Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Bottle Cap Detection and Classification Network(BCDC-Net)

Computer Vision Project

February 2021

Jaswanth Bandlamudi

Introduction

- Recognizing and Locating objects from images, a widely-used computer vision task
- Divided based on feature-extraction methods
  - Traditional
  - Deep learning based

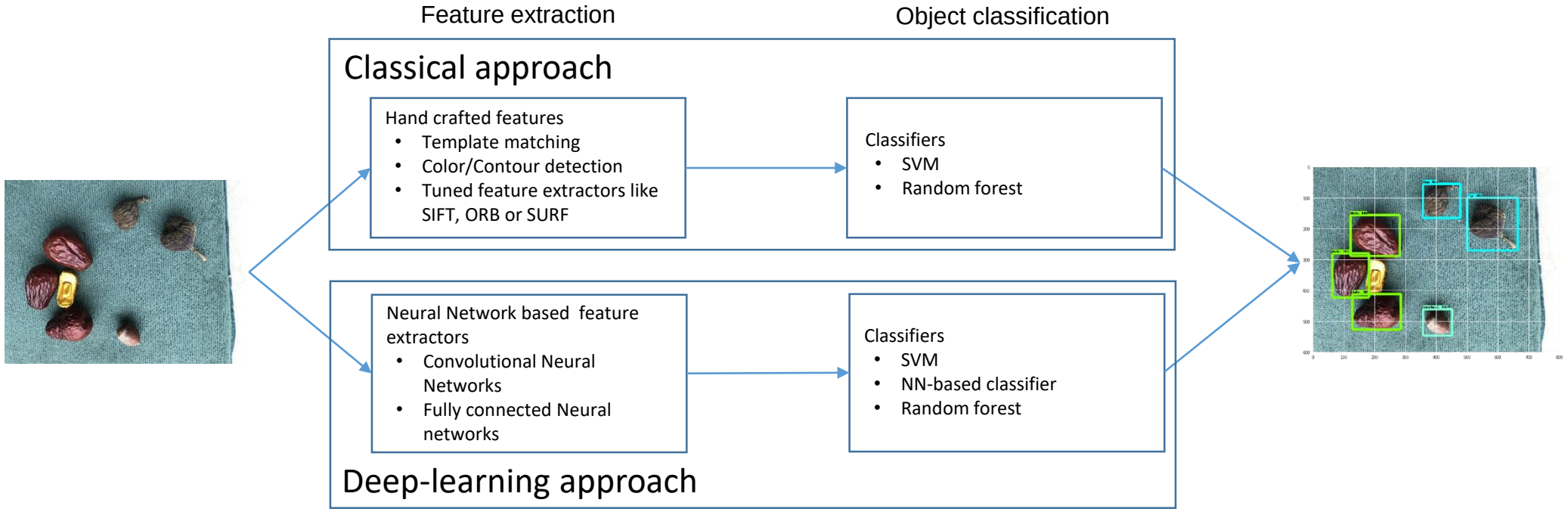


Figure 1: General information flow and processing in object detection [2, p.3]

Problem-Statement

- **Statement:** Localize crown caps in given stream of images and classify them into three different classes (Cap Face-up, Cap Face-down, Cap Deformed)
- Dataset: The dataset consists of video streams (.mp4) and the labels (.json)
- Object of interest: Cap face up, Cap face down, Cap deformed
- Distractors: Edible-Nuts, Coins, Paper-balls, Cubes , etc.

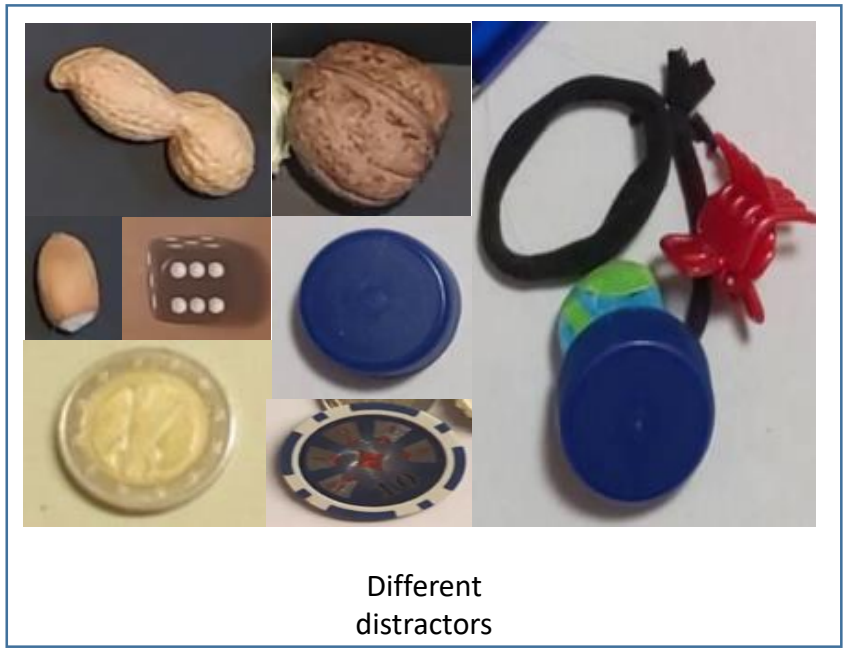


Figure 2: objects of interest and distractors provided in the dataset

Solution

- Steps:
  - Extract stable frame from video stream
  - Use a Single Shot Detector(SSD) [4] to localize and classify the caps
  - Post processing
    - Limit detections to inside of Tray
    - Compile detections into a *.csv* file

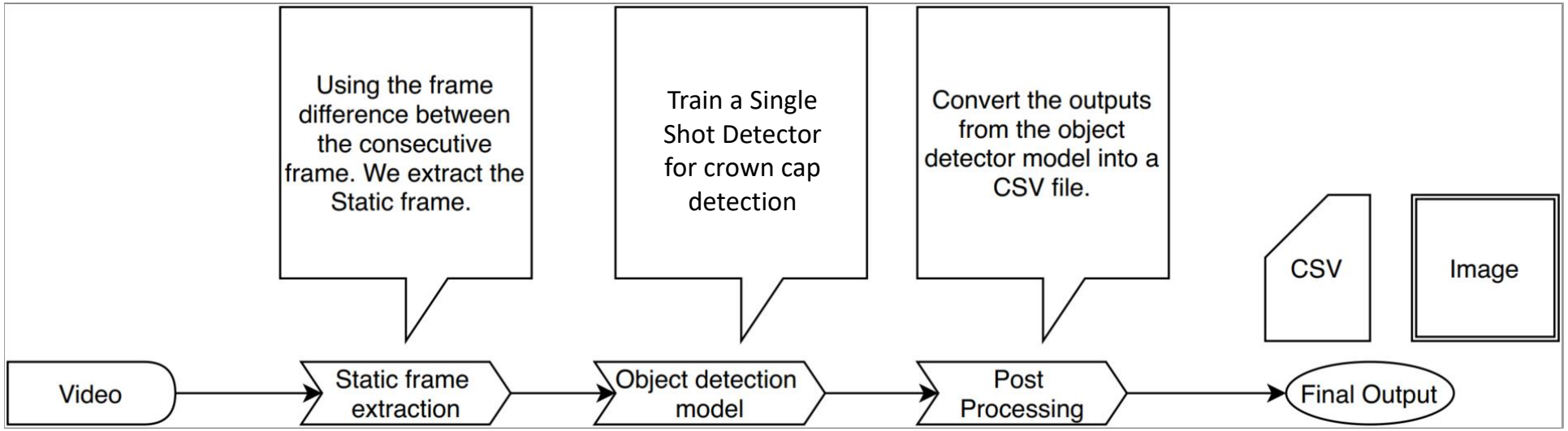


Figure 3: Pipeline for detecting the crown caps from a video stream.

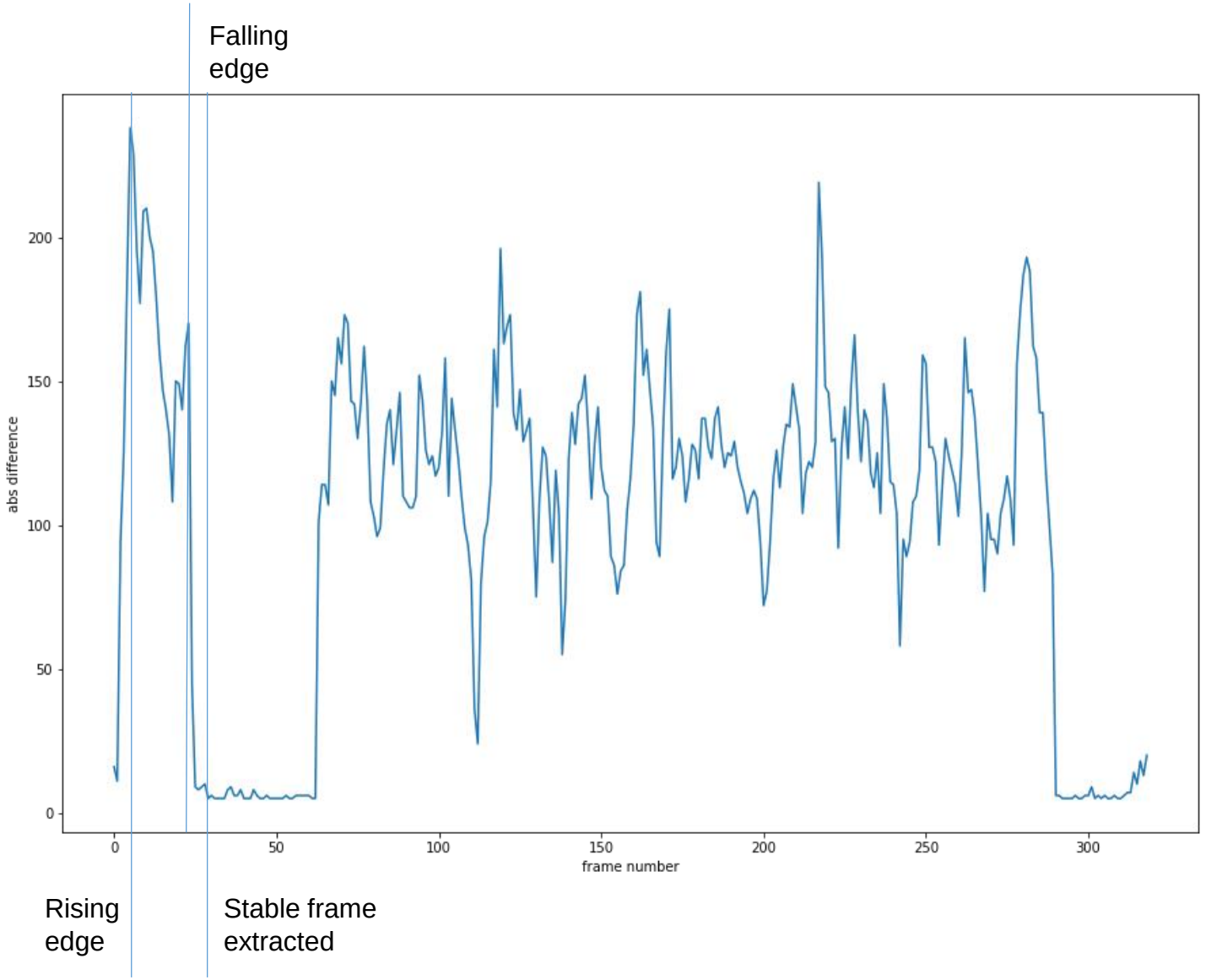
Implementation - 1

- Programming language: Python 3
- Tools used: OpenCV [3], Tensorflow Object Detection API [5]
- Stable Frame extraction
  - absolute difference between consecutive frames
  - Observe for raising and falling edges
  - Fallback: middle frame of video
- SSD object detector
  - Backbone: Inception V2 [1]
  - Classes: Tray, Cap facing-up, Cap facing-down, Cap deformed
  - Paradigm: Transfer learning
  - Augmentation: Random Horizontal Flip, Random Adjust Brightness, Random Adjust Contrast, Random Jitter Boxes, Random Black Patches, Random Pad Image

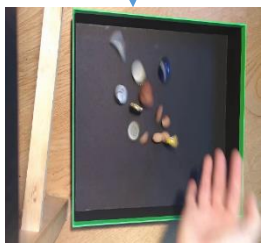
Implementation - 2

- Post-processing
  - Use Tray detection to filter out of Tray cap
    - If a Tray like object is not detected, find a largest rectangle in the image and use it as a Tray
  - Remove caps detected with less than 50% accuracy
  - Compile results in ***csv*** file

Results - Stable-frame extraction

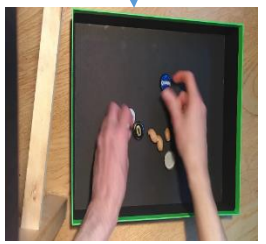


Frame before caps are thrown



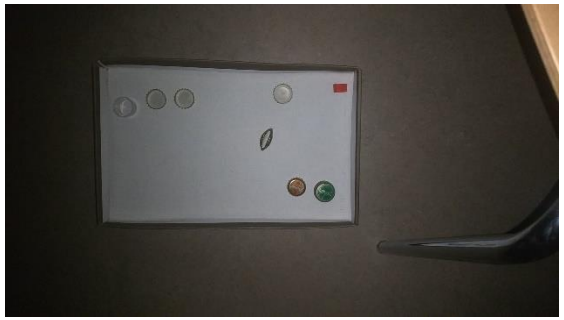
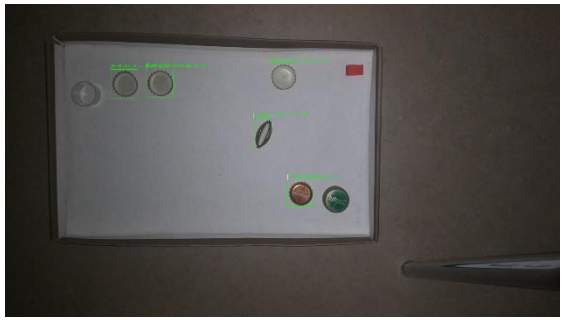
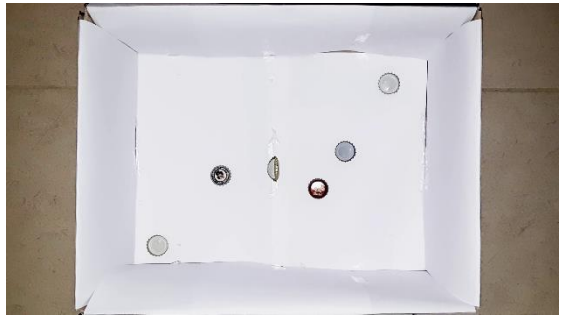
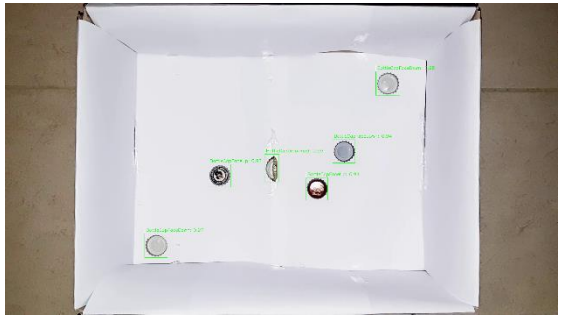
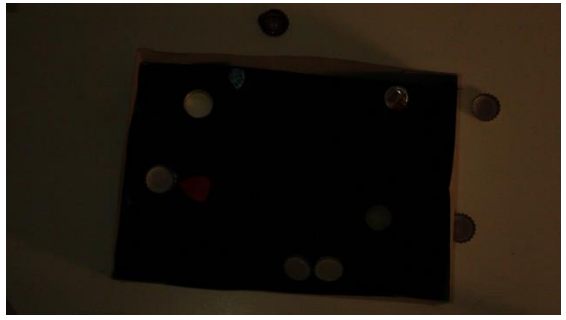
During the task of throwing coins into tray

Stable-frame extraction

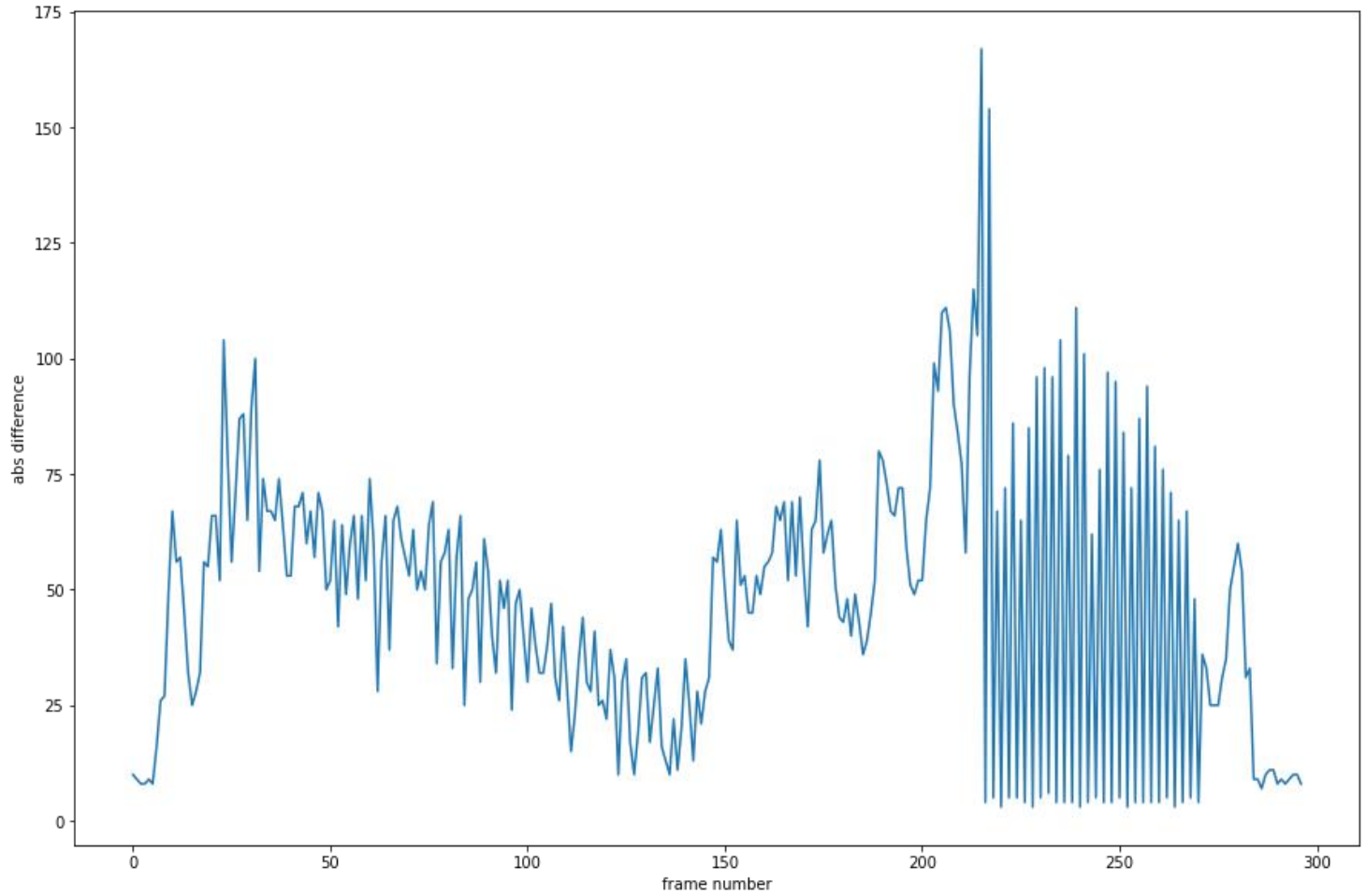


Retrieving from the Tray

Results - Cap Detection

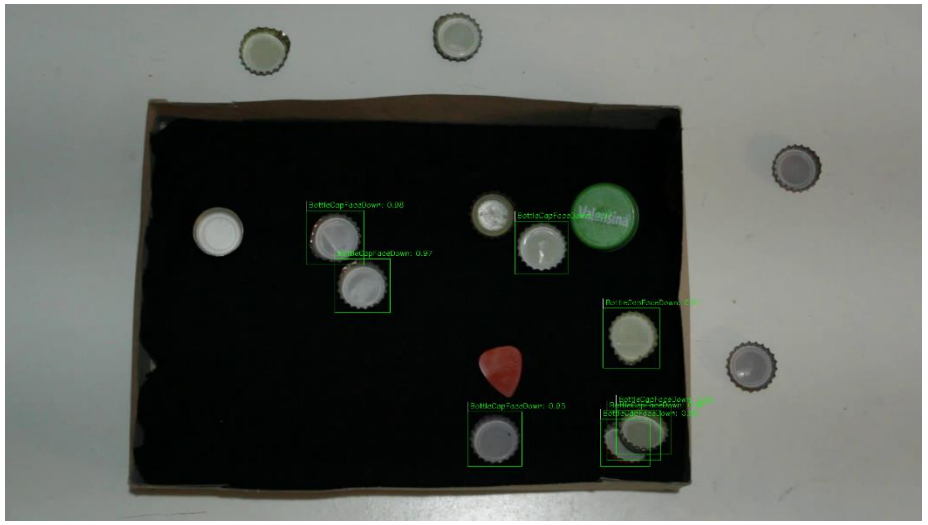
	Input	Output
Detection in a general scene with a tray and proper lighting		
Detection when using a Tray like object instead of a Tray		
Detection in poor lighting conditions		

Results – Stable-frame extraction Fails



Incorrect stable frame extracted because of the improper thresholds applied

Results - Cap Detection Fails



- Improper detection made in presence of an occlusion
- It resulted in detection of two caps and a detection for the caps combined

Solution:

- Non-Maximum Suppression
- Adversarial example training

References

[1] S. R. Bose and V. S. Kumar. Efficient inception v2 based deep convolutional neural network for real-time hand action recognition. *IET Image Processing*, 14(4):688–696, 2020.

[2] Dario Cazzato, Claudio Cimarelli, Jose Luis Sanchez-Lopez, Holger Voos, and Marco Leo. A survey of computer vision methods for 2d object detection from unmanned aerial vehicles. *Journal of Imaging*, 6(8), 2020.

[3] OpenCV, https://opencv.org.

[4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing.

[5] Tran D (2017) How to train your own Object Detector with TensorFlow's Object Detector API. URI: https://medium.com/towards-data-science/how-to-train-yourown-object-detector-with-TensorFlows-object-detector-api-bec72ecfe1d9. Cited January 25, 2021.

[6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015