

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011

## 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 2500 classes of cars comprised of images from google street view and other web sources and classified by car experts to account for even the most subtle of visual differences. We hope our work ushers in a new research area fusing fine-grained object detection and societal analysis.

## 1. Introduction

The artifacts with which we choose to surround ourselves 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. From a single picture such as fig. 1a of an individual with their car, we can guess that the person is rich because they own a tesla, and that they probably care about the environment. Similarly, it is evident that figs. 1 b and c are of poorer neighborhoods where as figs. 11 and b are that of wealthy ones. 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 big 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 [24] [5]. For example,

054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

Visual Census Using Cars

Anonymous CVPR submission

Paper ID \*\*\*\*



Figure 1. Cars on the street provide information about people and their neighborhoods. Without location information, we can guess that images from the first column are from wealthy areas where as those from the second column are from poorer neighborhoods.

recent work from [19] analyzed  $\sim 1M$  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 Torralba et al, Berg et al, and Hedalgo et al have recently started to apply visual scene analysis techniques to infer characteristics of neighborhoods and cities [26] [13] [20] [21]. 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 [4], and as seen in fig. 1 cars give a lot of information about individuals as well as neighborhoods.

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  $\sim 700k$  images from websites such as edmunds.com, cars.com, craigslist.com and street view (fig. 2). Our testing dataset for social analysis consists of more than 45 million google street view images from 200 of the largest American cities and 2519 zip codes



Figure 2. Left: examples of cars from craigslist, cars.com and edmunds.com, Right: examples of cars from streetview images. Cars from streetview images are blurred and occluded.

where an average of  $\sim 18k$  images is collected for each zip code.

Second, our paper offers a suite of interesting social analysis of American people and cities using a large scale fine-grained object detection system. This work is only possible because of the tremendous progress in object recognition: particularly object detection and fine-grained object recognition. We use DPM [8] [10] to detect cars due to scalability concerns and train a CNN [16] to classify them into one of 2567 classes. Using the result of our detections, we are able to offer insights into a number of very interesting questions related to our lives, communities and cities. Section 5 shows a series of analyses about how the types of cars we drive relate to our level of education, wealth and the neighborhoods and cities we live in and 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 [22] [21] [7] [26] [20] [13]. [22] created scores for perceptions of wealth, safety and uniqueness by asking people to rate images from 3 cities on a scale of 1–10 and [20] [21] predicted these scores using various global image features such as GIST and CNN. In another line of work analyzing cities, [26] perform city identity recognition after representing each city with higher level attributes. [7] 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 [25] [15] show impressive results by using part annotated datasets such as [23] [14] [15] and augmenting state of the art object detection algorithms such as RCNN [9]. However, it is difficult to evaluate fine-grained classification accuracy since there are no fine-grained datasets that match object classification datasets like imagenet [6] in the number of classes or images. Recent works such as [3] have introduced larger scale fine-grained datasets and [18] 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, [3] learns a spatio-temporal prior to improve bird classification and [18] uses some location information such as elevation to assist in car classification.

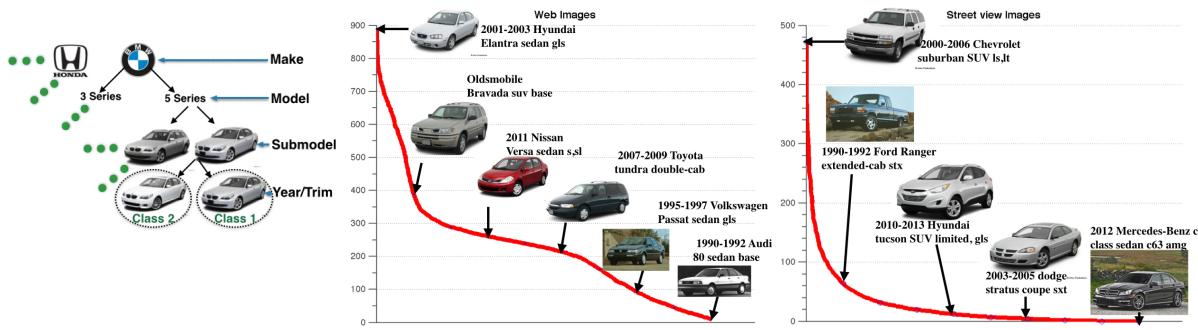
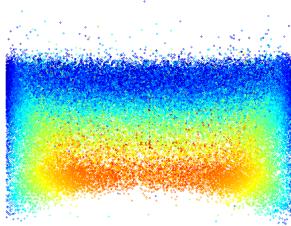
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226

Figure 3. Left: A hierarchy of car classes where images become more difficult to distinguish while travelling down the tree. The leaves (e.g. class 1 and class 2) are the fine-grained classes. Center: Class image distribution for web images. We have the most number of images for 2001-2003 Hyundai Elantra sedan glx and the least for 1990-1992 Audi 80 sedan base. Right: class image distribution for street view images.

231

232

## BBox Area



## BBox Position

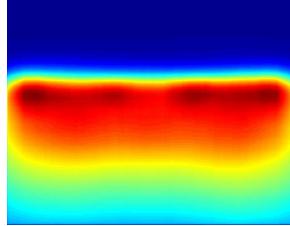


Figure 4. Left: Location/scale heatmap of streetview bounding boxes with red indicating larger sizes and blue smaller ones. Each point indicates the size of a bounding box centered at that point in a street view image.

Figure 5. Right: Heatmap of streetview bounding box locations. Most cars are located in a line slightly below the center of the image.

249

250  
251  
252  
253  
254  
255  
256  
257

Attribute	Train	Val	Test
No. streetview bboxes	34,712	6,915	27,865
No. web bboxes	313,099	-	-
No. web images	313,099	-	-
No. streetview images	199,666	39,933	159,732
No. Total images	512,765	39,933	159,732
No. classes	2657	-	-
No. car types	18017	-	-

Table 1. Street view data statistics

261  
262

## 3. Cars and Cities Dataset

263  
264  
265  
266  
267  
268  
269

We first 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 edmunds.com (all cars manufactured after 1990). To assemble our car dataset we obtained images of all cars in edmunds.com ( $\sim 18k$  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  $\sim 70k$  bounding boxes of cars from the street view images with fine-grained labels. We also labeled cars from craigslist and cars.com by parsing the car posting titles.

Table 1 summarizes the statistics of our dataset and fig. 2a shows some examples. We can see that cars from street view images are generally occluded and blurry where as those from other web sources have higher resolution. As seen in fig. 5, the street view bounding boxes are distributed in a line around the center of the image and bounding boxes near the center also occupy a larger space. Fig. 3a, shows a hierarchy of classes in our dataset where classes become increasingly visually indistinguishable while traveling down the tree. As the figure shows, the difference between leaf classes such as class 1 and class 2 is extremely subtle. Finally fig. 3a and b show the distribution of street view and web images for different classes.

## 4. Detecting and classifying cars

Although RCNN based fine-grained detection algorithms have reported state of the art results, [9] [25], 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  $\sim 20$ s per image using a machine with a single GPU. Thus our pipeline, instead, consists of using DPM [8] to detect cars and a CNN [16] to classify them. We present details in the sections below.

270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323

324  
325  
326  
327  
328  
329  
330  
331

Attribute	Acc %
Make	66.38
Model	51.83
Submodel	77.74
Price (5 bins)	61.61
Domestic/Foreign	87.71
Country	84.21

Table 2. Accuracy by car attribute

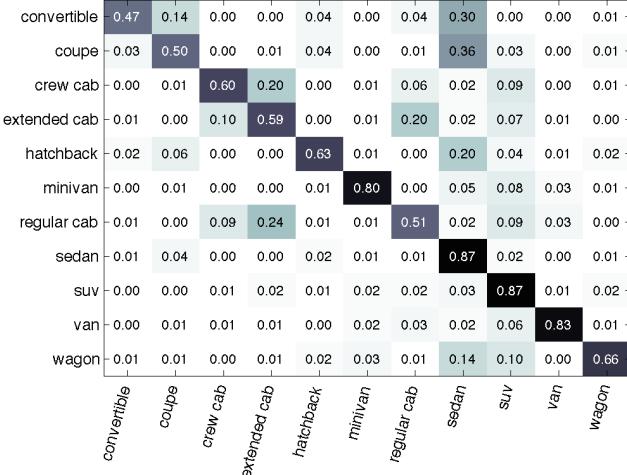


Figure 6. Confusion matrix for car submodel. As expected, the confusion is between similar types of submodels such as a sedan and coupe or a extended cab and a crew-cab.

## 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 [11] 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

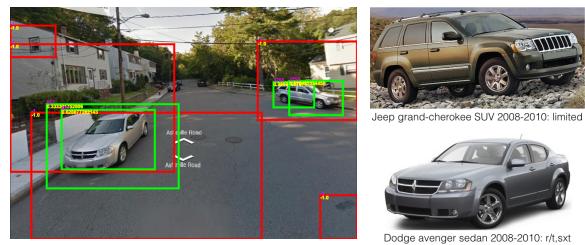


Figure 7. An example of our car detections and classifications. The red boxes are DPM bounding boxes that are not classified by our CNN and green boxes are boxes classified by the CNN. The two cars in the image are correctly classified to the classes shown on the right.

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 33.1% on the true positive DPM bounding boxes and 31.27% on the ground truth bounding boxes. Fig. ?? 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. 6 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 relationship 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. Societal analysis

In this section, we present societal analysis results from all 200 cities as well as case studies from those with available ground truth data. We divide our analysis into different sections below.

378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431

Census Variable	Car Attribute	r
Household income	No. 1990-1994 cars	-0.42
Household income	No. 1995-1999 cars	-0.40
Household income	No. 2000-2004 cars	0.21
Household income	No. 2005-2009 cars	0.46
Household income	% foreign cars	0.59
Household income	% German cars	0.57
Household income	% 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  $\sim 0$ .

## 5.1. What cars on the street tell us about people

### 5.1.1 Sanity Check: Cars on the street correlate well with registered cars

How do cars on the street relate to the cars that people drive? Specifically, can we learn about the registered cars in a zip code from our street view detections? We downloaded vehicle census data from Massachusetts, which is the only state to release extensive vehicle registration data, and found an extremely high Pearson correlation coefficient of 0.9 ( $p \sim 0$ ) between the number of cars we detected per zip code and the number of cars that are registered. The high correlation was only obtained after aggregating cars at the zip code level which shows that most people in MA drive within their zip code. After establishing a high correlation between the number of cars we detect and the number of registered cars, we also measured the correlation between the make of the detected and registered cars per zip code. As we can see in fig. 9 there is a high correlation for most of the makes. Thus the cars we detect from street view images contain useful information about the types of cars driven by people in a particular zip code.

### 5.1.2 What do rich/poor people drive?

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 median household income as well as education level per zip code. Fig. 8 shows a plot of median household income vs. average car price in a zip code. As expected, there is a high correlation between median household income and the average car price per zip code ( $r=0.49$ ,  $p \sim 0$ ). Our results also indicate that rich people prefer to drive foreign, especially German,

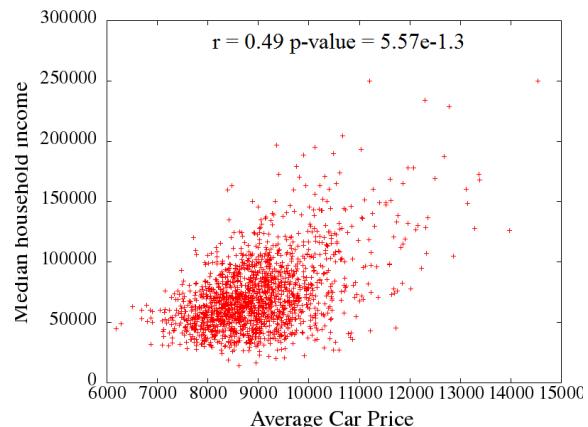


Figure 8. Scatter plot of median household income per zip code vs. average price of detected cars per zip code.

cars ( $r=0.59$ ). What is perhaps surprising is that there is a very high negative correlation ( $r=-0.55$ ,  $p \sim 0$ ) between the percentage of American cars in a zip code and median household income. So poor people live in places with many American cars.

Poor people also live near very old cars where as rich people live near newer ones. As table 3 shows, the correlation between median household income and the number of cars in 1990-1994 is very negative and increases to a high positive 0.59 for cars in the 2005-2009 range. Finally, a perhaps not surprising result is that poor people live near cars with low miles per gallon (MPG). This corroborates [12] study showing that poor people are more exposed to car pollution than rich people.

### 5.1.3 How does education relate to cars on the street?

As shown in 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 negligible difference in the price of cars driven by people who hold bachelors vs. graduate degrees.

## 5.2. What cars on the street tell us about neighborhoods

### 5.2.1 Which neighborhoods are wealthy/poor?

Intuitively, seeing many expensive cars on the street indicate that we are in a rich neighborhood and vice versa. Figure 10 A shows a heat map of the average price of detected

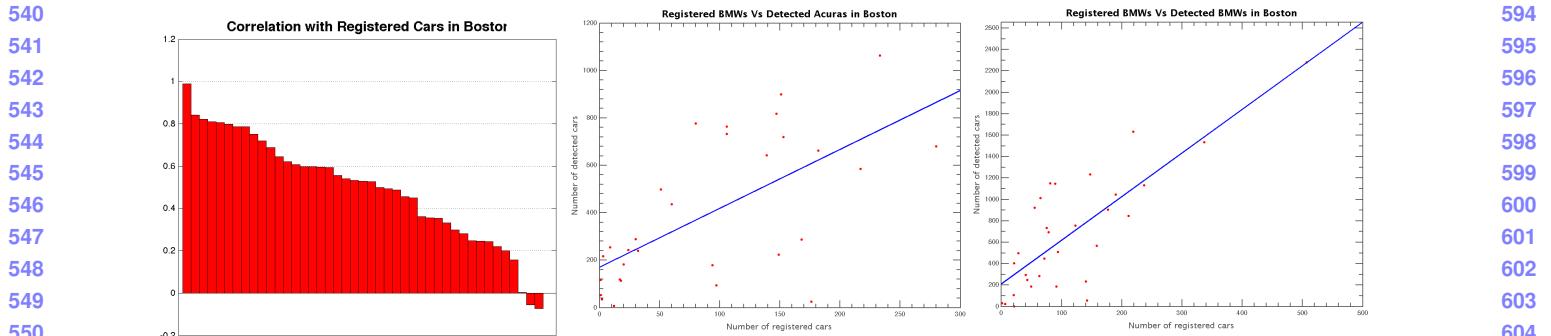


Figure 9. Left: Pearson correlation coefficient and p values between the number of detected and registered cars in Boston for each make. Center: scatter plot of detected cars per zip code vs. registered cars per zip code in Boston. Right: scatter plot of No. of detected Honda's per zip code vs. registered Honda's per zip code in Boston.

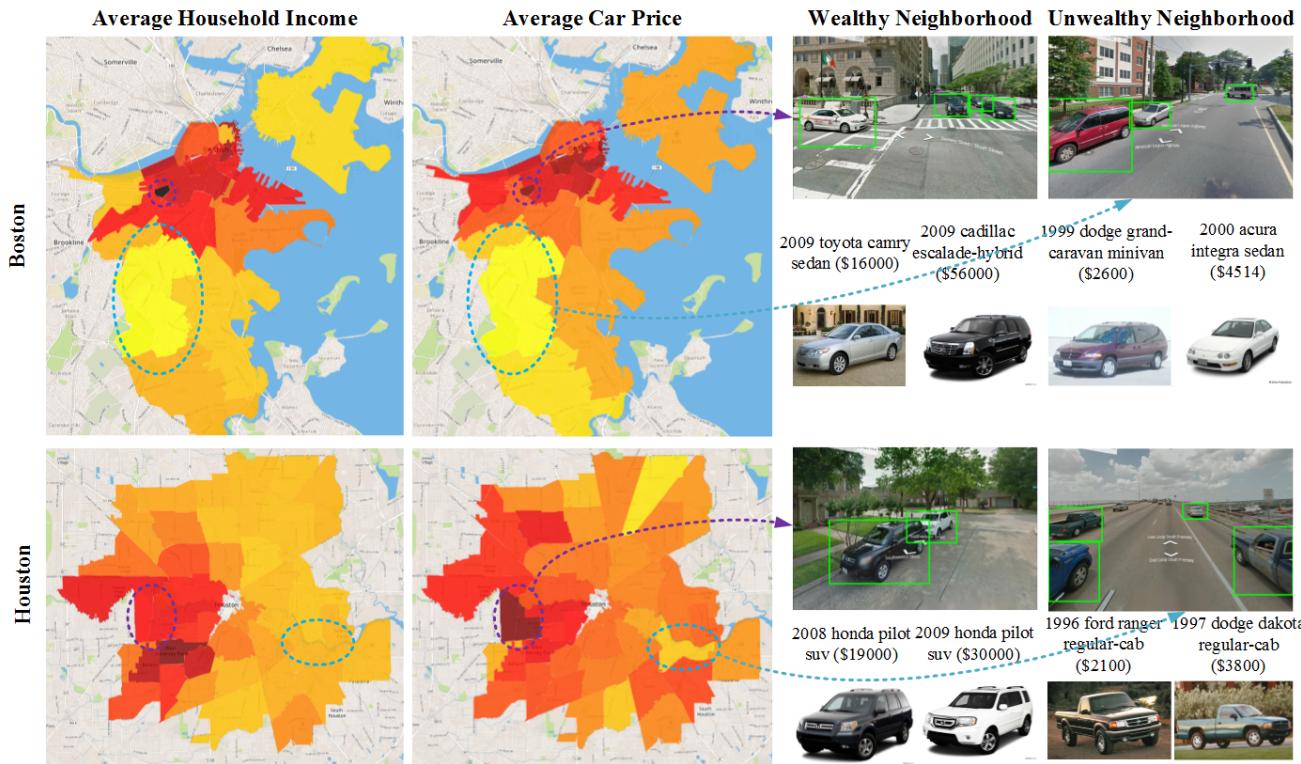


Figure 10. Top left: Heat map of average car price in Boston, center: heat map of median household income in Boston. Bottom left: Heat map of average car price in Houston, center: heat map of median household income in Houston. Left example of street view images and car detections corresponding to zip codes with high and low incomes. Low income areas have many cheap cars where as high income areas have expensive ones.

cars within a zip code and median household income in a zip code for Boston and figure 10 B shows the same visualization for Houston. We can see that in both cities, the average car price in a zip code is a very good predictor of wealthy/un-wealthy neighborhoods.

### 5.2.2 Which neighborhoods have high car pollution?

Can our street view detections tell us anything about which neighborhoods are affected by highly polluting cars? We plotted a heat map of the expected number of 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.

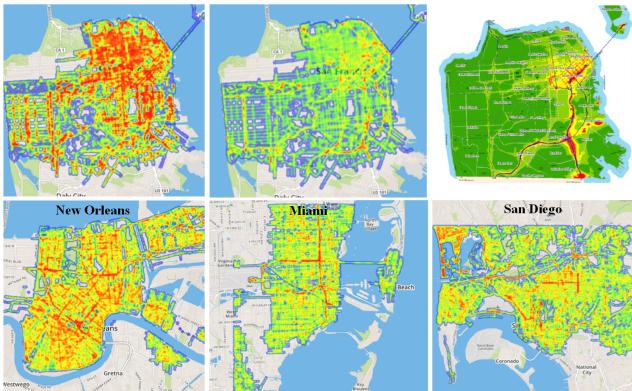


Figure 11. A. Density of cars in San francisco inversely weighted by their expected miles per gallon (MPG), B. The inverse of weighted average of car MPG in San Francisco. The weights are the expected number of cars in a GPS point, C. Ground Truth for Air quality (measured in annual particulate matter) in San Francisco from [2].

The first measure should give us a rough idea of 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. 11A shows the weighted density of cars in San Francisco and B shows the inverse weighted MPG. Although we could not find ground truth data of car pollution, Fig. 11C is a map of San Francisco air quality measuring annual average particulate matter concentration (MPG) from all sources [2]. To our surprise, their map seems to agree with Fig. 11B in most cases.

### 5.3. What cars on the street tell us about cities

#### 5.3.1 Which cities are more segregated?

Which cities show high clustering of similarly priced cars? Specifically, which cities have expensive cars clustered together with other expensive cars and cheap cars clustered with other cheap cars? Given the high correlation between median household income and average car price, the answer to this question should give us a good indication of the cities that are most and least segregated. Following the analysis of [22] 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. 12 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

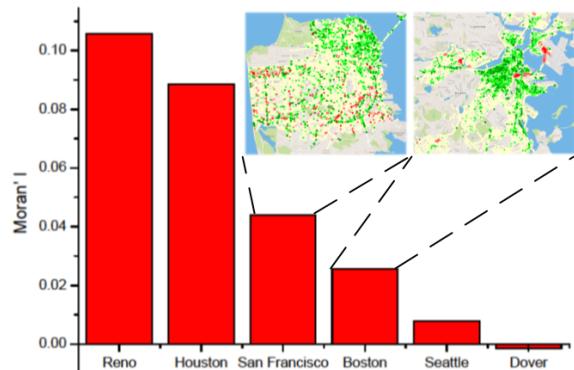


Figure 12. Moran I scores for car prices. The highest and lowest scoring cities are shown as well as 5 cities with scores in between. Reno exhibits the most amount of clustering by car price where as Dover exhibits very little clustering. The two maps for Boston and Dover show areas of high clustering where green indicates statistically significant clustering of expensive cars and red indicates statistically significant clustering of cheap cars.

shows the lowest.

#### 5.3.2 Which cities are more patriotic?

Which cities have the most number of domestic cars? As Fig. 14A shows the coastal cities have a high concentration of foreign made cars where as the midwest has a low concentration. This result agrees with [17] 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%).

#### 5.3.3 Which cities are wealthier?

Which city has the most expensive cars on average? Fig. 14B 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 such as Atlanta. We found the city with the most expensive cars to be New York with an expected price of  $\sim \$12k$  and the one with the least expensive cars to be El Paso ( $\sim \$7k$ ).

## 6. Using social priors to improve classification

This section explores the use of demographic census information as a prior to improve fine-grained detection and presents preliminary results run on our validation set. Intuitively, if a particular image was taken in a wealthy neighborhood, for example, one would expect the cars in that neighborhood to be expensive and viceversa. 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

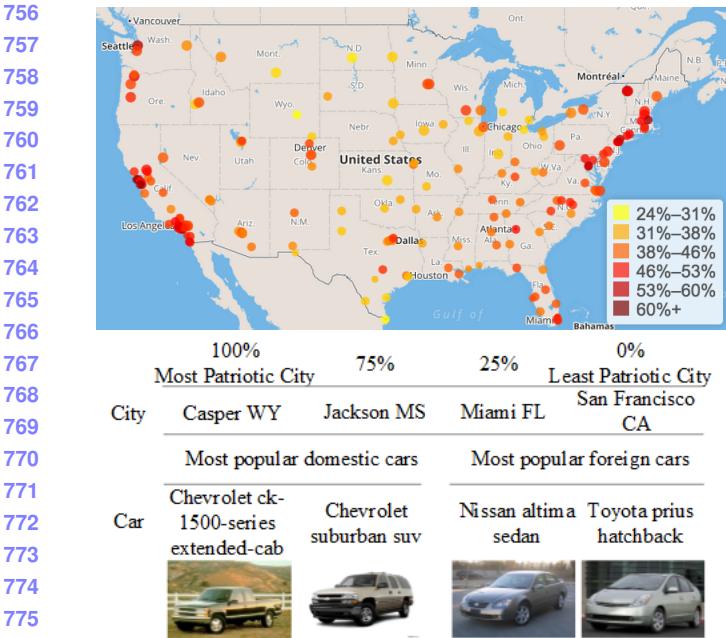


Figure 13. Map of the percentage of foreign cars in each city. San Francisco has the highest percentage with (61% and Casper the lowest with 21%.

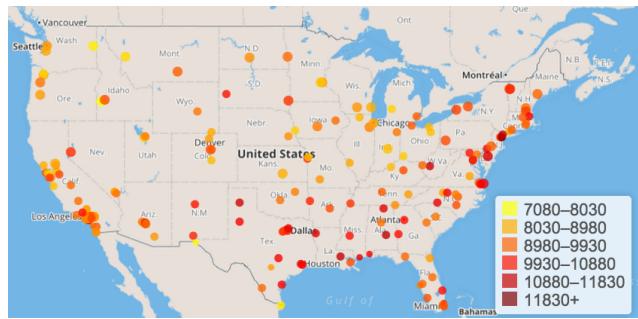


Figure 14. Map of the United States showing the expected car price in each city. New York has the highest expected car price where as El Paso has the lowest.

Attribute	Census Variable	Acc%
Price	Median hh income	31.60
Year	Median hh income	31.60
Make	No. ppl in management	31.57
Submodel	Median hh income	31.57
Domestic	No. ppl in management	31.54
Country	No. ppl in management	31.54

Table 4. Census variables resulting in the highest accuracy gain for each car attribute. Car price/year with median household income give the highest gain in accuracy.

an image and  $S \in \{S_1 \dots S_n\}$  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  $\sim 5\%$  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

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

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

In conclusion, by analyzing car detections from 45 million images across 200 cities, we have shown that cars detected from 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.

864

## References

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

- [1] S. Ardesir, 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] S. Bay Area Air Quality Management District. Average annual pm 2.5 concentration from all sources@ONLINE, June 2010. 7
- [3] 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
- [4] R. Chase. Does everyone in america own a car? @ONLINE, June 2009. 1
- [5] 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
- [6] 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
- [7] 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
- [8] 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. 2, 3
- [9] 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, 3
- [10] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. <http://people.cs.uchicago.edu/rbg/latent-release5/>. 2
- [11] 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
- [12] C. Katz. People in poor neighborhoods breathe more hazardous particles @ONLINE, June 2012. 5
- [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. 2, 3, 4
- [17] J. Levine. What do you drive? patterns in us car dealerships @ONLINE, June 2014. 7
- [18] 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
- [19] 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
- [20] 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
- [21] 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
- [22] 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
- [23] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2
- [24] 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
- [25] 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, 3
- [26] 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

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971