Exploring the Potential of Image-based Deep Learning in Insurance

Luisa F. Polanía Cabrera

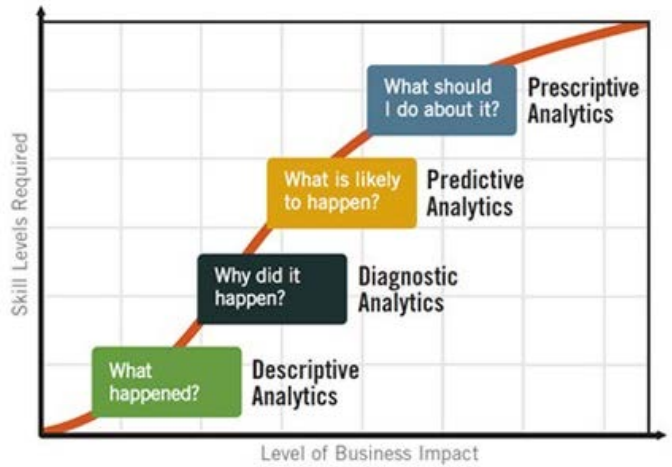
- Madison, Wisconsin – based American Family Insurance is the nation's third-largest mutual property/casualty insurance company .
- The company sells American Family-brand products, including auto, homeowners, life, business and farm/ranch insurance, through its agents in 19 states.
- American Family affiliates (The General, Homesite and AssureStart) also provide options for consumers who want to manage their insurance matters directly over the Internet or by phone.

Strategic Data & Analytics

In order to accelerate advancement of big data analytics capabilities and business transformation, AmFam created a center of excellence for predictive analytics

Mission

To create competitive advantage and economic value by bringing together data innovation, advanced analytics and business acumen to optimize or transform our business models.



Expertise

- Natural language processing
- Advanced modeling: Random forests, SVMs, etc.
- Probabilistic graphical models and Bayesian techniques
- Big data – parallel processing
- Deep learning

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

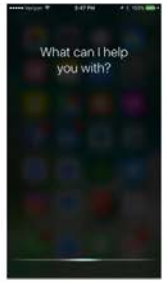
Deep learning

Near-human accuracy speech recognition and image recognition systems

From machines that could not beat a serious Go player, to beating a world champion.

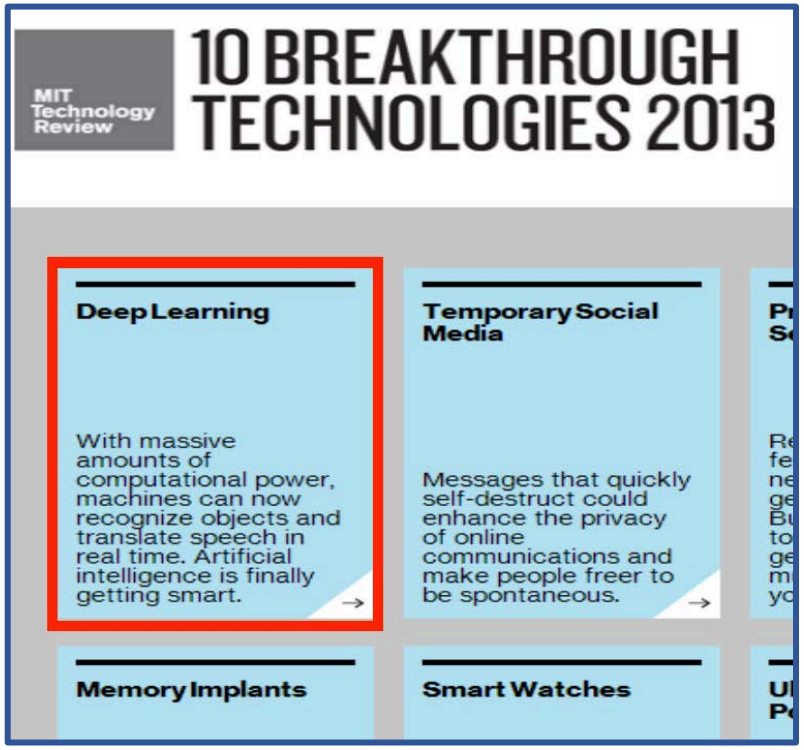
AMAZON'S ALEXA

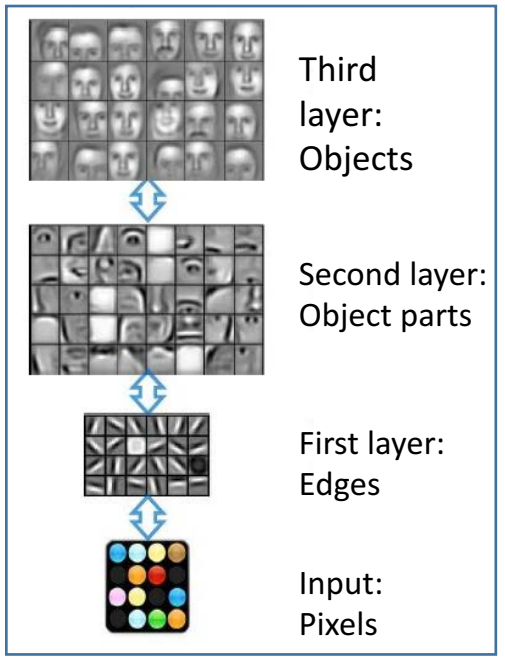
GOOGLE'S ASSISTANT



APPLE'S SIRI

MICROSOFT'S CORTANA





AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Image Processing and Computer Vision in Insurance

Roof Condition Prediction

BMI Classification using Selfie Images

Automatic odometer reading from images

Blurring of human faces from drone images.

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Roof Condition Prediction

Use case:

- There is a high likelihood that the roof age captured on the AmFam book of business, for property, is inaccurate (self-reported roof age), which results on risks not entirely assessed when insuring a property.
- There is a latent need to build predictive roof age/condition models.

Data from Millennium Inspection

Roof Condition

- A: Roof has 16+ years of remaining life.
- B: Roof has 11-15 years of remaining life.
- C: Roof has 6-10 years of remaining life.
- D: Roof has 0-5 years of remaining life.

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Challenges

- Uncontrolled illumination.
- Scale variation.
- Capture angle variation.
- Roof material variation.
- Noise artifacts, shadows, occlusions.

Roof Condition Prediction

Data from Millennium Inspection: In-the-wild data set (16M images)

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Filtering of Millennium Survey Images

Filtering with VGG: 1% of the total images are classified as roof images by VGG.

Additional Filtering

- We collected 1000 labels to classify images between good and bad images for roof image analysis. NEXT[1] was employed to collect the labels.

Examples of bad quality images for roof analysis


Low-illuminated roof area


Roof area is almost completely occluded by branches and leaves


Blurry roof region


Overexposed image


Roof area has low resolution (roof details are not clearly visible)


Small roof region, low-illuminated roof

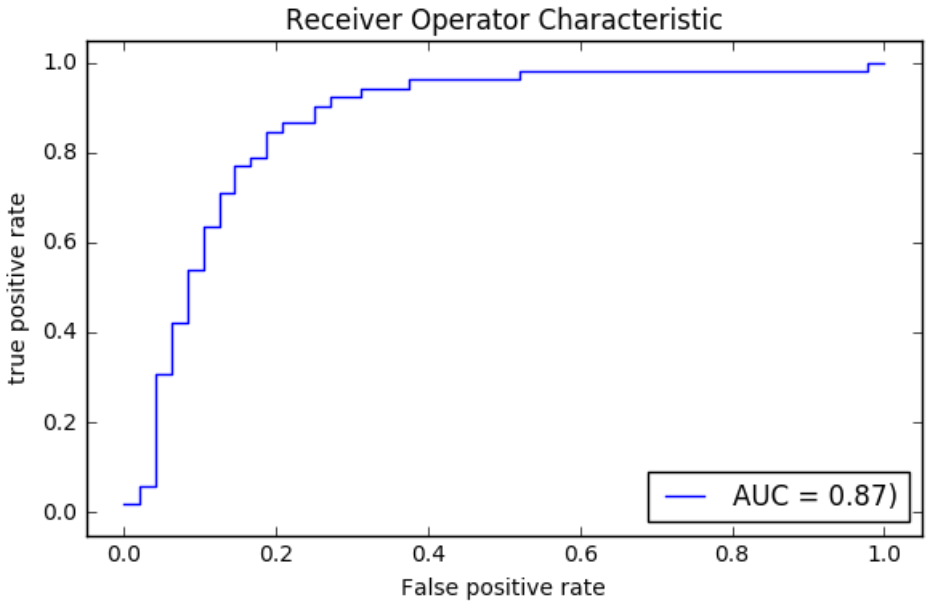

Roof covered by shadows


Non-roof images

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

[1] Jamieson, Kevin G., et al. "Next: A system for real-world development, evaluation, and application of active learning." *Advances in Neural Information Processing Systems*. 2015.

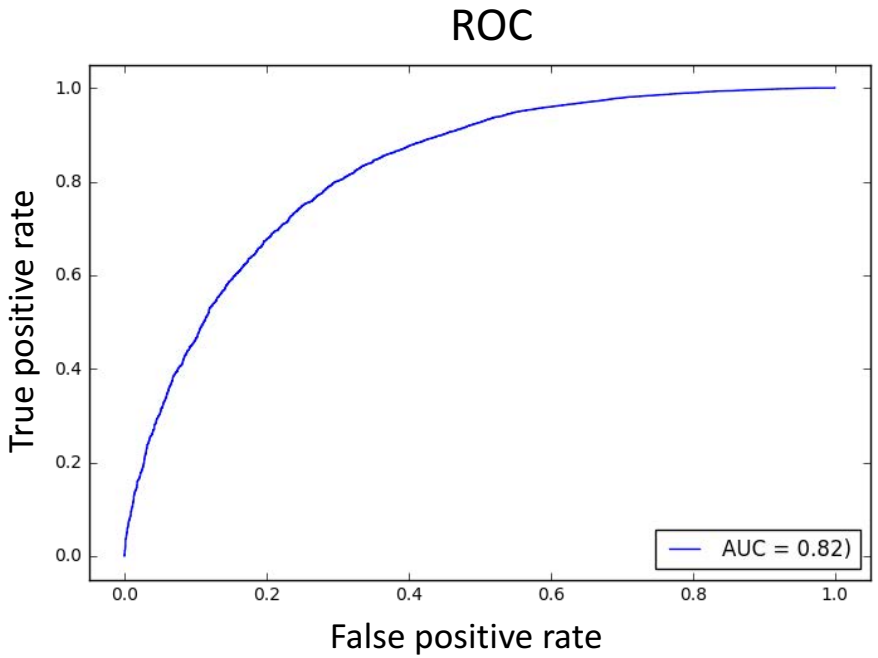
Additional Filtering



AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Roof Condition Prediction

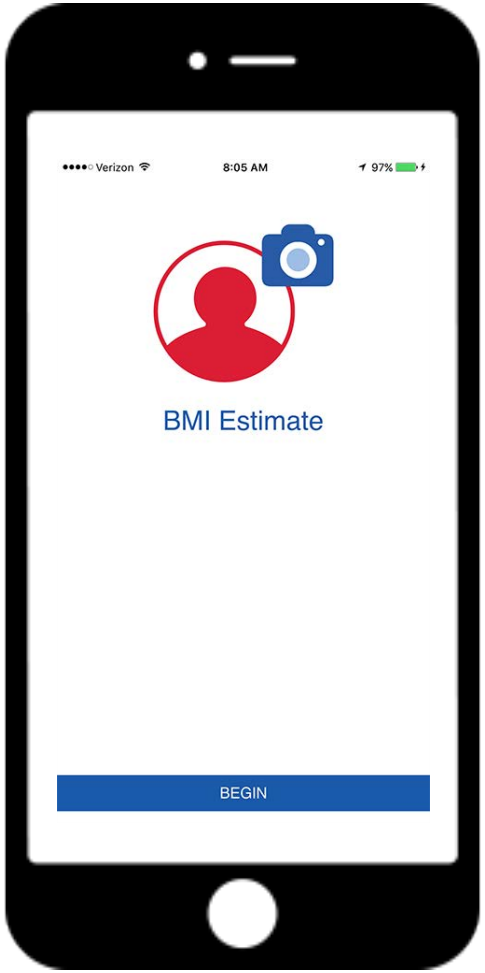
Filtered roof images → Deep neural network for roof condition prediction → Roof condition prediction



Next steps:

- Detection of signs of deterioration: missing shingles, curling shingles, lifting shingles, cupping, etc.

BMI Classification from Selfie Images



Use case:

- One of the factors that determine life insurance rates is body mass index (BMI).
- According to the Centers for Disease Control and Prevention, the correlation between a high BMI and the likelihood of developing dangerous health conditions is strong(1).
- Self-reported BMI is inaccurate.
- Challenge: Verify customer BMI without a paramedical exam.

(1) http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html - July 11, 2014

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

BMI Classification using Selfie Images

Mobile Data Capture App

- *Image guidance to acquire good quality frontal images.*
- *Image quality control: automatic rejection of blurry images, images with wrong head orientation, etc.*

Model for Predicting BMI that combines deep learning features with geometric features.

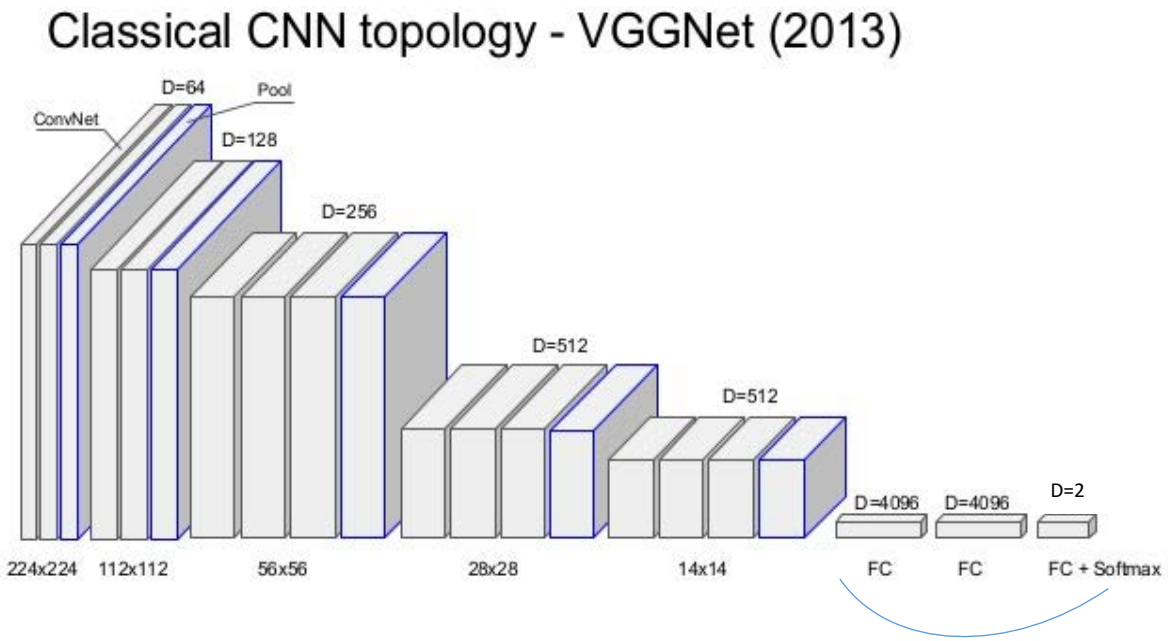
Binary classification: BMI<=30 and BMI>30

Fine-tuning of the VGG Face Model using images from the Florida Department of Corrections

Calculation of distances using facial landmarks

AMERICAN FAMILY INSURANCE® | STRATEGIC DATA & ANALYTICS

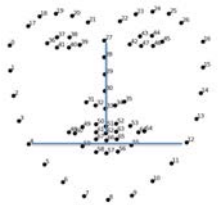
Fine-tuning of the VGG Face Model



- Thirteen convolutional layers and three fully-connected layers.
- Pre-trained model: 2.6M images.
- Dataset: 99096 images. BMI<30: 78%, BMI>=30: 22%.
- Training/Validation partitions: 70% / 30%

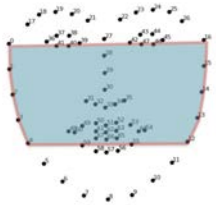
Geometric features

1



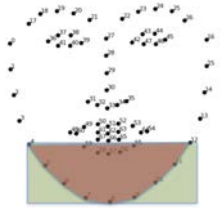
Eyes-to-Lips/Lower Width

4



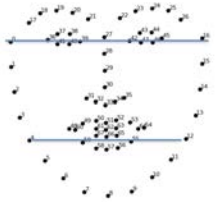
Perimeter Red/Area Blue

3



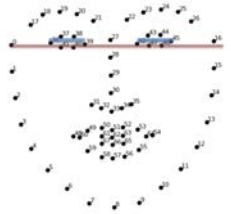
Area Green/Area Brown

0

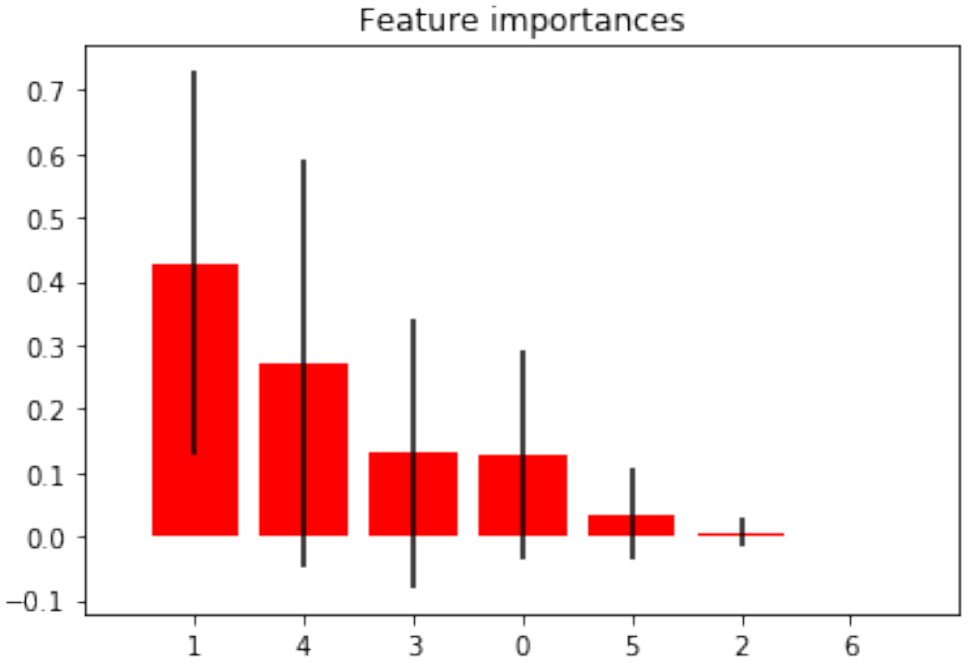


Upper Width/Lower Width

5



Average Eye Width/Upper Width



AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

BMI Classification

Selfie images

Preprocessing:

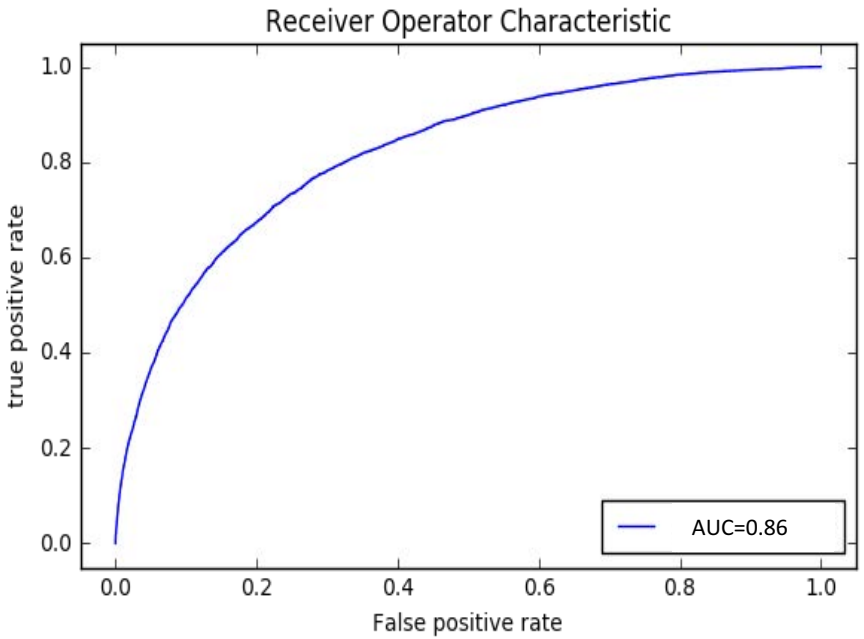
- Face detection and face cropping
- Landmark detection

Feature-level fusion:

- VGG features
- Gender information
- Geometric features

SVM classifier

Prediction



AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Automatic Odometer Reading

Use case:

- After a car accident, in addition to the pictures of the damaged areas, the adjuster also needs to take pictures of the license plate, VIN, and odometer. Then, they need to manually insert the information from these pictures into the claim system to receive a damage estimate.
- Automatic odometer reading would lead to time savings in the claim process and reduction in manual error annotations.

Output: 151018

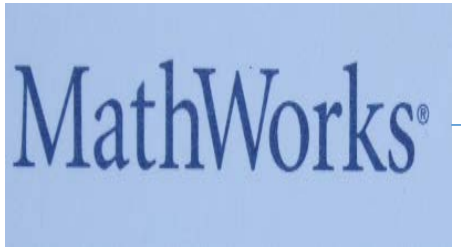
Dataset size: 6300 images

Examples of images in the dataset:

Challenges: Noisy images, non-uniform illumination, random noise, motion blur, odometer type variability (analog and digital).

AMERICAN FAMILY INSURANCE | STRATEGIC DATA & ANALYTICS

Character Recognition



Matlab OCR → MathWorks

Tesseract OCR → The MathWorks, Inc‘,3 Apple Hill DriveNatick, MA 01760-2098

Matlab OCR

Tesseract OCR

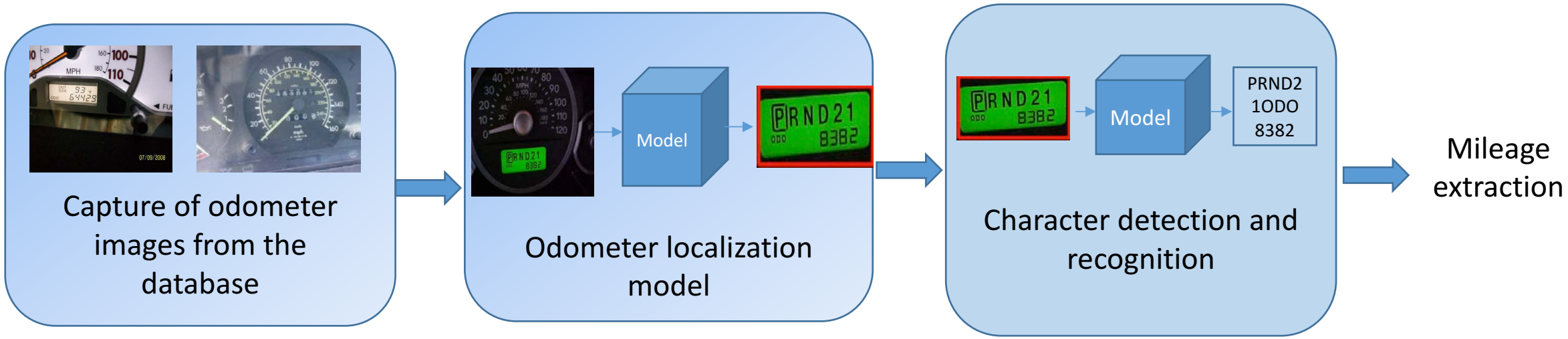
Matlab OCR → ':' ':c3M T7P ' '93.9’

Tesseract → 935391;? 93 ;:

Matlab → 4158035

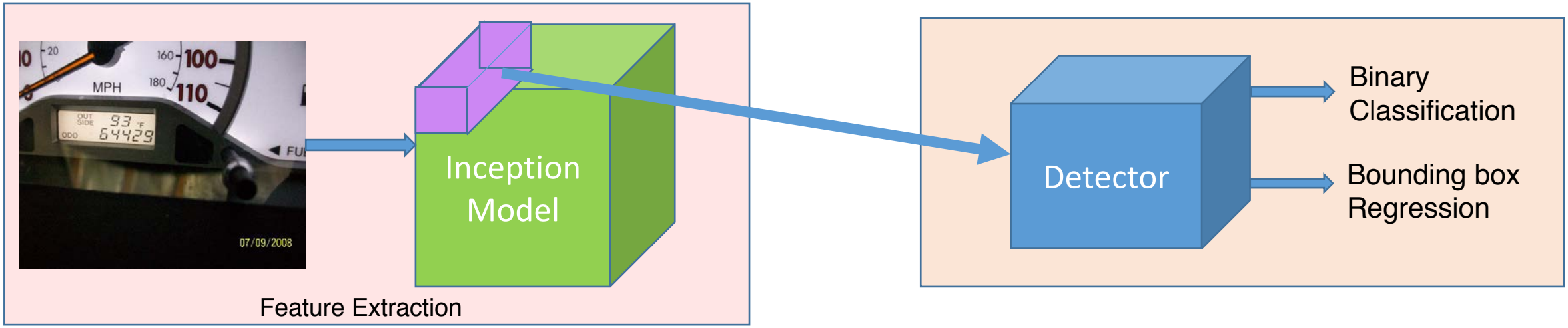
Tesseract → 415803.

Project Pipeline

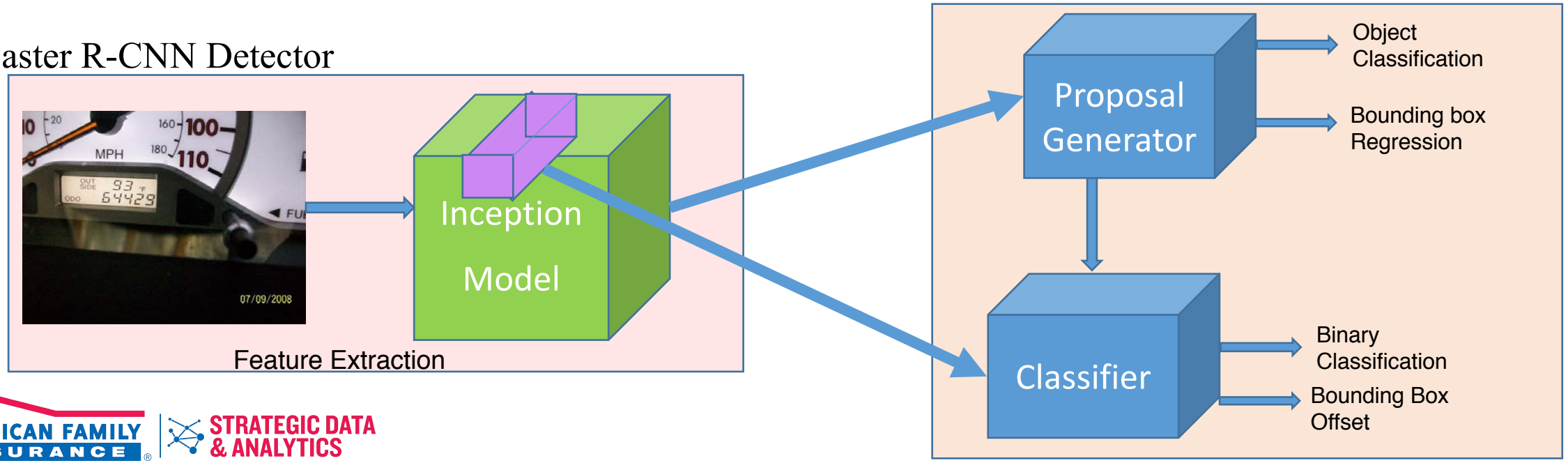


Models for Object Localization

Single Shot Multi Box Detector (SSD)



Faster R-CNN Detector



Odometer Localization Results

Results: 90% of the odometers in the images were correctly localized using Faster R-CNN.

True Positives

False Positives

Final Remarks

Deep learning is driving computing innovation as the insurance industry sets its sights on artificial intelligence.

In this talk, we described different insurance use cases for deep learning:

- Roof condition prediction from images for property underwriting.
- BMI classification from selfie images for life insurance.
- Automatic odometer reading from images to accelerate claim processing.


Q
&
A