the expected MPG of that sample. We also map the inverse of the weighted average of car MPG where the weights are the expected number of cars. The first measure indicates the location of highly polluting neighborhoods: for the same density of cars, areas with high MPG result in lower numbers than those with low MPG. For different densities of cars, the relative magnitude of the measure depends on both the density of cars and how efficient they are. The second measure, on the other hand, visualizes areas with a high concentration of low MPG cars. Fig. 10A and B show a heat map of the two pollution measures respectively. Although we could not find ground truth data for car pollution, Fig. 10C is a map of San Francisco air quality measuring annual average particulate matter concentration (MPG) from all sources [3]. To our surprise, their map seems to agree with Fig. 10B in most cases. Fig. 10 D-F show show our predicted pollution levels for three other cities. As expected, the highest values occur near areas with high congestion such as freeways.

5.3. What cars on the street tell us about cities

Our final analysis shows that we can use fine-grained car detection to compare city attributes such as level of segregation, and wealth.

Which cities have more income based segregation? We answer this question by finding cities which exhibit a high spatial clustering among similar car prices. Given the high correlation between median household income and average car price, we expect the most segregated cities to exhibit high clustering. Following the analysis of [24] we use the Moran I statistic to measure spatial autocorrelation where a value of 1 indicates perfect clustering of similar values, -1 indicates perfect dispersion and 0 indicates a random spatial arrangement (neither clustering nor dispersion). Fig. 11 plots the highest and lowest scoring cities as well as a few others in between. We can see that Reno shows the highest clustering where as Dover shows the lowest. An interesting observation is that expensive cars are more clustered together than cheap ones.

Which cities are more patriotic? We find the most and least patriotic cities in the US by comparing the percentage of foreign vs. domestic cars in each city. As Fig. 13A shows the coastal cities have a high concentration of foreign made cars where as the midwest has a low concentration. This result agrees with [18] who measured the ratio of American/foreign cars driven in the US. The city with the highest percentage of foreign cars was found to be San Francisco with 61% of foreign cars where as Casper Wyoming had the least percentage (21%). The most popular car make in San Francisco was found to be a Toyota Prius, which is a result consistent with conventional wisdom.

Which cities are wealthier? The high correlation between average car price and median household income indicates that the expected car price of a city can be a good predictor of t's wealth. Fig. 13B maps the average car price for each city. The map shows that many of the east coast cities have expensive cars as well as some cities in the south

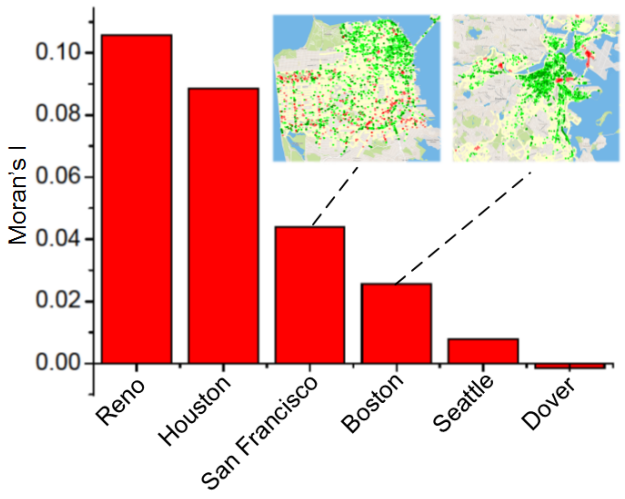

Figure 11. Moran I scores for car prices in different cities. The highest and lowest scoring cities are shown as well as 5 cities with scores in between. Reno, NV exhibits the most amount of clustering by car price while Dover, CO exhibits very little clustering. The two maps for San Francisco and Boston show areas of high clustering where green indicates statistically significant clustering of expensive cars and red indicates statistically significant clustering of cheap cars.

Attribute	Census Variable	Acc. Gain
Price	Median hh income	0.33%
Year	Median hh income	0.33%
Make	# ppl in management	0.30%
Submodel	Median hh income	0.30%
Domestic	# ppl in management	0.27%
Country	# ppl in management	0.27%

Table 4. Census variables resulting in the highest accuracy gain for each car attribute. "hh" is shorthand for house hold and "ppl" is shorthand for people. Using median household income as a prior for the car price or year results in the highest gain in accuracy.

such as Atlanta. We found the city with the most expensive cars to be New York with an expected price of ∼$12k and the one with the least expensive cars to be El Paso (∼$7k).

6. Using social priors to improve classification

As shown in sec:social fine-grained car detection results can be good predictors of demographic variables. In this section, we ask the reverse question: can knowledge of demographic variables help improve fine-grained car classification accuracy? Intuitively, if a particular image was taken in a wealthy neighborhood, for example, one would expect the cars in that neighborhood to be expensive. In order to use zip code level census variables as priors, we first calculated $P(C|I,S)$ where C is the fine-grained class, I is

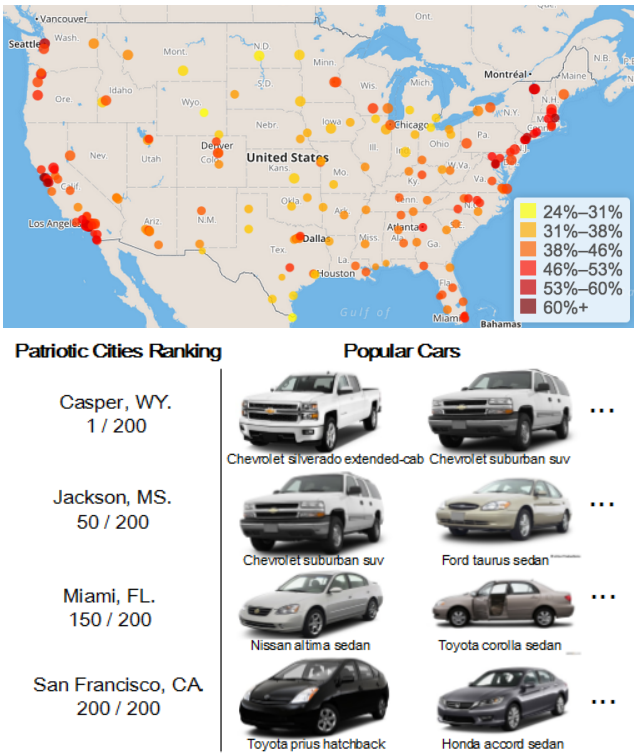

Figure 12. Map of the percentage of foreign cars in each American city. San Francisco, CA has the highest percentage with 61% and Casper, WY the lowest with 21%. The most popular cars in cities with the highest and lowest percentages as well as 2 others inbetween are listed.

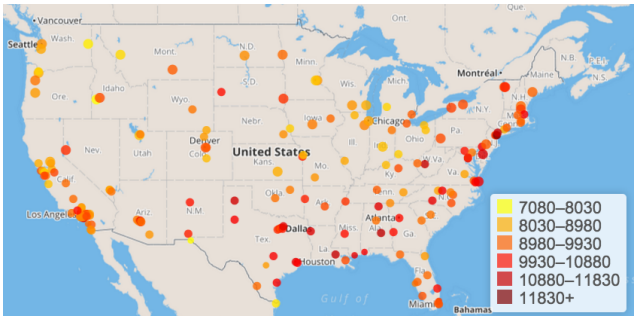

Figure 13. Map of the expected car price in each American city. New York, NY has the highest expected car price ($11,829) while El Paso,TX has the lowest ($7,079).

an image and $S \in \{S_1 \ldots S_n\}$is a particular zip code level census variable such as median household income. After applying Bayes' rule, assuming that the image and census data are conditionally independent given the class label, and applying Bayes' rule again we get:

$$P(C|I,S) \propto \frac{P(C|I)}{P(C)} P(C|S) \quad (1)$$

Where $P(C|I)$ is the output of our CNN classifier. A naive way of using census variables such as the one above reduces accuracy by ∼5% to 26.1%. This is not surprising given that we have over 2600 fine-grained classes, and many of them have similar attributes such as price. Thus, instead of using census variables directly as fine-grained class priors, we use them as priors for the car attributes. In order to incorporate these variables into our classification pipeline we can reformulate $P(C|S)$ in equation 1 as $P(C|A)P(A|S)$ where $A \in \{A_1 \dots A_n\}$ represents a car attribute such as price. After this modification, equation 1 can be written as

$$P(C|I,S) \propto \frac{P(C|I)}{P(C)} P(C|A)P(A|S) \qquad (2)$$

We calculate $P(C|I,S)$ for all car attributes and 30 different census variables, quantizing some car attributes and census variables such as car price and median household income into bins ranging from 2-20. Table 4 shows the highest accuracy numbers for various combinations of census variables and car attributes. It can be seen that using median household income and either car price or year give the highest accuracy gain. This result is to be expected since, as seen in section 5, there is a high correlation between median household income and car price and year. Although the accuracy gain at the fine-grained level is very slight (which is to be expected due to the large number of similar classes in our dataset). Our supplemental material discusses more extensive experiments and results in using census variables to improve our fine-grained classifications.

7. Conclusion

By analyzing car detections from 45 million images across 200 cities, we have shown that cars detected from Google Street View images contain predictive information about our neighborhoods, cities and their demographic makeup. To facilitate this work, we have collected the largest and most challenging fine-grained dataset reported to date. For our future work we plan to perform more extensive social analysis—such as crime prediction and predicting changes in cities across time—using fine-grained detection. We also hope to further explore the use of demographic data to assist in fine-grained detection.

References

[1] S. Ardeshir, A. R. Zamir, A. Torroella, and M. Shah. Gis-assisted object detection and geospatial localization. In *Computer Vision–ECCV 2014*, pages 602–617. Springer, 2014. 2

[2] H. Azizpour and I. Laptev. Object detection using strongly-supervised deformable part models. In *Computer Vision–ECCV 2012*, pages 836–849. Springer, 2012. 2

[3] S. Bay Area Air Quality Management District. Average annual pm 2.5 concentration from all sources@ONLINE, June 2010. 7

[4] T. Berg, J. Liu, S. W. Lee, M. L. Alexander, D. W. Jacobs, and P. N. Belhumeur. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2019–2026. IEEE, 2014. 2

[5] S. Branson, G. Van Horn, P. Perona, and S. Belongie. Improved bird species recognition using pose normalized deep convolutional nets. In *British Machine Vision Conference*, 2014. 2

[6] R. Chase. Does everyone in america own a car? @ONLINE, June 2009. 1

[7] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec. How community feedback shapes user behavior. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014. 1

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 2

[9] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes paris look like paris? *ACM Trans. Graph.*, 31(4):101, 2012. 2

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010. 4

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint arXiv:1311.2524*, 2013. 2, 4

[12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014. 4

[13] A. Khosla, B. An, J. J. Lim, and A. Torralba. Looking beyond the visible scene. 1, 2

[14] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011. 2

[15] J. Krause, J. Deng, M. Stark, and L. Fei-Fei. Collecting a large-scale dataset of fine-grained cars. 2

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 4

[17] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361, 1995. 2

[18] J. Levine. What do you drive? patterns in us car dealerships @ONLINE, June 2014. 7

[19] K. Matzen and N. Snavely. Nyc3dcars: A dataset of 3d vehicles in geographic context. In *Computer Vision (ICCV), 2013*

IEEE International Conference on, pages 761–768. IEEE, 2013. 2

[20] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, et al. Quantitative analysis of culture using millions of digitized books. *science*, 331(6014):176–182, 2011. 1

[21] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore–predicting the perceived safety of one million streetscapes. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 793–799. IEEE, 2014. 1, 2

[22] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001. 2

[23] V. Ordonez and T. L. Berg. Learning high-level judgments of urban perception. In *Computer Vision–ECCV 2014*, pages 494–510. Springer, 2014. 1, 2

[24] P. Salesses, K. Schechtner, and C. A. Hidalgo. The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, 8(7):e68400, 2013. 2, 7

[25] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2, 3

[26] R. West, H. S. Paskov, J. Leskovec, and C. Potts. Exploiting social network structure for person-to-person sentiment analysis. *arXiv preprint arXiv:1409.2450*, 2014. 1

[27] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Part-based r-cnns for fine-grained category detection. In *Computer Vision–ECCV 2014*, pages 834–849. Springer, 2014. 2, 4

[28] B. Zhou, L. Liu, A. Oliva, and A. Torralba. Recognizing city identity via attribute analysis of geo-tagged images. In *Computer Vision–ECCV 2014*, pages 519–534. Springer, 2014. 1, 2