Proofpoint
Logo Detection & Classification

18th March 2021

Duke

Agenda

- Logistics
- Training Classifier with Corrected Aspect Ratio
- KNN Based Open Set Learning
- Open Long Tail Recognition

Duke

Logistics

- Proofpoint Presentation on 4/22 at 12pm.
- Conclude exploratory work by 1st April (next Proofpoint meeting)
- Develop Unseen Test Set and Retrain models by 15th April.
- Submit deliverables to Proofpoint by 30th April.

Duke

Main Deliverables to Proofpoint

1. Code used to run predictions.
2. Code used to train models.
3. Code use for error/performance analysis.
4. Model Artifacts.
5. All images and annotations.
6. Write Ups.

Duke

Corrected Aspect Ratios

Augmentation Adjustments

- Resizing
  - Previously used standard resize in PyTorch
    - Did not maintain aspect ratio
- Color Jitter
  - Parameters were not set correctly → no impact upon inspect
  - Adjusted brightness, contrast, saturation, hue
- RandomResizeCrop and RandomRotation(90 degrees) maintained

Duke

Previous Resizing

Duke

New Augmentation Examples

Duke

1098 Classes Performance on Test Set

	Accuracy	Precision	Recall	F1-Score
Old Aug	0.9373	0.9780	0.9373	0.9532
New Aug	0.9451	0.9825	0.9451	0.9583

Duke

Closed Set (with Unknown) Performance

	Accuracy	Precision	Recall	F1-Score
Old Aug	0.8950	0.9120	0.8950	0.8978
New Aug	0.9210	0.9291	0.9210	0.9224

Duke

Open Set Learning

Duke

Terminology

Correct Unknowns: Percentage of true unknowns correctly classified as unknown.

Incorrect Unknowns: Percentage of known logos **incorrectly** called unknown.

Accuracy Open Set: Accuracy of model on known logos **given** that it was trained with the assumption that it would be deployed in the open world.

Duke

Open Set Performance (32 Known Classes)

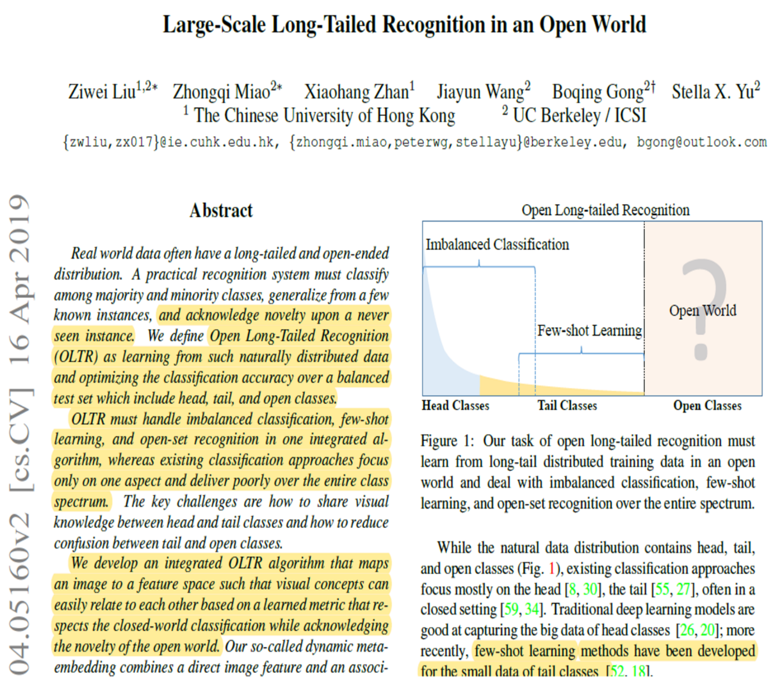
	Correct Unknowns	Incorrect Unknowns	Accuracy Open Set
Unknown Class	**84.4%**	**4.0%**	**94.5%**
OLTR 1	86.7%	16.4%	82.4%
OLTR 2	**88.9%**	**8.9%**	**90.7%**
KNN (1st)	92.2%	10.2%	89.5%
KNN (2nd)	93.3%	9.4%	90.4%
KNN (1st + 2nd)	**97.8%**	**9.4%**	**90.4%**

Duke

Open Long Tail Recognition - OLTR

Duke

Large-Scale Long-Tailed Recognition in an Open World



Large-Scale Long-Tailed Recognition in an Open World

Ziwei Liu[1,2*] Zhongqi Miao[2*] Xiaohang Zhan[1] Jiayun Wang[2] Boqing Gong[2†] Stella X. Yu[2]
[1] The Chinese University of Hong Kong [2] UC Berkeley / ICSI
{zwliu,zx017}@ie.cuhk.edu.hk, {zhongqi.miao,peterwg,stellayu}@berkeley.edu, bgong@outlook.com

Abstract

Real world data often have a long-tailed and open-ended distribution. A practical recognition system must classify among majority and minority classes, generalize from a few known instances, and acknowledge novelty upon a never seen instance. We define Open Long-Tailed Recognition (OLTR) as learning from such naturally distributed data and optimizing the classification accuracy over a balanced test set which include head, tail, and open classes.

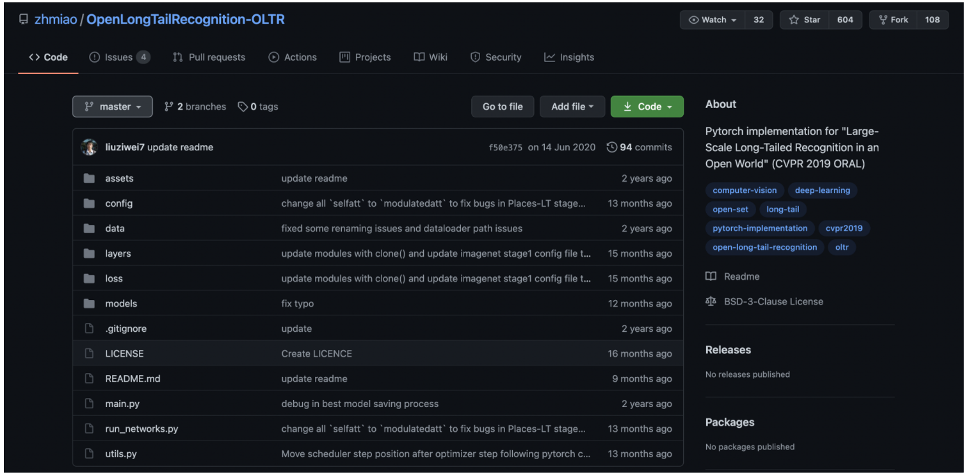
OLTR must handle imbalanced classification, few-shot learning, and open-set recognition in one integrated algorithm, whereas existing classification approaches focus only on one aspect and deliver poorly over the entire class spectrum. The key challenges are how to share visual knowledge between head and tail classes and how to reduce confusion between tail and open classes.

We develop an integrated OLTR algorithm that maps an image to a feature space such that visual concepts can easily relate to each other based on a learned metric that respects the closed-world classification while acknowledging the novelty of the open world. Our so-called dynamic meta-embedding combines a direct image feature and an associ-

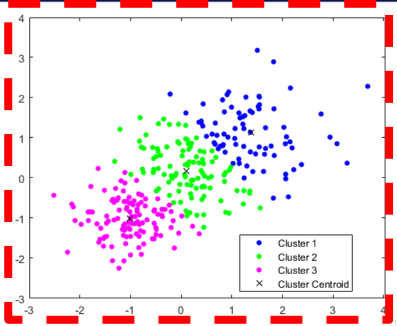
Open Long-tailed Recognition
Imbalanced Classification
Few-shot Learning
Open World
Head Classes
Tail Classes
Open Classes

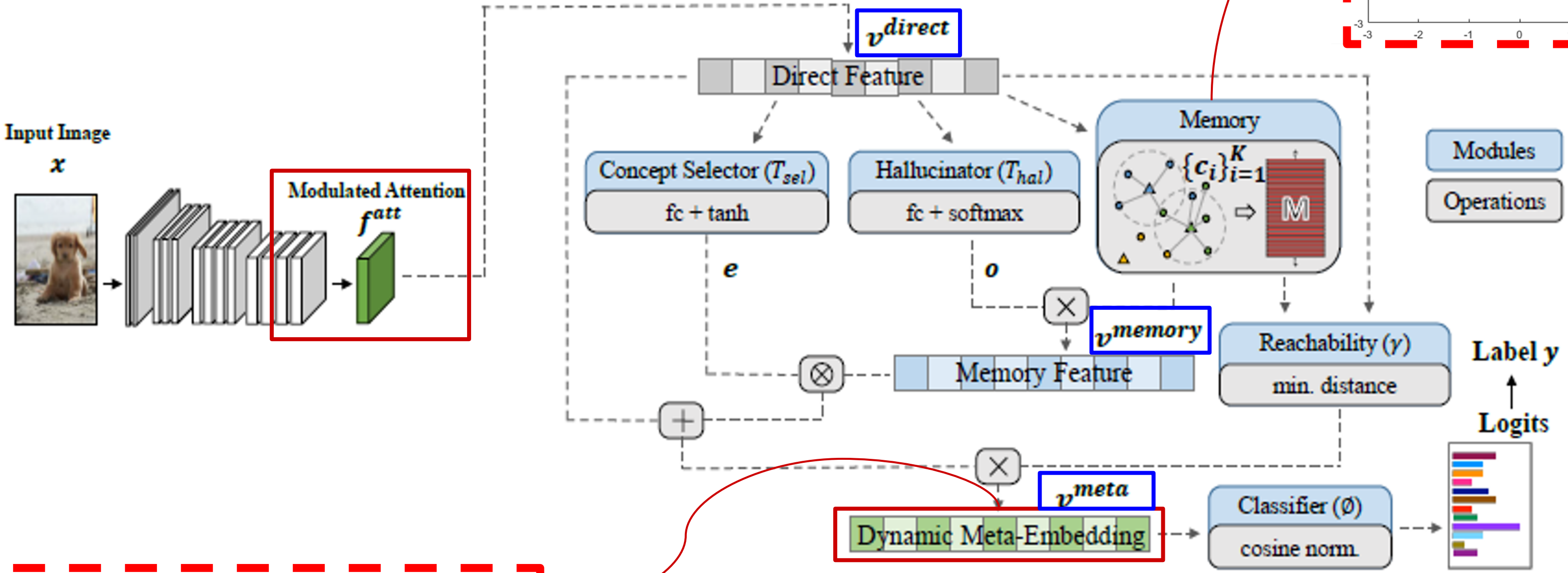
Figure 1: Our task of open long-tailed recognition must learn from long-tail distributed training data in an open world and deal with imbalanced classification, few-shot learning, and open-set recognition over the entire spectrum.

While the natural data distribution contains head, tail, and open classes (Fig. 1), existing classification approaches focus mostly on the head [8, 30], the tail [55, 27], often in a closed setting [59, 34]. Traditional deep learning models are good at capturing the big data of head classes [26, 20]; more recently, few-shot learning methods have been developed for the small data of tail classes [52, 18].



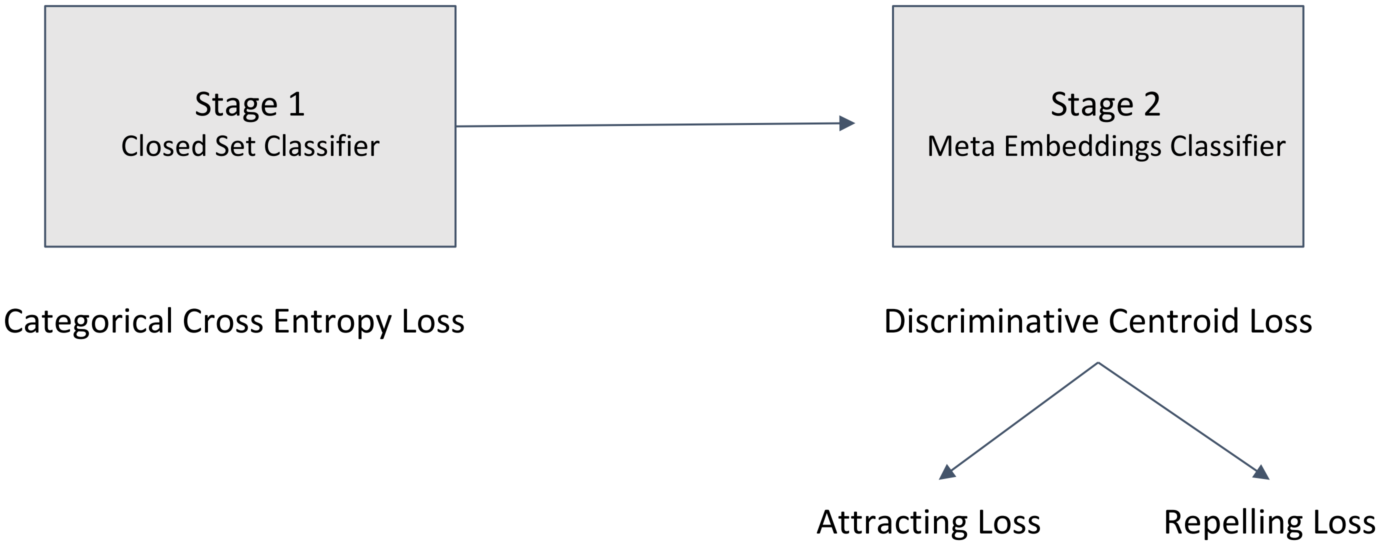
Duke





$$v^{meta} = (1/\gamma) \cdot (v^{direct} + e \otimes v^{memory})$$

Procedure



Stage 1
Closed Set Classifier
Stage 2
Meta Embeddings Classifier
Categorical Cross Entropy Loss
Discriminative Centroid Loss
Attracting Loss
Repelling Loss

Pros & Cons

Pros:

- Very quick during train and test time. Much faster than kNN
- No need to store and retrieve activation vectors

Cons:

- Not so great code
- Switching out architectures was complicated

KNN Based Open Set Learning

Duke

Procedure

1. Get second last activation vectors (AVs) of all train and validation observations.
2. Get distances of all true and unknown validation AVs from train AVs.
3. For each class, tune a distance threshold that maximizes accuracy for that class.
4. If no unknown is predicted into a given class:
   threshold = max(distances) + std(distances)
5. Use tuned thresholds for prediction on test observations and test unknowns.

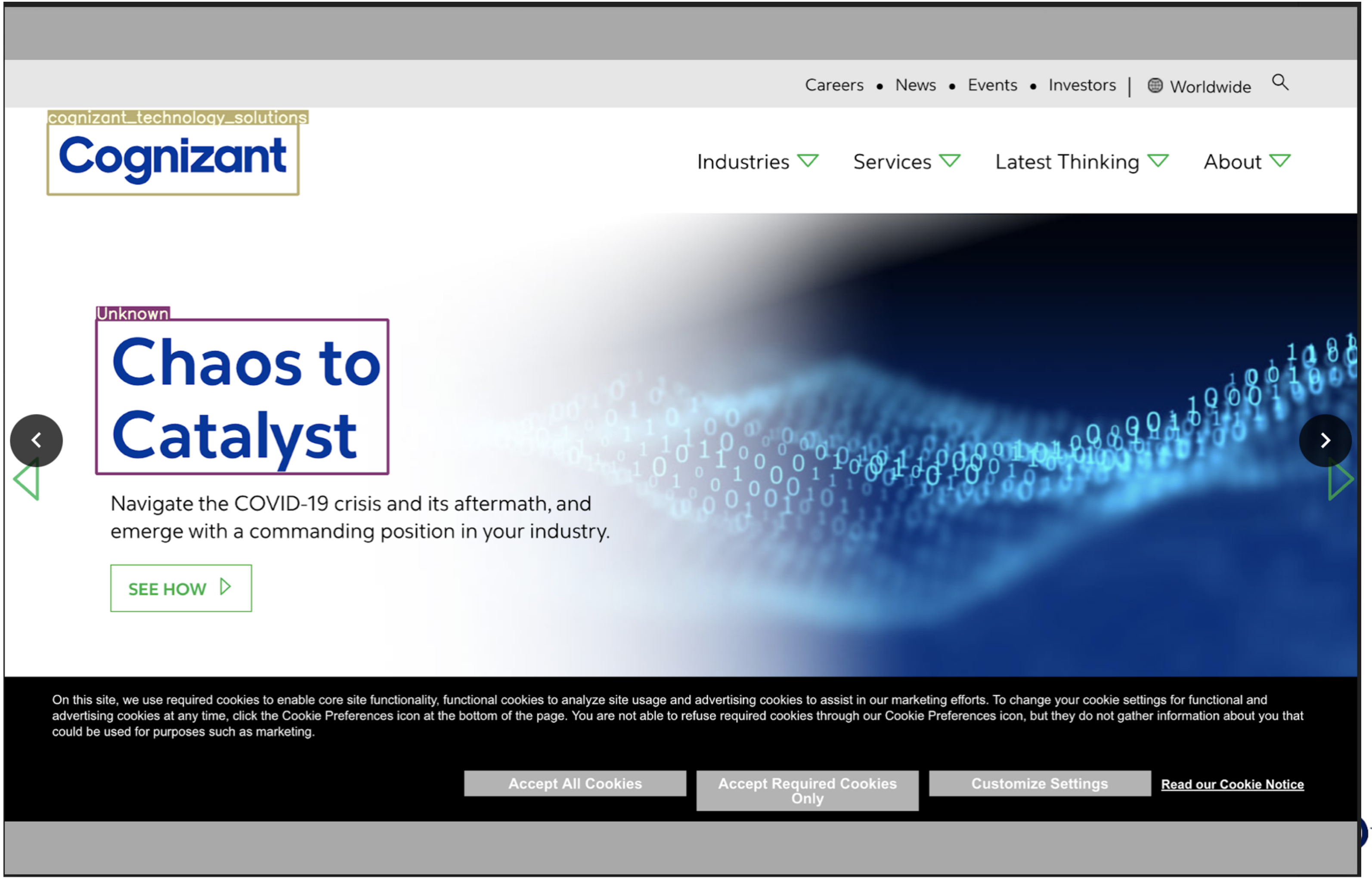
KNN Open Set Demo (1098 Known Classes)

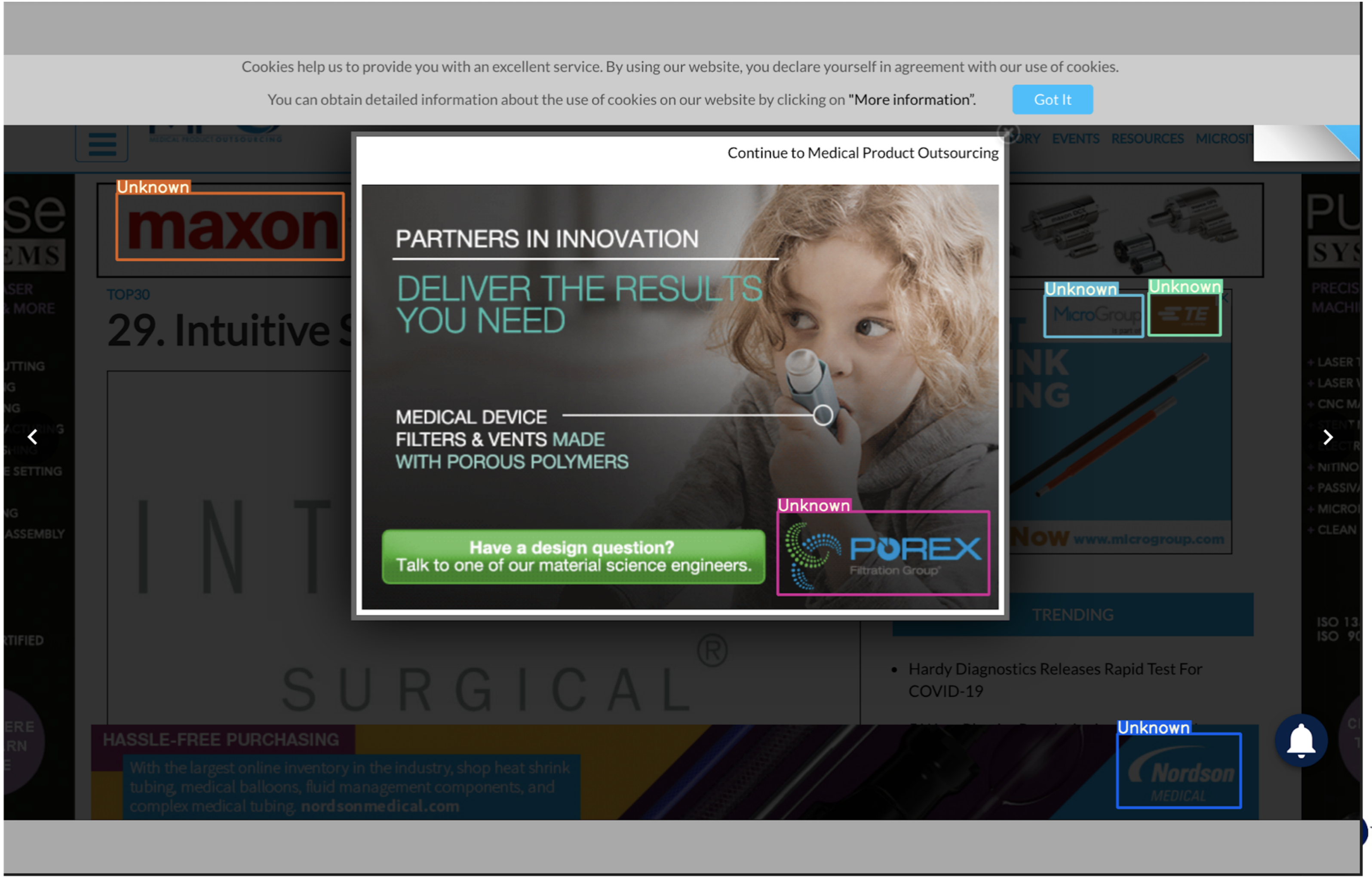
Correct Unknowns: 73.36%

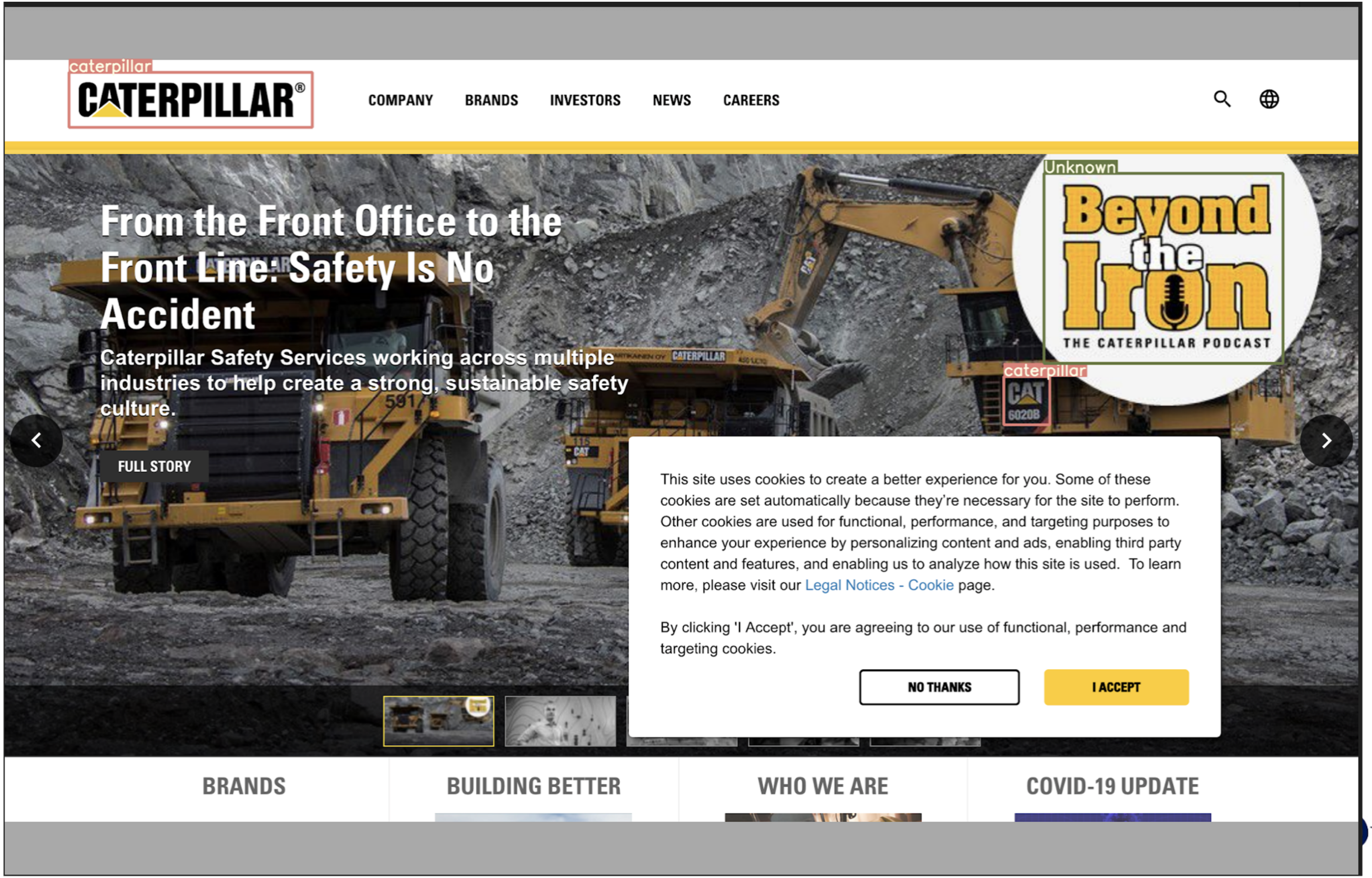
Incorrect Unknowns: 6.18%

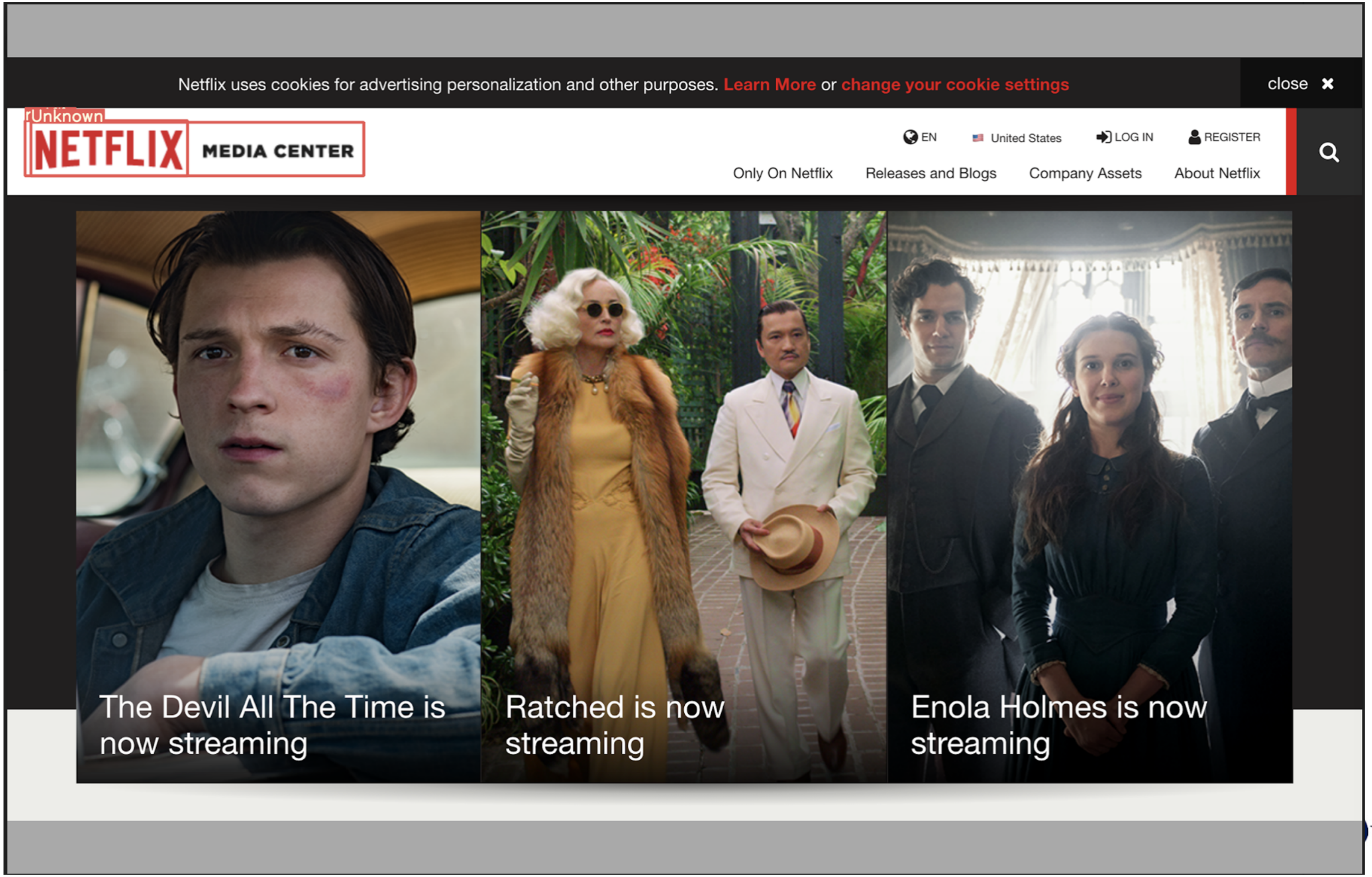
Accuracy on Open Set: 91.42%

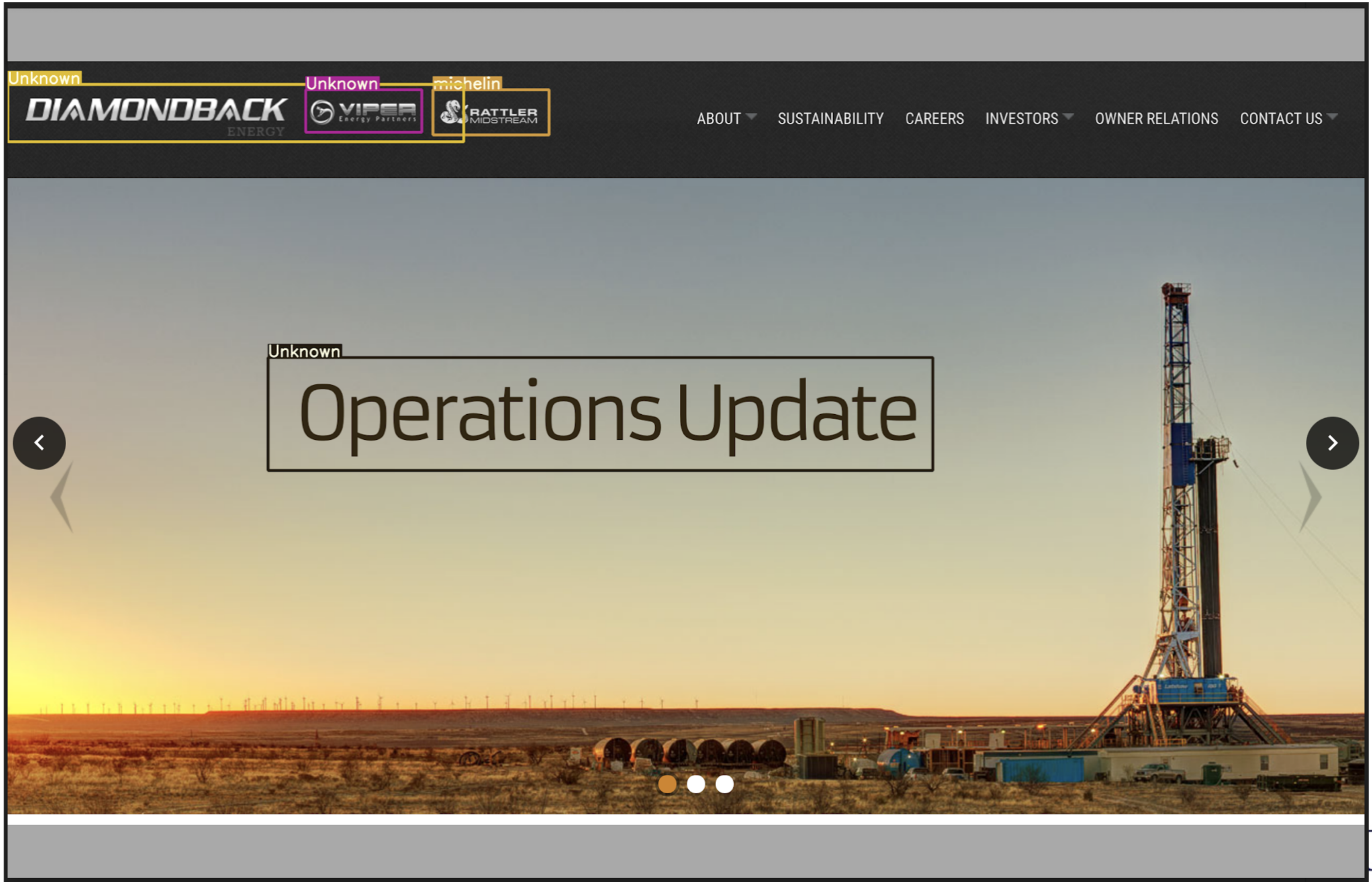
Demo:
https://colab.research.google.com/drive/12jBFlQVDchfblGxWrmb-UGRcjoq6L6pv?usp=sharing


Careers • News • Events • Investors | Worldwide
cognizant_technology_solutions
Cognizant
Industries
Services
Latest Thinking
About
Unknown
Chaos to Catalyst
Navigate the COVID-19 crisis and its aftermath, and emerge with a commanding position in your industry.
SEE HOW
On this site, we use required cookies to enable core site functionality, functional cookies to analyze site usage and advertising cookies to assist in our marketing efforts. To change your cookie settings for functional and advertising cookies at any time, click the Cookie Preferences icon at the bottom of the page. You are not able to refuse required cookies through our Cookie Preferences icon, but they do not gather information about you that could be used for purposes such as marketing.
Accept All Cookies
Accept Required Cookies Only
Customize Settings
Read our Cookie Notice

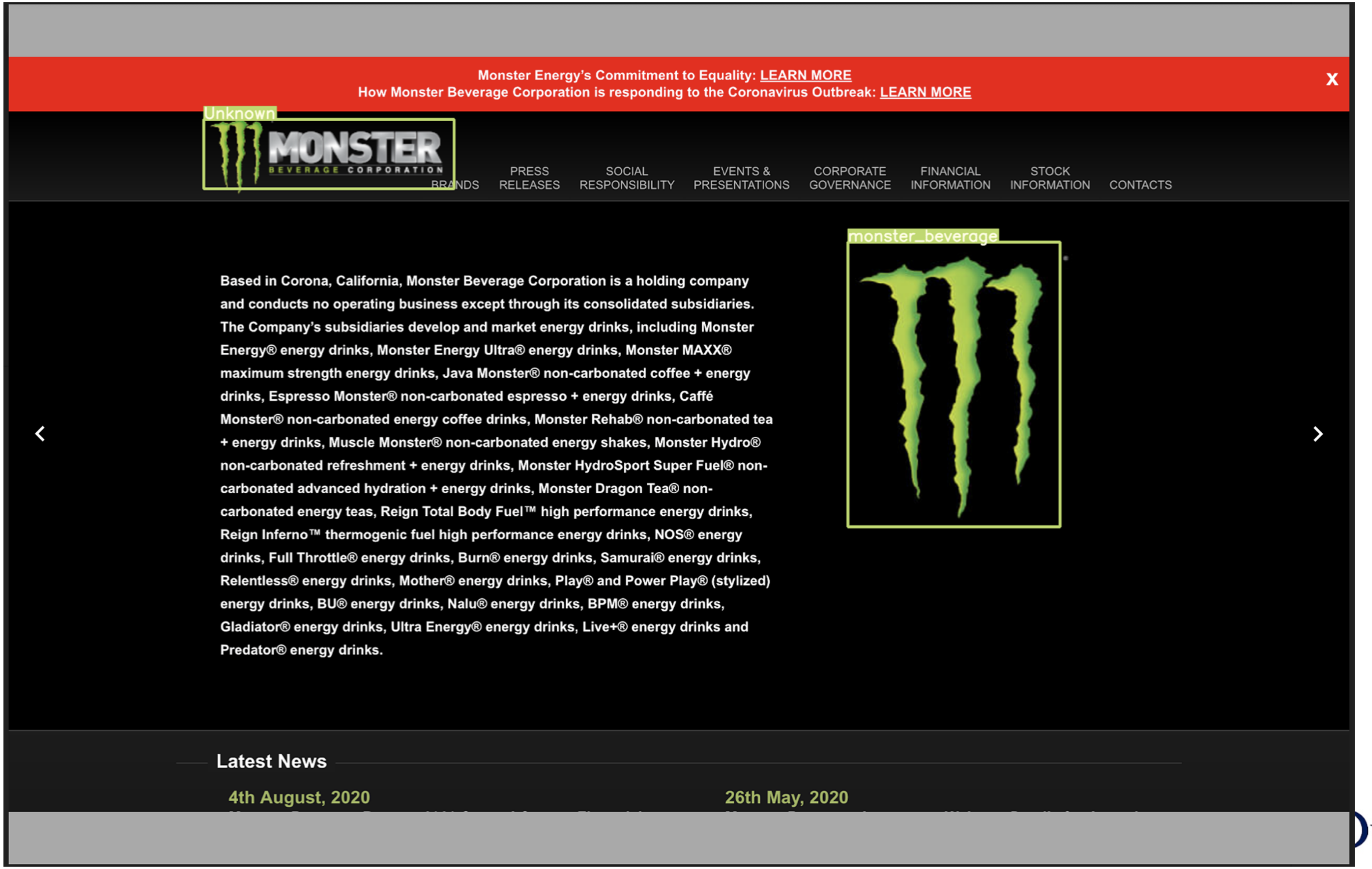

Cookies help us to provide you with an excellent service. By using our website, you declare yourself in agreement with our use of cookies.
You can obtain detailed information about the use of cookies on our website by clicking on "More information".
Got It
EVENTS
RESOURCES
Continue to Medical Product Outsourcing
Unknown
maxon
TOP30
29. Intuitive S
PARTNERS IN INNOVATION
DELIVER THE RESULTS YOU NEED
MEDICAL DEVICE
FILTERS & VENTS MADE
WITH POROUS POLYMERS
Have a design question?
Talk to one of our material science engineers.
Unknown
POREX
Filtration Group
Unknown
MicroGroup
Unknown
TE
www.microgroup.com
SURGICAL
TRENDING
Hardy Diagnostics Releases Rapid Test For COVID-19
HASSLE-FREE PURCHASING
With the largest online inventory in the industry, shop heat shrink tubing, medical balloons, fluid management components, and complex medical tubing. nordsonmedical.com
Unknown
Nordson
MEDICAL


caterpillar
CATERPILLAR®
COMPANY
BRANDS
INVESTORS
NEWS
CAREERS
Unknown
Beyond the Iron
THE CATERPILLAR PODCAST
caterpillar
CAT
From the Front Office to the Front Line: Safety Is No Accident
Caterpillar Safety Services working across multiple industries to help create a strong, sustainable safety culture.
FULL STORY
This site uses cookies to create a better experience for you. Some of these cookies are set automatically because they're necessary for the site to perform. Other cookies are used for functional, performance, and targeting purposes to enhance your experience by personalizing content and ads, enabling third party content and features, and enabling us to analyze how this site is used. To learn more, please visit our Legal Notices - Cookie page.
By clicking 'I Accept', you are agreeing to our use of functional, performance and targeting cookies.
NO THANKS
I ACCEPT
BRANDS
BUILDING BETTER
WHO WE ARE
COVID-19 UPDATE

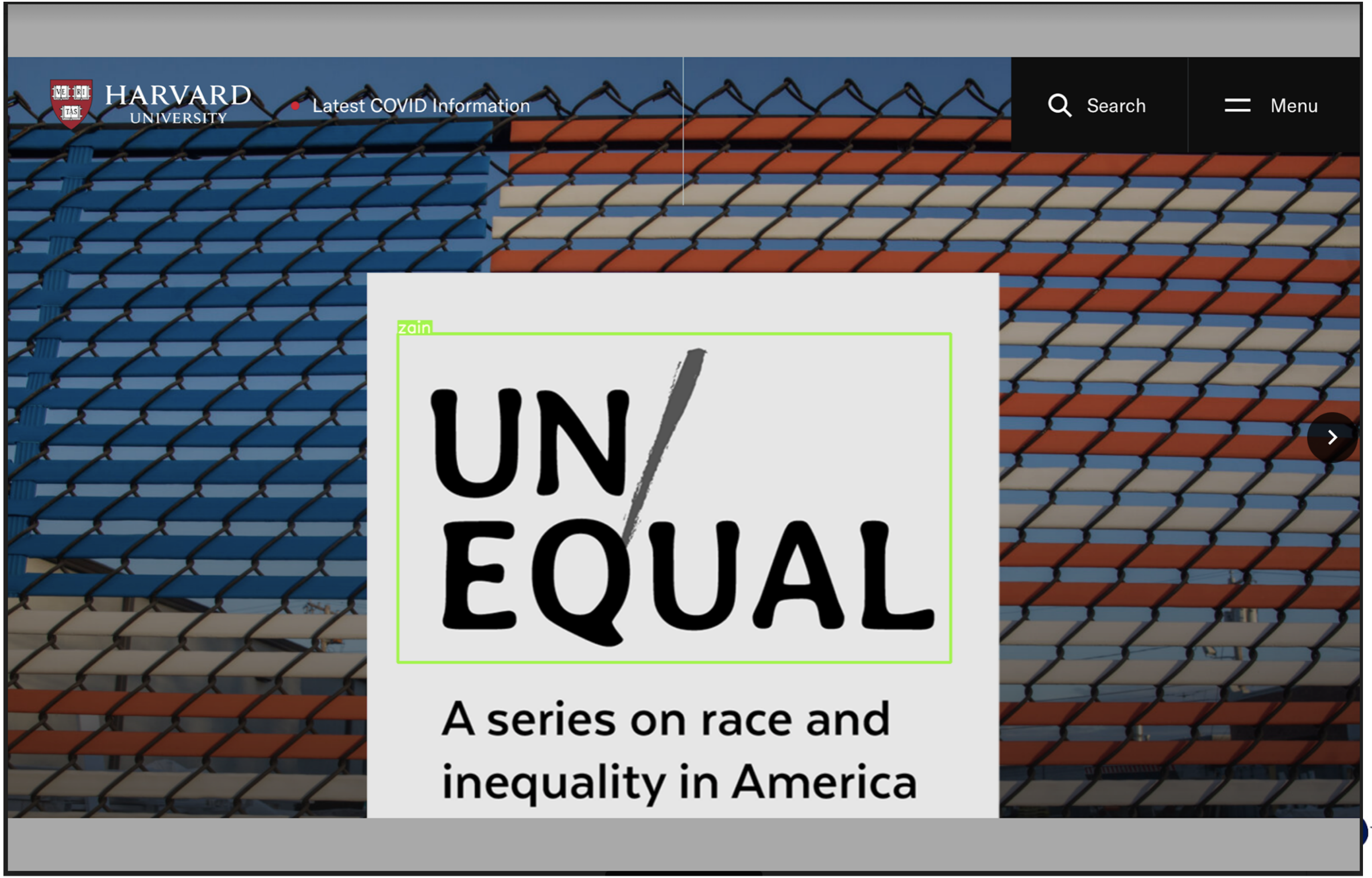

Netflix uses cookies for advertising personalization and other purposes. Learn More or change your cookie settings
close
rUnknown
NETFLIX MEDIA CENTER
EN
United States
LOG IN
REGISTER
Only On Netflix
Releases and Blogs
Company Assets
About Netflix
The Devil All The Time is now streaming
Ratched is now streaming
Enola Holmes is now streaming


Unknown
DIAMONDBACK
ENERGY
Unknown
VIPER
Energy Partners
michelin
RATTLER
MIDSTREAM
ABOUT
SUSTAINABILITY
CAREERS
INVESTORS
OWNER RELATIONS
CONTACT US
Unknown
Operations Update

Duke



Monster Energy's Commitment to Equality: LEARN MORE
How Monster Beverage Corporation is responding to the Coronavirus Outbreak: LEARN MORE
X
Unknown
MONSTER
BEVERAGE CORPORATION
BRANDS
PRESS RELEASES
SOCIAL RESPONSIBILITY
EVENTS & PRESENTATIONS
CORPORATE GOVERNANCE
FINANCIAL INFORMATION
STOCK INFORMATION
CONTACTS
monster_beverage
Based in Corona, California, Monster Beverage Corporation is a holding company and conducts no operating business except through its consolidated subsidiaries. The Company's subsidiaries develop and market energy drinks, including Monster Energy® energy drinks, Monster Energy Ultra® energy drinks, Monster MAXX® maximum strength energy drinks, Java Monster® non-carbonated coffee + energy drinks, Espresso Monster® non-carbonated espresso + energy drinks, Caffé Monster® non-carbonated energy coffee drinks, Monster Rehab® non-carbonated tea + energy drinks, Muscle Monster® non-carbonated energy shakes, Monster Hydro® non-carbonated refreshment + energy drinks, Monster HydroSport Super Fuel® non-carbonated advanced hydration + energy drinks, Monster Dragon Tea® non-carbonated energy teas, Reign Total Body Fuel™ high performance energy drinks, Reign Inferno™ thermogenic fuel high performance energy drinks, NOS® energy drinks, Full Throttle® energy drinks, Burn® energy drinks, Samurai® energy drinks, Relentless® energy drinks, Mother® energy drinks, Play® and Power Play® (stylized) energy drinks, BU® energy drinks, Nalu® energy drinks, BPM® energy drinks, Gladiator® energy drinks, Ultra Energy® energy drinks, Live+® energy drinks and Predator® energy drinks.
Latest News
4th August, 2020
26th May, 2020


HARVARD
UNIVERSITY
Latest COVID Information
Search
Menu
zain
UN
EQUAL
A series on race and inequality in America

Drawbacks of KNN Open Set Learning

- Speed - need to calculate distances of each predicted logos to 12,000+ activation vectors.

- Storage
  - Activation vectors (AVs) need to be stored and loaded
  - Generate new AVs for slight changes in models; slow to do model testing

- Hyper parameter tuning - requires all known classes to have some incorrect predictions for threshold tuning.

- Threshold flexibility is limited: percentage of correct unknowns vs incorrect unknowns are challenging to adjust.

Duke

Way Forward...