Deep Vision System for Vehicle Make/Model Recognition and Multi-Object Detection

ADSP 32023 Final Project

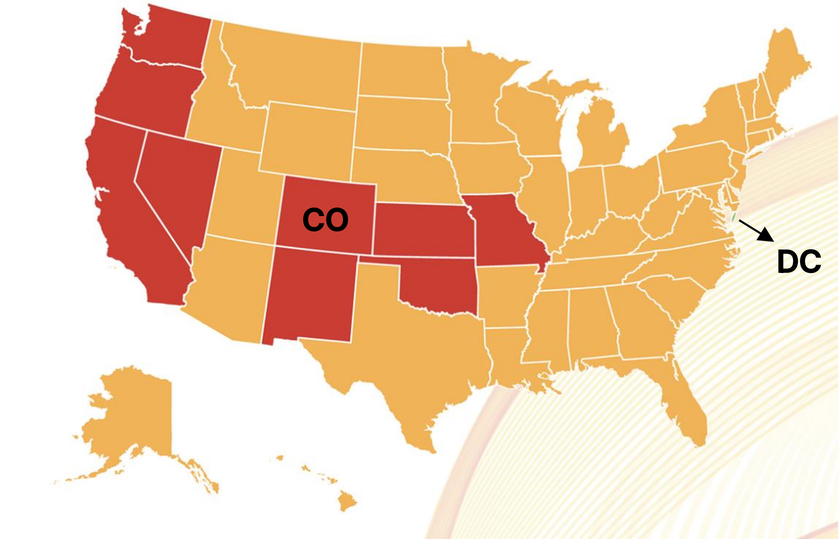
Presented By: Arnav Pillai, Daichi Ishikawa, Kshitiz Sahay, William DeForest

Table of Contents

Why do we need a Vehicle Identification system?

- Lack of automation in vehicle identification
- Inefficient law enforcement
- Insurance fraud
- Missed opportunities in marketing and advertising

NICB 'Hot Spots': Auto Thefts Up Significantly Across the Country



Washington, D.C. takes the top spot for 2020 with a theft rate of 562.98, a 40% increase over 2019. Colorado takes the 2nd spot.

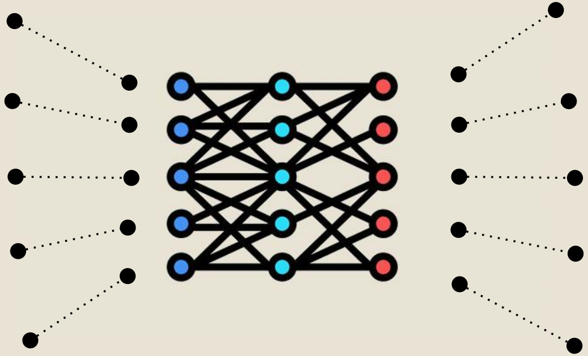
Market Potential: Why this matters?

- Stolen vehicle recovery and speed/toll violations for **Law Enforcement**
- Fraud detection and accident reconstruction in the **Insurance sector**
- Traffic management and security for **Smart Cities** and **Parking Management**

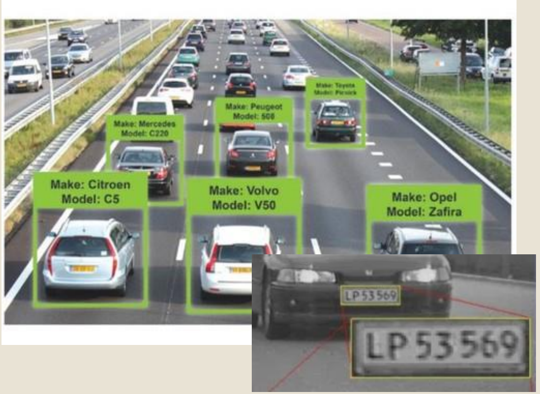
Introducing Deep Vision System for Vehicle Recognition

A computer vision system for vehicle make/model detection provides critical insights for diverse industries like insurance, law enforcement, and marketing, ensuring faster operations, increased accuracy, and significant cost savings.

Input Signals from Camera Sensors



Deep Vision Model



Vehicle Recognition, Identification, and License Plate Detection

Data Collection — Over 40K Vehicle Images

⬇

Exploratory Data Analysis — 391 Classes

⬇

Object Detection — Transfer Learning Methods

⬇

Vehicle Make/Model Recognition — Custom Computer Vision and DL Models

We collected data from Kaggle and Craigslist

Stanford Cars Dataset

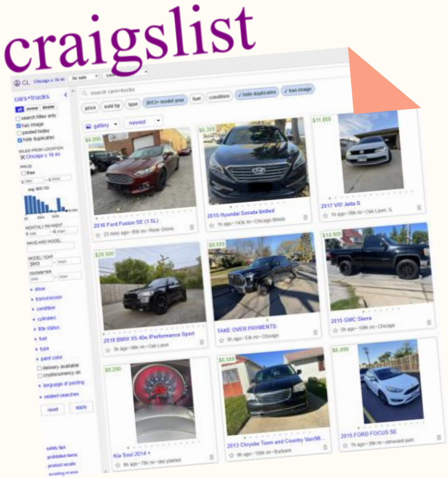
This classic CV dataset contains 16,185 images of 196 classes of cars at the level of Make, Model, Year. However, the dataset is **outdated - last updated in 2012.**

Scraped Craigslist Images

To update and improve the usability of the Stanford Cars Dataset, we scraped **~30,000 images of cars 2013 or newer from 20 cities** across the U.S.

Updated and Expanded Cars Dataset

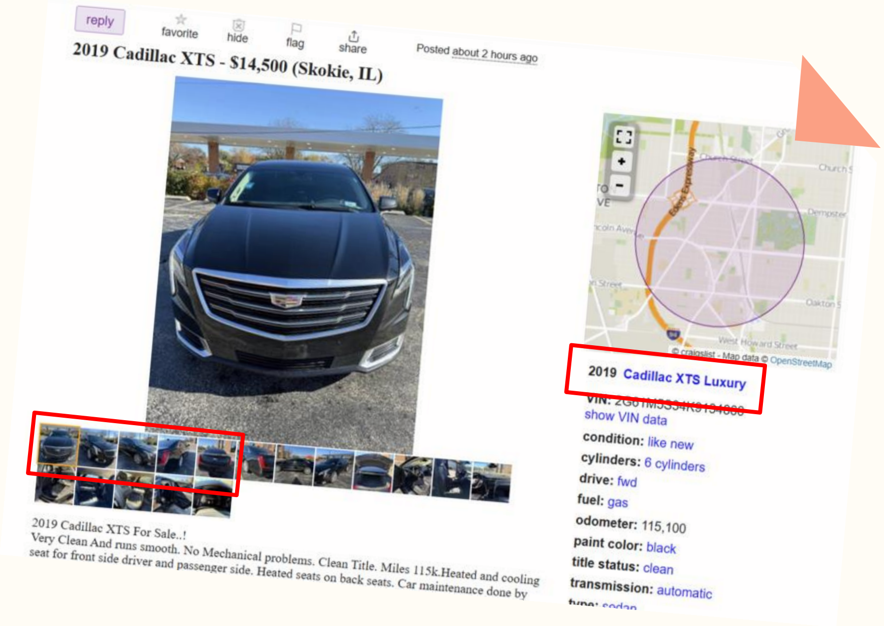
After cleaning and merging, our final dataset contains 29,126 images of 391 classes



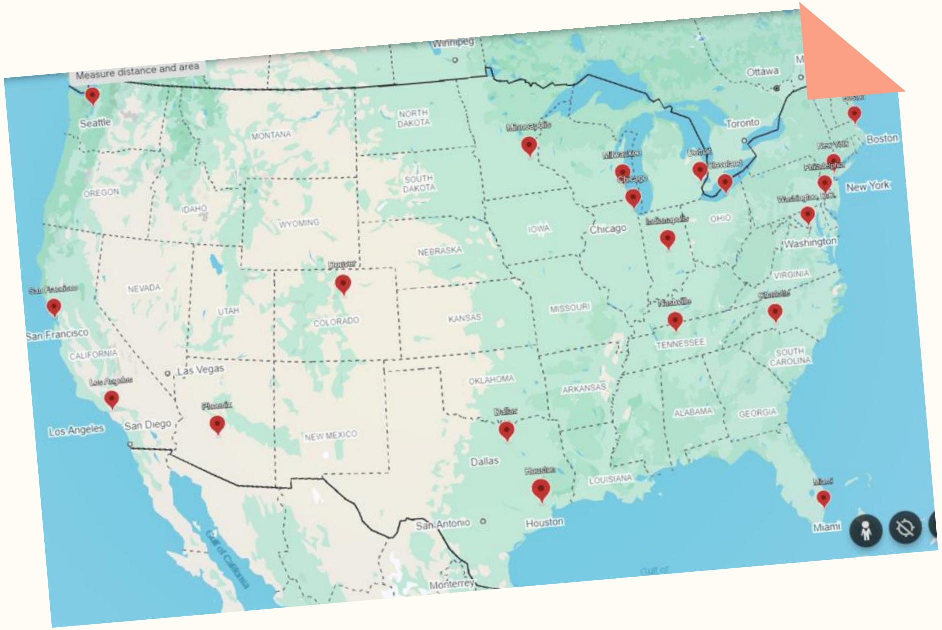
Updated & Expanded Dataset

Image Scraping Process

First 5 Images and Label Were Scraped from Craigslist Postings

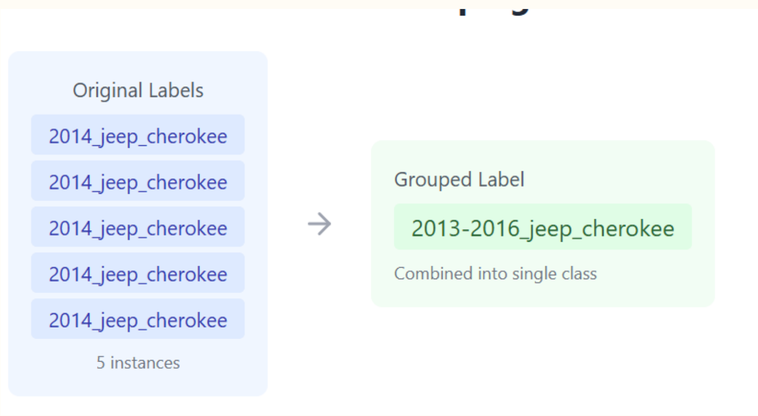


Postings were Collected from 20 Major Cities Across the U.S.



Exploratory Analysis - Labeling

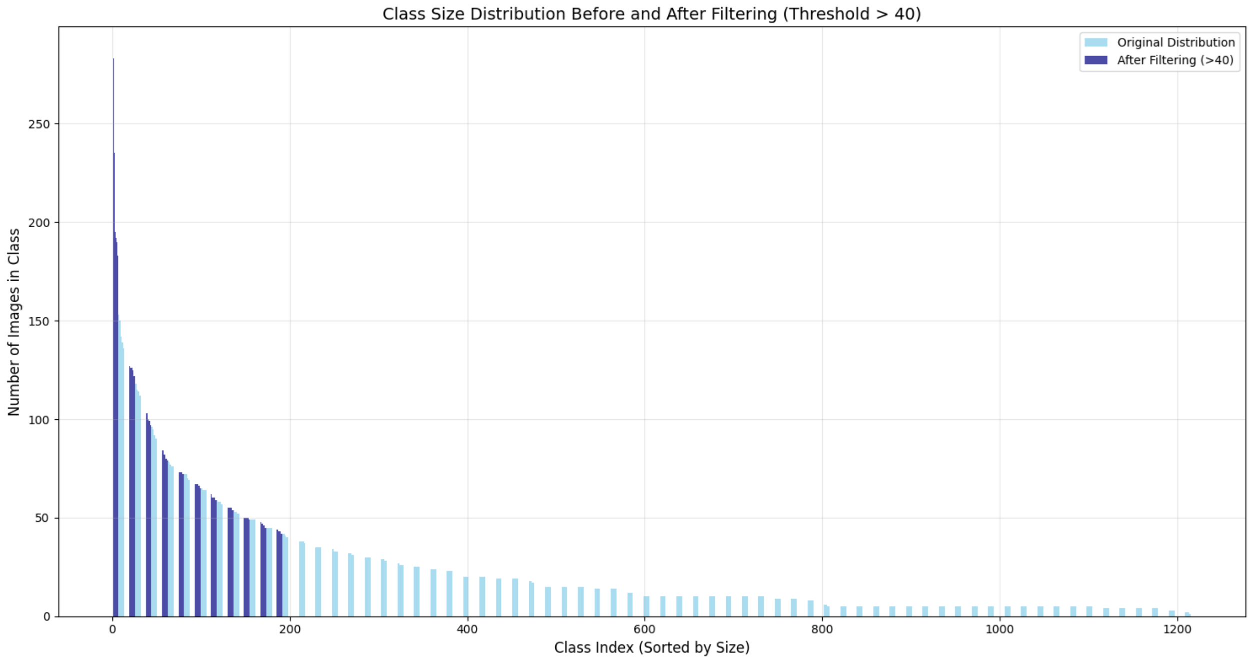
- Label Standardization Process:
  - Created official car database as reference to standardize messy Craigslist listings using fuzzy matching



- Temporal Grouping Strategy:
  - Grouped multi-year models into 4-year ranges to match typical vehicle update cycles, keeping single-year models separate
  - Note: Varying update schedules across manufacturers complicate standardization

Exploratory Analysis - Class & Image Distribution

Class size distribution before and after filtering.



Started with 1217 classes even after grouping by year.

- Filtered scraped data classes to 196 by filtering out classes with < 40 observations

Exploratory Analysis - Image Quality

Acura RL Sedan 2012 (stanford)

2013-16 BMW 3 Series (Scraped)

Problem: How to detect cars in the Craigslist data?

Scraped images from Craigslist are not always of the exterior of the car. Need a way to remove these non-car images.

YOLO (You Only Look Once) Model

- Real-time object detection system
- Extremely fast
- Detects multiple objects and their classes
- Learn generalizable representations

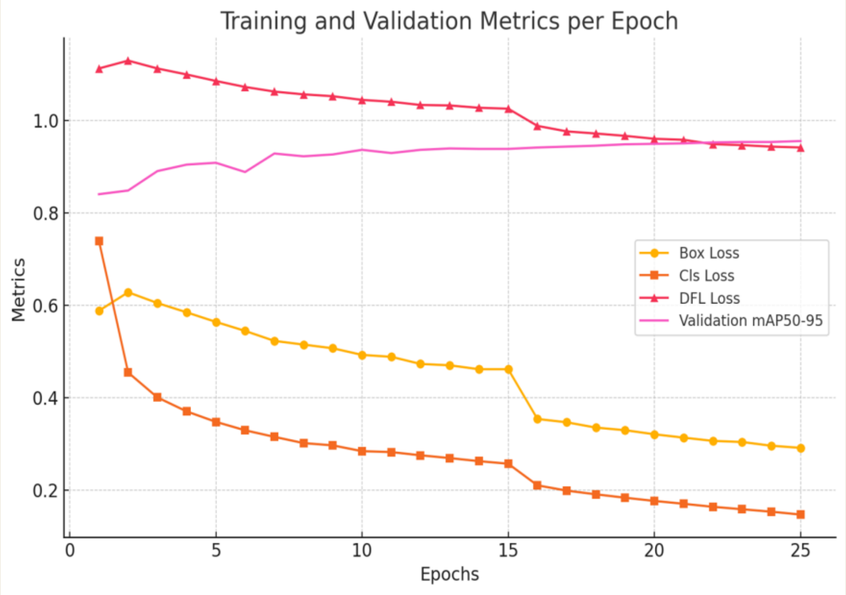
Fine-tuning YOLO

- **Pre-Trained YOLO (yolov8n.pt)**
  - Trained by COCO Dataset
  - A large-scale dataset containing 80 object categories
  - Designed for general-purpose object detection tasks

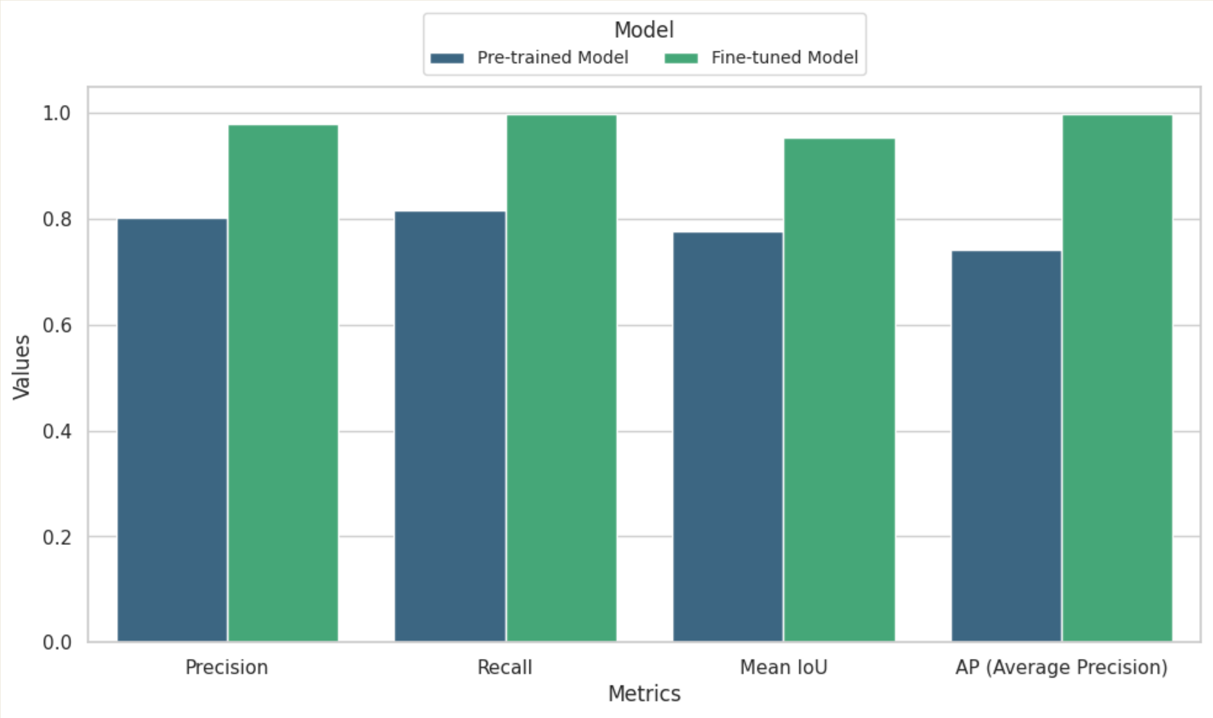
- **Fine-tuned YOLO**
  - Fine-tuned with Train Data from Stanford Cars Dataset
  - Specialize in car detection at the cost of general versatility
  - Tailored to this project's unique focus on images dominated by a single car

Fine-tuning YOLO

- Stanford dataset converted into train and validation set.
- Bounding box annotations (x1, y1, x2, y2) were converted to YOLO format (class_id, x_center, y_center, width, height), normalized by image dimensions.
- YOLO v8 Nano was fine-tuned due to its efficiency and suitability for custom tasks.
- The YOLO detection head was reconfigured to detect a single class (car).
- Model fine-tuned for 25 epochs using a custom data.yaml configuration and saved for inference.



YOLO Results: Pre-trained vs Fine-tuned Model



Fine-tuned YOLO significantly outperforms pre-trained YOLO on our vehicle test dataset.

YOLO Results: Pre-trained vs Fine-tuned Model

Pre-trained YOLO Output | Fine-tuned YOLO Output | Pre-trained YOLO Output | Fine-tuned YOLO Output

Fine-tuned YOLO works better when car model is rare/uncommon

Fine-tuned YOLO works better when there are other secondary cars

Fine-tuned YOLO works better when the image quality is poor

Fine-tuned YOLO works better when the car is viewed from the side/rear

Problem: Vehicle Classification

With the immense variety of vehicles worldwide (and their frequent similarities in appearance)—spanning brands, models, years, and configurations—this task becomes highly complex and non-trivial

2013-16 Honda Pilot

2013-16 Hyundai Sonata

2013-16 Chevy Traverse

Transfer Learning Modeling Approach

Resnet50: “Champion”

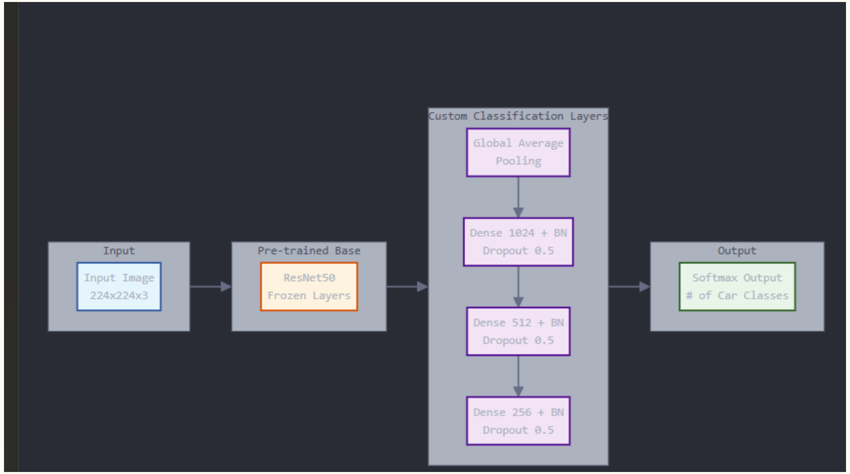
- Advantages
  - Superior accuracy on fine details
  - Deep feature extraction capability
  - Robust residual connections
  - Better at complex features
- Limitations
  - Higher computational cost
  - Larger memory footprint
  - Slower inference speed

MobileNet: “Challenger”

- Advantages
  - Fast inference speed
  - Lightweight deployment
  - Low memory footprint
  - Edge device compatible
  - Efficient depthwise separable connections
- Limitations
  - Lower accuracy on details
  - Basic feature extraction
  - Limited architecture depth

ResNet50 Best Model

- **Key Parameters:** Batch size (32), learning rate (0.001), Dropout (0.5), Batch Normalization, and early stopping to prevent overfitting
- **Transfer Learning:** Pre-trained on ImageNet with robust automotive feature extraction, preserving crucial brand-specific details
- **Optimization:** Adam optimizer with adaptive reduction, early stopping, and comprehensive augmentation (rotation, shift, zoom, flip)
- **Model Efficiency:** Global Average Pooling and skip connections enable deep feature learning while maintaining reasonable memory footprint




```
Total params: 26,351,754 (100.52 MB)
Trainable params: 2,760,458 (10.53 MB)
Non-trainable params: 23,591,296 (89.99 MB)
```

ResNet50 Results

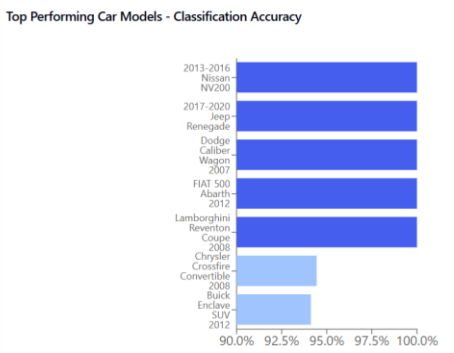
Model Performance Overview

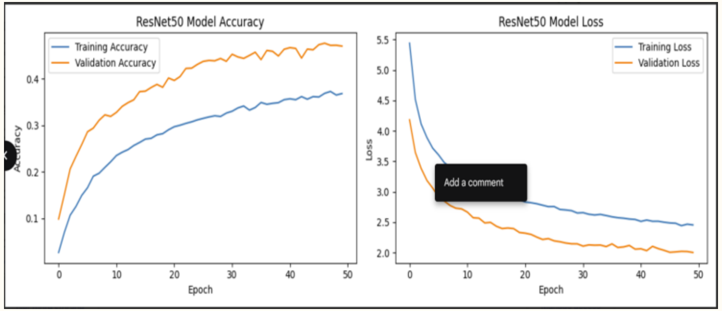
45.7%
Overall Accuracy

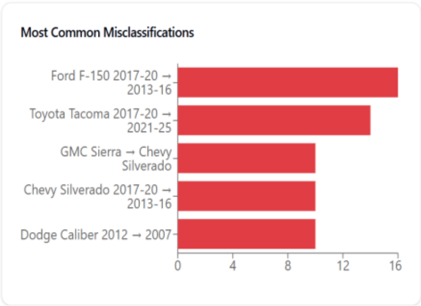
5,932
Total Samples

2,899
Correct Predictions

3,033
Incorrect Predictions







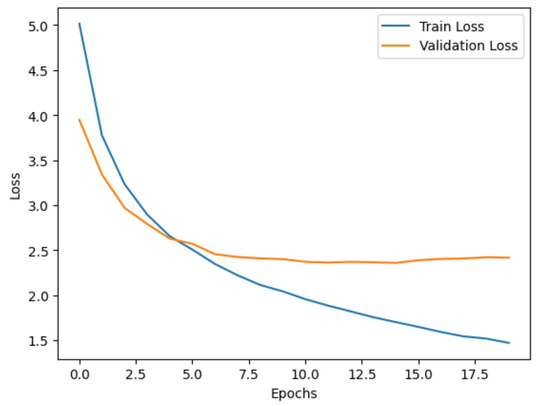
MobileNet Modeling

MobileNet's lightweight architecture is optimized for efficiency and speed which makes it ideal for real-time vehicle classification in the field

Base Model

Layer (type)	Output Shape	Param #
mobilenet_1.00_224 (Functional)	(None, 7, 7, 1024)	3,228,864
global_average_pooling2d_2 (GlobalAveragePooling2D)	(None, 1024)	0
dense_4 (Dense)	(None, 512)	524,800
dropout_2 (Dropout)	(None, 512)	0
dense_5 (Dense)	(None, 391)	200,583

Total params: 5,405,015 (20.62 MB)
Trainable params: 725,383 (2.77 MB)
Non-trainable params: 3,228,864 (12.32 MB)
Optimizer params: 1,450,768 (5.53 MB)



MobileNet (no classification layer) + Pooling, Dense, Dropout (0.3), Softmax (391 classes) → Test Loss: 2.126
Test Accuracy: 46.96%

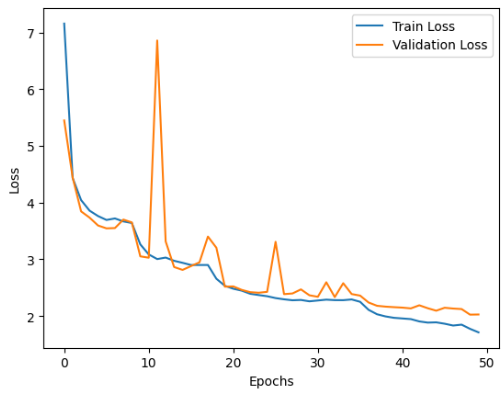
MobileNet Modeling

We optimized the base model to address the overfitting we observed

Advanced Model

Layer (type)	Output Shape	Param #
mobilenet_1.00_224 (Functional)	(None, 7, 7, 1024)	3,228,864
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1024)	0
dense (Dense)	(None, 512)	524,800
batch_normalization (BatchNormalization)	(None, 512)	2,048
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 391)	200,583

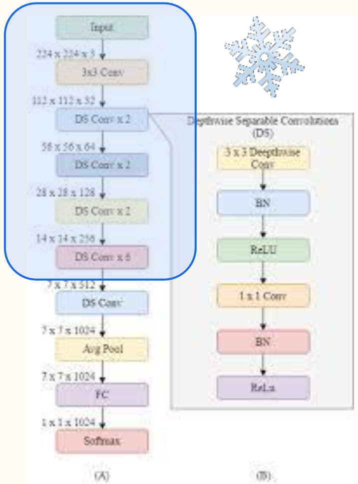
Total params: 5,409,111 (20.63 MB)
Trainable params: 726,407 (2.77 MB)
Non-trainable params: 3,229,888 (12.32 MB)
Optimizer params: 1,452,816 (5.54 MB)



Data Augmentation + Base Model + BatchNorm, l2 Reg. Dropout (0.4), LR scheduling, 50 epochs → Test Loss: 1.916 Test Accuracy: 60.0%

MobileNet Modeling

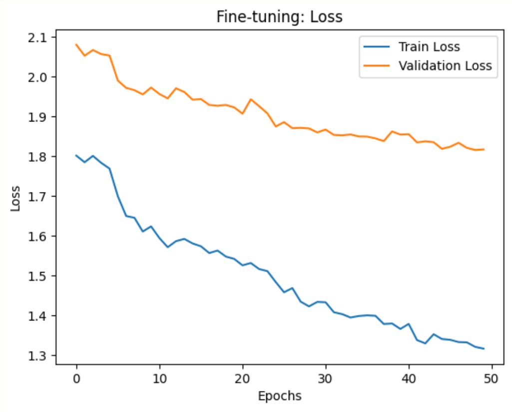
Lastly, we unfroze the last 16 layers of MobileNet (without classification head) to fine-tune the model and learn the finer features of the car images



Fine-tuned Advanced Model

Layer (type)	Output Shape	Param #
mobilenet_1.00_224 (Functional)	(None, 7, 7, 1024)	3,228,864
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1024)	0
dense (Dense)	(None, 512)	524,800
batch_normalization (BatchNormalization)	(None, 512)	2,048
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 391)	200,583

Total params: 5,409,111 (20.63 MB)
Trainable params: 726,407 (2.77 MB)
Non-trainable params: 3,229,888 (12.32 MB)
Optimizer params: 1,452,816 (5.54 MB)



MobileNet with last 16 layers unfrozen

Advanced Model (w/ lowered LR)

Test Loss: 1.701
Test Accuracy: 62.84%

Modeling Results: Stanford vs Scraped Classes

ResNet50

62.65% Stanford Overall Accuracy	**32.01%** Scraped Overall Accuracy
195 Stanford Total Classes	**196** Scraped Total Classes
3,264 Stanford Total Samples	**2,668** Scraped Total Samples

MobileNet

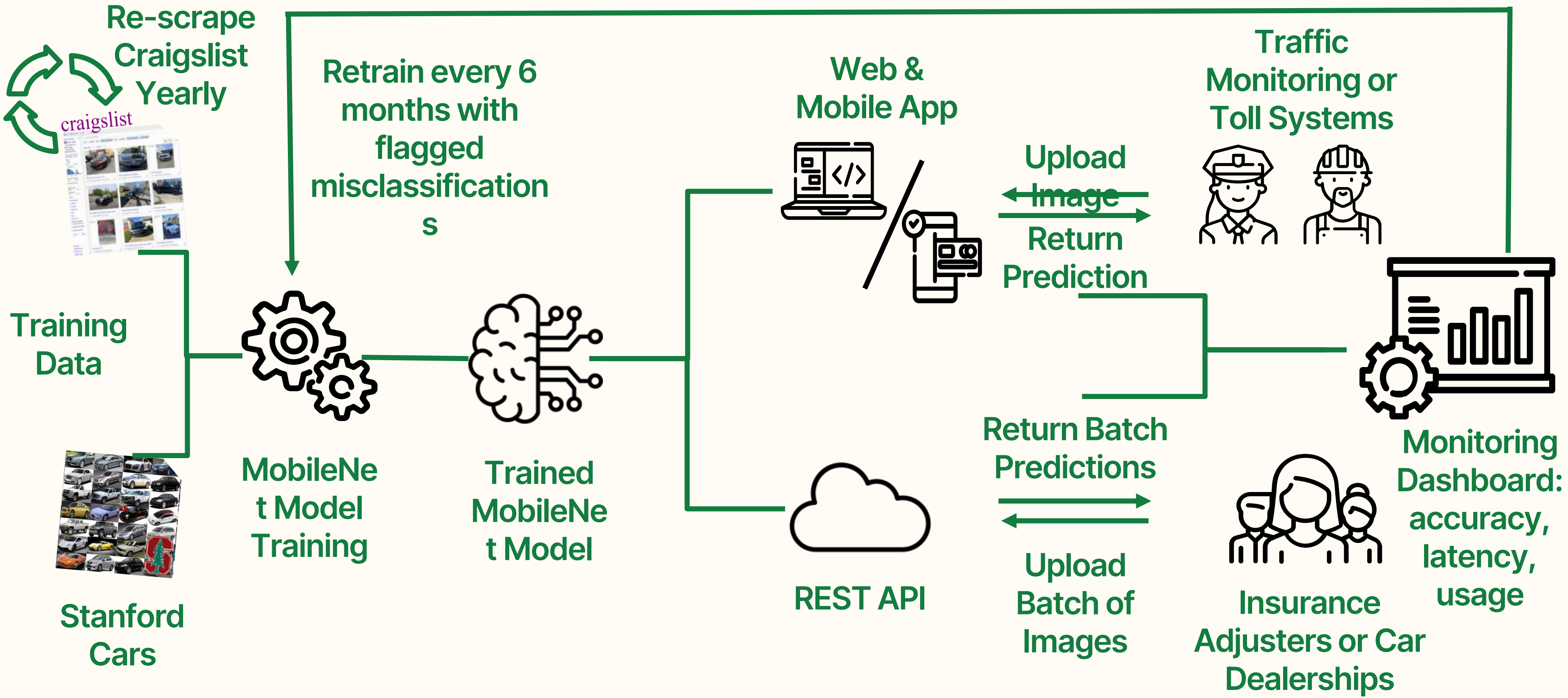
74.72% Stanford Overall Accuracy	**48.44%** Scraped Overall Accuracy
195 Stanford Total Classes	**196** Scraped Total Classes
3,264 Stanford Total Samples	**2,668** Scraped Total Samples

Live Demo

We created a simple Streamlit App to give a better understanding of how users can utilize our models

Link to App

Deployment & Maintenance Plan



Future Work

To improve classification accuracy and robustness, future efforts will focus on:

Dataset Improvement

- Double the number of images
  - 30,000 -> 60,000
- Address class imbalances
  - Min. 30 images per class
- Use single year groupings
- Diversify sources

Facebook Marketplace

cars & bids

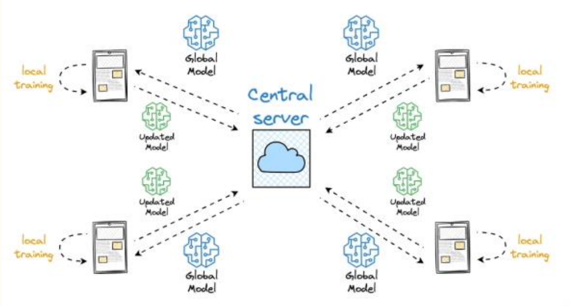
ebay

New Model Architectures

- Vision Transformers
  - As dataset grows, ViT architectures become more viable

Integrate user data

- Federated Learning for privacy
  - Incorporate user-uploaded images into the training set while maintaining privacy



Wrap-Up: Key Takeaways

Many Use Cases for Vehicle Detection & Classification

- Law Enforcement:
  - Identify vehicles of interest using traffic cams.
- Insurance Fraud:
  - Automate damage assessments for faster claims
- Toll Systems:
  - Recognize vehicle make/model for automated billing

Dataset Expansion Efforts

- Data Collection:
  - Doubled the Stanford dataset by scraping 30,000+ images from Craigslist
- Updated Data:
  - Expanded to include modern cars

Model Training & Deployment Achievements

- Models:
  - Fine-tuned YOLO for object detection
  - MobileNet for classification
- Performance:
  - Achieved 62% classification accuracy
- Deployment:
  - Developed a Streamlit app for real-world use

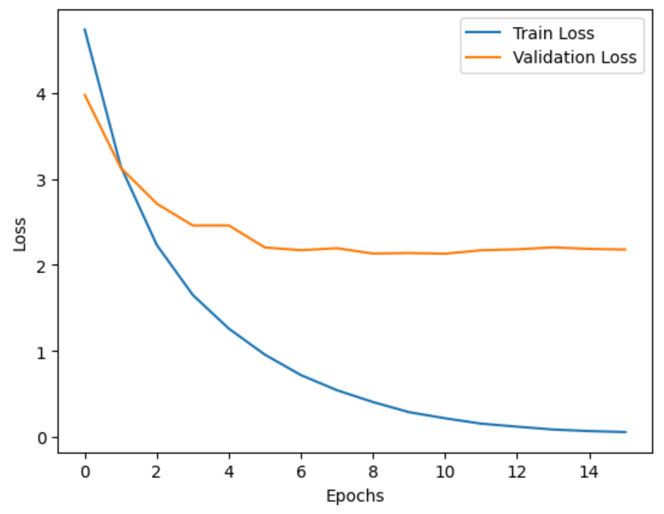
References

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3D object representations for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 554–561). IEEE. https://doi.org/10.1109/ICCVW.2013.77

Appendix

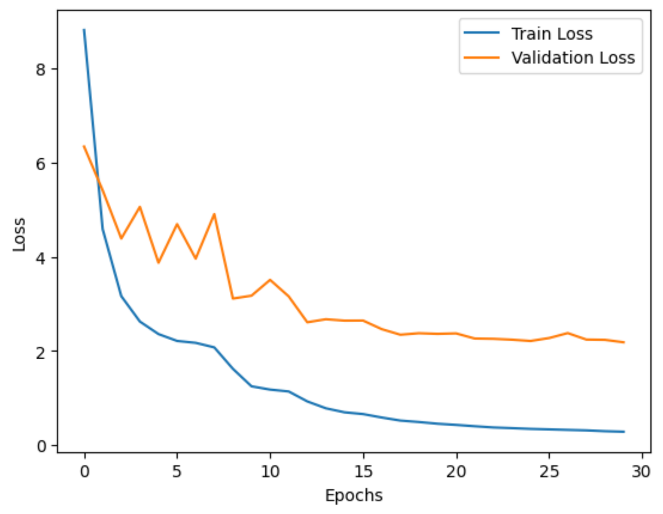
Stanford Dataset MobileNet Modeling

Base Model



Test Loss: 2.07
Test Accuracy: 48.70%

Advanced Model

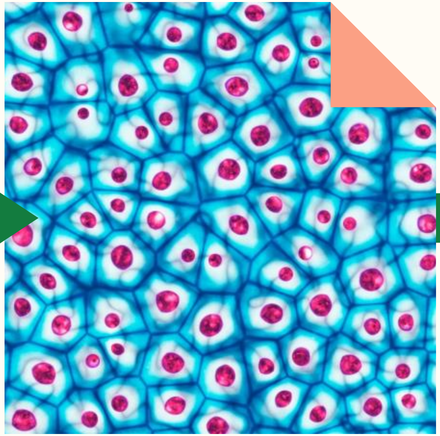


Test Loss: 2.15
Test Accuracy: 53.05%

ResNet50 Results: Stanford vs Scraped

62.65% Stanford Overall Accuracy	**32.01%** Scraped Overall Accuracy
195 Stanford Total Classes	**196** Scraped Total Classes
3,264 Stanford Total Samples	**2,668** Scraped Total Samples

Amino Acids & Proteins



Tissue

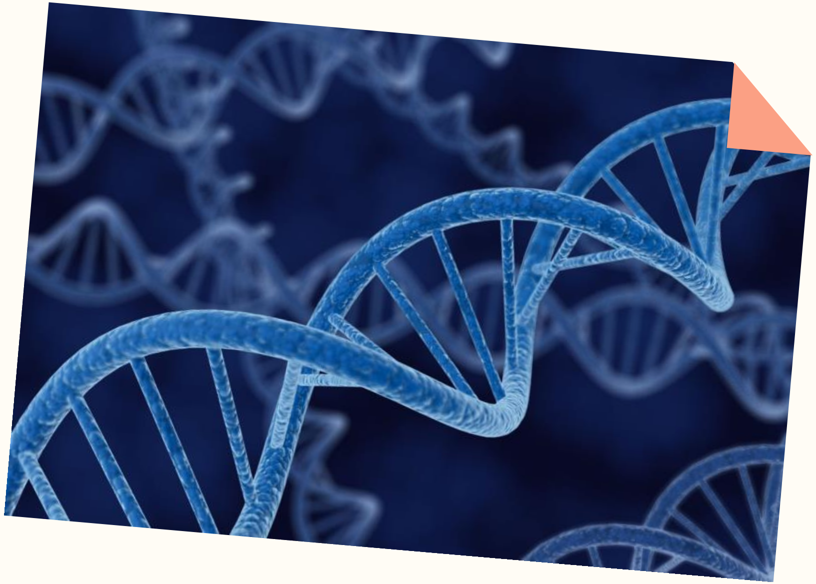
Organs

Living Creatures

DNA & Proteins

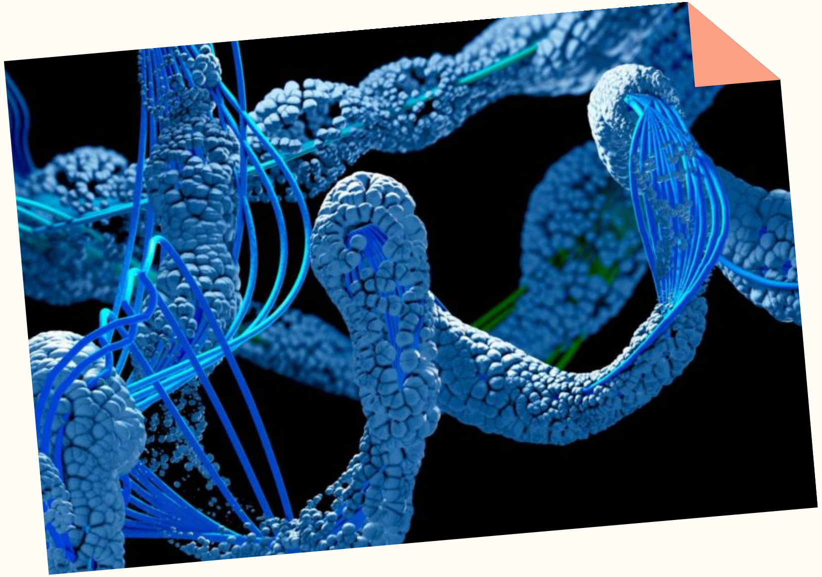
DNA

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.



Proteins

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.



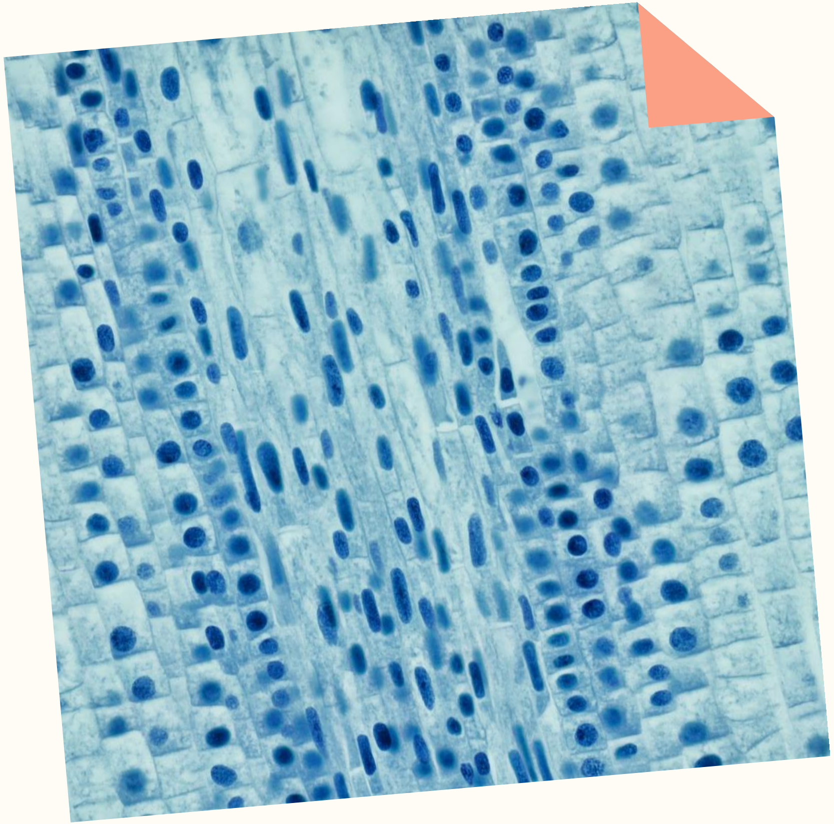
Recap

- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet

- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet
- Lorem ipsum dolor sit amet

Understanding DNA

Lorem ipsum dolor sit amet, consectetur adipiscing elit?	yes	no
Lorem ipsum dolor sit amet, consectetur adipiscing elit?	yes	no
Lorem ipsum dolor sit amet, consectetur adipiscing elit?	yes	no
Lorem ipsum dolor sit amet, consectetur adipiscing elit?	yes	no



Next Lesson

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam.