

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053054
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 ****

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 Fig. 1 is that of a wealthy one. Traditionally, the most prevalent method for gathering such personal and demographic 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 insights by analyzing massive user texts and so-



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.

cial networks [?] [?]. For example, recent work from [?] 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 [?] [?] [?] [?]. 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. XX% of Americans own cars, 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



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.

from 200 of the largest American cities and XXX zip codes where an average of XXX 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 [?] to detect cars due to scalability concerns and train a CNN [?] 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 [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 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 [?] [?] [?] 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 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.

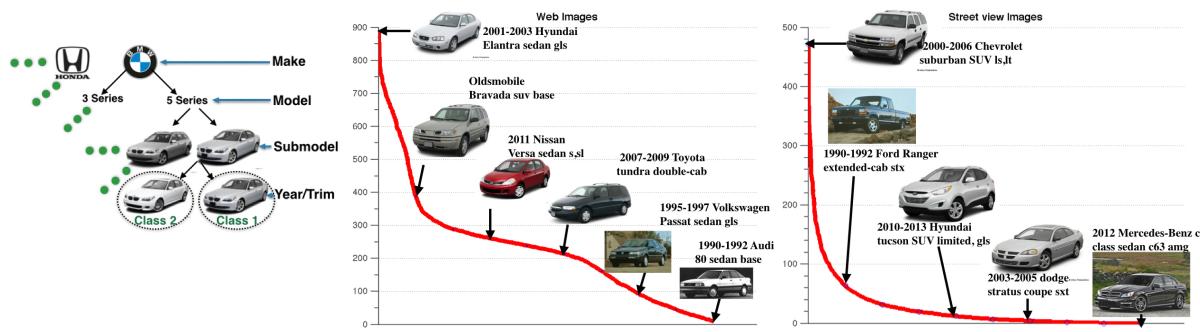
216
217
218
219
220
221
222
223
224
225
226

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

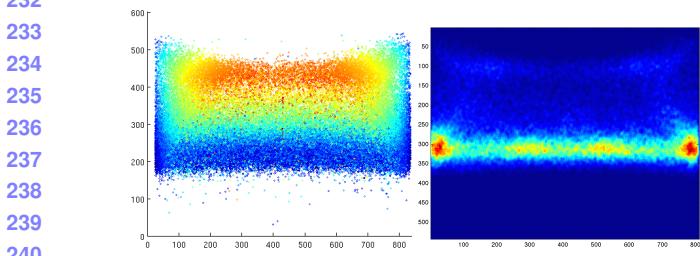


Figure 4. Left: Location/scale heatmap of streetview bounding boxes with red indicating larger sizes and blue smaller ones. Each point indicates the size of a bounding box centered at that point in a street view image.

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

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. Accuracy by car attribute

258

259

260

261

262

3. Cars and Cities Dataset

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



Figure 3. Right: A line graph of Street view Images distribution for various car models. We have the most number of images for 2000-2006 Chevrolet suburban SUV ls,lt and the least for 2012 Mercedes-Benz c-class sedan c63 amg.

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. ??, 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. ??a, 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. ??a 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, [?] [?] [?], 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 ~ 20 s per image using a machine with a single GPU. Thus our pipeline, instead, consists of using DPM [?] to detect cars and a CNN [?] to classify them. We present details in the sections below.

324
325
326
327
328
329
330
331
332

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

Table 2. Accuracy by car attribute

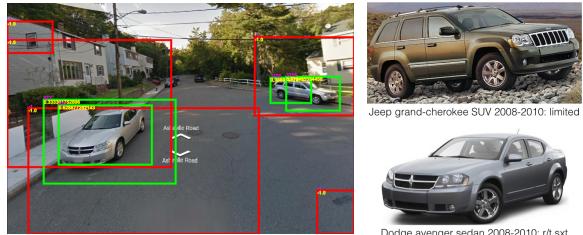


Figure 6. An example of our car detections and classifications. The red boxes are DPM bounding boxes that are not classified by our CNN and green boxes are boxes classified by the CNN. The two cars in the image are correctly classified to the classes shown on the right.

343

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.

362
363

4.2. Car Classification

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 these images during training to liken them to street view images. We give details of training the CNN in our supplementary material. At test time we take the top 10% scoring DPM bounding boxes and classify them. This results in a 10x increase in speed but only a ~2% drop in AP as compared to using all the bounding boxes detected by DPM. We achieve an accuracy of 33.1% 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. ?? shows a few of our classification mistakes. It can be seen that some of the errors are between highly similar classes.

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 ?? lists accuracies by various car attributes such as those in fig. ?? 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.

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 ($p < 0.01$) 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. 8 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.

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

Correlation with Registered Cars in Boston

Car Model	Pearson Correlation Coefficient (Red Bar)	P Value (Green Bar)
Smart	~0.95	~0.90
Suzuki	~0.85	~0.80
Land Rover	~0.75	~0.35
BMW	~0.70	~0.25
Fiat	~0.65	~0.20
Mitsubishi	~0.60	~0.15
Chrysler	~0.55	~0.10
Mercedes-Benz	~0.50	~0.05
Aston	~0.45	~0.02
Honda	~0.40	~0.01
Toyota	~0.35	~0.00
Subaru	~0.30	~0.00
Nissan	~0.25	~0.00
BMW	~0.20	~0.00
Honda Accord	~0.15	~0.00
Honda Fit	~0.10	~0.00
Honda CR-V	~0.05	~0.00
Honda Fit	~0.00	~0.00
Volkswagen	~0.00	~0.00
Jaguar	~0.00	~0.00
Dodge	~0.00	~0.00
BMW X5	~0.00	~0.00

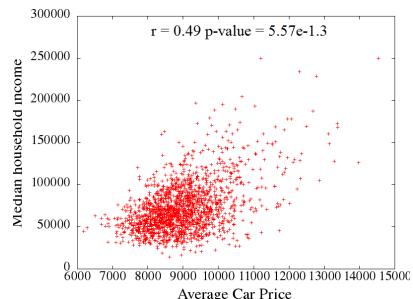
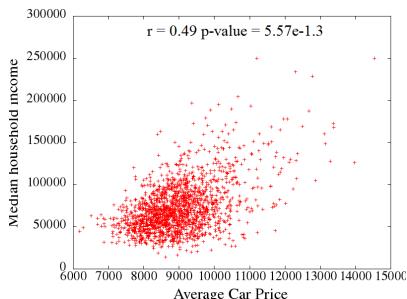


Figure 8. Left: Pearson correlation coefficient and p values between the number of detected and registered cars in Boston for each make. Center: scatter plot of detected cars per zip code vs. registered cars per zip code in Boston. Left: scatter plot of No. of detected Honda's per zip code vs. registered Honda's per zip code in Boston.

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 ~ 0 .

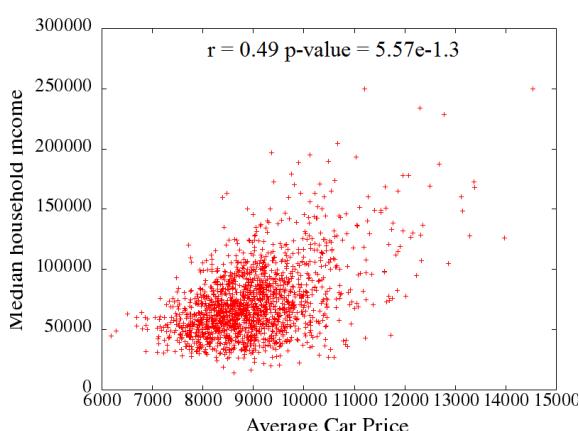


Figure 7. Scatter plot of median household income per zip code vs. average price of detected cars per zip code.

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 3 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 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 [?] 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.

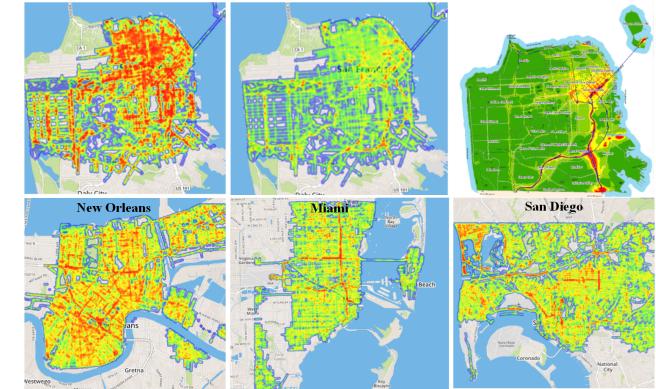
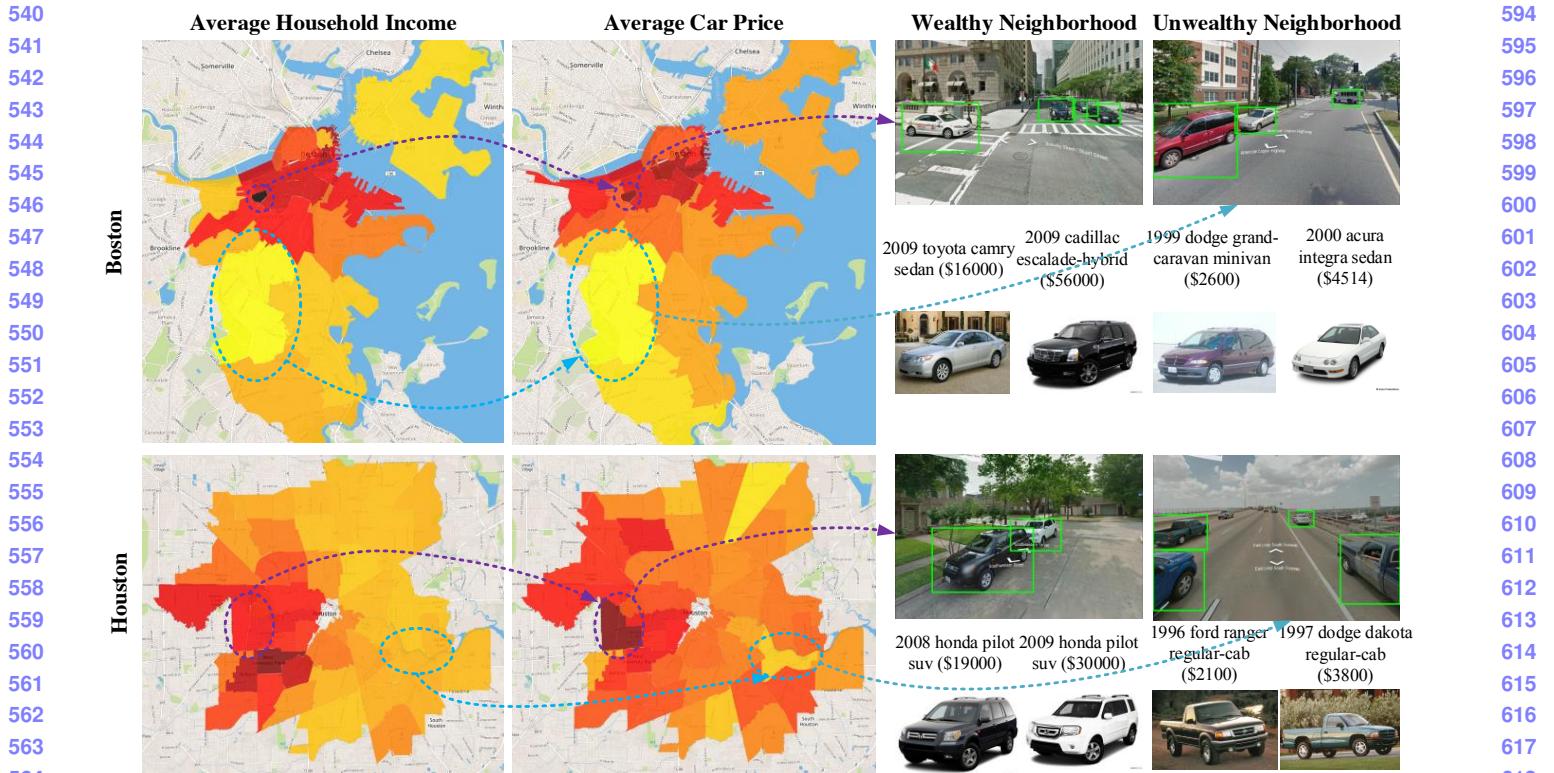


Figure 10. A. Density of cars in San francisco inversely weighted by their expected miles per gallon (MPG), B. The inverse of weighted average of car MPG in San Francisco. The weights are the expected number of cars in a GPS point, C. Ground Truth for Air quality (measured in annual particulate matter)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

648 **5.2.2 Which neighborhoods have high car pollution?** 702
649

650 Can our street view detections tell us anything about which
651 neighborhoods are affected by highly polluting cars? We
652 plotted a heat map of the expected number of cars per sample
653 inversely weighted by the expected MPG of that sample.
654 We also map the inverse of the weighted average of car
655 MPG where the weights are the expected number of cars.
656 The first measure should give us a rough idea of the location
657 of highly polluting neighborhoods: for the same density of cars,
658 areas with high MPG result in lower numbers than those with low MPG.
659 For different densities of cars, the relative magnitude of the measure depends on both the density
660 of cars and how efficient they are. The second measure,
661 on the other hand, visualizes areas with a high concentration
662 of low MPG cars.

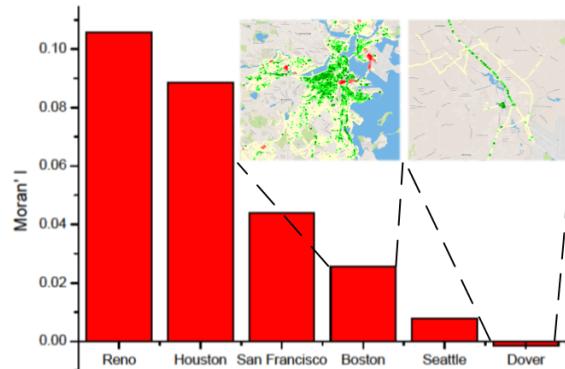
663 Fig. 10A shows the weighted density of cars in San Francisco
664 and B shows the inverse weighted MPG. Although we
665 could not find ground truth data of car pollution, Fig. 10C
666 is a map of San Francisco air quality measuring annual
667 average particulate matter concentration (MPG) from all
668 sources [?]. To our surprise, their map seems to agree with
669 Fig. 10B in most cases.

670 **5.3. What cars on the street tell us about cities**671 **5.3.1 Which cities are more segregated?**

672 Which cities show high clustering of similarly priced cars?
673 Specifically, which cities have expensive cars clustered
674 together with other expensive cars and cheap cars clustered
675 with other cheap cars? Given the high correlation between
676 median household income and average car price, the answer
677 to this question should give us a good indication of the cities
678 that are most and least segregated. Following the analysis
679 of [6] we use the Moran I statistic to measure spatial auto-
680 correlation where a value of 1 indicates perfect clustering
681 of similar values, -1 indicates perfect dispersion and 0 indicates
682 a random spatial arrangement (neither clustering nor
683 dispersion). Fig. 11 plots the highest and lowest scoring
684 cities as well as a few others in between. We can see that Reno
685 shows the highest clustering where as Dover shows the lowest.

686 **5.3.2 Which cities are more patriotic?**

687 Which cities have the most number of domestic cars? As
688 Fig. 13A shows the coastal cities have a high concentration
689 of foreign made cars where as the midwest has a low con-
690 centration. This result agrees with [?] who measured the
691 ratio of American/foreign cars driven in the US. The city
692 with the highest percentage of foreign cars was found to be
693 San Francisco with 61% of foreign cars where as Casper
694 Wyoming had the least percentage (21%).



695 Figure 11. Moran I scores for car prices. The highest and lowest
696 scoring cities are shown as well as 5 cities with scores in between.
697 Reno exhibits the most amount of clustering by car price where
698 as Dover exhibits very little clustering. The two maps for Boston
699 and Dover show areas of high clustering where green indicates
700 statistically significant clustering of expensive cars and red indicates
701 statistically significant clustering of cheap cars.

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

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

705 **5.3.3 Which cities are wealthier?**

706 Which city has the most expensive cars on average?
707 Fig. 13B maps the average car price for each city. The map
708 shows that many of the east coast cities have expensive cars
709 as well as some cities in the south such as Atlanta. We found
710 the city with the most expensive cars to be New York with
711 an expected price of $\sim \$12k$ and the one with the least ex-
712 pensive cars to be El Paso ($\sim \$7k$).

713 **6. Using social priors to improve classification**

714 This section explores the use of demographic census in-
715 formation as a prior to improve fine-grained detection and
716 presents preliminary results run on our validation set. Intui-
717 tively, if a particular image was taken in a wealthy neigh-
718 borhood, for example, one would expect the cars in that
719 neighborhood to be expensive and viceversa. In order to
720 use zip code level census variables as priors, we first cal-
721 culated $P(C|I, Sk)$ where C is the fine-grained class, I is
722 an image and $Sk \in \{S1 \dots Sn\}$ is a particular zip code level
723 census variable such as median household income. After
724

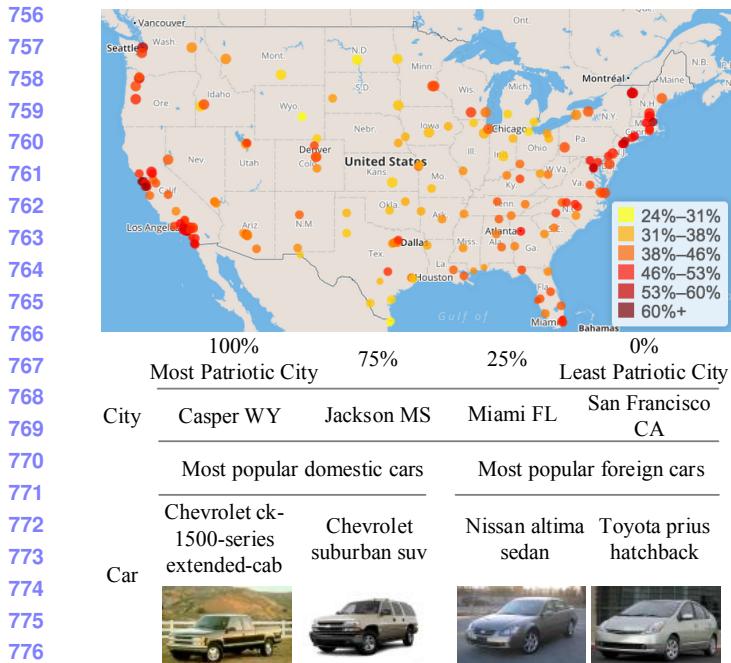


Figure 12. Map of the percentage of foreign cars in each city. San Francisco has the highest percentage with (61% and Casper the lowest with 21%.

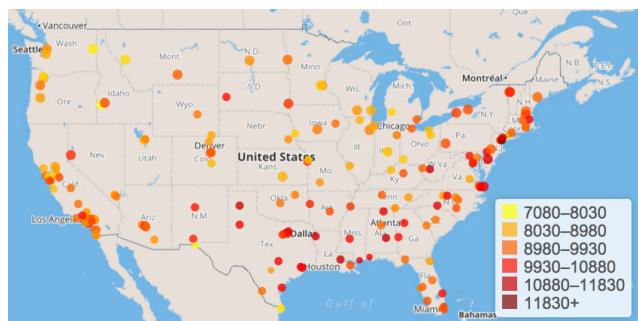


Figure 13. Map of the United States showing the expected car price in each city. New York has the highest expected car price where as El Paso has the lowest.

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% 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 $P(C|Sk)$ 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). 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.

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. 2
- [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] 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
- [4] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo. Streetscore—predicting the perceived safety of one million

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

864	streetscapes. In <i>Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on</i> , pages 793–	918
865	799. IEEE, 2014.	919
866		920
867	[5] V. Ordonez and T. L. Berg. Learning high-level judgments	921
868	of urban perception. In <i>Computer Vision–ECCV 2014</i> , pages	922
869	494–510. Springer, 2014. 1, 2	923
870	[6] P. Salesse, K. Schechtner, and C. A. Hidalgo. The collabor-	924
871	ative image of the city: mapping the inequality of urban	925
872	perception. <i>PLoS one</i> , 8(7):e68400, 2013. 2, 7	926
873	[7] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie.	927
874	The caltech-ucsd birds-200-2011 dataset. 2011.	928
875	[8] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Part-	929
876	based r-cnns for fine-grained category detection. In <i>Computer</i>	930
877	<i>Vision–ECCV 2014</i> , pages 834–849. Springer, 2014.	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