VISUAL CENSUS OF STREET-LEVEL IMAGERY

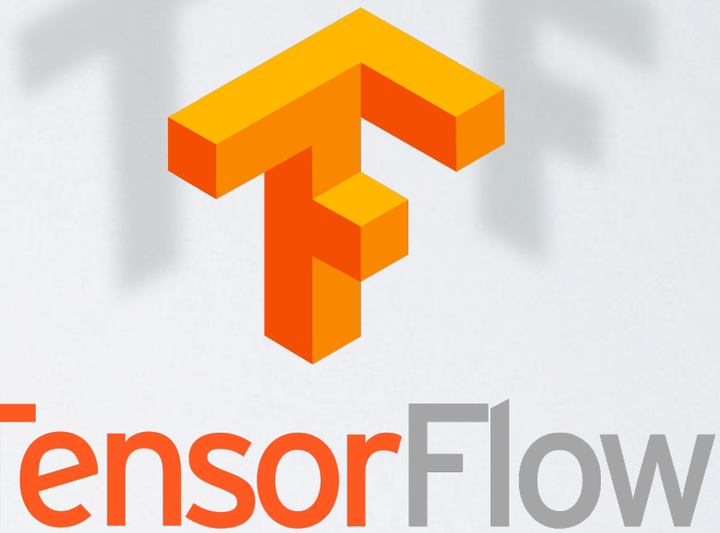
USING A FINE-GRAINED CAR-TYPE IMAGE CLASSIFIER

KATERINA SCHULZ | GALVANIZE DATASCIENCE IMMERSIVE
SF, CA 2019

A VISUAL CENSUS OF STREET-LEVEL IMAGERY

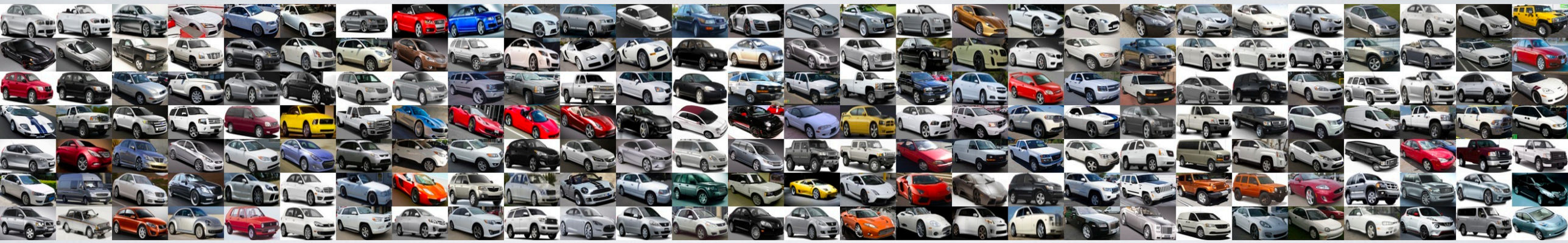
OVERVIEW

- Build a Fine-Grained Image Classifier for Cars
- Use TensorFlow 2.0 and Pre-Trained Classifier
- Train on Stanford Cars Dataset[1]
- Extract Car-Type object instances from street-level images[2]
- Classify Car-Type objects in each scene
- Predict socio-economic status of scene location



A VISUAL CENSUS OF STREET-LEVEL IMAGERY

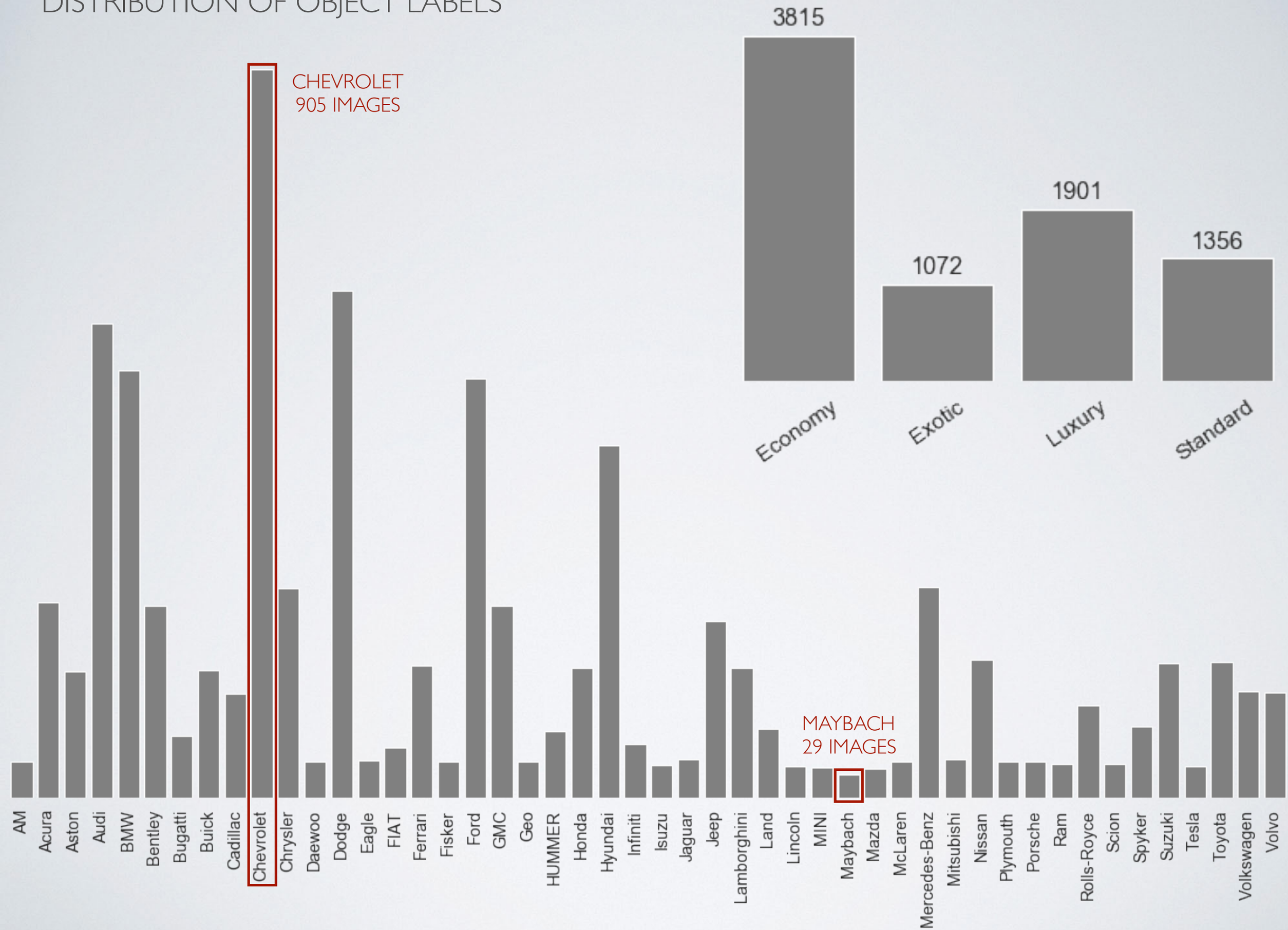
THE DATA



Stanford Cars Dataset

- 8,144 image training set
- 196 labels of car make-model-year
- Multiple views for each model
- Models range from Geo to Bugatti

DISTRIBUTION OF OBJECT LABELS



DISTRIBUTION OF OBJECT LABELS

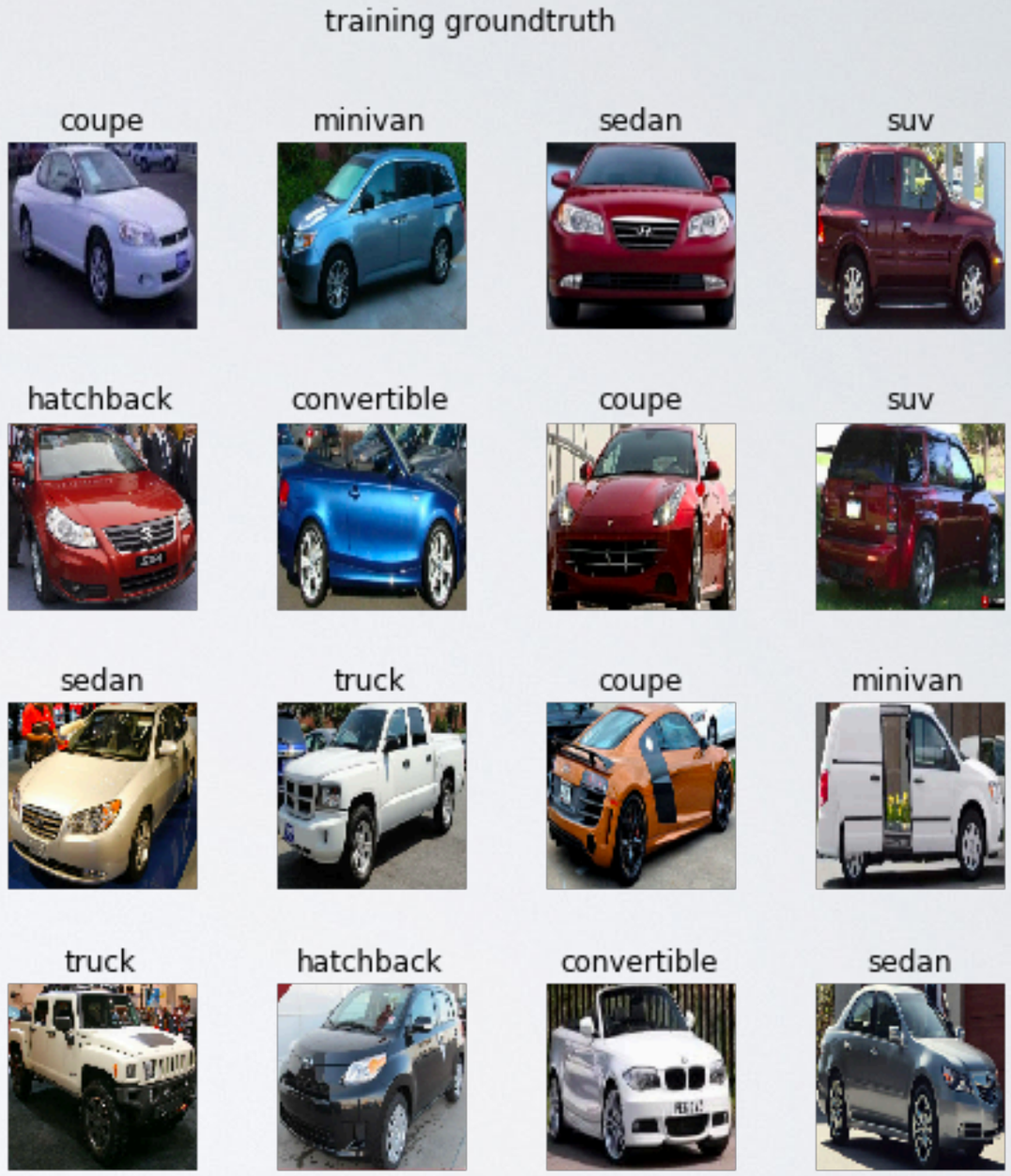
CHEVROLET
905 IMAGES

MAYBACH
29 IMAGES

A VISUAL CENSUS OF STREET-LEVEL IMAGERY

STEP 1: PROCESS THE DATA

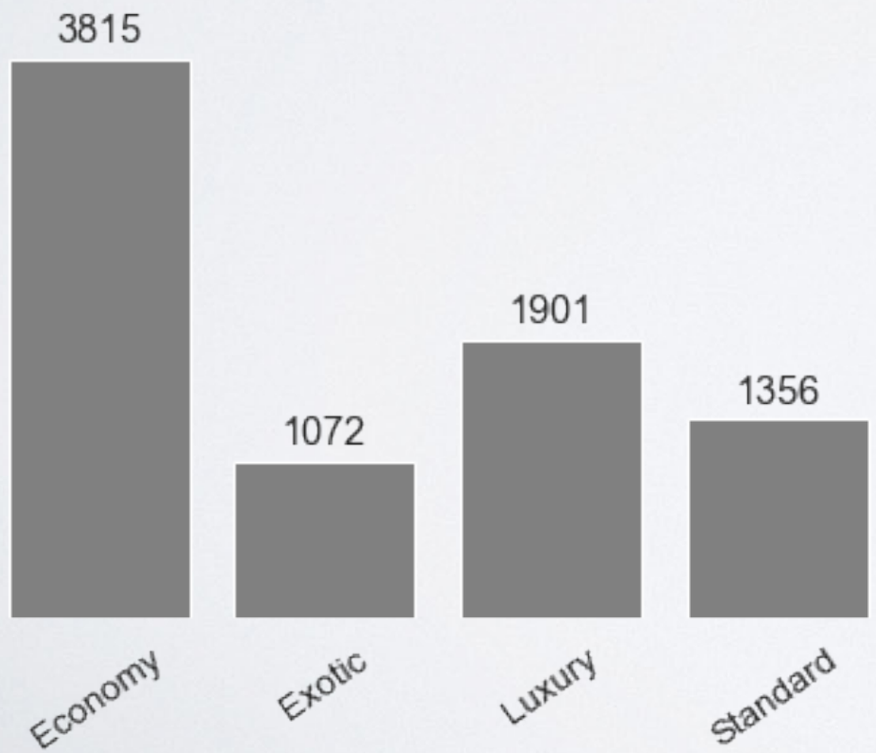
- Metadata written to **tfRecords**
- Metadata includes raw images, object bounding boxes, and labels
- Choose labels to train on (Type, Value, or Manufacturer)
- tfRecords get parsed; raw images read, cropped, and resized to 224 by 224

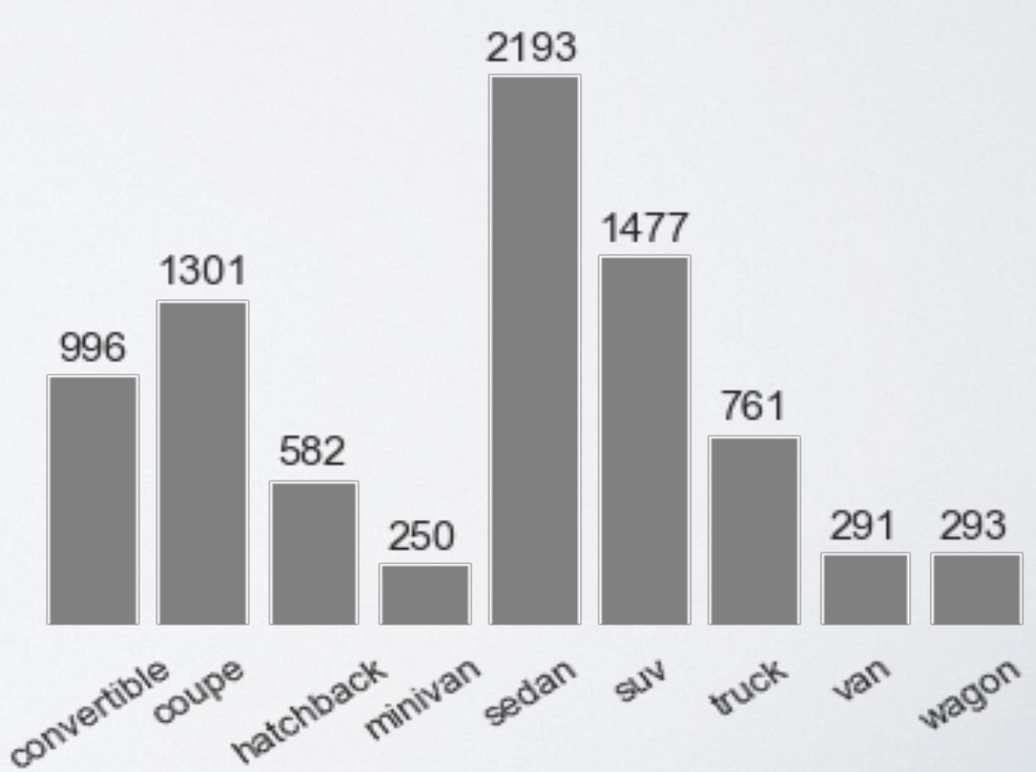


A VISUAL CENSUS OF STREET-LEVEL IMAGERY

STEP 2: BUILD & TRAIN A FINE-GRAINED IMAGE CLASSIFIER

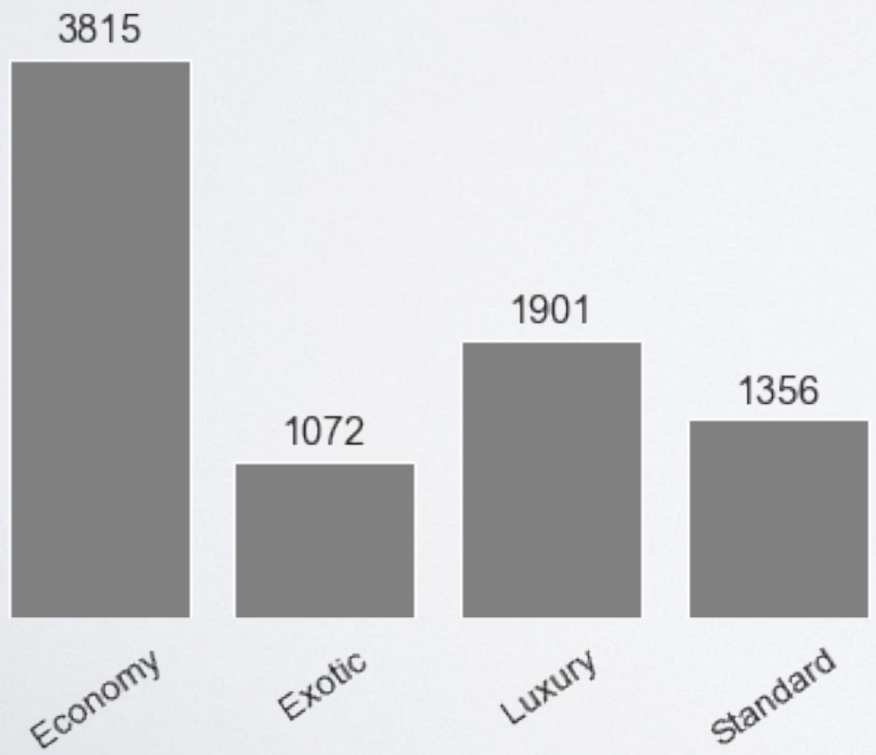
- Built on Pre-Trained Classifier
- Trained on 6,400 images [1]
- 200 batches of 32 images per epoch
- Validation Accuracy: 66% (Type)

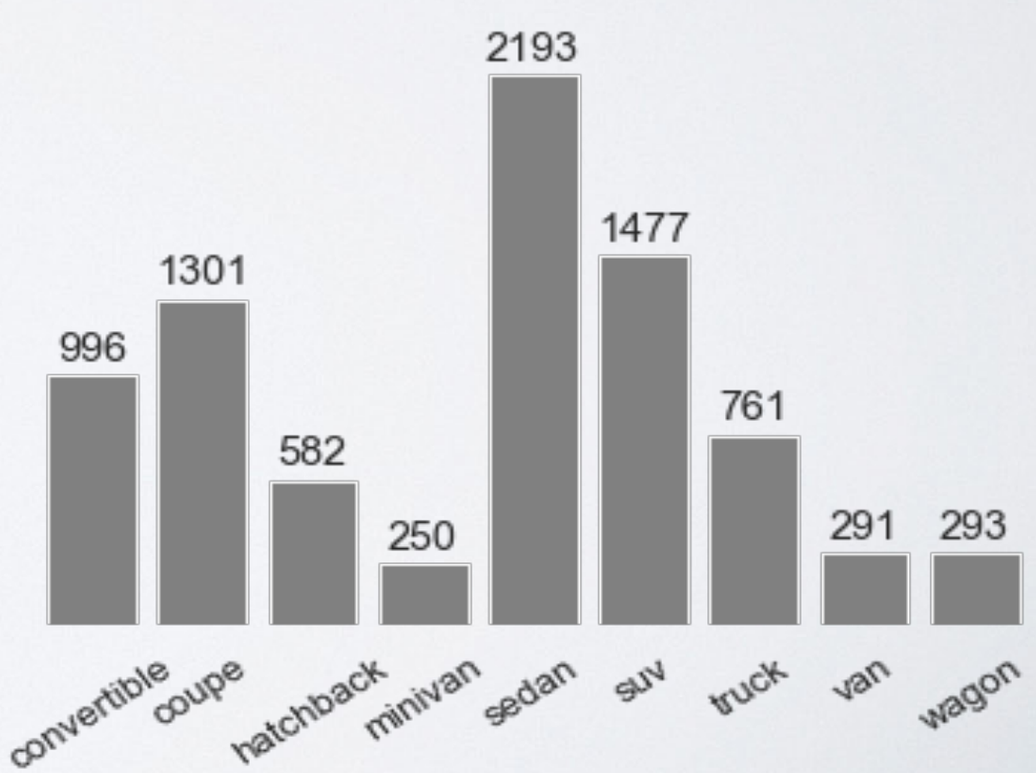




STEP 2: BUILD & TRAIN A FINE-GRAINED IMAGE CLASSIFIER

- Built on Pre-Trained Classifier
- Trained on 6,400 images [1]
- 200 batches of 32 images per epoch
- Validation Accuracy: 68% (Value)



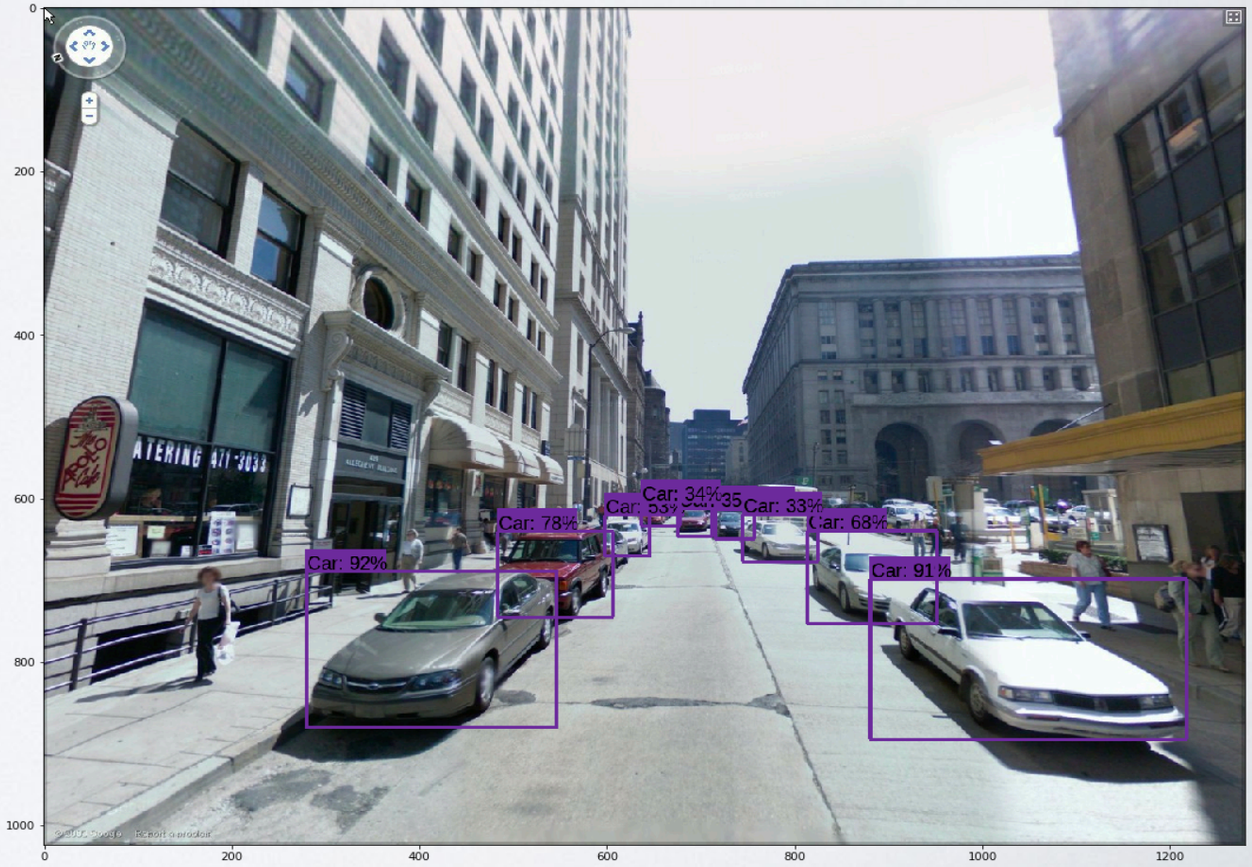


A VISUAL CENSUS OF STREET-LEVEL IMAGERY

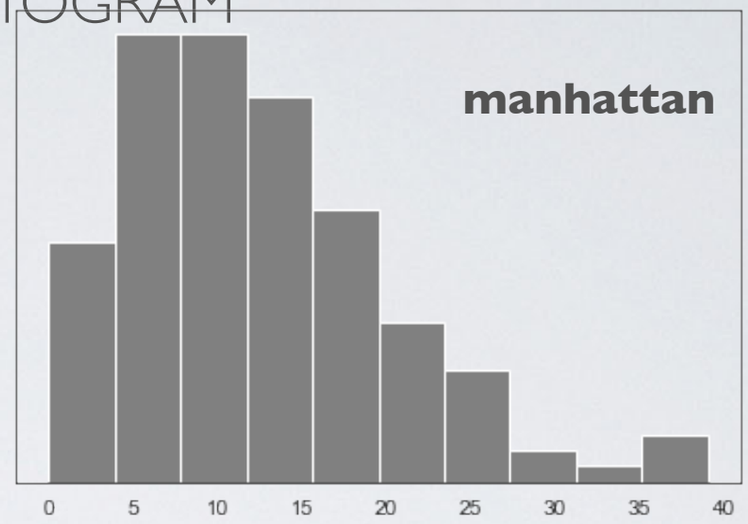
STEP 3: OBJECT DETECTION FROM STREET-LEVEL IMAGES

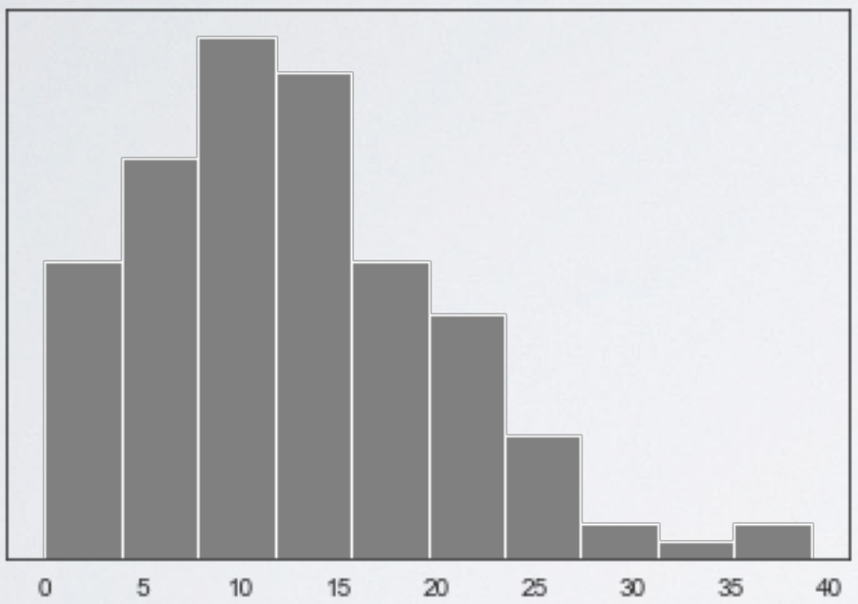
UCF Google StreetView Dataset

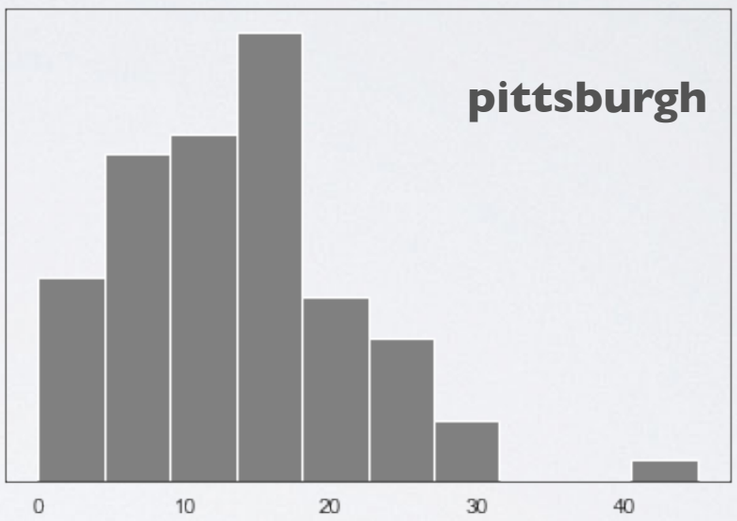
- Images covering Orlando, Pittsburgh, Manhattan
- GPS coordinates for each image
- **TF-Hub module** trained to perform object detection (**ImageNet** detector)
- Cropped car-type objects over detection confidence threshold

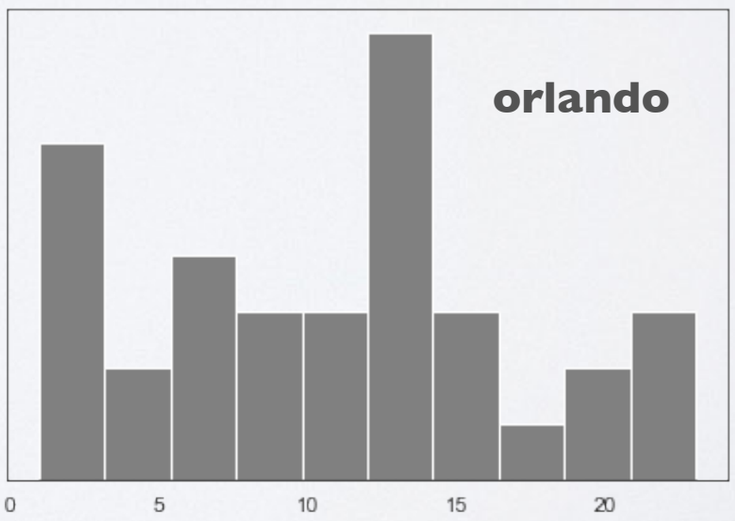


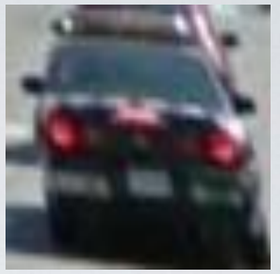
DISTRIBUTION OF OBJECT INSTANCES: HISTOGRAM

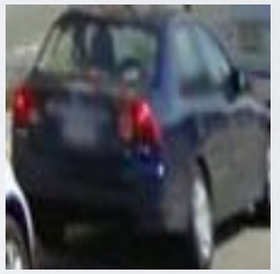






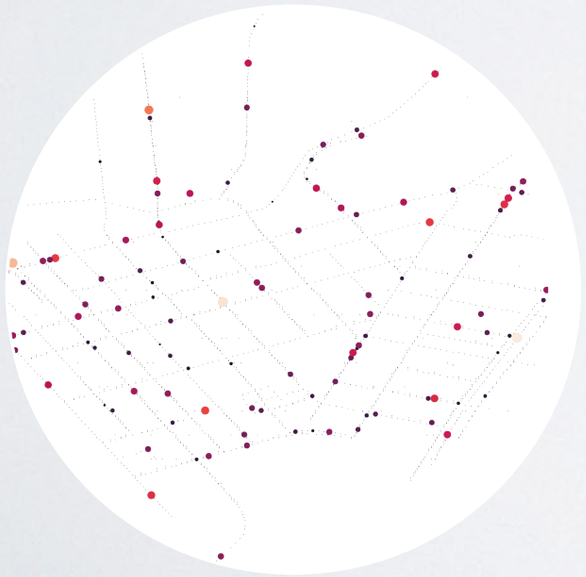




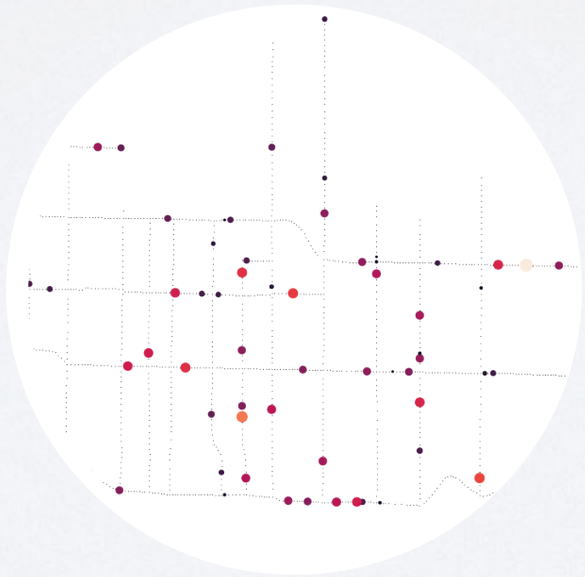


DISTRIBUTION OF OBJECT INSTANCES: GEOSPATIALLY

- Ran object detector over 25% of scenes in the dataset
- Many objects detected were not large enough to be classified
- Classified approximately 1,500 individual vehicles



pittsburgh



orlando

manhattan

A VISUAL CENSUS OF STREET-LEVEL IMAGERY

THE RESULTS

- Maximum validation accuracy on Value classification (68%)
- Achieved 66% validation accuracy on Type classification
- Results and visual analysis show lack of diversity in car Types and Value
- Current model not robust enough to convey socioeconomic status based on vehicles
- Street level dataset car-type objects may be too sparse

GITHUB

github.com/kittyschulz/visual_census

DATA

[1] **3D Object Representations for Fine-Grained Categorization**
Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei
4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.

[2] **Image Geo-localization Based on Multiple Nearest Neighbor Feature Matching using Generalized Graphs**
Amir Roshan Zamir and Mubarak Shah
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2014