

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 and education levels. Finally we show that just as fine-grained classes provide demographic information, societal cues can assist in fine-grained classification and improve accuracy. To facilitate our work, we have collected the largest and most challenging fine-grained dataset reported to date consisting of 3147 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. 1 of an individual with their car, we can guess that the person is rich because they own a tesla, that they probably care about the environment and live in a mostly liberal and wealthy area such as San Francisco. Traditionally, the most prevalent method for gathering such personal information is through surveys: in 2010, the US government spent XXX on the 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

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. some pull figure

societal insights by analyzing massive user texts and social networks [?] [?]. Although there has been no work in computer vision matching the depth of analysis by works such as [?] [?], 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 [?] [?] [?] [5]. All of the computer vision techniques used in these works use global scene level features to analyze cities.

In this work, we are also interested in analyzing cities and neighborhoods and the demographic makeup of their inhabitants. However, our approach uses fine-grained object recognition as opposed to global scene features. In particular, we are interested in using cars as a lens into understanding our society. Fig. 1 shows that cars say a lot about us, and XX% of Americans own cars. Furthermore, due to the ubiquity of geotagged images such as street view, cars are more visible and easier to obtain images for compared to more personal items. Thus, to facilitate our work, we gathered a dataset for 3147 car types consisting of all cars produced after 1990. Our dataset consists of labeled cars for training car detection and classification models, mined from specialty websites such as edmunds.com, cars.com and craigslist as well as street view images. In order to perform societal analysis, we obtained more than 45 million google street

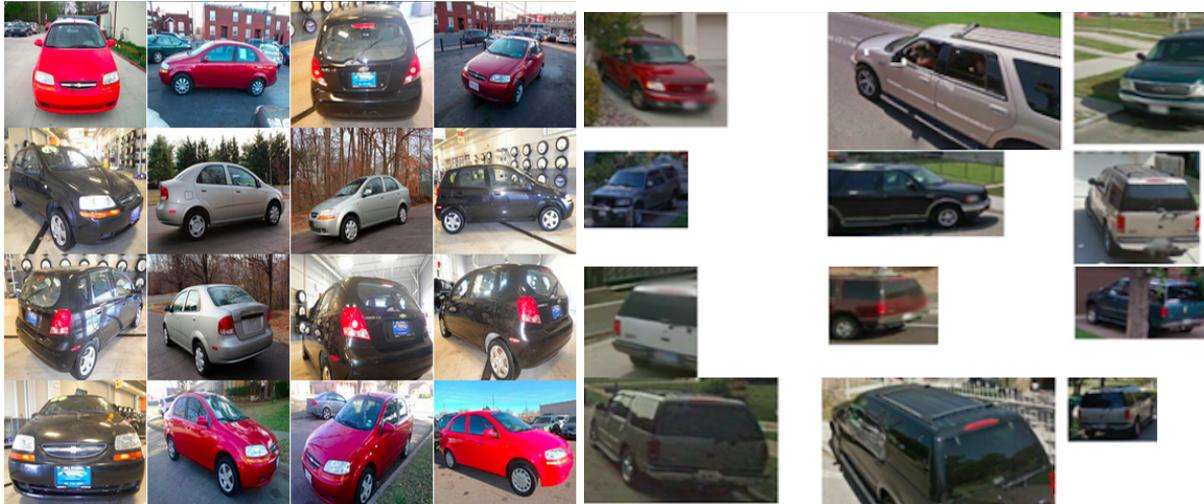


Figure 2. Left: examples of cars from craigslist, cars.com and edmunds.com, Right: examples of cars from streetview images. Cars from streetview images are blurred and occluded.

view images from 200 American cities. One might ask if street view car detections can tell as an accurate story of our neighborhoods and cities, since after all, these pictures were taken at different times, and cars are mobile objects. Our analysis in Section 5 shows that surprisingly, results using these detections correlated highly with demographic data obtained from ACS as well as government sources. This work is only possible because of the tremendous progress in object recognition: particularly object detection and fine-grained object recognition. We use DPM [?] to detect cars due to scalability concerns and train a CNN [?] to classify them into one of 2567 classes. Section 4 discusses the tradeoffs we considered for car detection and classification. 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. Finally, section 6 explores the use of social knowledge such as median household income in a zip code to improve fine-grained car classification.

1. We present the largest fine-grain car dataset reported to our knowledge.
2. We perform mass fine-grained detection across 200 cities
3. We perform a large scale social analysis of cities using street view car detections.

## 2. Related Work

**City analysis via image features.** There has been recent interest in using images to characterize

cities [6] [5] [?] [?] [?] [?]. [6] 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 [?] [5] predicted these scores using various global image features such as GIST and CNN. In another line of work analyzing cities, [?] perform city identity recognition after representing each city with higher level attributes. [?] identify unique qualities of cities such as Paris and Prague and [?] 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 where as [5], for example, showed that their method is not easily generalizable to other cities (e.g. training on Boston images does not give good predictions for San Francisco).

**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 [?] [?] [?] show impressive results by using part annotated datasets such as [?] [?] [?] and augmenting state of the art object detection algorithms such as RCNN [?]. However, it is difficult to evaluate fine-grained classification accuracy since there are no fine-grained datasets that match object classification datasets like imagenet [?] in the number of classes or images. Recent works such as [2] have introduced larger scale fine-grained datasets and [?] has introduced a 3D car dataset annotated with metadata such as location information. We introduce a geotagged

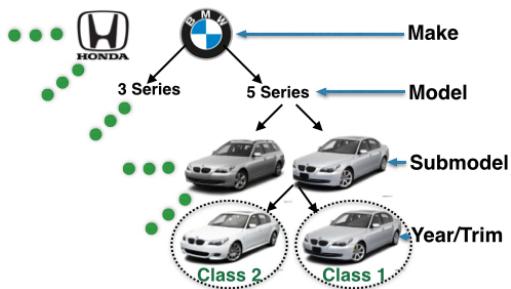


Figure 3. Car class hierarchy. Classes are usually more visually similar when we travel down the tree.

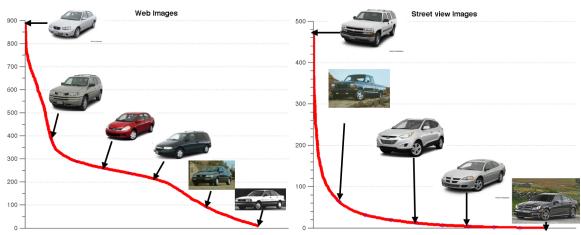


Figure 4. Distribution of images for web and street view classes.

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, [2] learns a spatio-temporal prior to improve bird classification and [?] uses some location information such as elevation to assist in car classification.

### 3. Data

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 gather 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 gathered 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). In order to gather this dataset, we first created a class list using edmunds.com by obtaining example images of all the cars they have (18017 cars) and grouped the cars into visually indistinguishable classes using a series of amazon mechanical turk tasks as well as manual labor by the authors. This resulted in 3147 visually indistinguishable classes. We then collected additional images

of cars from cars.com and craigslist and used AMT to obtain 313,099 bounding boxes of cars for these images and 543,641 bounding boxes for our street view images.

Finally, we hired car experts to label 69,492 bounding boxes of cars from the street view images, classifying them into one of the 3147 classes that we created. We also labeled cars from craigslist and cars.com by parsing the car posting titles. The final car dataset consists of 781,922 images where 468,823 are from street view and 313,099 are images from craigslist.com, cars.com and edmunds.com. Our training, validation and test sets roughly consist of 50%, 10% and 40% of our street view images and images from all other sources are also included in the training set. Although we are interested in classifying street view images, having car images from other sources in our training set increases the amount of data we have without paying expensive car experts to annotate street view images. Fig. 1 shows some examples of cars from our dataset. We can see that cars from street view images are generally occluded and blurry where as cars from other web sources have higher resolution and are not occluded.

Fig. 3 shows one of a few possible car hierarchies. The leaves are the fine-grained classes and the parents are composite classes consisting of the children. Cars become more and more visually indistinguishable while traveling down the tree, with the most similar classes, in most cases, being the leaves sharing a parent. As the figure shows, the difference between class 1 and class 2 is extremely subtle. Given the difficulty of our dataset, we were very surprised by the 31.27% accuracy of our classifier at the leaf level.

## 4. Detecting and classifying cars

Although RCNN based fine-grained detection algorithms have reported state of the art results, [?] [?] [?], its computational and memory requirements make it impractical for use in a large scale detection setting such as ours. Training an RCNN detector on our data would require  $\sim$ XXGB of memory and XXX secs and detecting cars would take  $\sim$ 20s per image and XXX days for  $\sim$ 45 million images. Thus our pipeline, instead, consists of using DPM [?] to detect cars and a CNN [?] 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 uses 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  $\sim$ 22 secs per image. Detailed plots and timing measurements of other

324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335

Attribute	Acc %
Make	X
Model	X
Submodel	X
Year (5 bins)	X
Price (5 bins)	X
Domestic/Foreign	X
Country	X

Table 1. Accuracy by car attribute

336  
337

DPMs we have trained are discussed in our supplementary material.

338  
339

## 4.2. Car Classification

340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358

We use a standard CNN from [?] with [?] 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. ?? 61% of our positive training images are composed of cars from other sources such as craigslist. Thus we add deformations such as blurring to make the non-street view images look like 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 0.X% drop in AP as compared to using all the boxes detected by DPM. We achieve an accuracy of 33.X% 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.

359  
360  
361

## 4.3. Analyzing Hierarchical Classification Accuracy

362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377

Some types of classification mistakes are more costly than others for the task of social analysis. For example, an error misclassifying 2001 Honda Accord Ix to 2001 Honda Accord dx is not significant. However, misclassifying a 2012 BMW 3-series to a 1996 Honda Accord 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. Similarly, misclassifying an SUV to a sedan would confound an analysis estimating the cities with the most number of large cars. In order to gain more insight into the types of errors our classifier makes, we measure the accuracy of classifying different car attributes. Table ?? lists accuracies by various car attributes such as those in fig. 3 and others like car price, year etc... We can see that the accuracy is much higher after aggregating by different attributes.

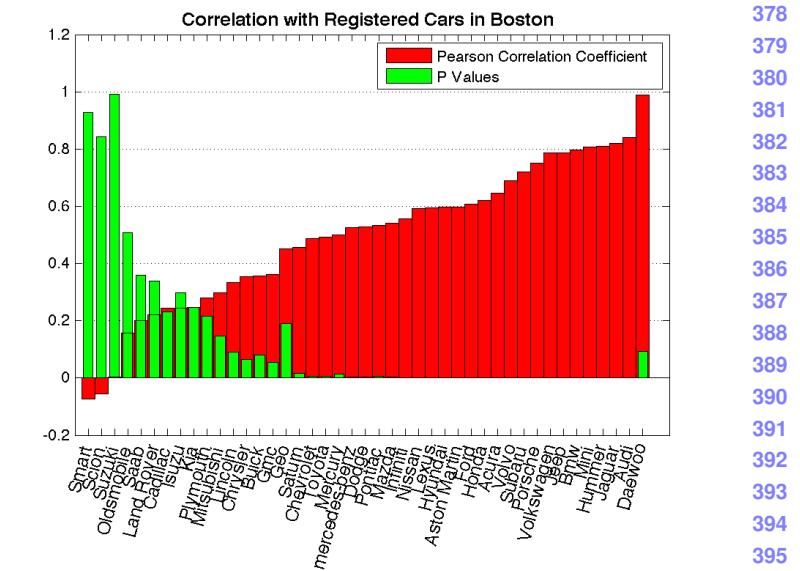


Figure 6. Pearson correlation coefficient and p values between the number of detected and registered cars in Boston for each make.

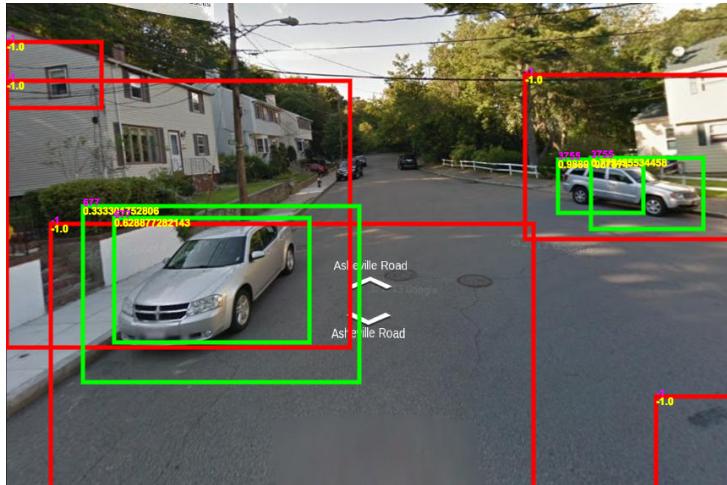
## 5. Societal analysis

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.

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

#### 5.1.1 Sanity Check

The first question we asked is 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 value=XXX) 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. 6 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.



Jeep grand-cherokee SUV 2008-2010: limited



Dodge avenger sedan 2008-2010: r/t,sxt

Figure 5. An example of our car detections and classifications. The red boxes are DPM bounding boxes that are not classified and green boxes are classifications. The two cars in the image are correctly classified to the classes shown on the right.

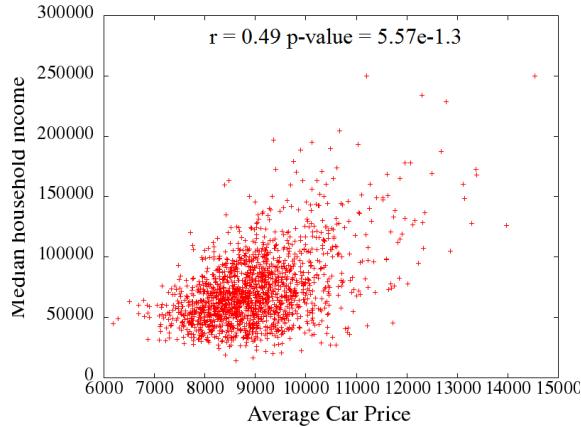


Figure 7. Median household income per zip code vs average price of detected cars per zip code.

### 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 2 shows correlation values between various attributes of the detected cars and median household income as well as education level per zip code. Fig. 7 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, 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 2 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 [?] 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 2 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 as we go from high school to college educated, the jump from college to graduate school is very low ( $r=0.31$  for college educated and 0.39 for graduate school). This tells us that there is a very low difference in the price of cars driven by people who only hold bachelors as opposed to graduate degrees.

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

### 5.2.1 Which neighborhoods are wealthy/poor?

We ask the question: what can our street view car detections tell us about the wealth of a neighborhood? Specifically, can

486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539

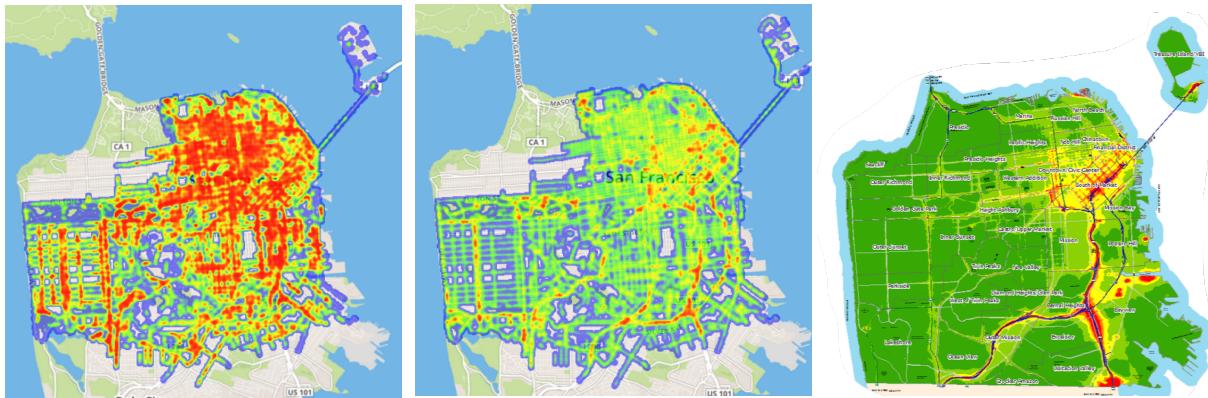


Figure 9. A. Density of cars in San francisco inversely weighted by their expected MPG, B. The weighted average of car MPG in San Francisco. The weights are the expected number of cars, C. Ground Truth for Air quality (measured in annual particulate matter) in San Francisco.

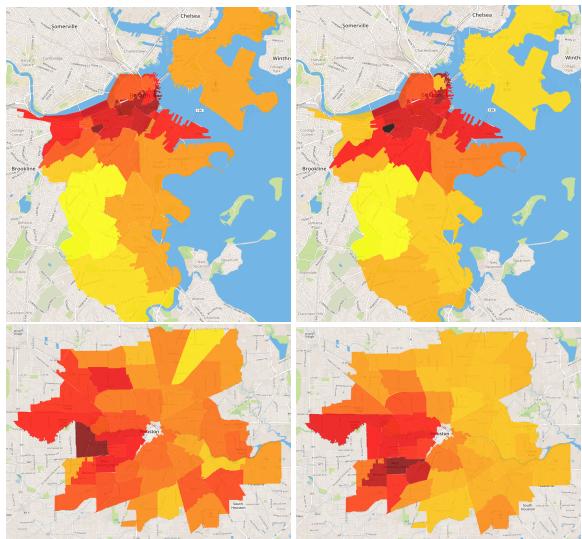


Figure 8. Top left: Heat map of average car price in Boston, right: heat map of median household income in Boston. Bottom left: Heat map of average car price in Houston, right: heat map of median household income in Houston

we predict which neighborhoods are wealthy or un-wealthy using our detections? Intuitively, if we see many expensive cars on the street, we suspect that we are in a rich neighborhood and vice versa. Figure 8 A shows a heat map of the average price of detected cars within a zip code and median household income in a zip code for Boston and figure 8 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.

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 2. Pearson correlation coefficient between various census variables and detected car attributes. All p values are  $\sim 0$ .

### 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? To answer this question, we plotted a heat map of the expected number of cars per sample inversely weighted by the expected MPG of that sample as well as the weighted average of car MPG where the weights are the expected number of cars. 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. 9A shows the weighted density of cars in San Francisco and B shows the weighted MPG. Although we could not find ground truth data of car pollution, Fig. 9C is a map of San Francisco air quality measuring annual average par-

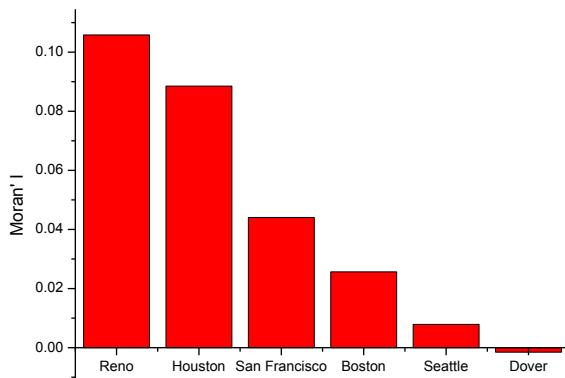


Figure 10. Moran I scores for car prices. The highest and lowest scoring cities are shown as well as 5 cities with scores in between. Reno exhibits the most amount of clustering by car price where as Dover exhibits no clustering.

ticulate matter concentration (MPG) from all sources. To our surprise, their map seems to agree with Fig. 9B in most cases.

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

#### 5.3.1 Which cities are more segregated?

In this section we ask the question: 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 [6] 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. 10 plots the highest and lowest scoring cities as well as well as a few others in between. We can see that Reno shows the highest clustering where as Dover shows the lowest.

#### 5.3.2 Which cities are more patriotic?

Which cities have the most number of domestic cars? As Fig. 11A shows the coastal cities have a high concentration of foreign made cars where as the midwest has a low concentration. This result is to be expected since the coasts also have a higher number of immigrants as well as a higher concentration of wealthy people. And as we showed in our previous analyses, wealthy people tend to drive foreign cars. The city with the highest percentage of foreign cars was found to be San Francisco where as Casper Wyoming had the least percentage.

Attribute	Acc %
Make	X
Model	X
Submodel	X
Year (5 bins)	X
Price (5 bins)	X
Domestic/Foreign	X
Country	X

Table 3. Gain in fine-grained accuracy with ground truth attributes.

Attribute	Census Variable	acc-fg	acc-m
Price	Median hh income	X	X
Year	Median hh income	X	X
Make	No. ppl in Management	X	X
Submodel	No. ppl 6-17 years	X	X
Domestic	Median hh income	X	X
Country	Median hh income	X	X

Table 4. Census variables resulting in the highest accuracy gain for each car attribute. Acc-fg is the fine-grained accuracy and acc-m is the make level accuracy. Car price/year with median household income give the highest gain in accuracy.

#### 5.3.3 Which cities are wealthier?

In our final analysis, we ask which city has the most expensive cars on average? Fig. 11B maps the average car price for each city. We can see that many of the east coast cities have expensive cars as well as some cities in the south like Atlanta. We found the city with the most expensive cars to be New York and the one with the least expensive cars to be El Paso. The fact that our results predict New York may not be surprising given that some of the wealthiest people in the United States live there and those who are less wealthy tend to take public transportation.

## 6. Using social priors to improve classification

In section 5 we showed that street view car detections give a lot of information about people, neighborhoods and cities. In this section we ask the reverse question: does our knowledge of neighborhood demographics teach us something about the cars in that neighborhood? Intuitively, if we know that a particular image was taken in a wealthy neighborhood, for example, we would expect the cars in that neighborhood to be expensive. Can we use this information to improve our car classifications?

To answer this question, we first tried to use zip code level census variables as priors, calculating  $P(C|I, Sk)$  where  $C$  is the fine-grained class,  $I$  is an image and  $Sk \in \{S1 \dots Sn\}$  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 condition-

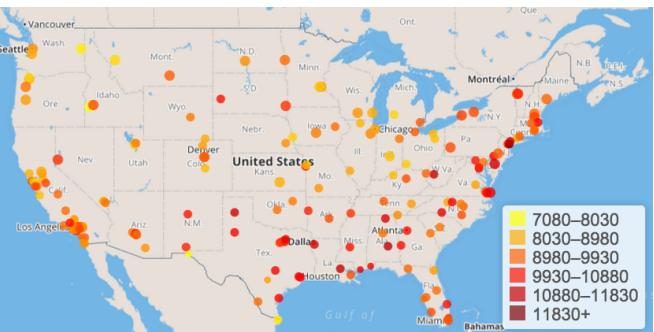
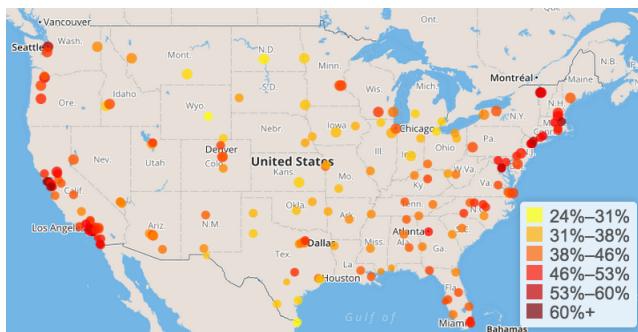


Figure 11. A. Map of the percentage of foreign cars in each city. San Francisco has the highest percentage and Casper the lowest. B. Map of the United States showing the expected car price in each city. New York has the highest expected car price where as El Paso has the lowest.

ally independent given the class label, and applying Bayes' rule again we get:

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

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.x% to 26.x%. This is not surprising given that we have over 2600 fine-grained classes, and many of them have similar attributes such as price. We expect census variables to be good indicators of cheap vs. expensive or old vs. new cars but not to give us much information about the presence/absence of cars with similar attributes at the fine-grained level such as 2001 Honda Accord lx vs. 2001 Honda Accord dx. With this insight, instead of using census variables directly as fine-grained class priors, we use them as priors for the car attributes. Table 3 lists fine-grained accuracy gains after plugging in ground truth attributes. These values give an upper bound for the gain in accuracy that can be achieved after perfectly predicting each attribute. With this in mind, to incorporate the census variables into our classification pipeline we can reformulate  $P(C|Sk)$  in equation 1 as  $P(C|A_j)P(A_j|Sk)$  where  $A_j \in \{A_1 \dots A_n\}$  represents a car attribute such as price. After this modification, equation 1 can be written as

$$P(C|I, Sk) \propto \frac{P(C|I)}{P(C)} P(C|A_j)P(A_j|Sk) \quad (2)$$

We calculate  $P(C|I, Sk)$  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), we get more significant gains in accuracy at higher levels of the hierarchy such as car make.

## 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 can teach us a lot 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. Finally, we showed that demographic knowledge can also help improve fine-grained classification. For our future work we plan to perform even 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. 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. 3
- [2] 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, 3
- [3] 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. 3
- [4] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore—predicting the perceived safety of one million streetscapes. In *Computer Vision and Pattern Recognition*

864	Workshops (CVPRW), 2014 IEEE Conference on, pages 793–	918
865	799. IEEE, 2014.	919
866	[5] V. Ordonez and T. L. Berg. Learning high-level judgments	920
867	of urban perception. In <i>Computer Vision–ECCV 2014</i> , pages	921
868	494–510. Springer, 2014. 1, 2	922
869	[6] P. Salesse, K. Schechtner, and C. A. Hidalgo. The collabora-	923
870	tive image of the city: mapping the inequality of urban	924
871	perception. <i>PLoS one</i> , 8(7):e68400, 2013. 2, 6	925
872	[7] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie.	926
873	The caltech-ucsd birds-200-2011 dataset. 2011.	927
874	[8] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Part-	928
875	based r-cnns for fine-grained category detection. In <i>Computer</i>	929
876	<i>Vision–ECCV 2014</i> , pages 834–849. Springer, 2014.	930
877		931
878		932
879		933
880		934
881		935
882		936
883		937
884		938
885		939
886		940
887		941
888		942
889		943
890		944
891		945
892		946
893		947
894		948
895		949
896		950
897		951
898		952
899		953
900		954
901		955
902		956
903		957
904		958
905		959
906		960
907		961
908		962
909		963
910		964
911		965
912		966
913		967
914		968
915		969
916		970
917		971