

# DRIVEN DATA

Intelligent Vehicle  
Identification for Market  
Data Generation



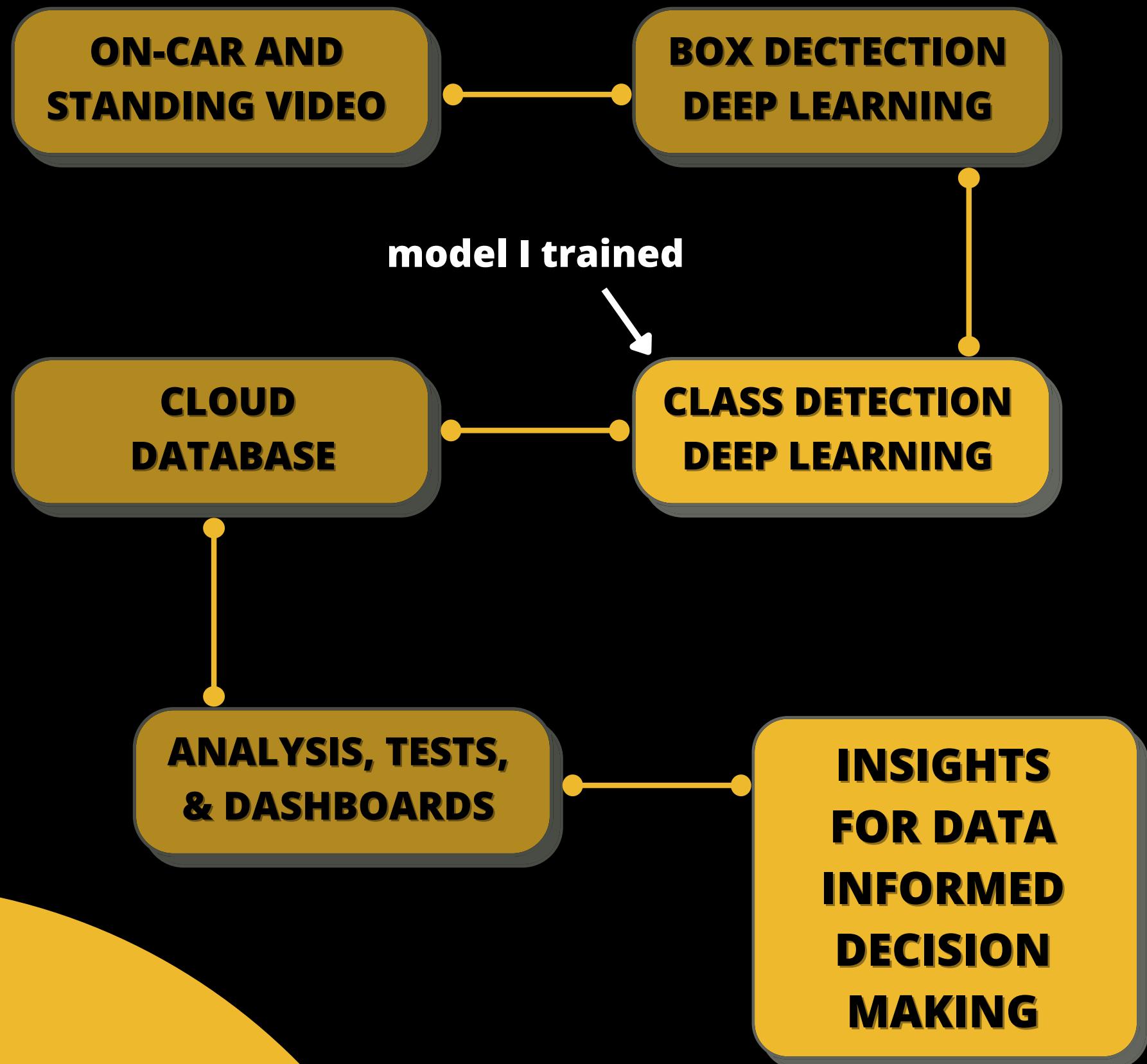
Chase Kent-Dotson

# THE INFORMATION CHALLENGE

**HOW DOES A COMPANY OBTAIN ON-ROAD MARKET DATA FOR ALL BRANDS IN THE MARKET?**

- Sales data, customer address data, and vehicle GPS provide some insights on a company's own vehicles
- Information on competing brands is not publicly available and difficult to generate via traditional means
- This presents a perfect opportunity for a machine vision solution with a big data backbone





# THE SOLUTION

- Use deep learning to identify vehicle make, model, year, etc. from on-car camera images and record location data
- Send this data using vehicle Wi-Fi to a national database stored in a commercial cloud computing service
- Data engineers and data scientists can work together to extract valuable insights from the database
- These insights help drive data informed decision making

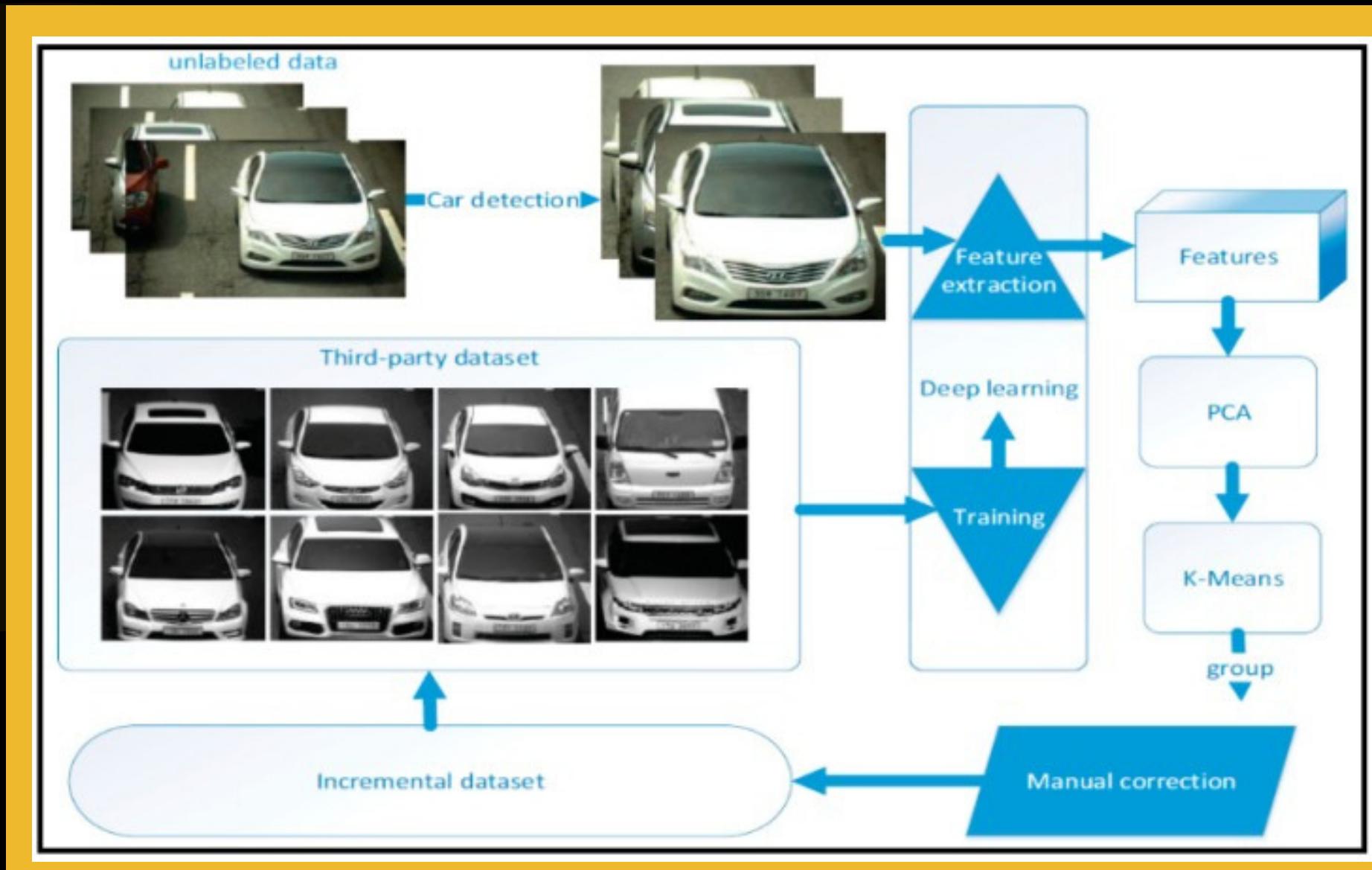
# EXISTING WORK

## MODEL TYPES

- One model is needed to find cars within images
- One model is needed to determine what kind of cars the image contains

## STATE OF THE ART SOLUTIONS

- Lots of images are needed for production scale model training
- Other ML techniques on top of deep learning can reduce the labeling labor requirements



[Image and information source here](#)

# THE DATASET

## STANFORD AI DATASET

- 16,000 vehicle images
- 196 classes, in the form "Make Model Type Year," e.g. "Audi S5 Sedan 2012"
- Half train data, half test data
- Classes are well balanced for model training
- Very accurate labels, zero mistakes were found throughout the course of the project



[Data source here](#)

# MODEL TRAINING

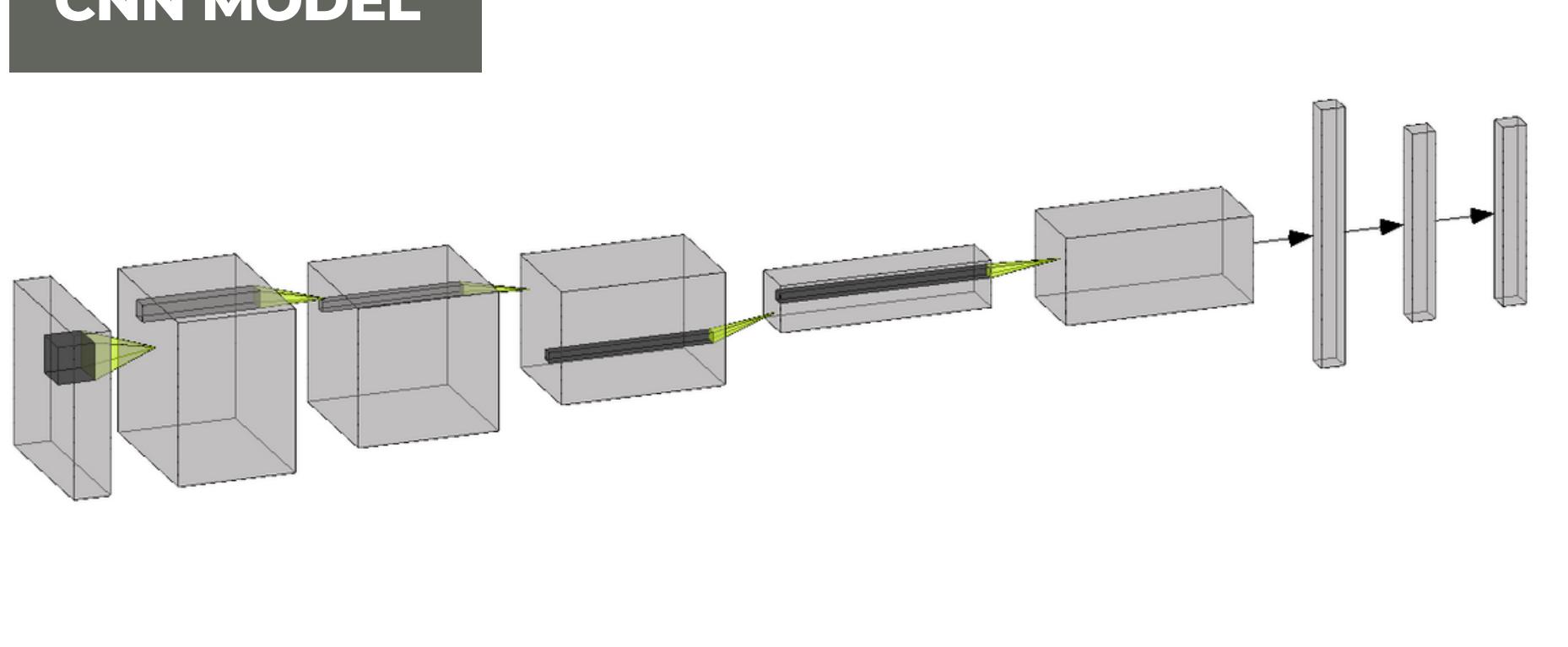
## DATA PREPARATION

- all images were cropped to their bounding box limits to minimize impact of environment imagery on the model
- Scaling, normalization, and random augmentation were used to achieve the best model results

## MODELS ATTEMPTED

- Initially CNNs assembled from scratch were trained and optimized
- In search of greater predictive power for generating market data, the final solution uses transfer learning with pre-trained weights

## CNN MODEL



# RESULTS EVALUATION

## STRENGTHS & WEAKNESSES

- The model very precisely identifies boxy-shaped vehicles with unique features
- The model is least precise at identifying round-shaped vehicles with generic features
- Color may also play a role

## FINAL ACCURACY SCORES

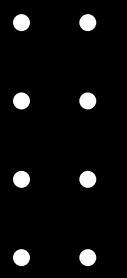
**77%** make, model, and year identification

**86%** vehicle make identification

**90%** body type identification

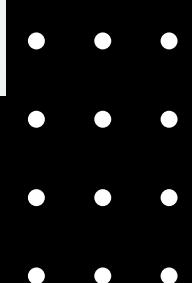


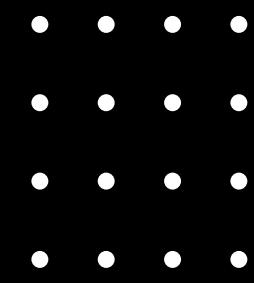
[Data source here](#)



# NEXT STEPS

- Original goal of creating a model capable of predicting vehicle make, model, and year with greater than 70% accuracy was achieved
- Placing the model I trained at the heart of a system to label additional training data in an automated way would improve accuracy drastically
- The production quality model could then be coupled with a vehicle bounding box finding model to begin producing data
- The data would eventually be analyzed and used to help the company make smarter data informed decisions





# THANK YOU

## DRIVEN DATA

by Chase Kent-Dotson  
[chasekd@gmail.com](mailto:chasekd@gmail.com)

