CoNNOR: Convolutional Neural Network for Outsole Recognition

Miranda Tilton

CSAFE, Iowa State University

November 30, 2018

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Background: Footwear and Forensic Analysis

- Shoe prints and outsole characteristics are pattern evidence
- May want to identify print at crime scene as made from a certain type of shoe or by a given suspect
- Also useful to determine the frequency of a given shoe (or features of the shoe) in a local population
- **Goal:** Develop an efficient method to automatically identify footwear outsole characteristics

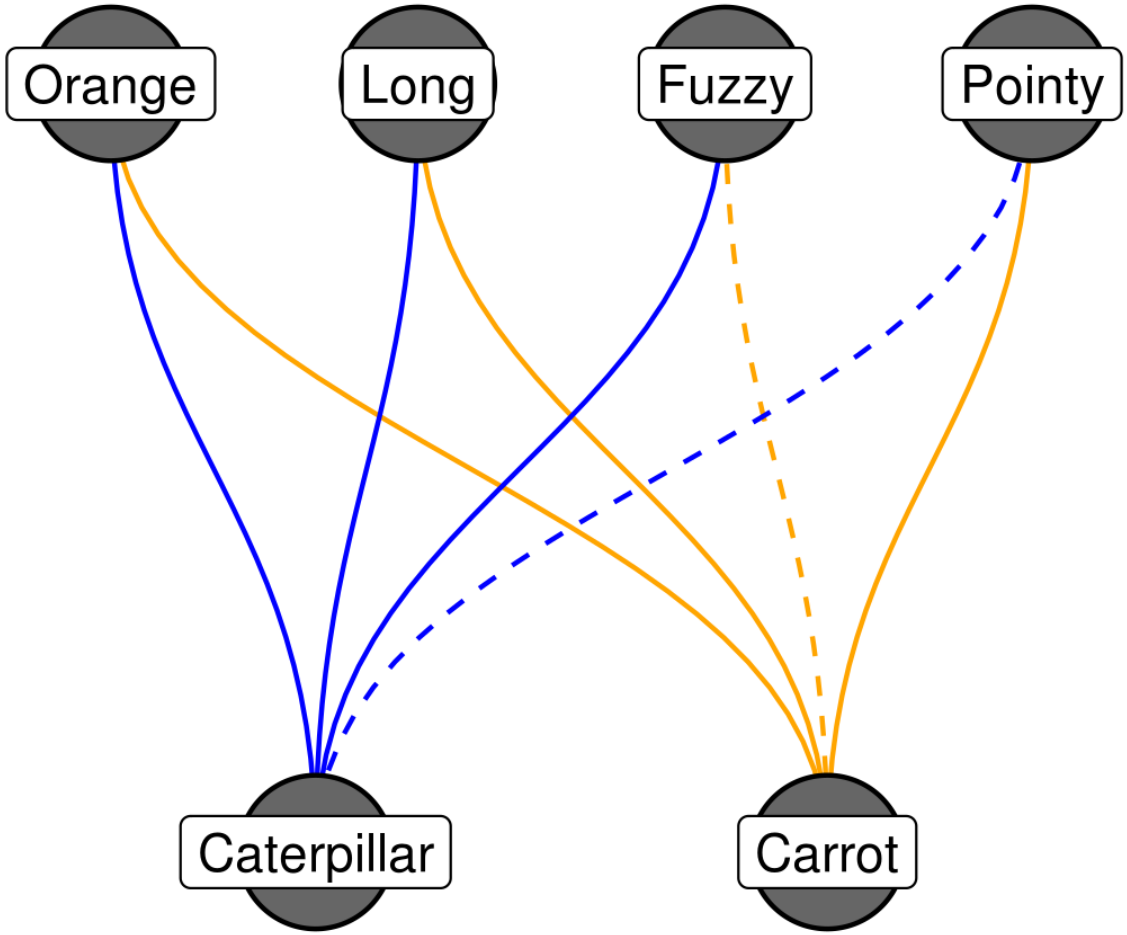
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Classification

- What features does your brain use to classify objects?

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Classification



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Classification

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Convolutional Neural Networks (CNNs)

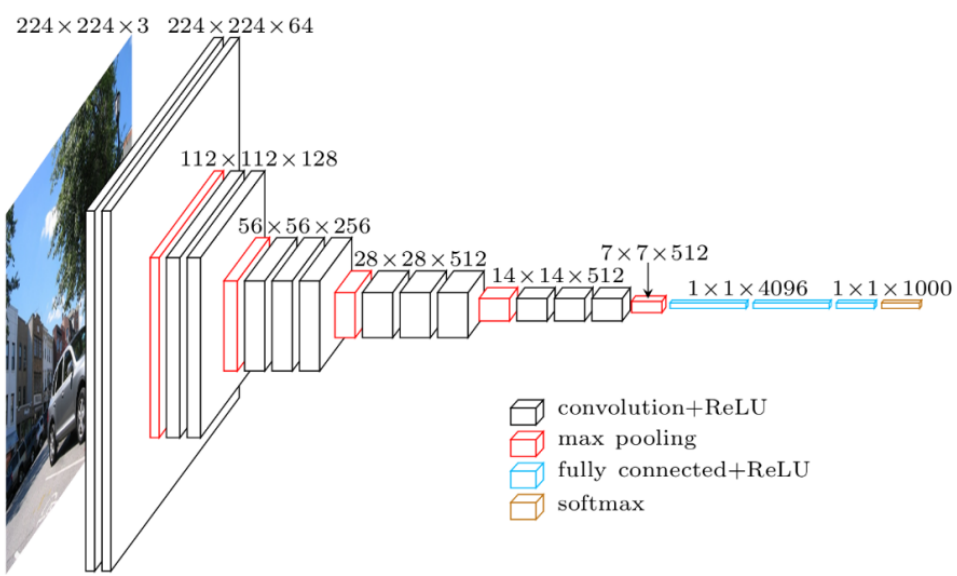
- Tool for supervised deep-learning, well-suited to image analysis
- Translation-invariant pattern detection
- Analogous to how our eyes detect features and our brain weights them for classification
  - Eyes $\rightarrow$ Brain $\rightarrow$ Classification
- Use combinations of convolutional and pooling layers to filter raw information into features

Pre-trained CNNs

- Large, complex networks trained to identify a subset of images from ImageNet
  - $\sim$1.2 million images spanning 1,000 categories
- Examples: LeNet (1998), AlexNet (2012), GoogLeNet/Inception (2014), **VGG** (2014), ResNet (2015)
  - More on how they compare here and here
- Contain both “eyes" to detect features and “brain" to weight those features for classification

VGG16 Architecture

- 16 “functional" layers – 13 convolution and 3 fully connected
- 5 max-pooling layers, summarize features and structure model



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Filters and Convolution

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input

1	0	1
0	1	0
1	0	1

Filter / Kernel

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Filters and Convolution

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

Input x Filter

4		

Feature Map

Filters and Convolution

1	1x1	1x0	0x1	0
0	1x0	1x1	1x0	0
0	0x1	1x0	1x1	1
0	0	1	1	0
0	1	1	0	0

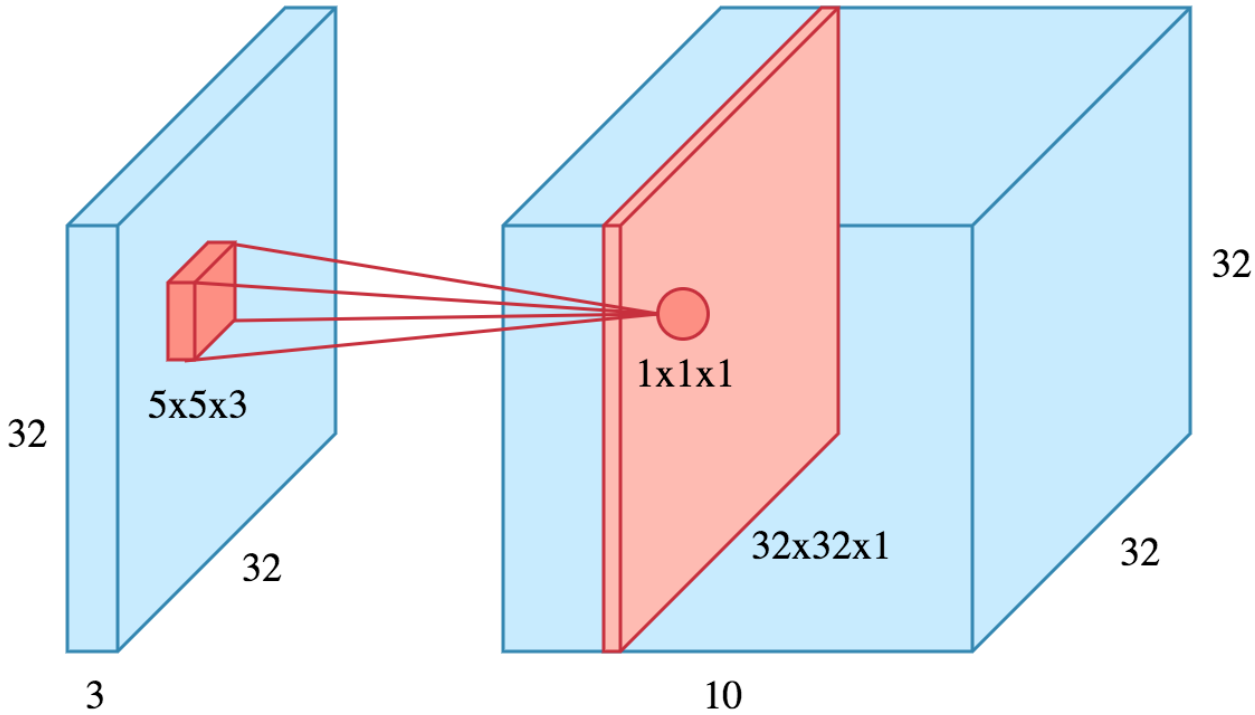
Input x Filter

4	3	

Feature Map

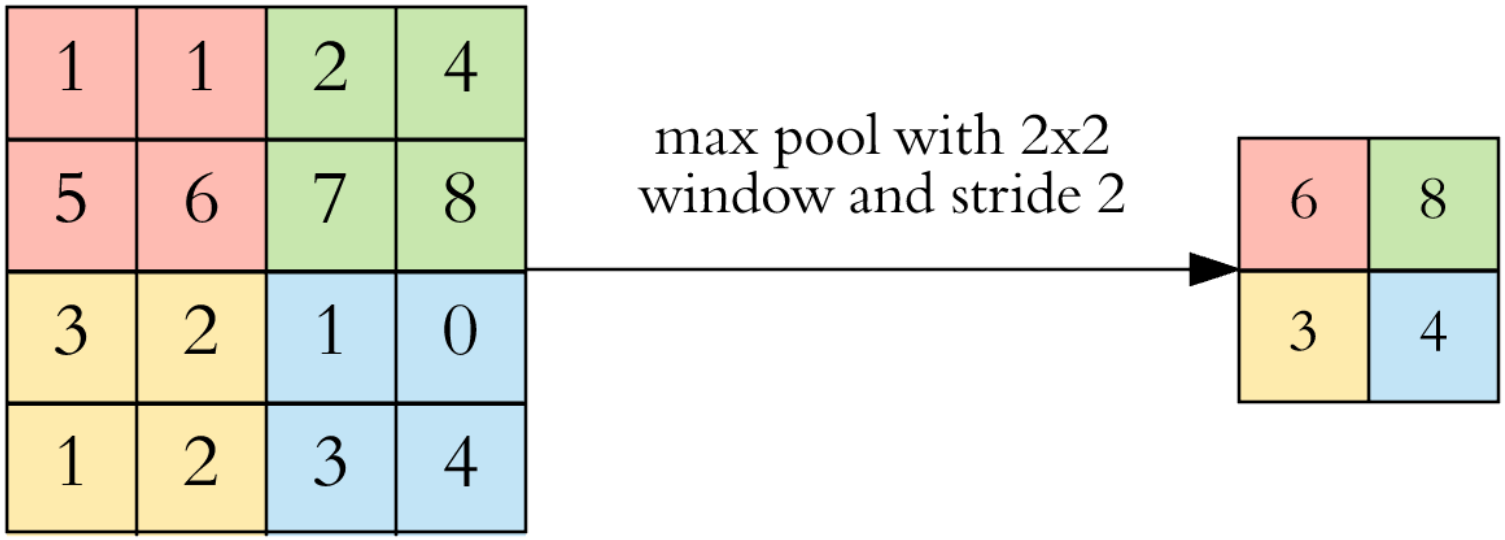
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Filters and Convolution

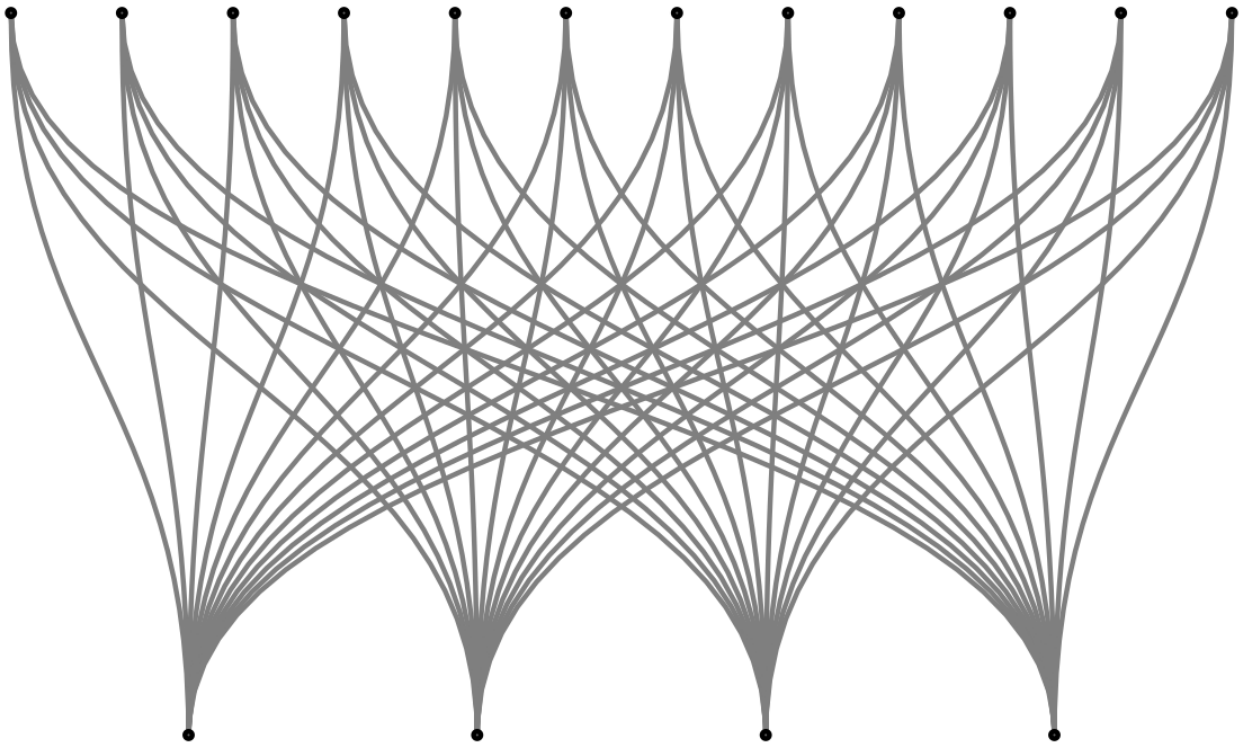


csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Max Pooling



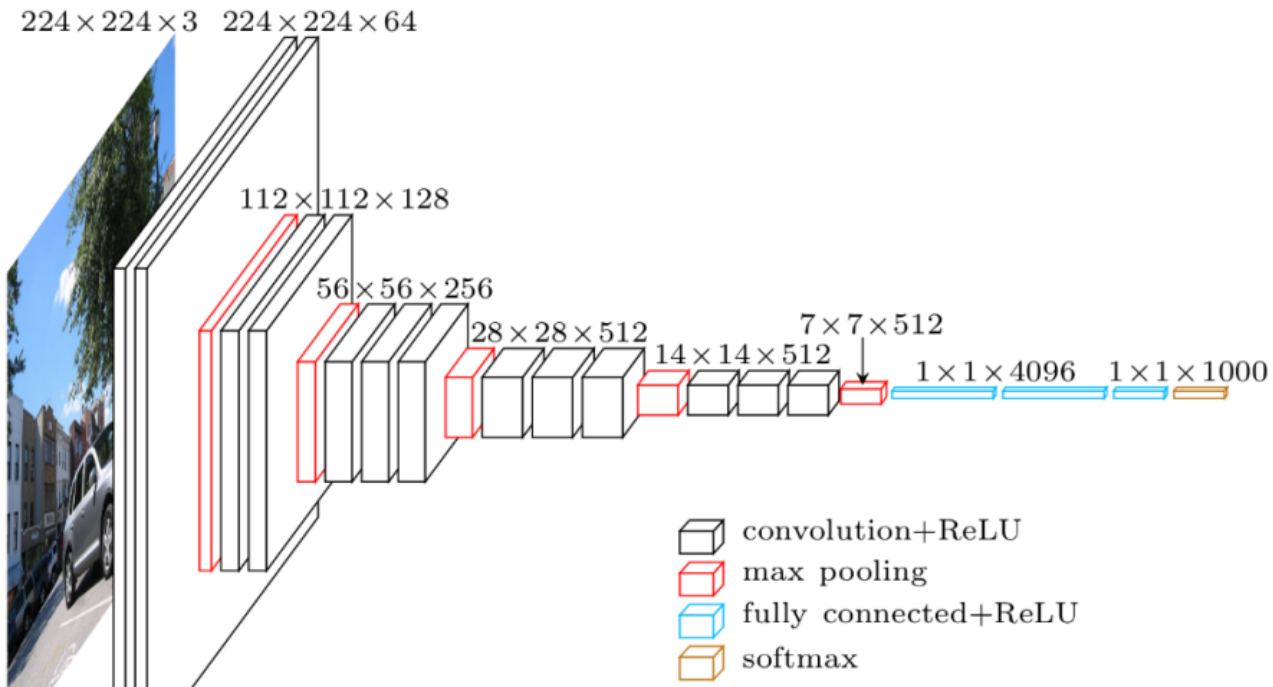
Densely Connected Layers



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

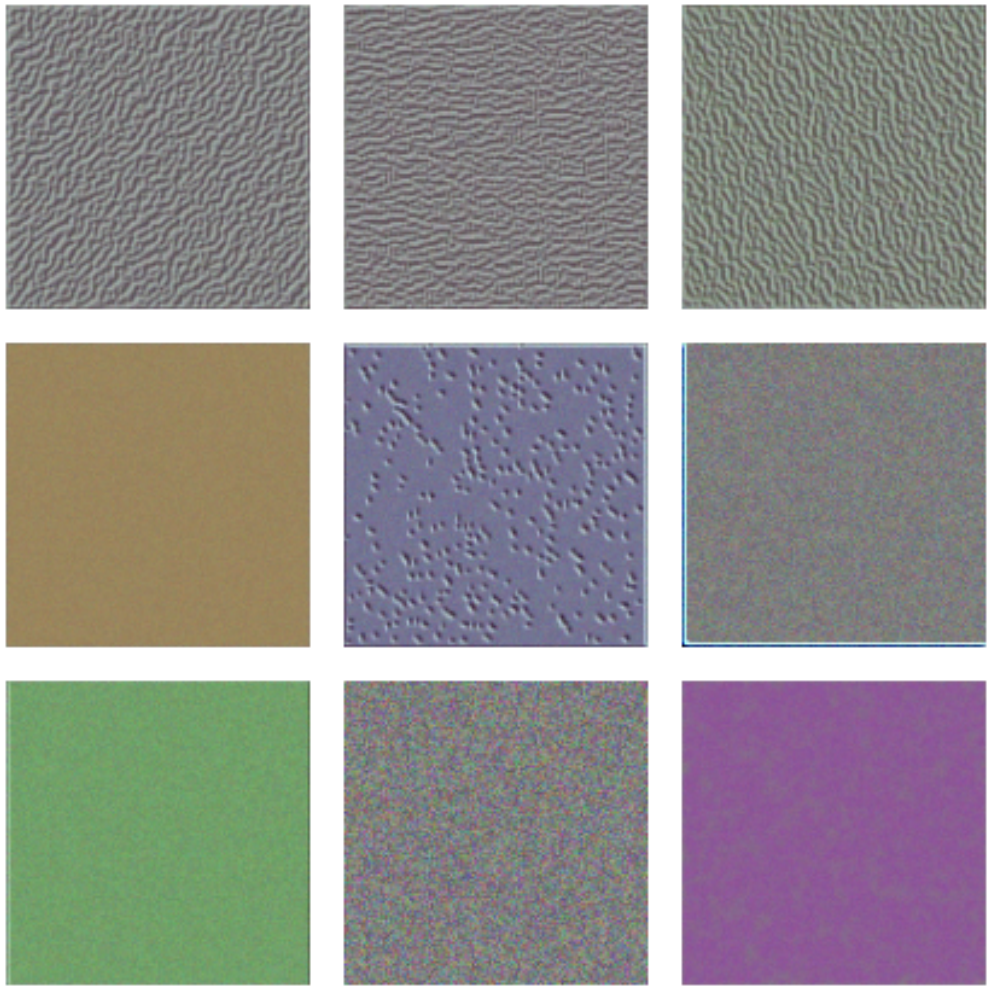
VGG16 Architecture

- Rectified Linear Units (ReLU) $= x * \mathrm{I}_{\{x>0\}}$



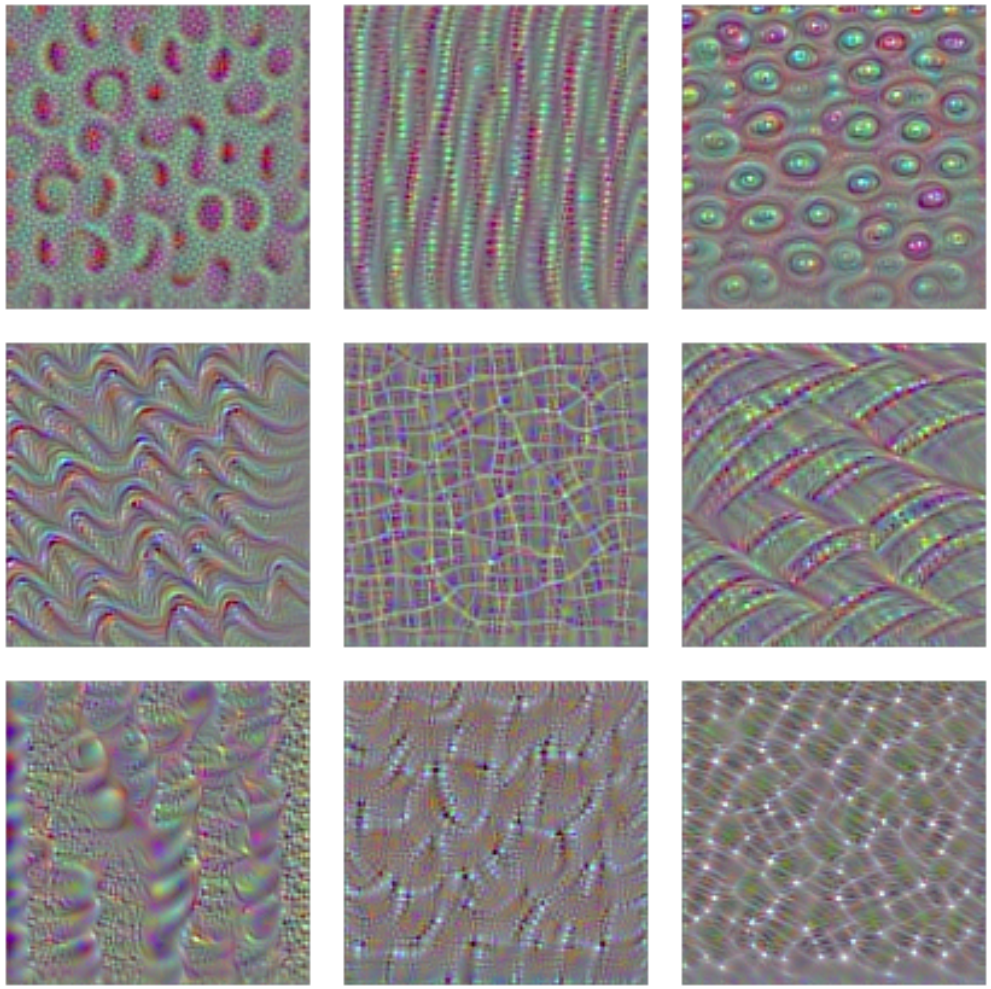
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

VGG16 Filters - First Block



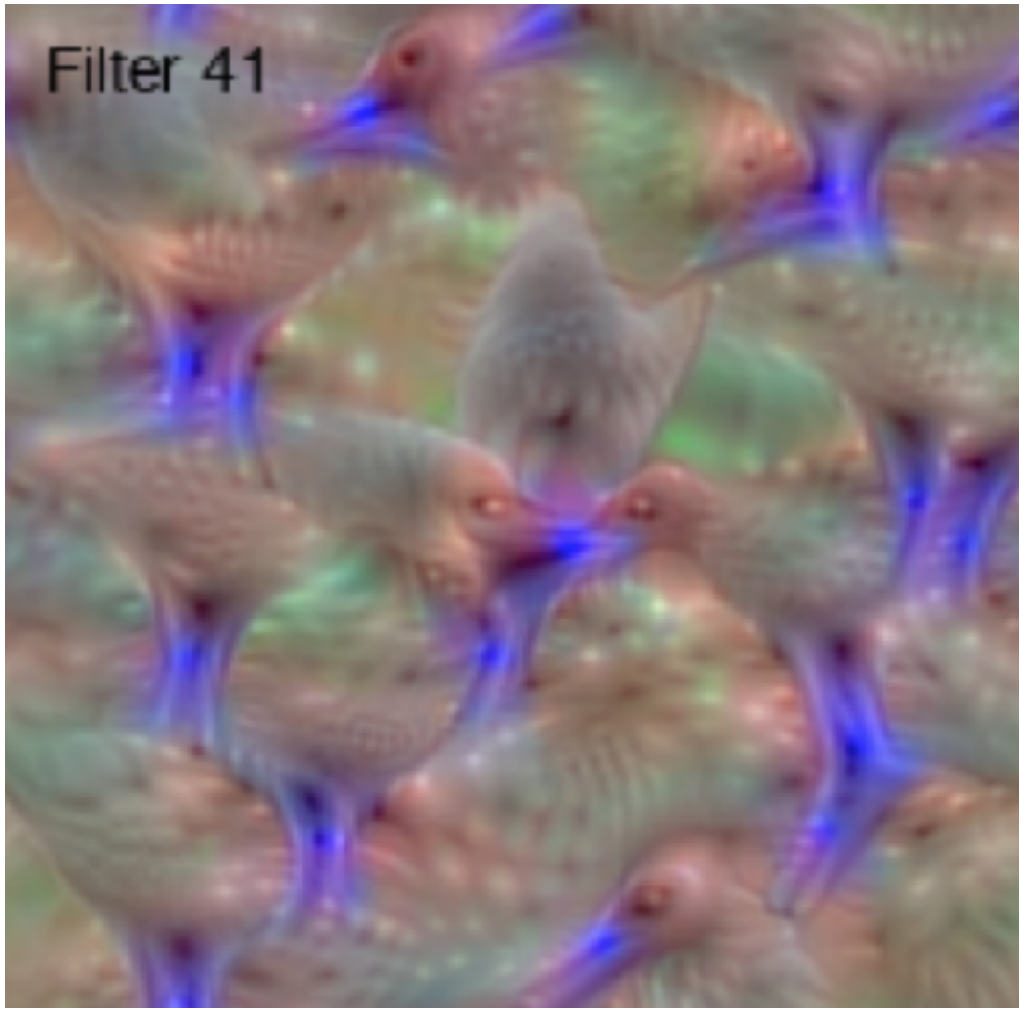
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

VGG16 Filters - Fourth Block



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

VGG16 Filters - Fifth Block



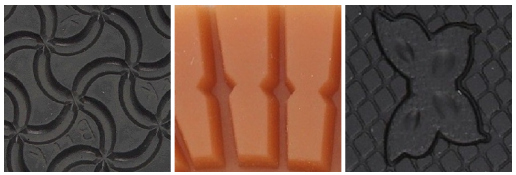
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Using a Pre-trained CNN for a New Task

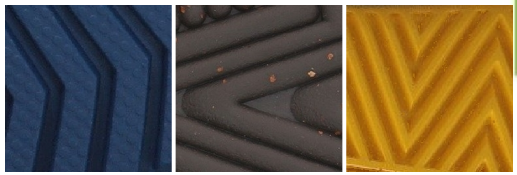
- Large CNNs with many layers have millions of parameters to train
  - VGG16 “eyes" have 14,714,688 trainable parameters
- Using the existing “eyes" and training only a “brain" brings number of images required from millions to only thousands
  - “Eyes" = convolutional and pooling layers
  - “Brain" = fully connected layers

Outsole Class Characteristics

Bowtie



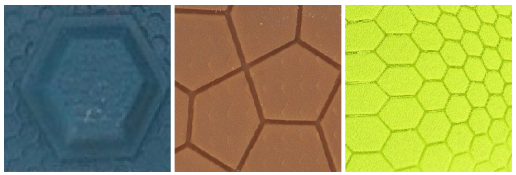
Chevron



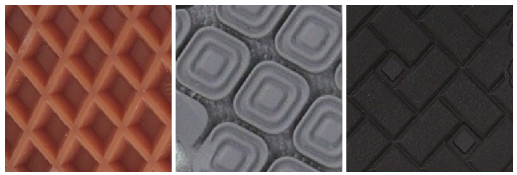
Circle

Line

Polygon



Quad



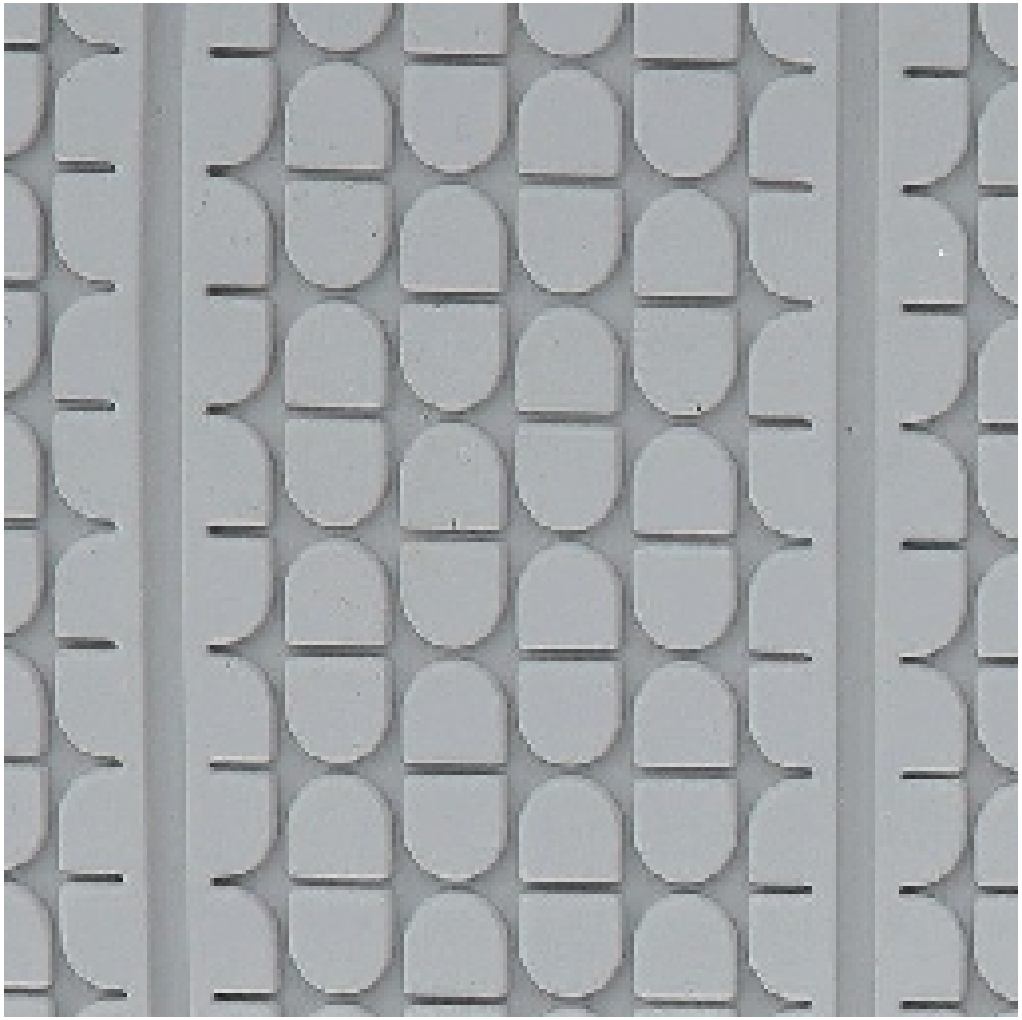
Star

Text

Triangle

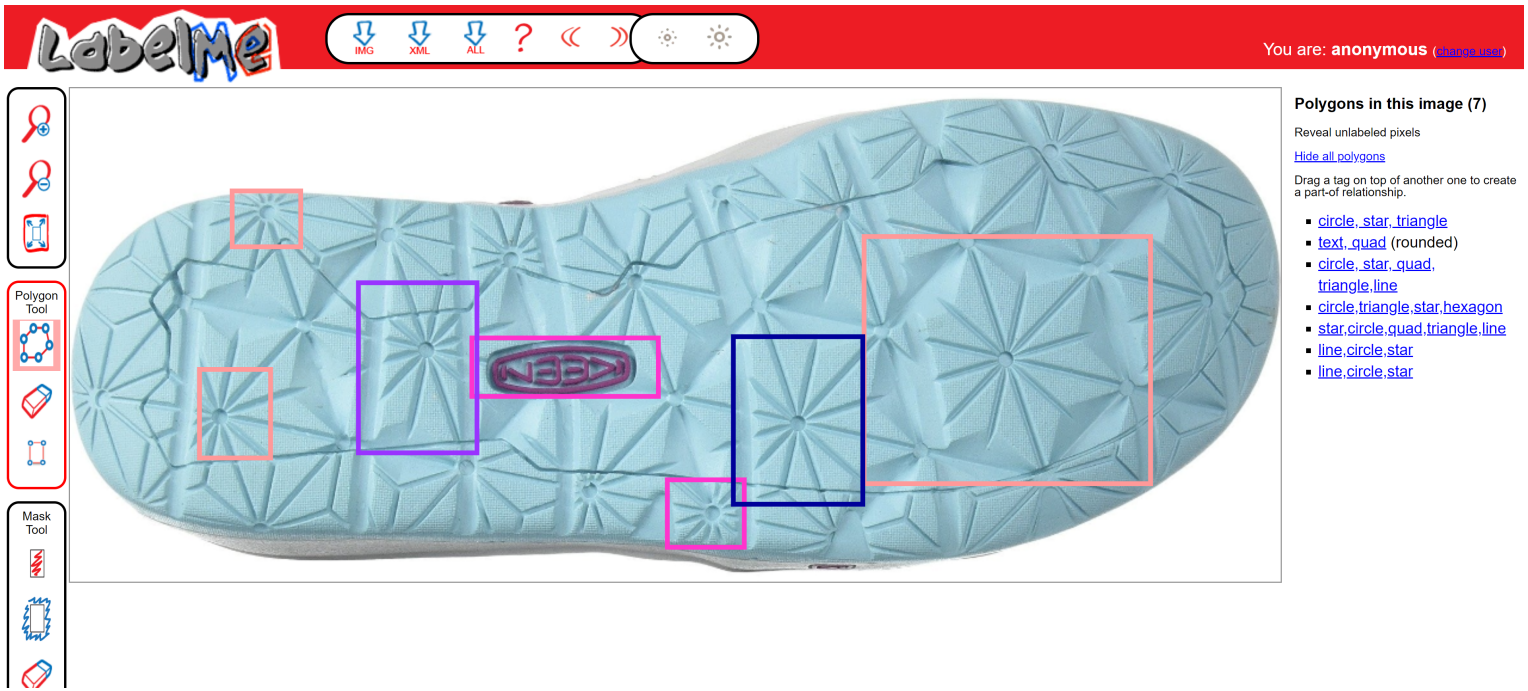
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Circle or Quadrilateral?



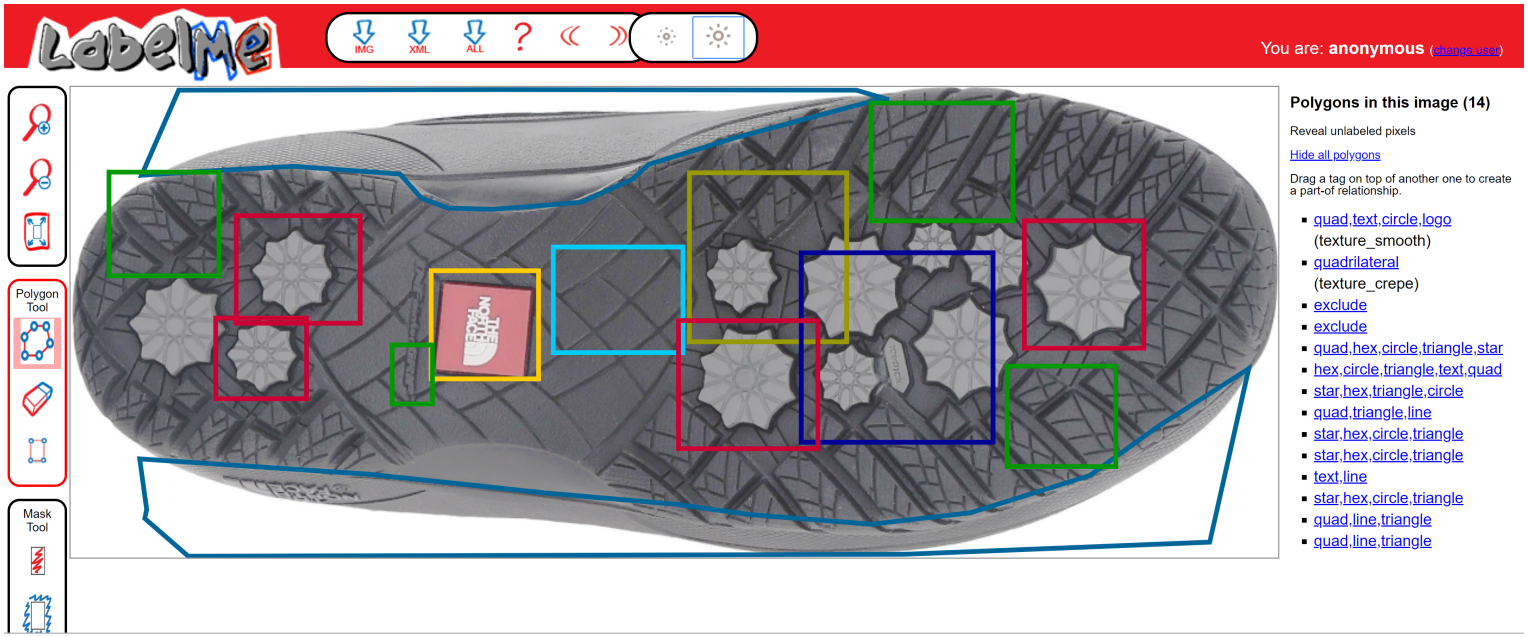
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Data Collection



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

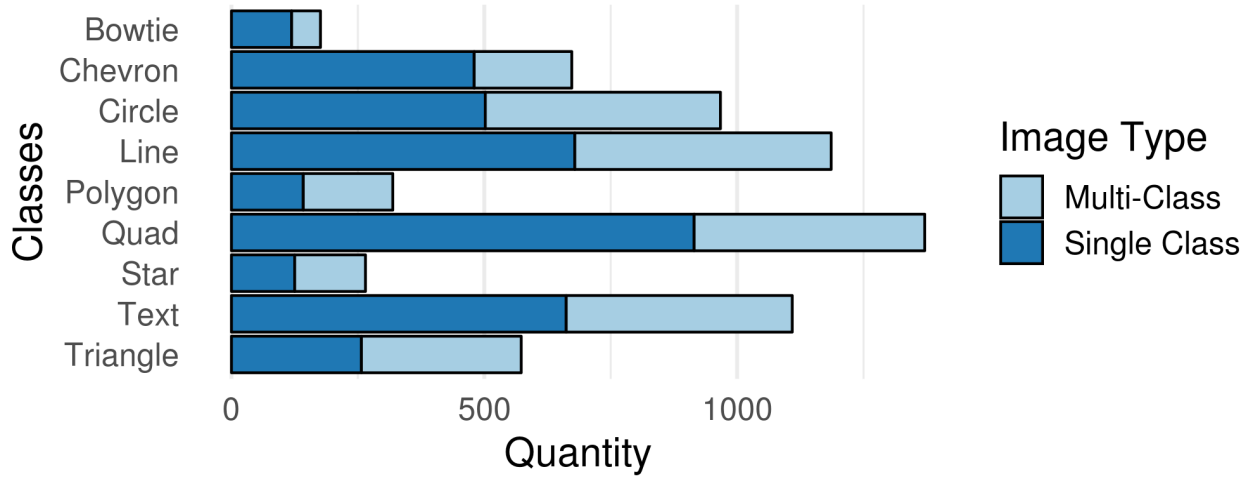
Data Collection



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Training a Neural Network – Data Structure

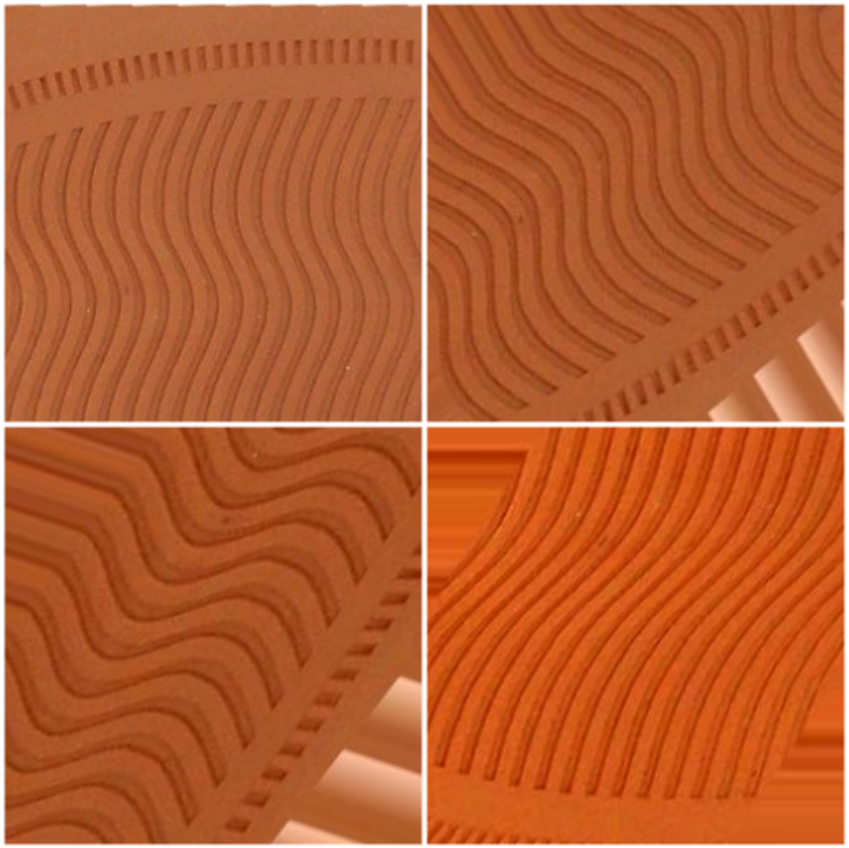
- 24,000 **multi-label** images from 2,200 shoes
  - Outsole images web-scraped from Zappos.com
- Partition existing data
  - 50% for training*
  - 25% for validation, used to regulate training
  - 25% for testing, used to evaluate model performance



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Training Technique – Augmentation

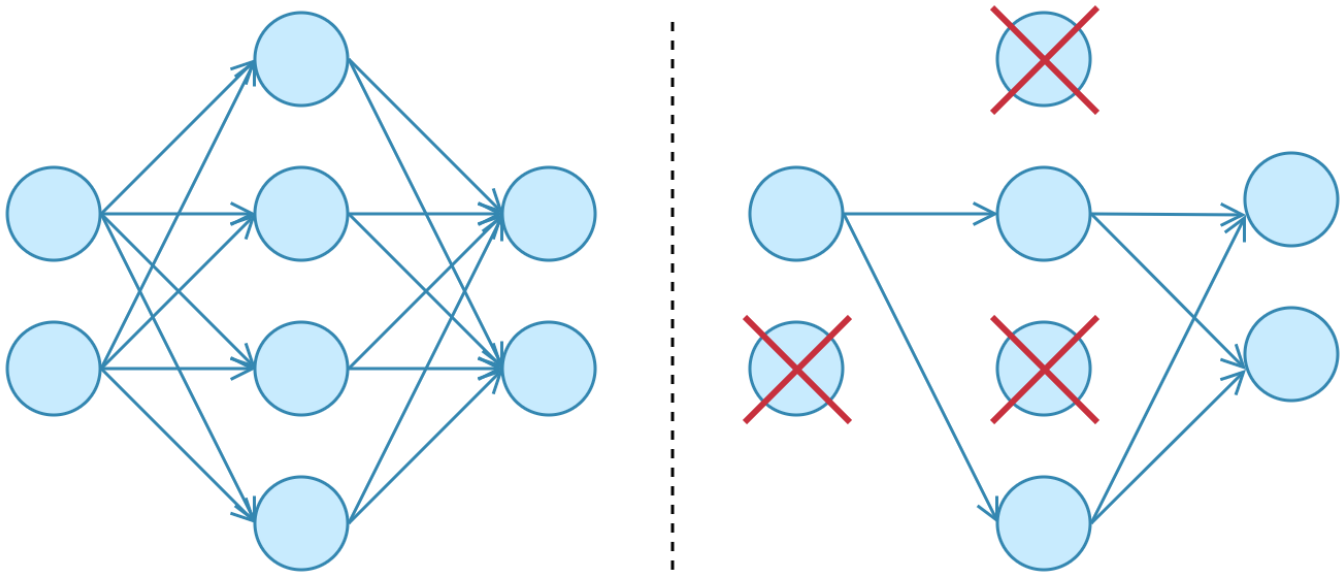
- More training data, helps prevent over-fitting



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Training Technique – Drop-out

- Ensures no neuron is too specialized, helps prevent over-fitting



No Dropout

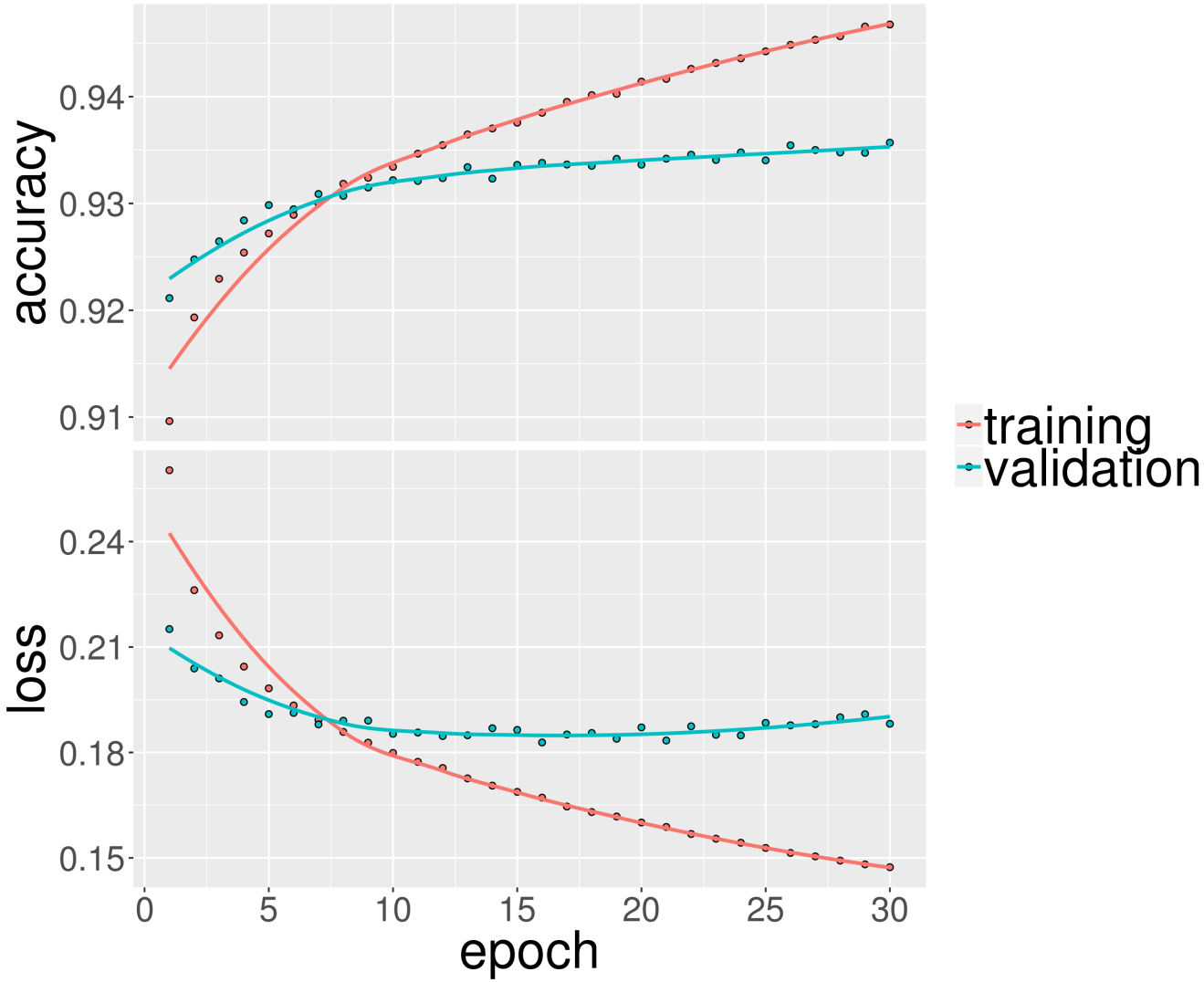
With Dropout

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Training a Neural Network – Computation

- Keras: a neural network API, written in Python
- Supports **TensorFlow**, CNTK, or Theano as computational backend
- `Keras`: an R interface to Keras
  - For more info, see Deep Learning with R
- Note: Keras requires GPU (or *very* good CPU)
  - Seriously, don't try this on your personal machine

Training a Neural Network



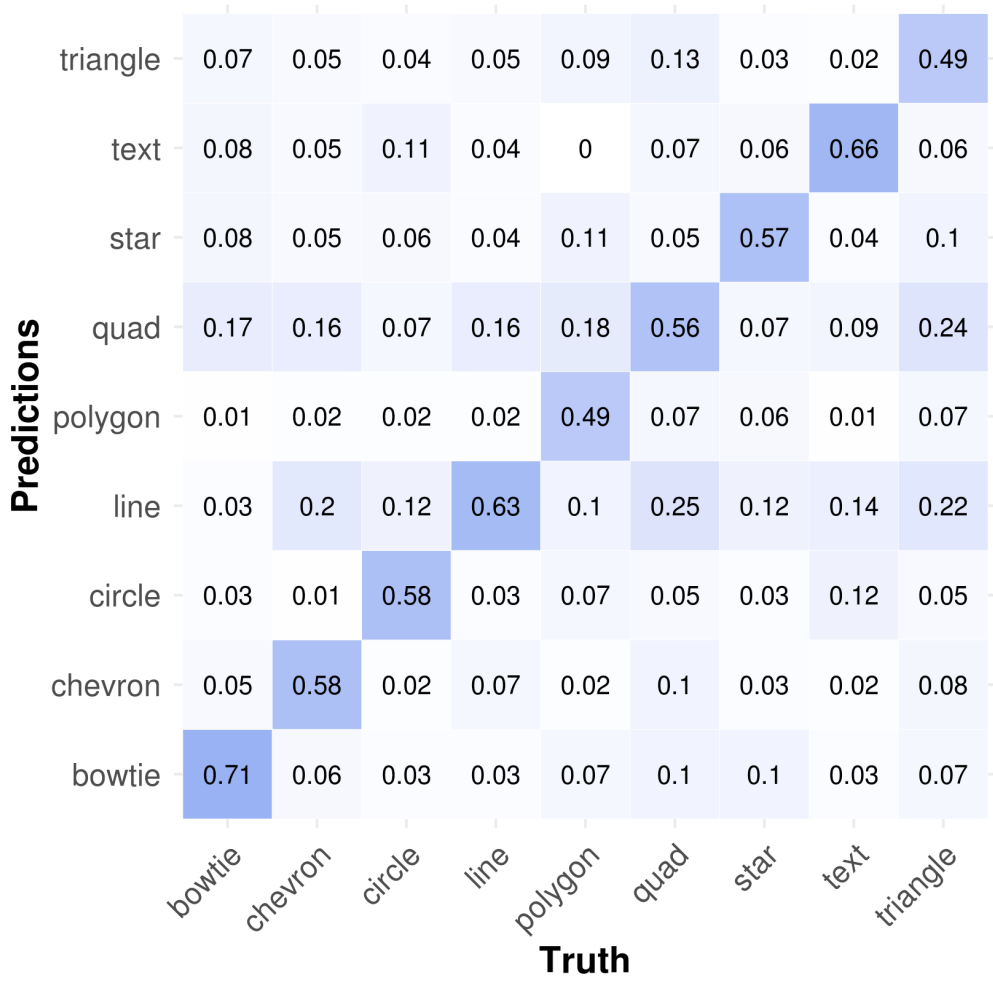
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Visualizing Model Predictions

- Shiny App

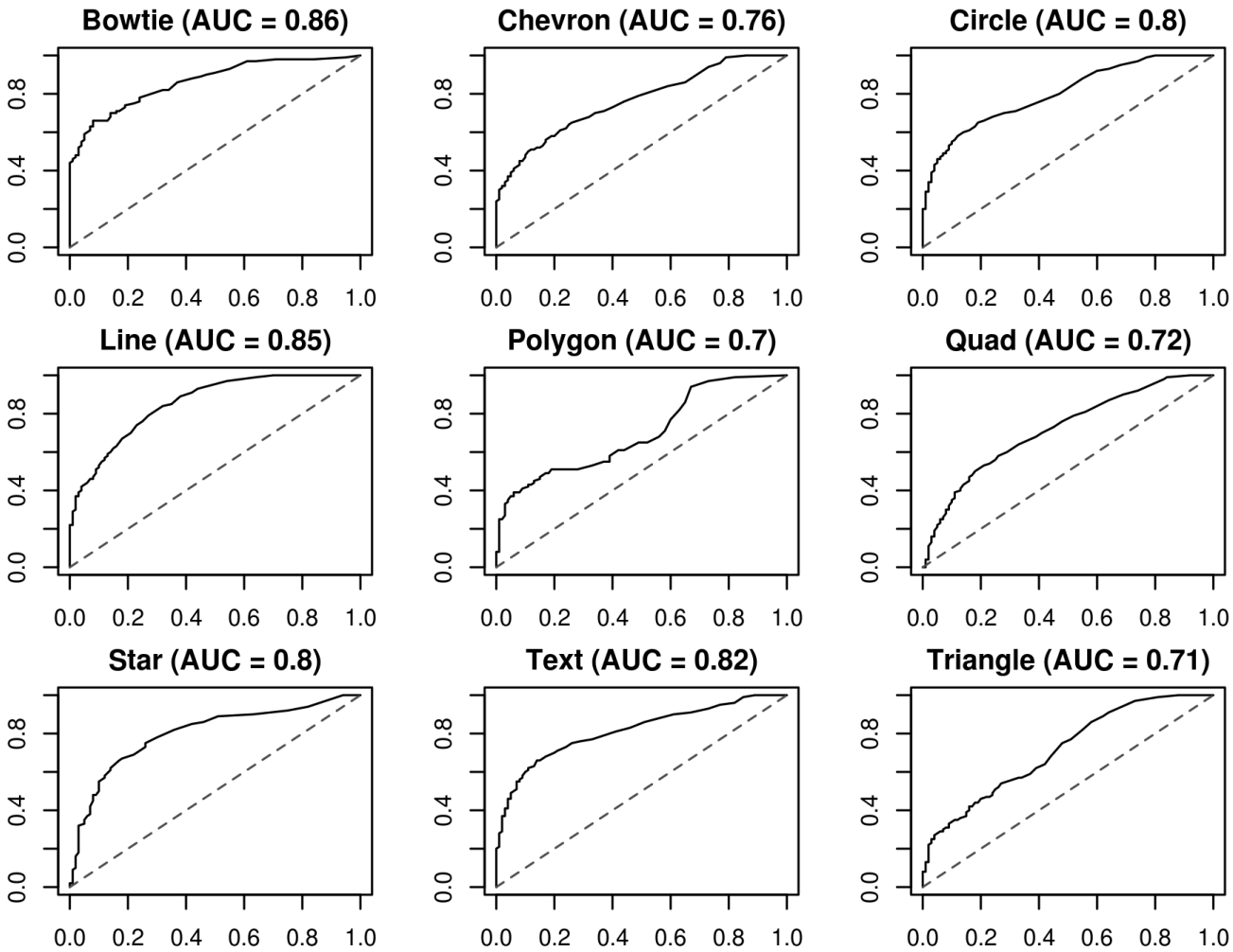
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Prediction Accuracy



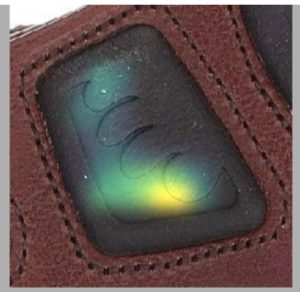
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

ROC



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Heatmaps



Circle: 0.506

Text: 0.312

Quad: 0.302

Triangle: 0.077

Line: 0.07

Polygon: 0.066

Chevron: 0.019

Star: 0.009

Bowtie: 0.003

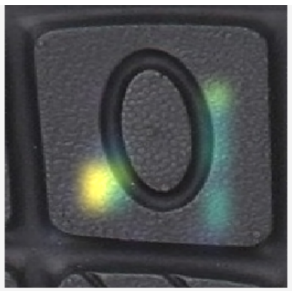
csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Heatmaps

Circle: 0.783

Quad: 0.543

Line: 0.075



Text: 0.059

Polygon: 0.022

Triangle: 0.017

Chevron: 0.012

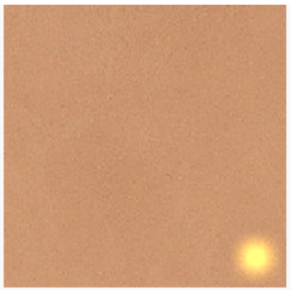
Star: 0.004

Bowtie: 0.001

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Heatmaps

Line: 0.044



Quad: 0.04

Text: 0.036

Circle: 0.032

Star: 0.031

Triangle: 0.025

Bowtie: 0.022

Chevron: 0.015

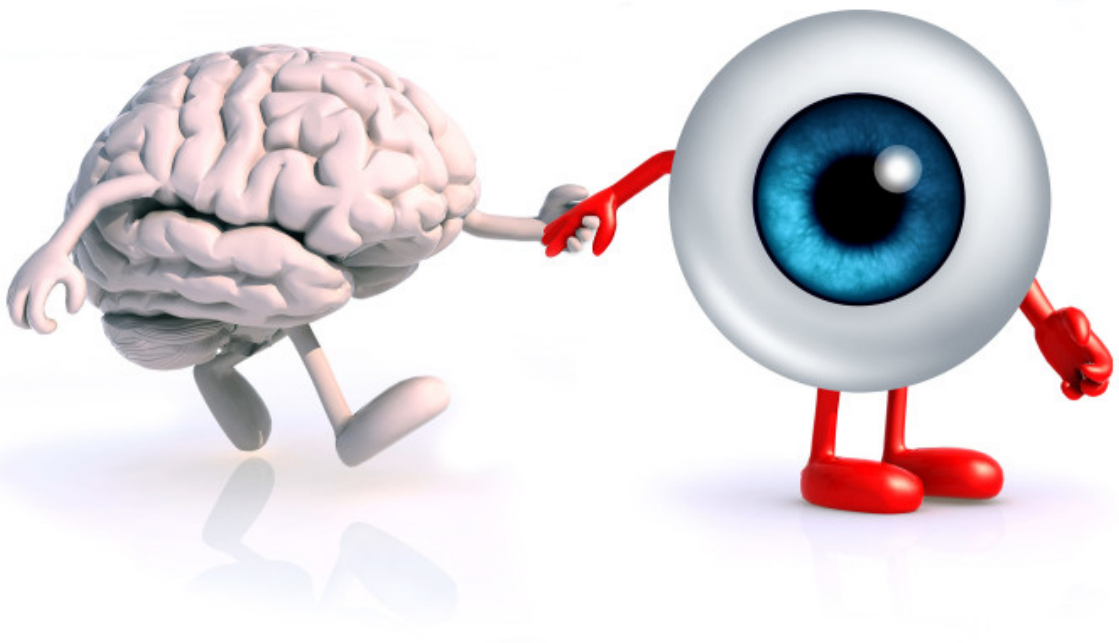
Polygon: 0.013

csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

Summary

- CNNs view images as collections of features, then use those features to make classifications
- Pre-trained CNNs detect complex features and classify images from thousands of categories by learning from large image databases
- Retraining the classifier on a pre-trained CNN greatly reduces the required number of training images for a new classification task
- CNNs are well-suited to outsole classification, but the biggest challenge is getting a large amount of consistent data

Questions?



csafe
Center for Statistics and Applications in Forensic Evidence
ForensicStats.org

References I

[1] Susan Gross et al. "The Variability and Significance of Class Characteristics in Footwear Impressions". 2013.

[2] Sheida Hancock et al. "The Interpretation of Shoeprint Comparison Class Correspondences". 2012.

[3] Alex Krizhevsky et al. "ImageNet Classification with Deep Convolutional Neural Networks". 2012.

[4] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". Sept. 2014.