

# Proofpoint Logo Detection & Classification

18th February 2021

Duke

# Agenda

- Pipeline Updates
- Pipeline Demo
- Performance / Speed Comparisons
- Classifier Updates
- Next Steps

# Unified Pipeline

- Developed a unified pipeline for detection and classification.
- Detection and classification happen consecutively.
  - No additional storage
  - No additional saving and loading time
- Use multiple models with single command line input.
- Given trained weights, users can choose from among:
  - 3 (possibly 6) YOLO models
  - 8 Faster-RCNN models
  - EfficientNet B0-4
- Can plot results and/or save as JSON or text files.
- Makefile manages all requirements.
- Currently limited to Linux with CUDA 10.1

# Pipeline Overview

--source : location of images for detection  
--detector : yolo or rcnn  
--detector-weights : associated weights for detector  
--rcnn-arch : specific architecture if detector is Faster-RCNN  
--detect-thres : object detector confidence threshold  
--classifier : efficientnet b0-4  
--classifier-weights : associated weights for classifier  
--name : folder to store predictions ./runs/detect/<name>

# Demo

```
python detect.py --source ../images/test --detector yolo  
    --detector-weights weights/yolov3_spp.pt  
    --classifier efficientnet_b0  
    --classifier-weights weights/efficientnet_b0.pt  
    --detect-thres 0.1 --name detect_test
```

# Comparing Detectors

Duke



## Checking

## Savings

## Credit Cards

## Home Loans

## Auto Loans

Merrill Lynch I:7

Better Money Habits®

How can we help you?



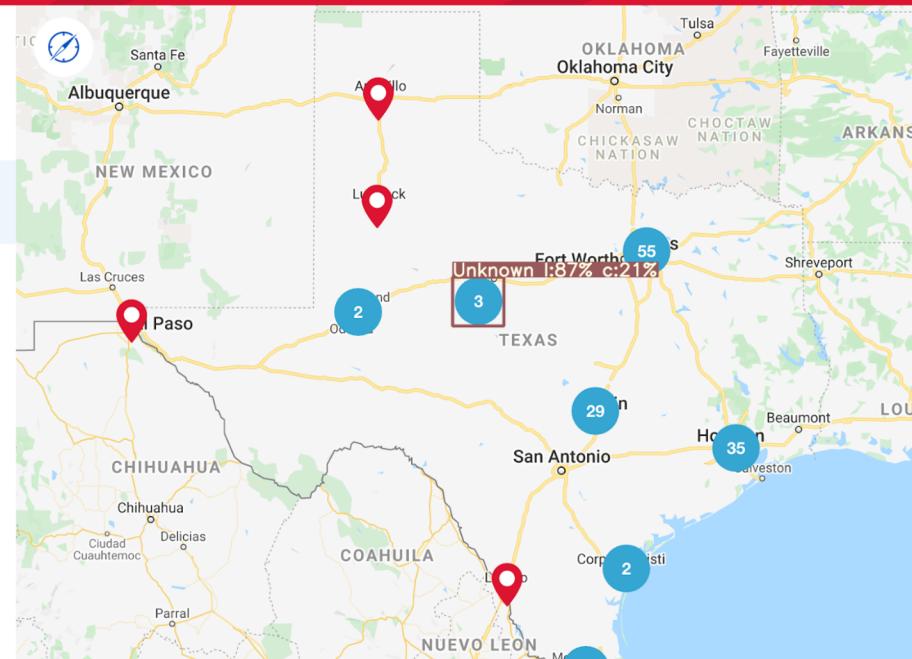
## Financial centers and ATMs in Texas

Enter address, ZIP code or landmark 

## Financial centers and ATMs in Texas

**⚠** Financial centers will be closed on Monday, October 12, in observance of Columbus Day.

- A Abilene
  - Allen
  - Alvin
  - Amarillo
  - Aransas Pass
  - Arlington
  - Austin
  - Azle  
  - B Balch Springs
  - Baytown



Duke



Checking

Savings

Credit Cards

Home Loans

Auto Loans

Investing

Better Money Habits®

How can we help you?



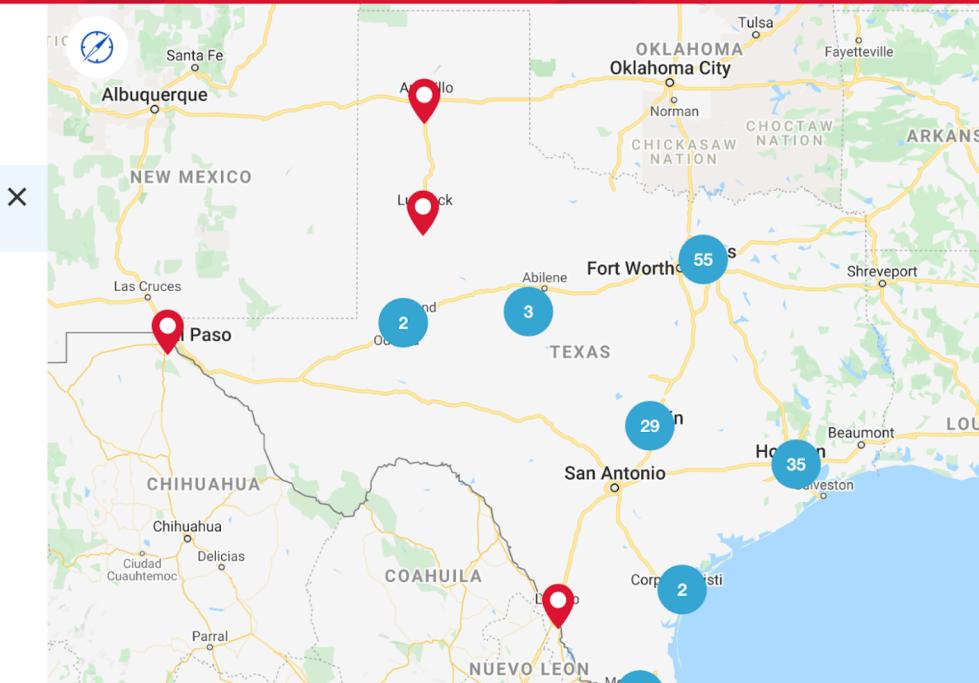
## Financial centers and ATMs in Texas

Enter address, ZIP code or landmark



### Financial centers and ATMs in Texas

**A** Financial centers will be closed on Monday, October 12, in observance of Columbus Day.



**A** Abilene

Allen

Alvin

Amarillo

Aransas Pass

Arlington

Austin

Azle

**B** Balch Springs

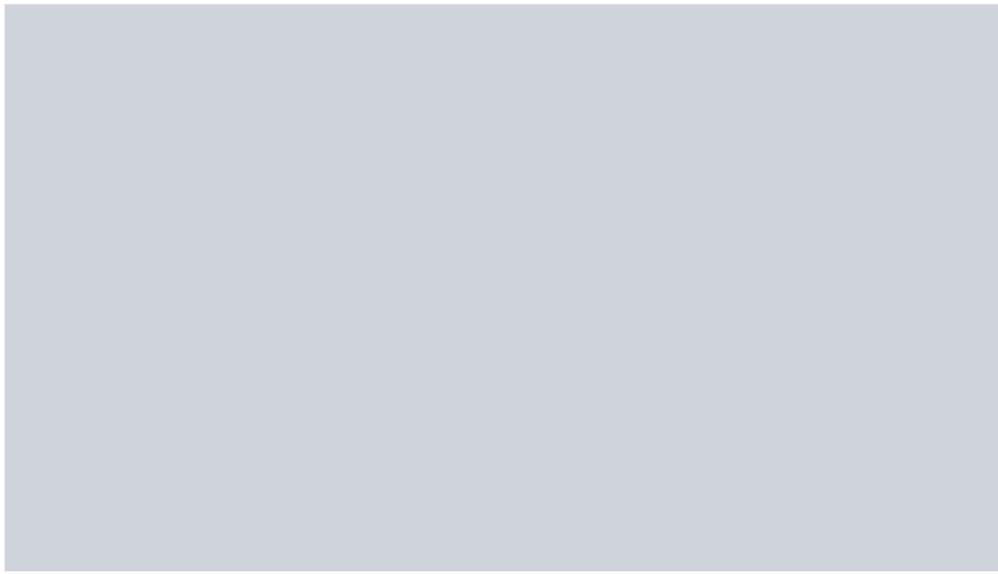
Baytown

uke

Esurance I:100% c:66%



Sections ▾ Top 50 ▾ Featured Partners ▾ Events ▾ Webinars Best Practices Resources ▾ Job Board



Logo via Cigna

#### FEATURED PARTNERS

## Cigna Covers Cost of Coronavirus

#### Follow us

Face Twitte Insta YouTube I:95% c:98%



Abbott I:99% c:98%



Search jobs at  
these companies



Wells Fargo I:96%

uke

Esurance d:82% c:69%



Sections ▾

Top 50 ▾

Featured Partners ▾

Events ▾

Webinars

Best Practices

Resources ▾

Job Board



Logo via Cigna

#### FEATURED PARTNERS

## Cigna Covers Cost of Coronavirus

#### Follow us

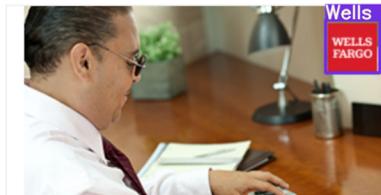
FaceTwitterInstagram d:76% c:97%15%



Abbott d:47% c:98%



Search jobs at  
these companies



Wells Fargo d:3

Duke

We use cookies to give you a better experience. By continuing to use alamy.com you accept our [use of cookies](#). X

Unknown l:100% c:22%



223,616,283 stock photos, vectors and videos

Lightboxes

Cart

Account

All images ▾



Logo of public company Honeywell Int'l Inc. displayed on a smartphone. Grey background. Credit: PIXDUCE - Image ID: 2A8RH1N



Enlarge

Download preview



Save to lightbox



Add to cart



Available for editorial and personal use only. Get in touch for commercial uses

Buy this stock image now...

Personal use \$ 19.99

Presentation or newsletters \$ 19.99

Editorial website \$ 49.99

Magazines and books \$ 69.99

Choose another rights-managed license

Download

Duke

We use cookies to give you a better experience. By continuing to use alamy.com you accept our [use of cookies](#). X

Unknown d:70% c:18%



223,616,283 stock photos, vectors and videos

+ Lightboxes

Cart

Account

All images ▾



Logo of public company Honeywell Int'l Inc. displayed on a smartphone. Grey background. Credit: PIXDUCE - Image ID: 2A8RH1N



Available for editorial and personal use only. Get in touch for commercial uses [②](#).

Buy this stock image now...

Personal use [②](#)

\$ 19.99

Presentation or newsletters [②](#)

\$ 19.99

Editorial website [②](#)

\$ 49.99

Magazines and books [②](#)

\$ 69.99

Choose another rights-managed license ▼

Download

Download preview



Save to lightbox



Add to cart



Duke

Home

Vimeo d:99% c:57%



Duke

Home



uke

Unknown d:99% c:21%



Top News Sections ▾ News By Category ▾

All News ▾

Search Newsroom...

Twitter Facebook Telegram d:87% c:21%

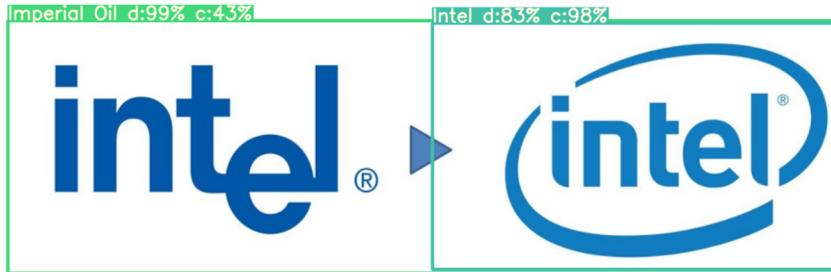
## Editorial

March 31, 2014

Contact Intel PR

# Tech Company Logos Reinvented

Since 1968, Intel Has Had Two Company Logos



Reinvention is the mantra of Silicon Valley from individual entrepreneurs and engineers to startups and the even the large enterprises that form the foundation of the tech industry. The evolution of tech company logos offers a glimpse at how companies have changed over the years — some have changed radically and some barely at all.

Intel, for one, has had only two corporate logos since it was founded in July of 1968. The company founders, Robert Noyce

Unknown d:73% c:11%



Latest News: Intel Free Press

Intel d:98% c:95%



October 19, 2015  
Descending into the Biggest Cave on Earth

Intel d:99% c:95%



September 3, 2015  
Deaf Engineer Hears and Speaks Through Others



August 13, 2015  
Thunderbolt 3 and USB Type-C the 40 Gbps

uke



## Editorial

March 31, 2014

Contact Intel PR

# Tech Company Logos Reinvented

Since 1968, Intel Has Had Two Company Logos

SAP d:26% c:49%



intel d:58% c:96%



Reinvention is the mantra of Silicon Valley from individual entrepreneurs and engineers to startups and the even the large enterprises that form the foundation of the tech industry. The evolution of tech company logos offers a glimpse at how companies have changed over the years — some have changed radically and some barely at all.

Intel, for one, has had only two corporate logos since it was founded in July of 1968. The company founders, Robert Noyce



Latest News: Intel Free Press



October 19, 2015

Descending into the Biggest Cave on Earth



September 3, 2015

Deaf Engineer Hears and Speaks Through Others



August 13, 2015

Thunderbolt 3 and USB Type-C: the 40 Gbps

Duke

YouTube d:99% c:100%

YouTube Search  SIGN IN

Up next AUTOPLAY

Unknown d:99% c:17%

APPLIED MATERIALS

Uber d:85% c:32%

Career Opportunities at Applied Materials

1,239 views • Apr 29, 2020

Applied Materials 4.67K subscribers

WebMoney d:96% c:32%

What will you make possible today? At Applied Materials, our innovations make possible the

SUBSCRIBE

Up next

Maydan Technology Center Overview Video  
Applied Materials 23K views • 3 years ago

Celebrating 40 Years in Japan  
Applied Materials 1.7K views • 11 months ago

Inside the mind of a master procrastinator | Tim Urban  
TED 31M views • 4 years ago

4th Dimension - Tesseract, 4th Dimension Made Easy - Carl...  
AstronomyCorner 5.9M views • 9 years ago

10 ways to have a better conversation | Celeste Headlee  
TED 10M views • 4 years ago

Making Memory Chips – Process Steps  
MicronTechnology 308K views • 3 years ago

uke

YouTube d:44% c:99%

☰ YouTube Search ⌂ ⌂ ⌂ SIGN IN

# APPLIED MATERIALS

0:00 / 1:11

© Applied Materials, Inc. 2019

Career Opportunities at Applied Materials

1,239 views • Apr 29, 2020

Unknown d:34% c:22%

Applied Materials 4.67K subscribers

What will you make possible today? At Applied Materials, our innovations make possible the

Up next

AUTOPLAY

 Maydan Technology Center Overview Video  
Applied Materials  
23K views • 3 years ago

 Celebrating 40 Years in Japan  
Applied Materials  
1.7K views • 11 months ago

 Inside the mind of a master procrastinator | Tim Urban  
TED   
31M views • 4 years ago

 4th Dimension - Tesseract, 4th Dimension Made Easy - Carl...  
AstronomyCorner  
5.9M views • 9 years ago

 10 ways to have a better conversation | Celeste Headlee  
TED   
10M views • 4 years ago

 Micron d:10% c:64% Making Memory Chips - Process Steps  
MicronTechnology  
308K views • 3 years ago

uke

# Current Performance and Speed

- Tesla V100-SXM2-16GB
- YOLO: 0.1 threshold; 0.6 IoU
- Faster-RCNN: 0.7 threshold; 0.6 IoU
- **0.0125s** per logo

Model	Precision	Recall	mAP @ 0.5	Speed
Faster RCNN	0.48	<b>0.88</b>	<b>0.872</b>	0.149 s
YOLO v3 - SPP	<b>0.778</b>	0.867	0.790	<b>0.044s</b>
Classifier	0.916	0.901	-	0.050s

# Error Analysis

- Empirical Observations:
  - RCNN struggles with files with no logos or unexpected dimensions
  - YOLO struggles with large logos
- Run through:
  - Regular test set
  - Data set with no logos
  - Data set of PDF documents
  - Test set with reduced quality
  - Test set with cropped dimensions
- Develop script to assess common mistakes:
  - Average area of false negatives
  - Average area of false positives
  - General location of missed logos
- Compare and contrast Faster-RCNN and YOLO

# Classifier

Duke

# More Data for the Classifier

Alphabet



is for Google

As Sergey and I wrote in the original founders letter 11 years ago,  
"Google is not a conventional company. We do not intend to  
become one." [more](#)

*Larry Page*

Larry Page

Train Set Image

Investors



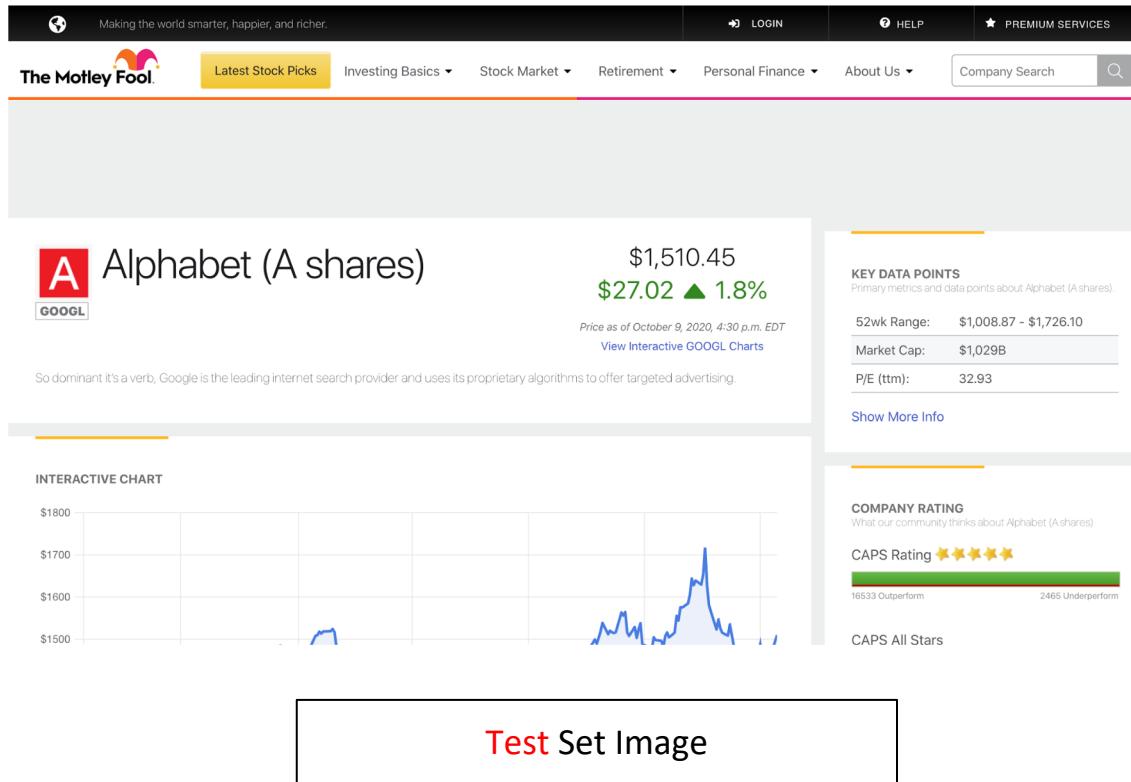
# Alphabet



Our Annotations

Duke

# More Data for the Classifier



Model Predicted Annotations

Duke

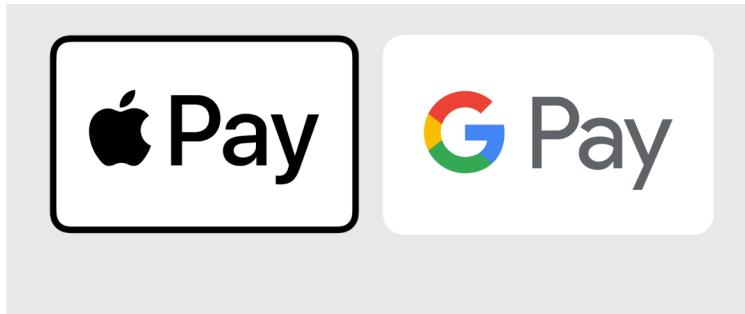
# A quick note....

- Decimal Prediction [34.267, 25.362, 894.213, 15.736]
- I had to round it up
- Cannot crop images with decimals

# Improvements to Data of Logo Classifier

- Types of images removed
  - Non-logos
  - Two or more logos of different brands in the same image
- Classes removed
  - Contained almost no representative logos of the brand
  - Duplicates (PwC and PricewaterhouseCoopers)
- Added companies/brands from technology, energy, industrial, software sections in 1000logos.net (and some others)
  - Increased from 251 to 683 classes
- Collected 12 logos per class, cleaned, and split into train/val
  - May need some more sample variety → use extracted annotation dataset

# Sample Removed Logos

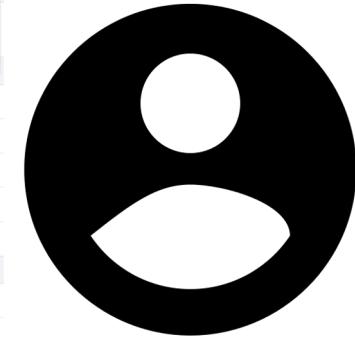


secured by

sage pay



MasterCard Verified by  
SecureCode. VISA

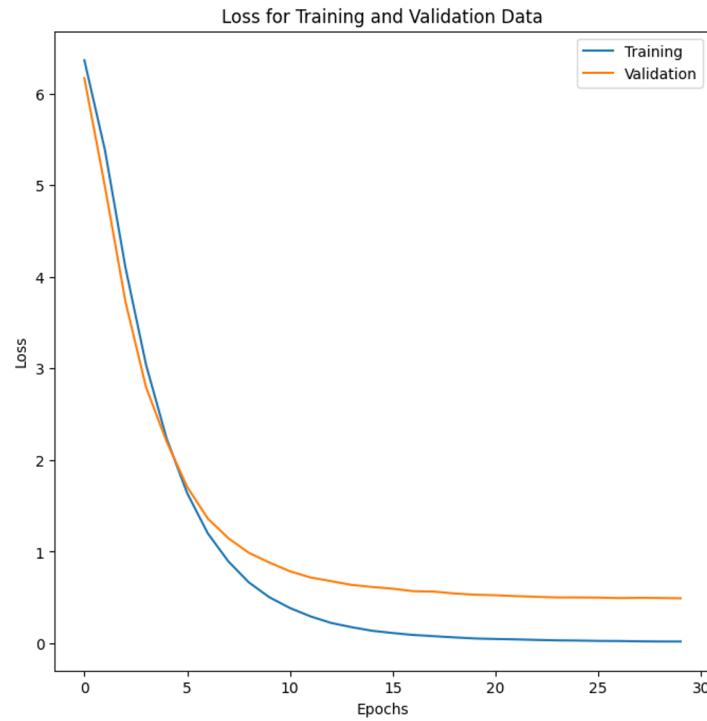
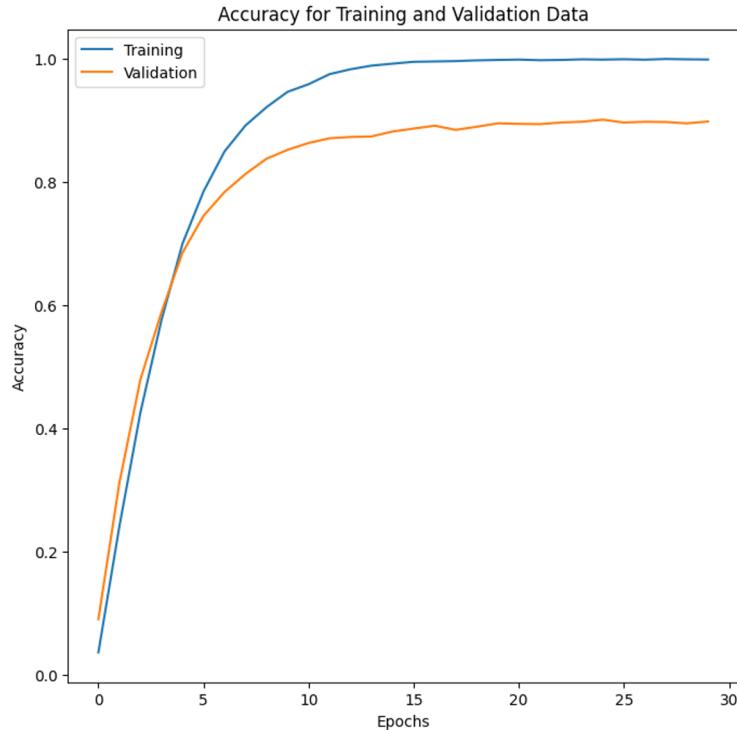


Duke

# Logo Classifier Iterations (All EfficientNetB0)

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
Classes	261	251	683
Number of Samples	Train: 3822 Val: 1273	Train: 3338 Val: 1117	Train: 6963 Val: 2354
Accuracy	<b>Train: 98.51%</b> <b>Val: 81.3%</b>	<b>Train: 99.6%</b> <b>Val: 90.6%</b>	<b>Train: 99.8%</b> <b>Val: 90.06%</b>

# Accuracy and Loss Curves of Most Recent Model



Duke

# Notes on Classifier/Data

- Some imbalance of classes
  - Twitter, Google, Facebook have many sample train images
  - Smaller companies may only have a few
  - May be representative of few-shot/single-shot
- Images generally high quality/resolution
  - Struggles on smaller detected logos
- Needs some more in-class variation
  - Improve accuracy
  - Include “words” next to logo
- Metrics on Validation Set
  - Accuracy: 90.06%
  - Precision: 91.57%
  - Recall: 90.06%

# Need for Additional Sample Variation



Duke

# Next Steps

- Error Analysis
- Data
  - Annotations from training data
    - Add samples to increase confidence and improve performance
    - Additional classes (bing downloader not working correctly)
  - Merge train/val and resplit after adding new images
- Training
  - EfficientNetB4
    - Compare model performance
  - Additional data augmentation and hyperparameter tuning
- Open-set Learning
  - Kmeans on mean activation vector vs. softmax threshold
  - Group unknowns or simply classify as unknown?