

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-grain 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 crime statistics. Finally we show that just as fine-grain classes provide demographic information, societal cues can assist in fine-grain classification and improve accuracy. To facilitate our work, we have collected the largest and most challenging fine-grain 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 ubiquity of street view images has jumpstarted a new line of computer vision research focused on understanding cities through images [6] [4] [5]. For example, [6] showed that crime predictions can be improved by incorporating human perceptions of neighborhoods' images rather than using census data such as income alone [6] and [5] and [4] learn these perceptions using computer vision techniques. However, in order to extend these methods to other cities, extensive annotations of millions of images from each city would be required since, as [5] showed, algorithms trained on images of Boston, for example, cannot predict safety or wealth on images from San Francisco. We explore the question of learning social priors using large scale fine-grain classifications of cars and show that many neighborhood statistics such as income and crime rate can

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 \*\*\*\*

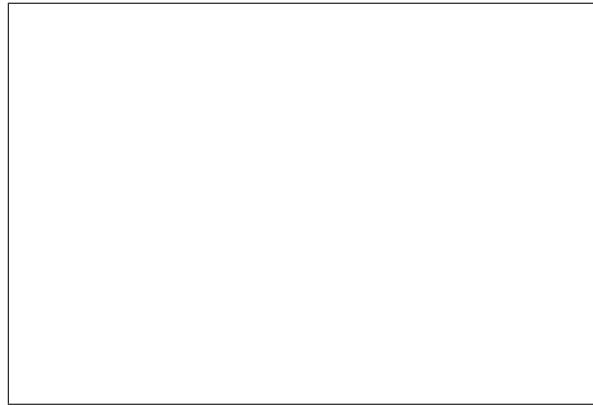


Figure 1. some pull figure

be predicted from car detections. Furthermore, using our detections in conjunction with census data, we can answer questions like what types of cars do rich/poor people drive?

Finally we show that we can use the answers to these questions to help improve fine-grained 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. This metadata can be especially important in fine grain classification. For example, just as detecting a large number of expensive cars in one area can give us a hint that we are in the vicinity of a wealthy neighborhood, knowing that we are in a wealthy neighborhood can also increase our likelihood of detecting expensive cars. Similarly, knowing that we are in a farm area increases our likelihood of detecting farm related cars and seeing many family households with young children increases our likelihood of detecting SUVs. We show that this information can be leveraged to improve fine-grain classification. Although there has been previous work on learning spatio-temporal priors for fine-grain classification [2] and exploiting street view geometry and GIS systems to improve object detection [3, 1] to our knowledge this is the first time census data and other social cues have been used to assist in fine-grain classification.

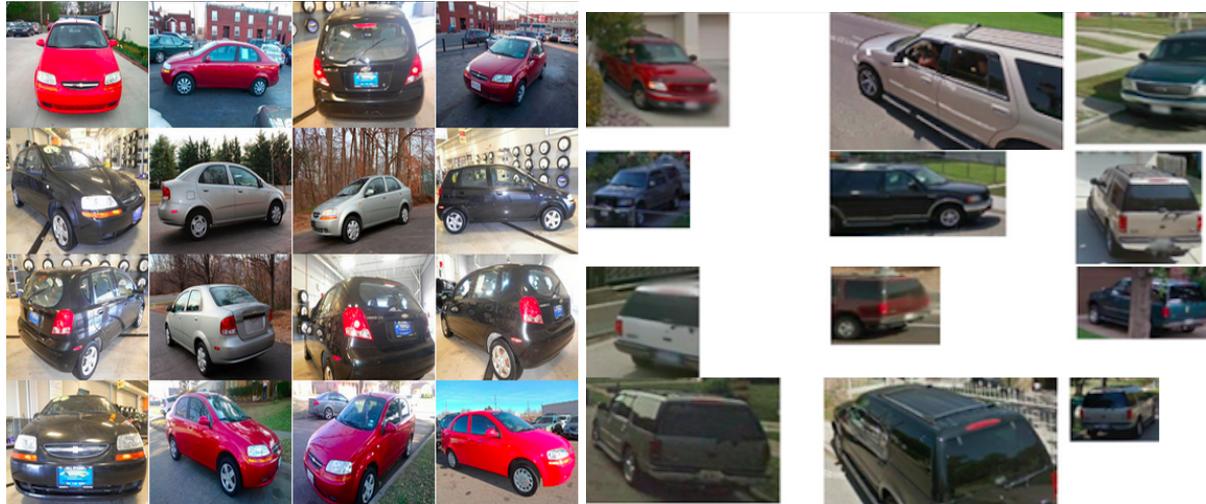


Figure 2. Left: examples of cars from craigslist, cars.com and edmunds.com, Right: examples of cars from streetview images. Cars from streetview images are blurred and occluded.

Summarizing our contributions:

1. We perform a large scale analysis of cities using our car detections and present intuitive as well as interesting insights
2. We show that using social cues extracted from census data can improve fine-grain classification accuracy
3. We present the largest fine-grain car dataset reported to our knowledge, complete with geotags and class as well as geography metadata
4. We include a larger set of 45 million street view images with car detections and fine-grain class predictions

## 2. Related Work

### Analysis of cities using images.

1. Plos one journal from MIT asking people to predict whether an area is safe/wealthy etc... after looking at the images
2. streetscore MIT paper predicting safety wealth scores etc.. just from images [4]
3. tamara Berg's paper on safety on ECCV [8] [7]

### Using GPS data to improve object detection.

4. Amir's work in GIS assisted object detections. For objects like streetlamps and trashcans, uses GIS to reproject objects to a plane and reduce the search space for object detection .

5. NYC 3D uses geographic elevation data to create view-point aware detectors and extract ground planes for them
6. Birdsnap fine-grain dataset with spatio-temporal prior for birds

## 3. Data

In order to find out how much we can learn about cities by analyzing cars detected from street view images we need to:

1. Collect a very dense number of street view images from the cities that we are interested in. If the sampling is not dense enough, our conclusions may not be valid or arise because of sampling error.
2. Gather a large dataset of cars that contains most of the cars that we would find in the US so that we can train a car detector and classifier.
3. Gather census and demographic data for the location of each street view image.

We briefly discuss 1. and 2. in this section. We first collected 45 million images from 8 million lat, longs by sampling roads every 25m. For each lat, long we have 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 create this dataset, we first created a class list using edmunds.com by obtaining example images of all the

216 cars they have (18017 cars) and grouped the cars into visually  
 217 indistinguishable classes using a series of amazon mechanical  
 218 tурk tasks as well as manual labor by the authors.  
 219 This resulted in 3147 visually indistinguishable classes. We  
 220 then collected additional images of cars from cars.com and  
 221 craigslist. Using AMT, we labeled 400k of these images  
 222 and 200k of our streetview images with bounding boxes for  
 223 cars.

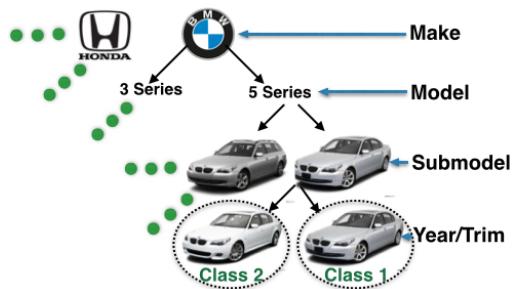
224 Finally, we hired car experts to label the cars in 100k  
 225 of our street view images, classifying them into one of the  
 226 3147 calsses that we created. We also labeled the cars from  
 227 craigslist and cars.com by parsing the car posting titles. The  
 228 final car dataset consists of 500k images consisting of 200k  
 229 street view images and 300k images from craigslist.com,  
 230 cars.com and edmunds.com. We split our data into training,  
 231 validation and test sets where our training set contains  
 232 199,666 streetview images and 313,099 web images,  
 233 and the validation and test sets contain 39,933 and 159,732  
 234 streetview images respectively. Although we are interested  
 235 in classifying street view images, having web images in our  
 236 training set increases the amount of data we have without  
 237 paying expensive car experts to annotate street view im-  
 238 ages. Fig. ?? shows some examples of cars from our dataset.  
 239 We can see that cars from street view images are generally  
 240 occluded and blurry where as cars from other web sources  
 241 have higher resolution and are unoccluded.

## 242 4. Detecting and classifying cars

243 After collecting images and creating a dataset, we have  
 244 to train a car detection and classification algorithm. Al-  
 245 though we would have ideally used the recent RCNN  
 246 from [?] due to its state of the art results, the computa-  
 247 tional and memory requirements of RCNN make it imprac-  
 248 tical for use in such a large scale detection setting such as  
 249 ours. Since our training set consists of XXX images and  
 250 we want to detect cars in  $\sim$ 45 million images, scalabil-  
 251 ity—not only accuracy—becomes an imortant considera-  
 252 tion. Specifically, training an RCNN detector on our data  
 253 would require roughly XXGB of memory and XXX sec-  
 254 onds. And detecting cars during test time would take  $\sim$ 20s  
 255 per image which is impractical for detecting multiple cars  
 256 in  $\sim$ 45 million images. Thus our pipeline, instead, consists  
 257 of using DPM [?] to detect cars and a CNN [?] to classify  
 258 them. We present details in the sections below.

### 259 4.1. Car Detection

260 As mentioned above, we not only have to worry about  
 261 accuracy but also scalability. Thus, we experimented with  
 262 different DPMs to evaluate accuracy/speed trade off. A sin-  
 263 gle component DPM with 8 parts gives us the best tradeoff  
 264 between accuracy and speed and our supplemental material  
 265 discusses many other configurations that we have tried. For  
 266 example, a DPM with 1 component and 8 parts results in



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

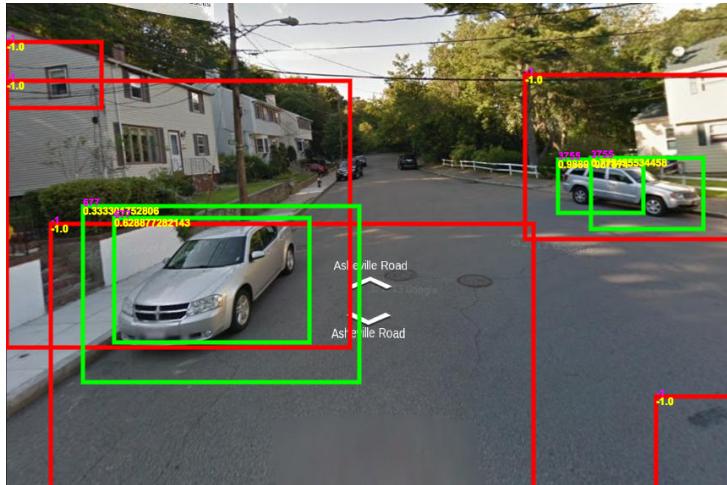
Attribute	Acc %
Make	55
Model	60
Submodel	70
Year (5 bins)	40
Price (5 bins)	50
Domestic/Foreign	80
Country	80

269 Table 1. Accuracy by car attribute

270 an average precision (AP) of 64.2% and takes 5 seconds per  
 271 image at test time where as a DPM with 3 components and  
 272 8 parts results in 68% AP and takes  $\sim$ 15 sec per image. The  
 273 highest AP that we measured was a value of 68.7% with a 5  
 274 component 8 parts model that takes  $\sim$ 22 sec per image.

### 275 4.2. Car Classification

276 After doing car/no car detection, we use a standard CNN  
 277 from [?] with [?] to classify the cars into one of 2657  
 278 fine-grained classes. Although we want to classify cars  
 279 from street view images, as mentioned in sec. 3 XX% of  
 280 our training images consist of cars from other sources such  
 281 as craigslist due to the fact that annotating street view im-  
 282 ages is much more expensive. Thus we add deformations  
 283 such as blurring in an attempt to make the non-street view  
 284 images look like street view images. We give details of  
 285 training the CNN in our supplementary material. At test  
 286 time we take the top 10% scoring DPM bounding boxes  
 287 and classify them. This gives a 10x increase in speed but  
 288 only a 0.X% drop in AP as compared to using all the boxes  
 289 detected by DPM. We achieve an accuracy of 33.X% on  
 290 the true positive DPM bounding boxes and 31.27% on the  
 291 ground truth bounding boxes. Fig. ?? shows an example  
 292 of our streetview detections. Surprisingly, the cars are de-  
 293 tected and classified correctly even though blurriness and  
 294 occlusions in street view images makes this a difficult task.

324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339

Jeep grand-cherokee SUV 2008-2010: limited



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

378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397

Figure 4. 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.

343

### 4.3. Analyzing Hierarchical Classification Accuracy

344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359

Although we are classifying to one of 2657 fine-grained classes, some types of mistakes are more costly for our task of social analysis than others. For example, we might not care about misclassifying 2001 Honda Accord lx to 2001 Honda Accord dx but a mistake classifying a 2012 BMW 3-series to a 1996 Honda Accord would create large errors in our analysis if we, for example, want to estimate the average price or age of a car in a zip code and measure how it is related to median household income. Similarly, misclassifying an SUV to a sedan would confound an analysis estimating the cities with the most number of large cars. Inorder to gain insight into our classification errors we perform a hierarchical accuracy analysis.

360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372

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 travelling down the tree, with the most similar classes,in most cases, being the leaves sharing a parent. Can you tell the difference between class 1 and class 2? Given the difficulty of our dataset, we were very surprised by the 31.27% accuracy of our classifier at the leaf level. Table ?? lists accuracies by various car attributes such as those in 3 and others like car price, year etc... We can see that the accuracy is much higher after aggregating by different attributes.

373  
374  
375  
376  
377

## 5. Societal analysis

After collecting data, training fine-grained car detectors and classifying cars in all of our images, we now have all the components to perform social analysis. We show some

general results from the entire united states as well as case studies from cities we could obtain ground truth data for. We divide our analysis into different sections below.

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

#### 5.1.1 Cars on the street versus cars people drive

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? In order to answer this question we downloaded the vehicle census from Massachusetts, which is the only state to release extensive vehicle registration data. Surprisingly, we 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 per zip code in the three cities we analyzed in Massachusetts: Boston, Worcester and Springfield. The high correlation was only obtained after aggregating cars at the zip code level which tells us that most people in these cities must 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. 5 there is a high correlation for most of the makes. This shows us that the cars we detect from street view images can actually tell us a lot about the types of cars that people in a particular zip code drive.

#### 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

427  
428  
429  
430  
431

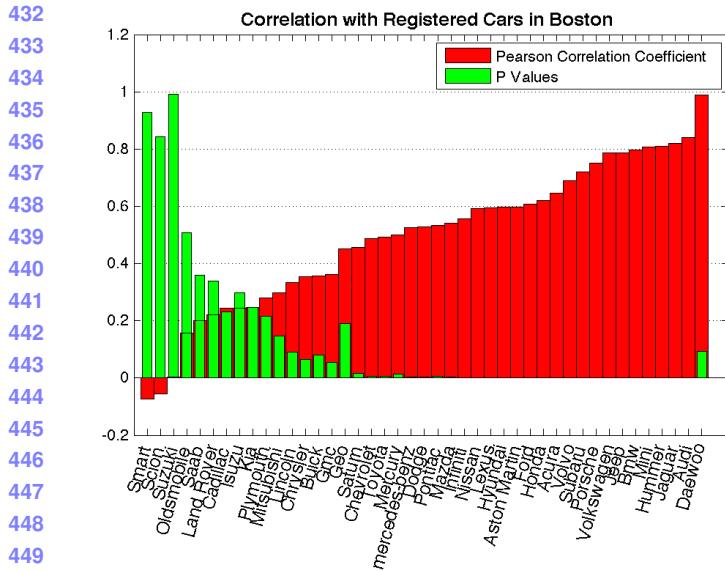


Figure 5. Pearson correlation coefficient and p values between the number of detected and registered cars in Boston for each make.

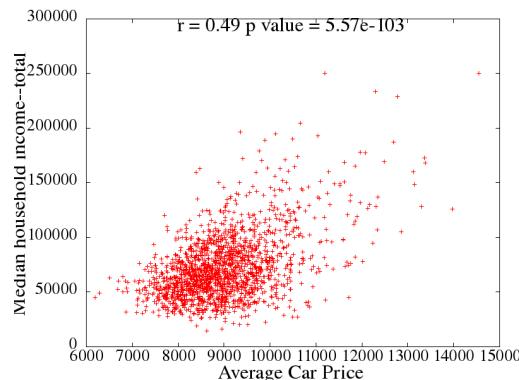


Figure 6. Median household income per zip code vs average price of detected cars per zip code.

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. 6 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 ( $r=0.49$ ,  $p \sim 0$ ). This makes sense since rich people mostly live around places with expensive cars and tend to drive expensive cars. Our results also indicate that rich people prefer to drive foreign, especially German, cars ( $r=0.59$ ). This observation is also consistent with our expectations since expensive cars such as BMWs are German. What is perhaps surprising is that there is a very high nega-

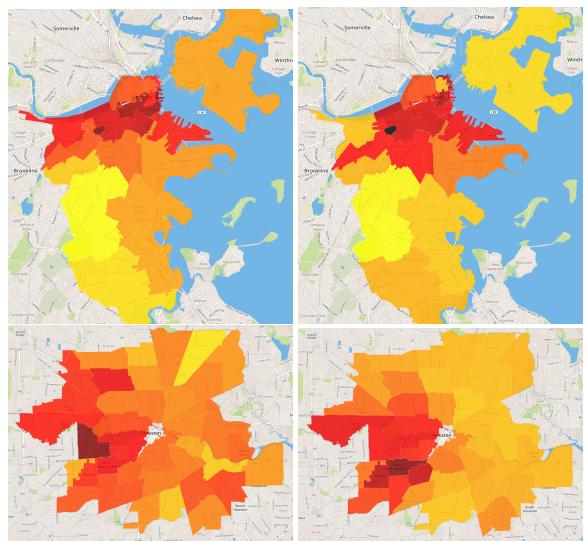


Figure 7. Top left: Heat map of average car price in Boston, right: heat map of median household income in Boston. Bottom left: Heat map of average car price in Houston, right: heat map of median household income in Houston

tive 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?

One would probably guess that 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. Indeed, we found that this is the case ( $r=0.3$   $p \sim 0$ ). 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.

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

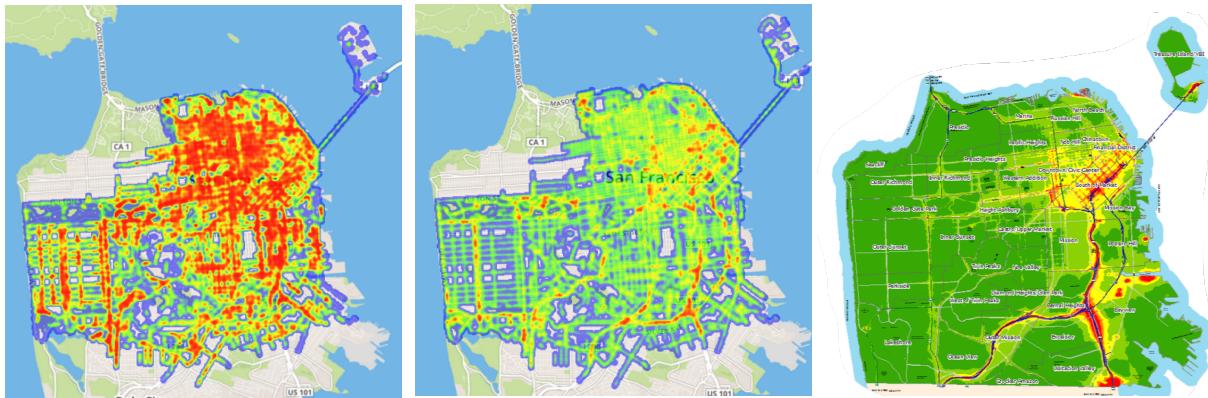


Figure 8. A. Density of cars in San francisco inversely weighted by their expected MPG, B. The weighted average of car MPG in San Francisco. The weights are the expected number of cars, C. Ground Truth for Air quality (measured in annual particulate matter) in San Francisco.

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. 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 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. However, the correlation between car prices and neighborhood wealth is not going to be perfect because we are not necessarily detecting the cars that are registered by residents. Figure 7 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 7 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? 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. 8A 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. 8C is a map of San Francisco air quality measuring annual average particulate matter concentration (MPG) from all sources. To our surprise, their map seems to agree with Fig. 8B 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

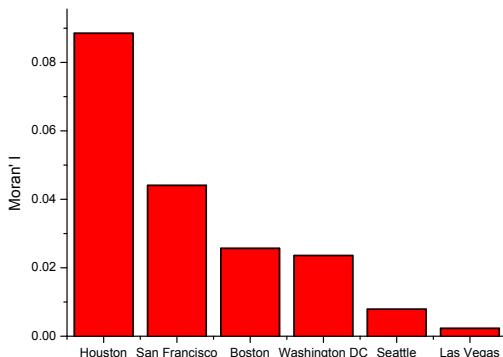


Figure 9. Moran I scores for 5 cities. Houston has the highest score where as Las Vegas has the lowest

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. 9 plots the highest and lowest scoring cities as well as well as few others in between. We can see that Houston shows the highest clustering where as Las Vegas shows the lowest.

### 5.3.2 Which cities are more patriotic?

Which cities have the most number of domestic cars? As Fig. 10A 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.

### 5.3.3 Which cities are wealthier?

In our final analysis, we ask which city has the most expensive cars on average? Fig. 10B 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

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

702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755

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 gine-grained accuracy and acc-m is the make level accuracy. Car price/year with median household income give the highest gain in accuracy.

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 conditionally 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 2500 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 Ix 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

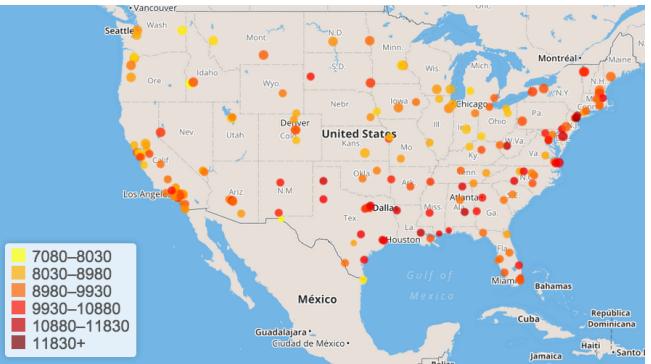
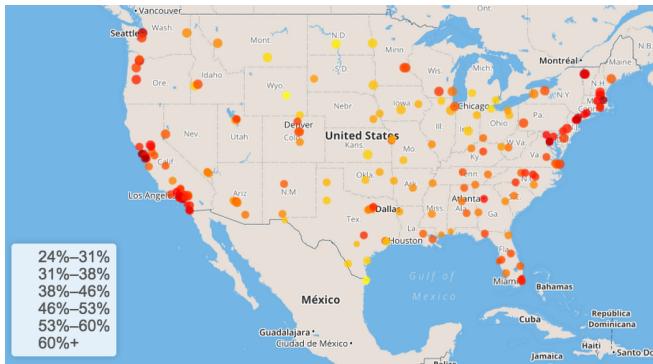


Figure 10. A. Map of the percentage of foreign cars in each city. San Francisco has the highest percentage and Casper the lowest. B. Map of the United States showing the expected car price in each city. New York has the highest expected car price where as El Paso has the lowest.

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  $P(C|Aj)P(Aj|Sk)$  where  $Aj \in \{A1\dots An\}$  represents a car attribute such as price. After this modification, equation 1 can be written as

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

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

## 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. 1
- [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. 1

- [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. 1
- [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 Workshops (CVPRW), 2014 IEEE Conference on*, pages 793–799. IEEE, 2014. 1, 2
- [5] V. Ordonez and T. L. Berg. Learning high-level judgments of urban perception. In *Computer Vision–ECCV 2014*, pages 494–510. Springer, 2014. 1
- [6] 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. 1, 6
- [7] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2
- [8] 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