Logos Finder : Computer Vision Project

AY 2021/2022
Davide Basile   1810355
Valerio Goretti 1811110
Marco Raffaele 1799912

Good Eleven

The project

An Artificial Intelligence based system that alows the user to control the presence of logos in images and videos
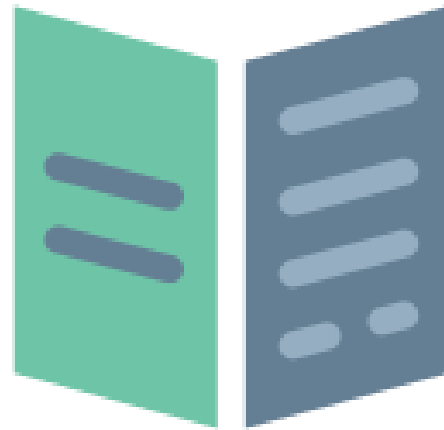
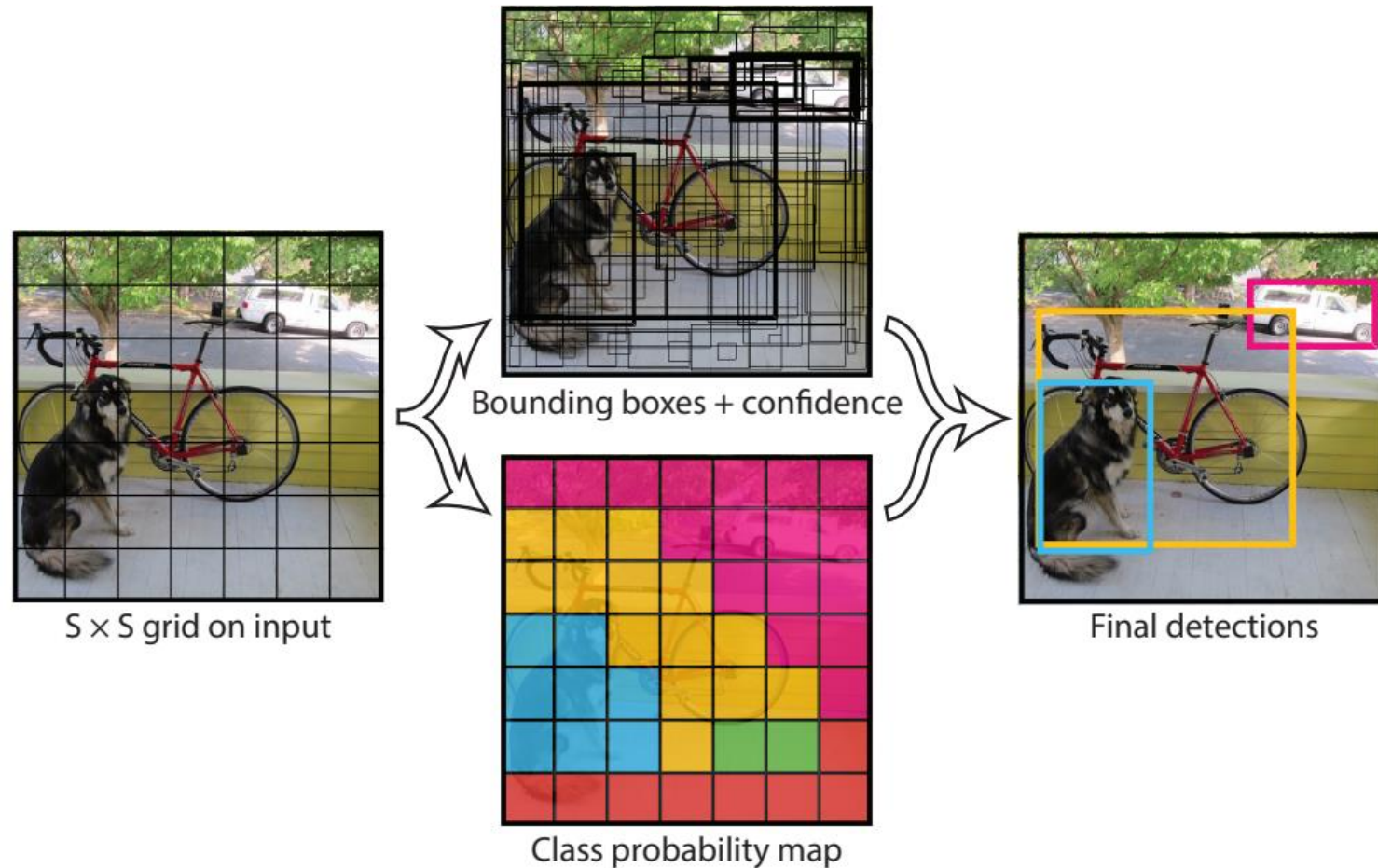Good Eleven

Use case scenario

The state of the art

<<You Only Look Once: Unified, Real-Time Object Detection>>

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. (2016).

Bounding boxes + confidence

S × S grid on input

Class probability map

Final detections

Good Eleven

LINK: https://arxiv.org/abs/1506.02640

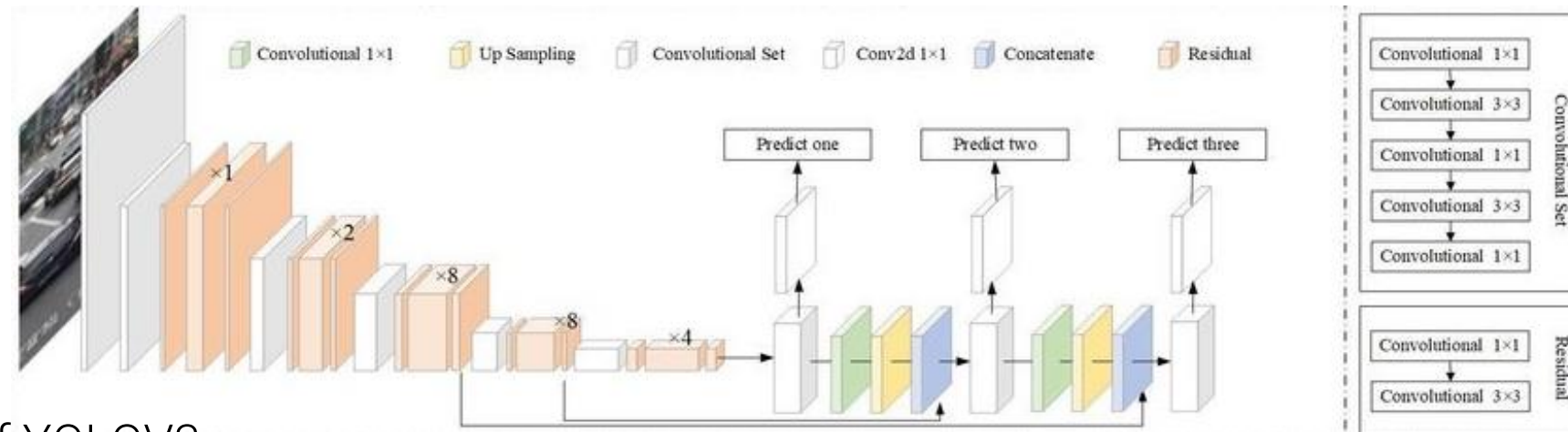# <<YOLOv3: An Incremental Improvement>>

Joseph Redmon, Ali Farhadi. (2018).

Blue rectangle: Image
Red rectangle: Cell
Yellow Rectangles: Anchor Boxes

Architecture of YOLOV3

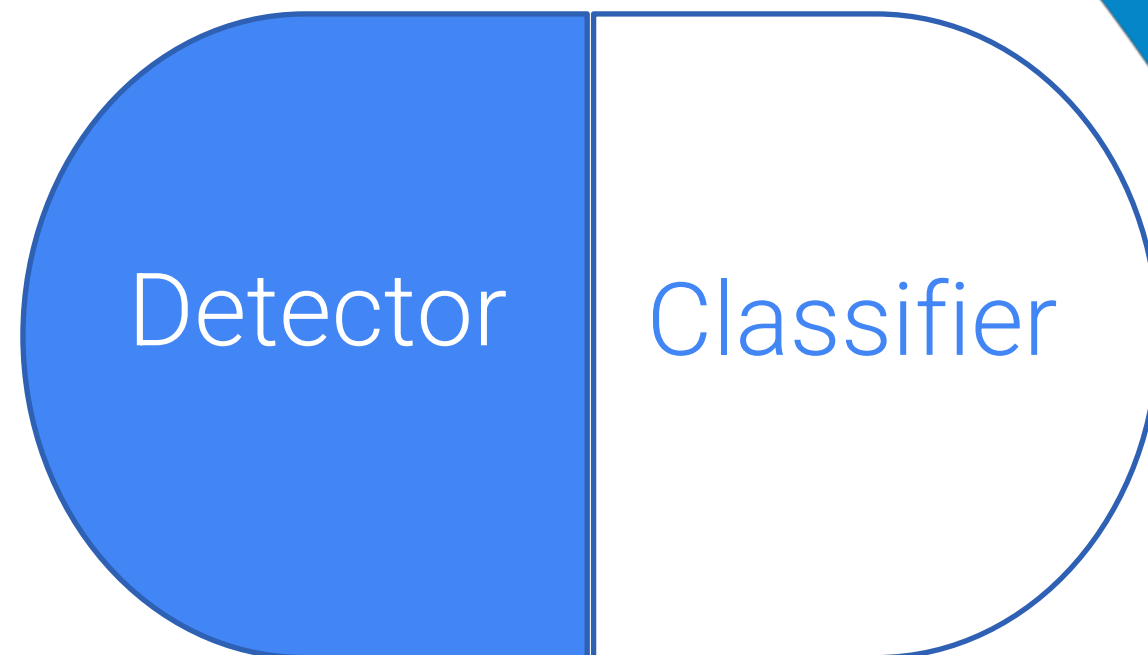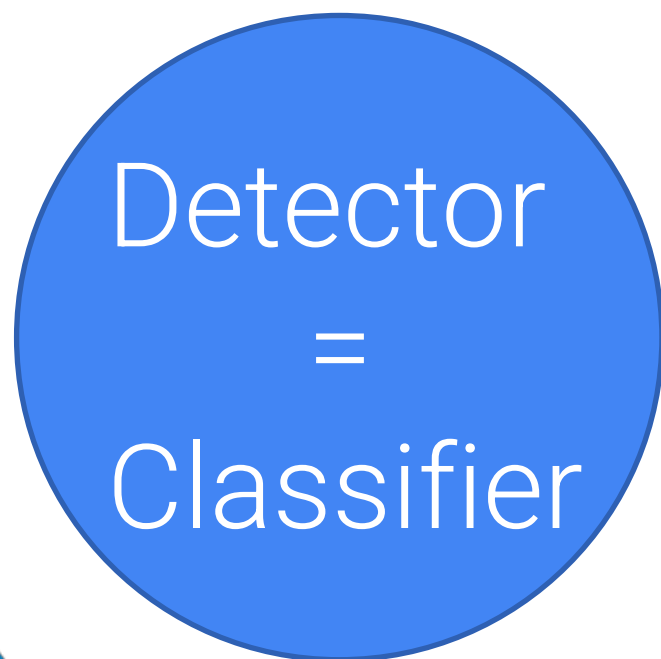LINK: https://arxiv.org/abs/1804.02767

Good Eleven

<<YOLOV3 BASED SHIP DETECTION IN VISIBLE AND INFRARED IMAGES>>

Lena Chang; Yi-Ting Chen; Ming-Hung Hung; Jung-Hua Wang; Yang-Lang Chang. (2021).

LINK: https://ieeexplore.ieee.org/document/9555032

Good Eleven

Divide et impera solution

Good Eleven

The AI workflow

Input Image

Preprocessed Image

Logos Hypothesis

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & .... & a_{1j} & .... & a_{1n} \\ .... & .... & .... & .... & .... & .... & .... \\ a_{i1} & a_{i2} & a_{i3} & .... & a_{ij} & .... & a_{in} \\ .... & .... & .... & .... & .... & .... & .... \\ a_{m1} & a_{m2} & a_{m3} & .... & a_{mj} & .... & a_{mn} \end{bmatrix}$$

Detector

Output

Classifier
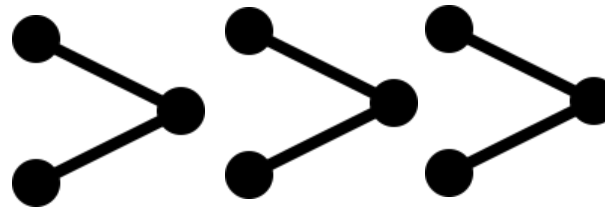
Preprocessed Cropped Image

FedEx class

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & .... & a_{1j} & .... & a_{1n} \\ .... & .... & .... & .... & .... & .... & .... \\ a_{i1} & a_{i2} & a_{i3} & .... & a_{ij} & .... & a_{in} \\ .... & .... & .... & .... & .... & .... & .... \\ a_{m1} & a_{m2} & a_{m3} & .... & a_{mj} & .... & a_{mn} \end{bmatrix}$$

Good Eleven
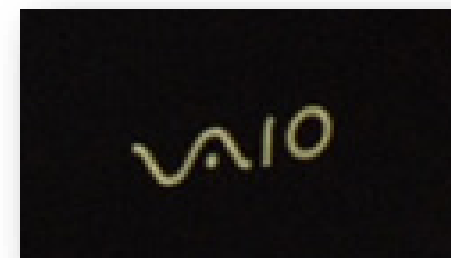
The detector development
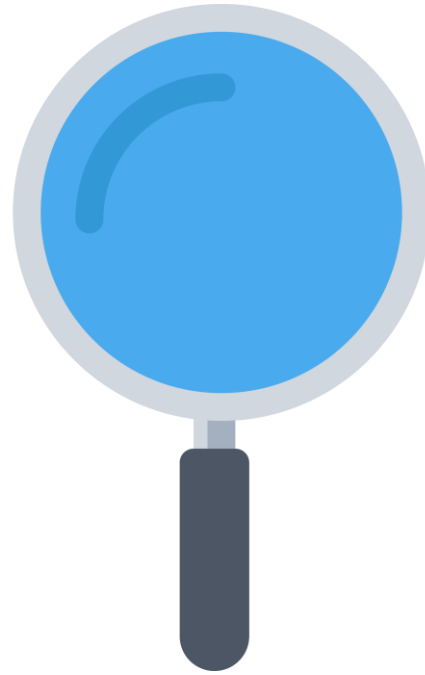
Good Eleven

The goal

Input image

Detector
(YOLOV3)

Hypothesis
Coordinate

Good Eleven

The datasets

LogosInTheWild

11054 images
871 classes
32.850 ROI's
Voc-format and ROI format

LogosDet3k

158.652 images
3000 classes
194.261 ROI's
Only voc-format

Good Eleven

Typical Voc-format image

Typical ROI format image

Good Eleven

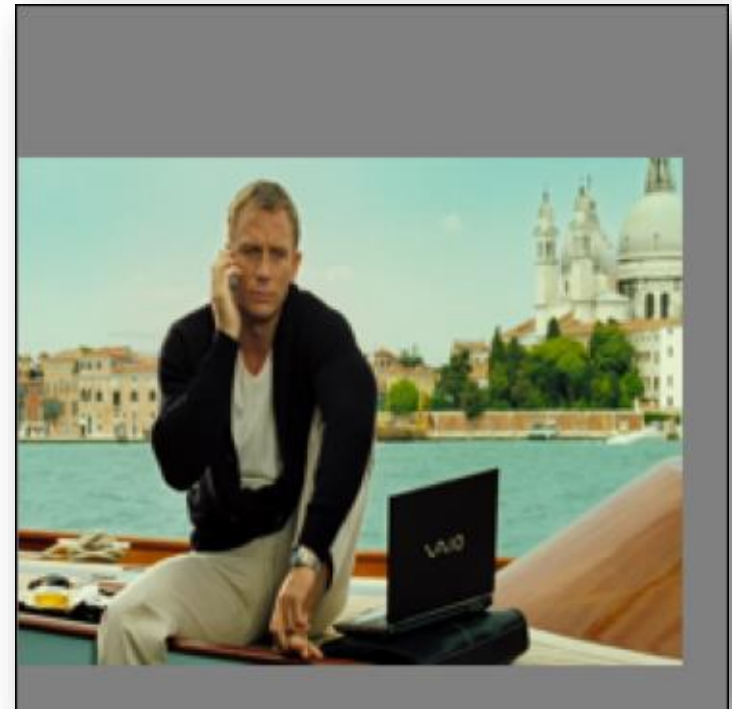# Image Preprocessing

Preprocessing Flow

1. Resize

2. Random Flip

3. Transposition to HSV

4. Random Distortion

5. Random Noise

Good Eleven

Image Preprocessing

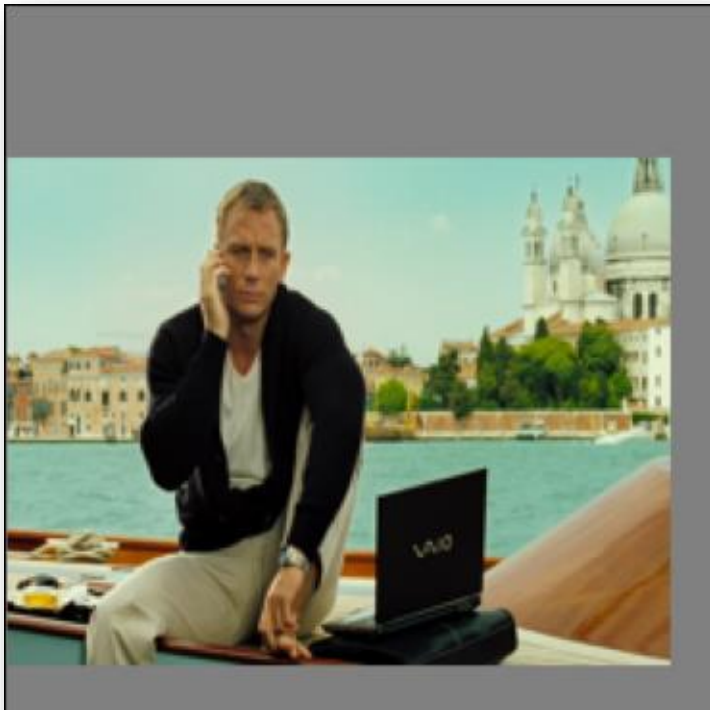1. Resize to 224 x 224

Good Eleven

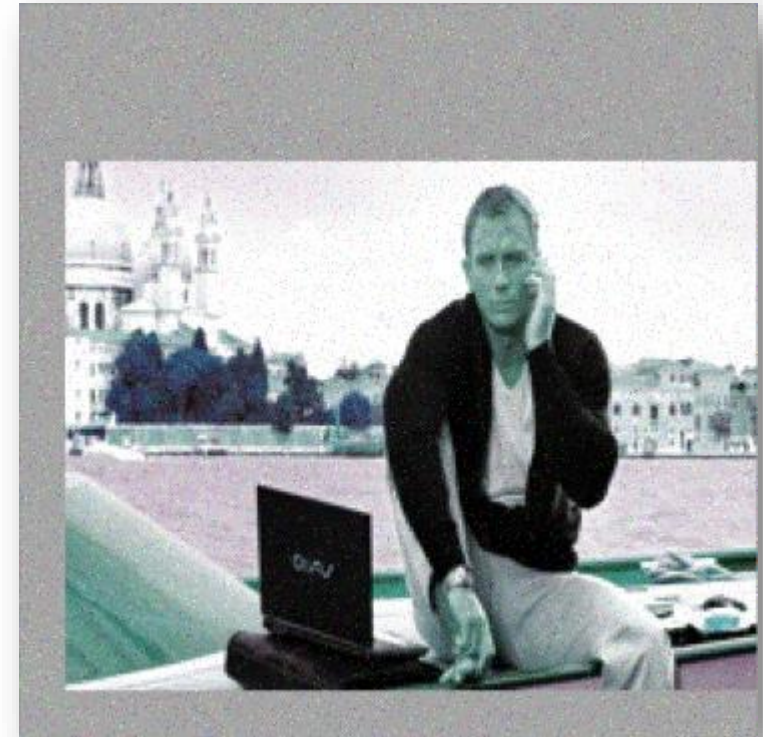# Image Preprocessing

2. Random flip

Image Preprocessing

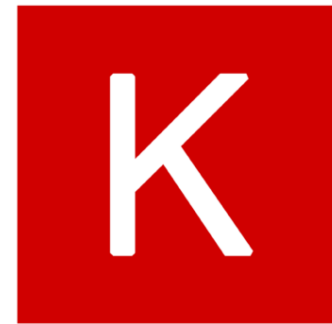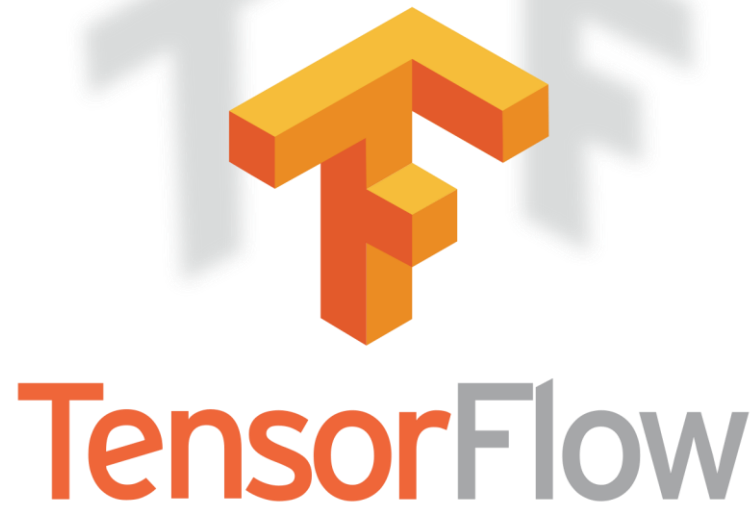## 3-4. Transposition to HSV space and random distortion

Image Preprocessing

5. Random noise
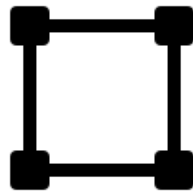

Good Eleven

Training of YoloV3

Good Eleven

Training of YoloV3

X: Preprocessed Image

Y: Ground Truth Box, Class (Logo)

Training YoloV3

Logo Knowledge
(Model Weights)

Training Infos

1 Class (Logo)
11.000 image samples
150 Epochs

Good Eleven

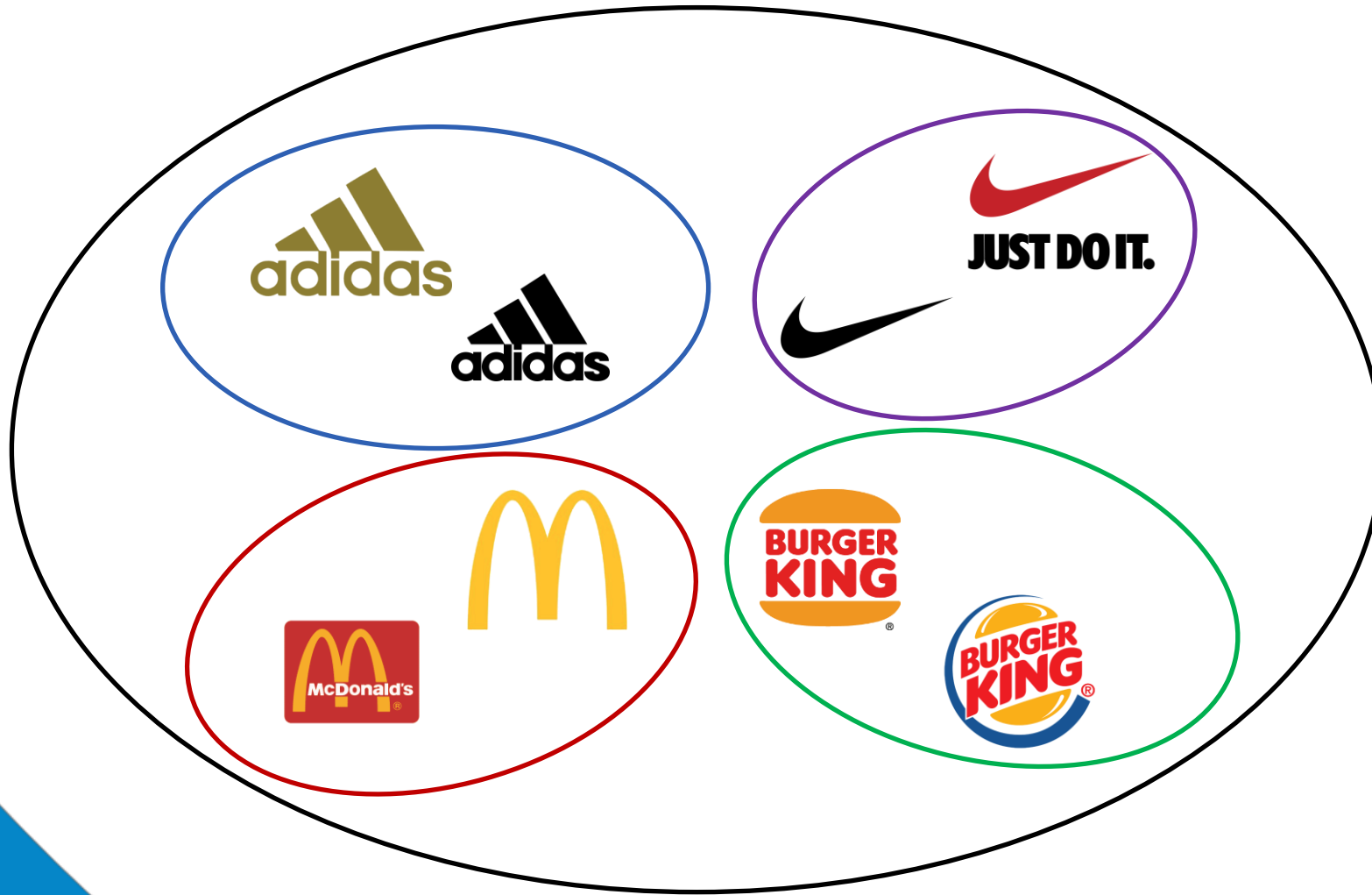# Evaluation of the trained model

Good Eleven

Evaluation of the trained model

```
Accuracy of the test: 0.87425
Mean IOU of the test: 0.2363496496710042
```

Good Eleven

The classifier development

Good Eleven

Custom training approach

INPUT   **CONVOLUTION + RELU**   **POOLING**   **CONVOLUTION + RELU**   **POOLING**   **FLATTEN**   **FULLY CONNECTED**   **SOFTMAX**

FEATURE LEARNING   **CLASSIFICATION**

— CAR
— TRUCK
— VAN
— BICYCLE
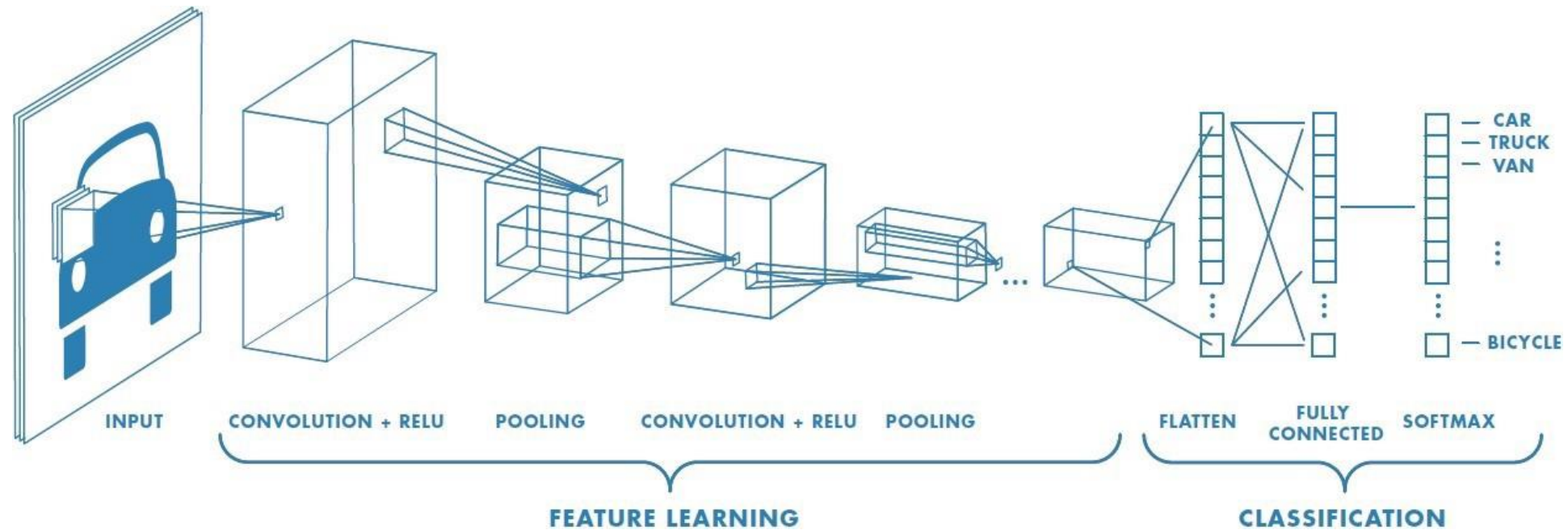
Typical CNN architecture

Problem

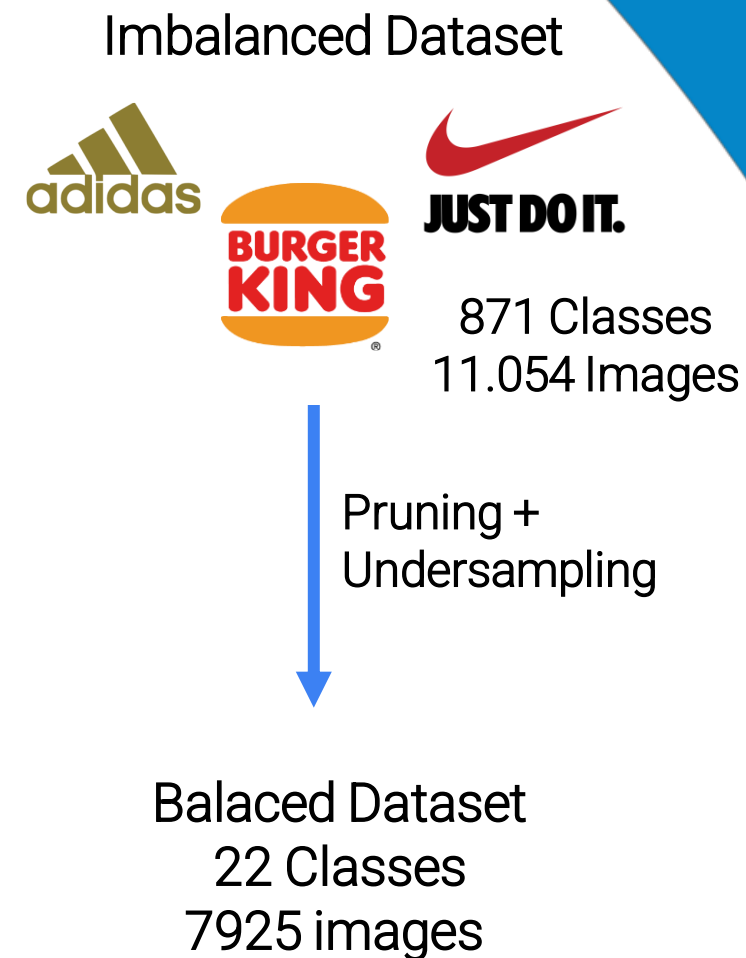Both the datasets have too much classes and each of them has a too variable number of samples

Solutions

- Balancing Operations: Undersampling + Pruning

- Class weights dictionary

Good Eleven

Balancing Operations

Pruning: Delete the classes that have too few samples

Undersampling: Delete random samples from the classes that have too much images

Imbalanced Dataset

871 Classes
11.054 Images

Pruning +
Undersampling

Balaced Dataset
22 Classes
7925 images

Good Eleven

Class weight dictionary

$$\text{Weight(C)} = \frac{\text{Number of samples in the dataset}}{\text{Number of classes} \times \text{Number of samples of C}}$$

Good Eleven

CNN Architectures

Model	Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
InceptionV3	0.937	23,851,784	159	42.25	6.86
VGG16	0.901	138,357,544	23	69.50	4.16
Resnet50V2	0.930	25,613,800	-	45.63	4.42

Good Eleven

Evaluation of the trained models
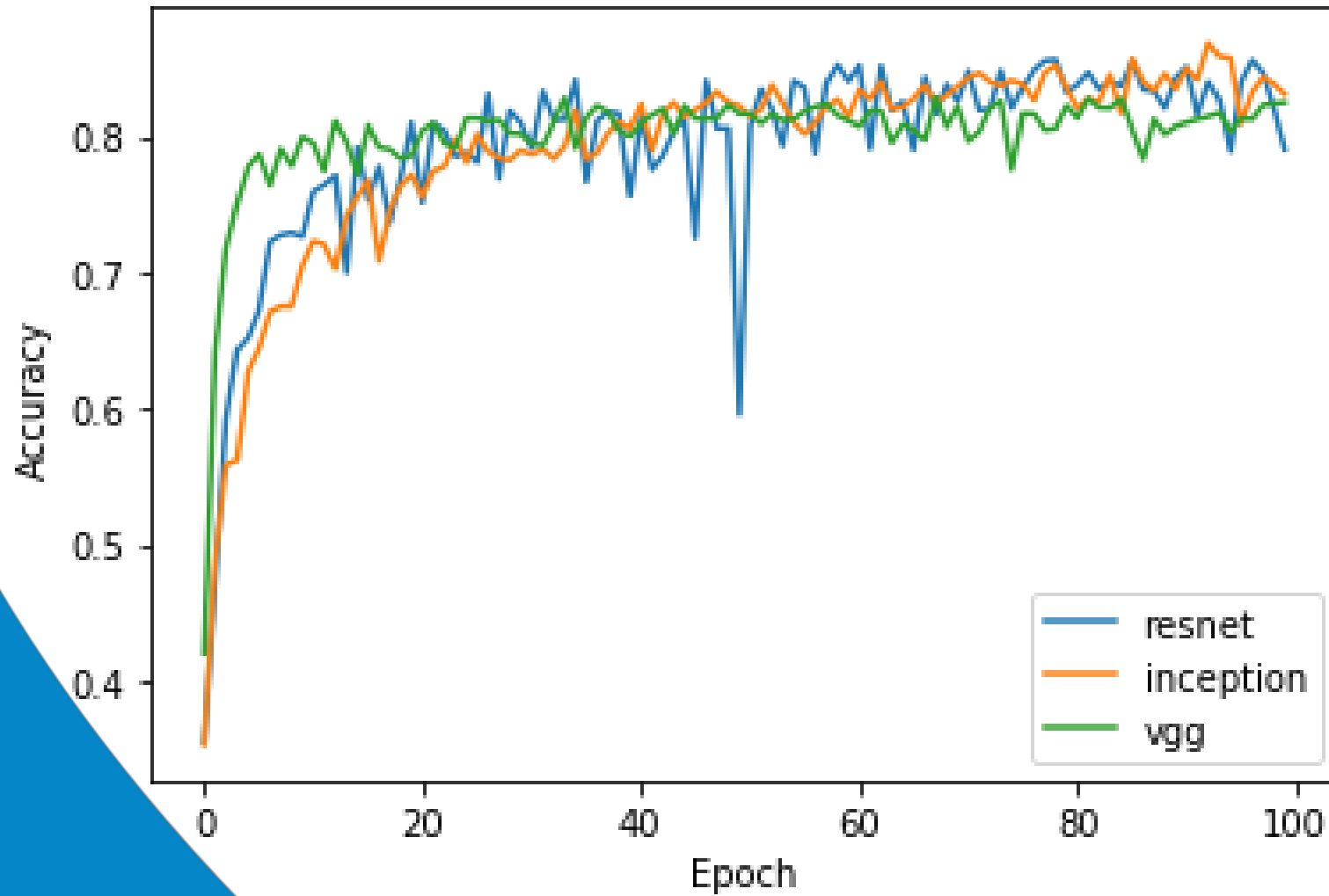
Accuracies

InceptionV3	0.8307
Resnet50V2	0.7905
VGG16	0.8245

Good Eleven

Transfer Learning Approach

Good Eleven
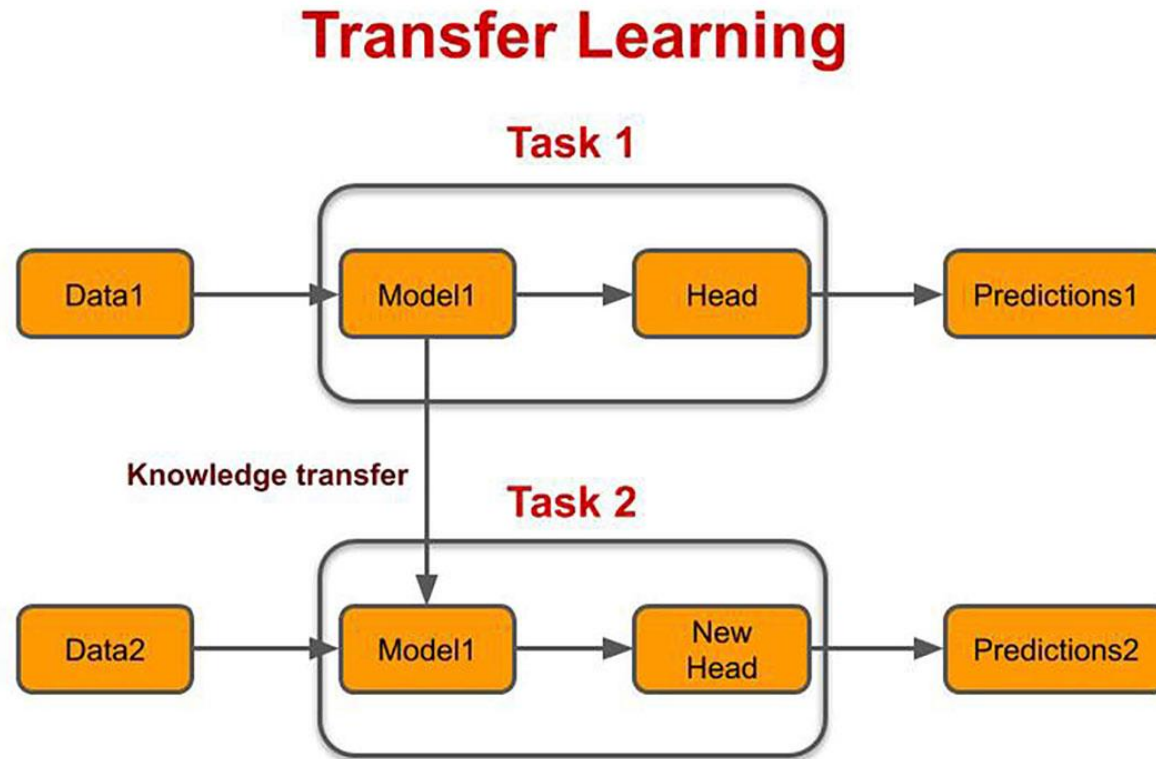
Our Architecture
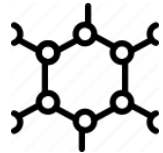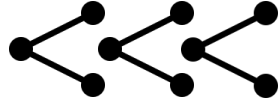
InceptionV3 feature extractor
or
VGG16 feature extractor
or
Resnet50V2 feature extractor

Pretrained Feature extractor

Fully Connected Layer	258 Units	Relu Activation
Fully Connected Layer	258 Units	Relu Activation
Fully Connected Layer	22 Units	Softmax Activation

Classifier

Good Eleven

Transfer Learning training routine

1. Freeze all the layers of the pretrained network

2. Train only the top layers of the network

3. Unfreeze al the layers

4. Fine tuning: Retrain the whole network with a very low learning rate
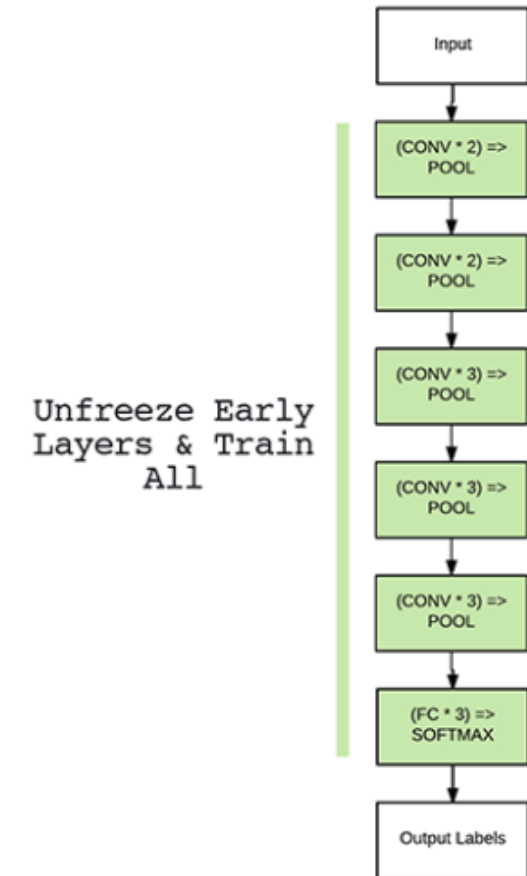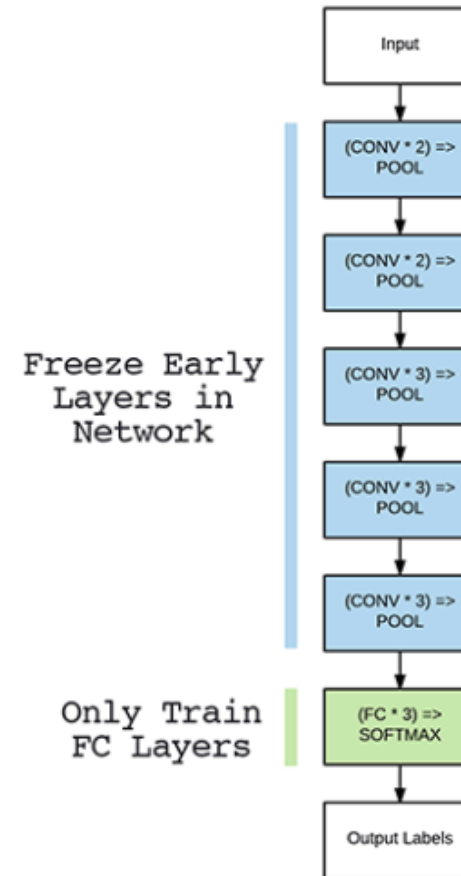
Good Eleven

Performance Improvements

	Custom Training	Transfer Learning
InceptionV3	0.8307	⬆ 0.9113
Resnet50V2	0.7905	⬆ 0.9056
VGG16	0.8245	⬆ 0.8695

Good Eleven

Application: LogosFinder

Good Eleven

Modes of use

Detector only

Full process

Good Eleven

Applications

Logos Finder: Image Blurring

Upload image

Applications

Logos Finder: Image Blurring

Upload image

Detection and
recognition of logos

Good Eleven

Applications

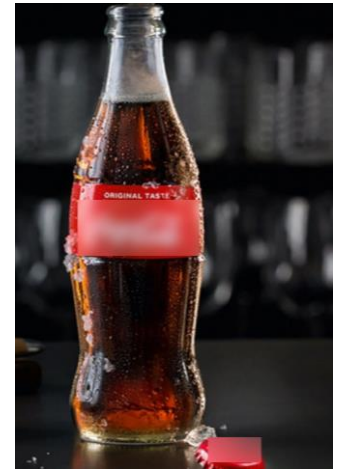## Logos Finder: Image Blurring

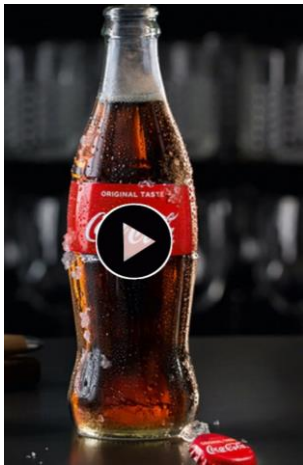Upload image

Detection and recognition of logos

Blur logos in blacklist

Good Eleven

Applications

LogosFinder: Image Blurring
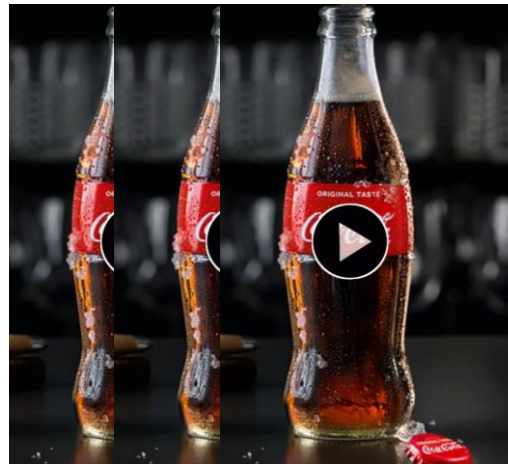
LogosFinder: Offline Video Blurring

Upload video

Good Eleven

Applications

LogosFinder: Offline Video Blurring

Upload video

Split video by frames

Good Eleven

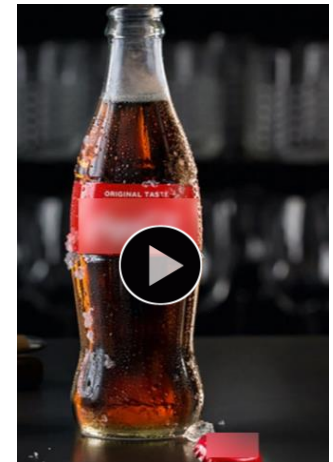# Applications

LogosFinder: Offline Video Blurring

Upload video

Split video by frames

Watch the video with the blurred logos

Good Eleven

Applications

LogosFinder: Live Video Blurring

Good Eleven

Applications

LogosFinder: Image Blurring

LogosFinder: Offline Video Blurring

LogosFinder: Live Video Blurring

Good Eleven

Demonstration of use of the application

Good Eleven