EDIBLE NUT DETECTION

COMPUTER VISION PROJECT

FINAL PRESENTATION

Mihir Mulye

Table of Contents

Problem Statement

- Detect objects (edible nuts) in a video depicting a roll of objects
- Categories of nuts- Peanuts, Walnuts and Hazelnuts
- Requirements-
  - Desired category of nuts should be detected in a stationary frame (that frame where the configuration first attains a stationary state)
- Observed variations in collected data-
  - Partial occlusions
  - Variable lighting conditions
  - Distractors (objects not of interest) outnumbering the desired number of objects
  - Variable camera configurations resulting in different perspectives

Assumptions

- Distractors do not outnumber the nuts
- Complete occlusion is not dealt with

Required materials

- TensorFlow 1.14
- OpenCV 3.4
- Lab recordings
- LabelMe (annotation tool)

Possible approaches (1 of 2)

- Object detection can be dealt with using either classical vision techniques or by deep learning methods
- Classical methods:
  - Template matching [3]
  - SIFT [4]
- Deep learning methods:
  - SSD [1]
    - Lightweight network able detect objects without much processing requirement

Possible approaches (2 of 2)

- Classical methods pros & cons:
  - Fast to implement as no training needed
  - Classical methods need hand crafting of features
  - Fail to generalize and have difficulties handling variation in data
- Deep learning pros & cons:
  - Able to generalize relatively better than classical approaches
  - Need annotated data
  - Might overlearn if improper data is used for training
- Based on the pros and cons, best of both approaches can be combined to get good results

Proposed Pipeline

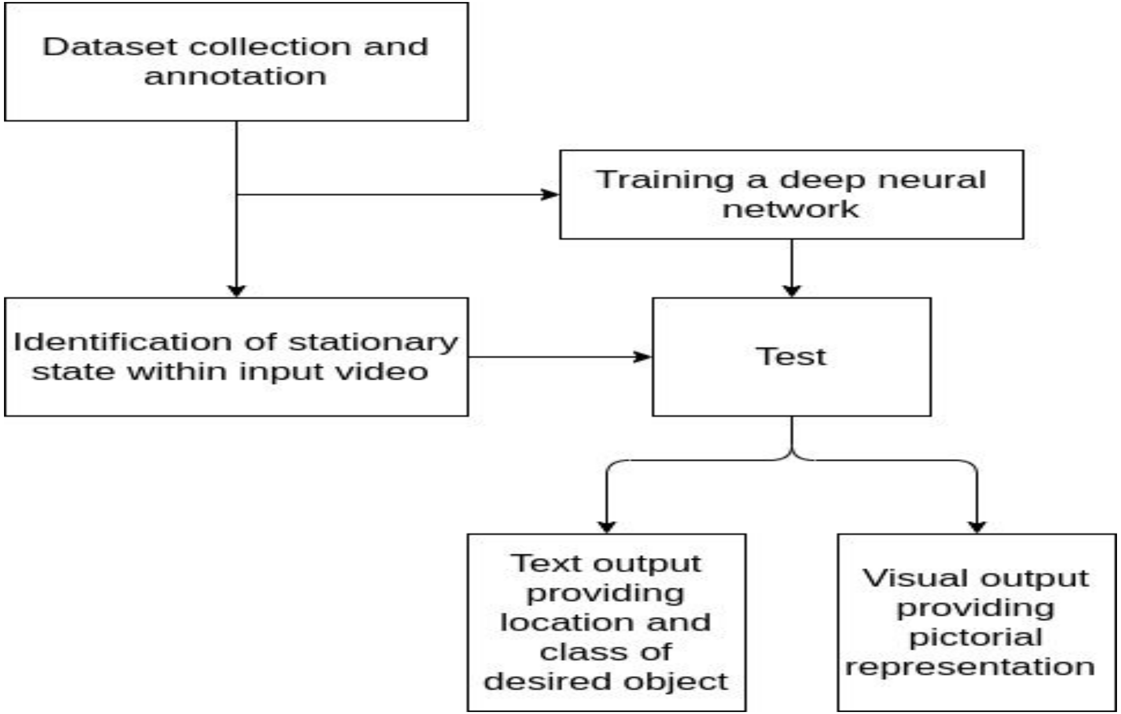

Figure 1 : Overview of proposed pipeline

Implementation (1 of 3)

- Combination of classical Computer Vision and Deep Learning has been employed to accomplish the tasks
- Identification of stationary state within input video:

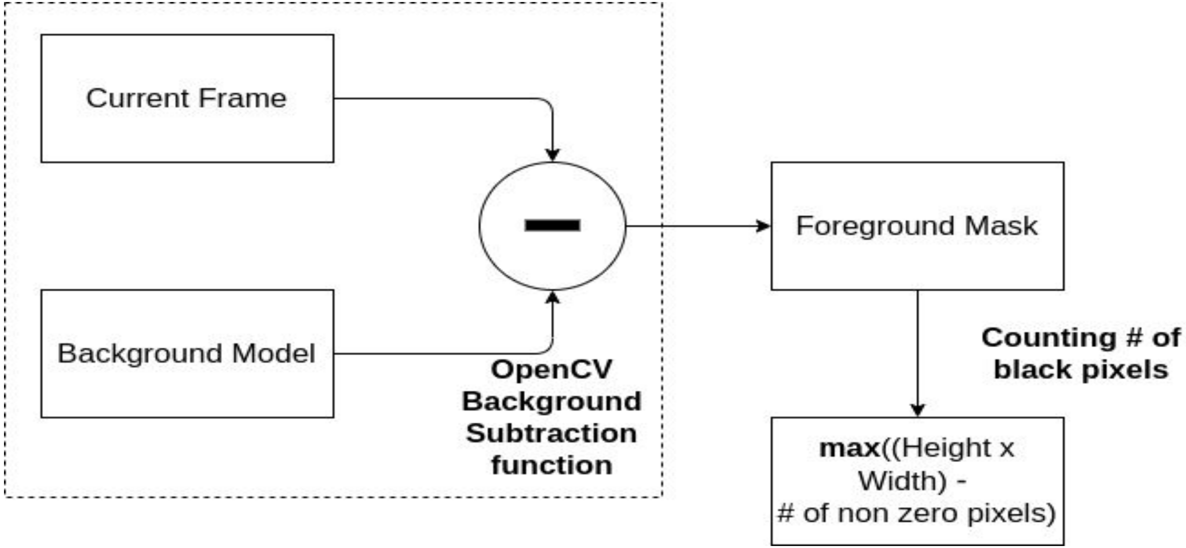

Figure 2 : Index of stationary frame is extracted at the end of this process [2]

Implementation (2 of 3)

- SSD Mobilenet [1] selected for the object detection

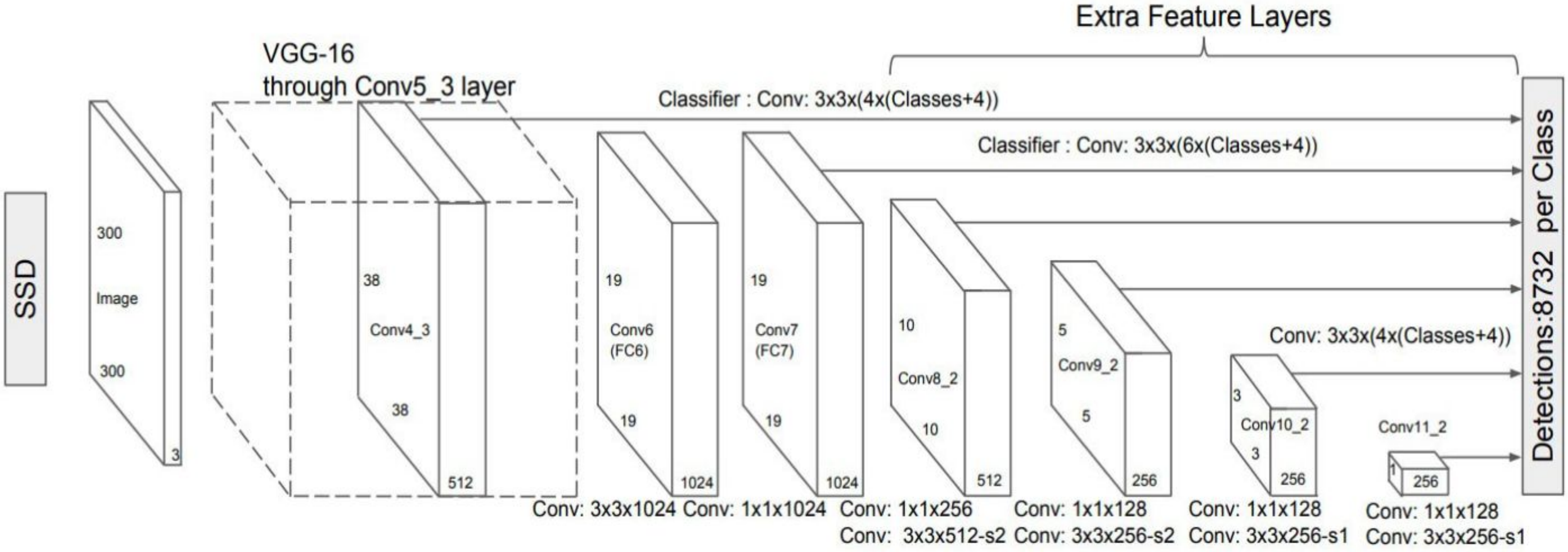

Figure 3 : Single Shot Multibox Detector overview [1]

Implementation (3 of 3)

- Overview of SSD working
  - Initial phase extracts feature from input image (edges, colors)
  - Final phase classifies object based on extracted features
- Output from the ‘stationary state identification’ fed as input to SSD
- SSD framework yields class and the bounding box coordinates of the edible nuts
- Centroid of the bounding box coordinates calculated and output written to a csv file

Results 1 of 6:

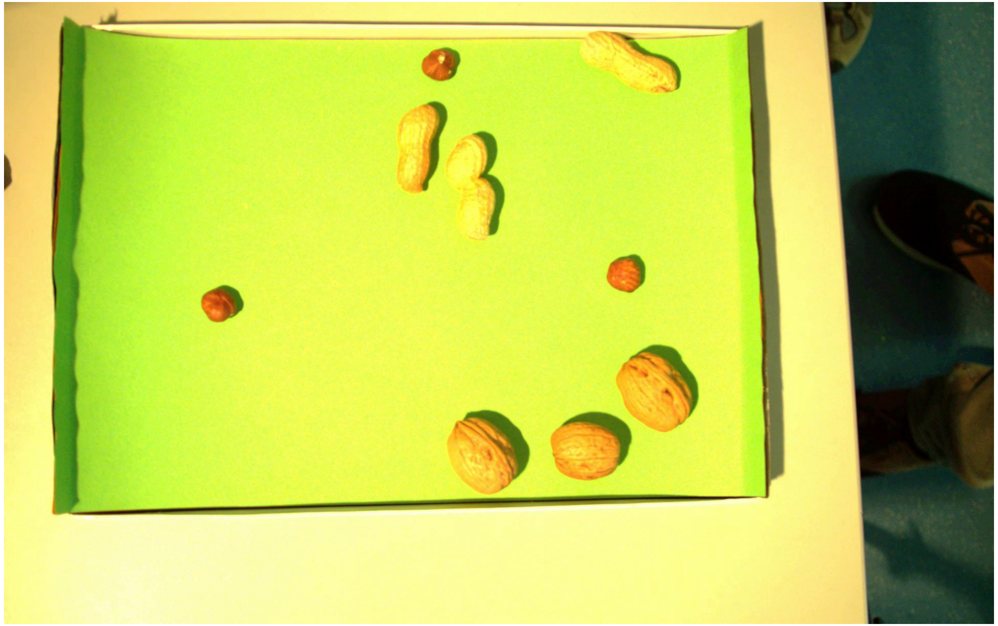

Figure 4 : Stationary frame
(Proper Illumination condition)

Results 2 of 6:

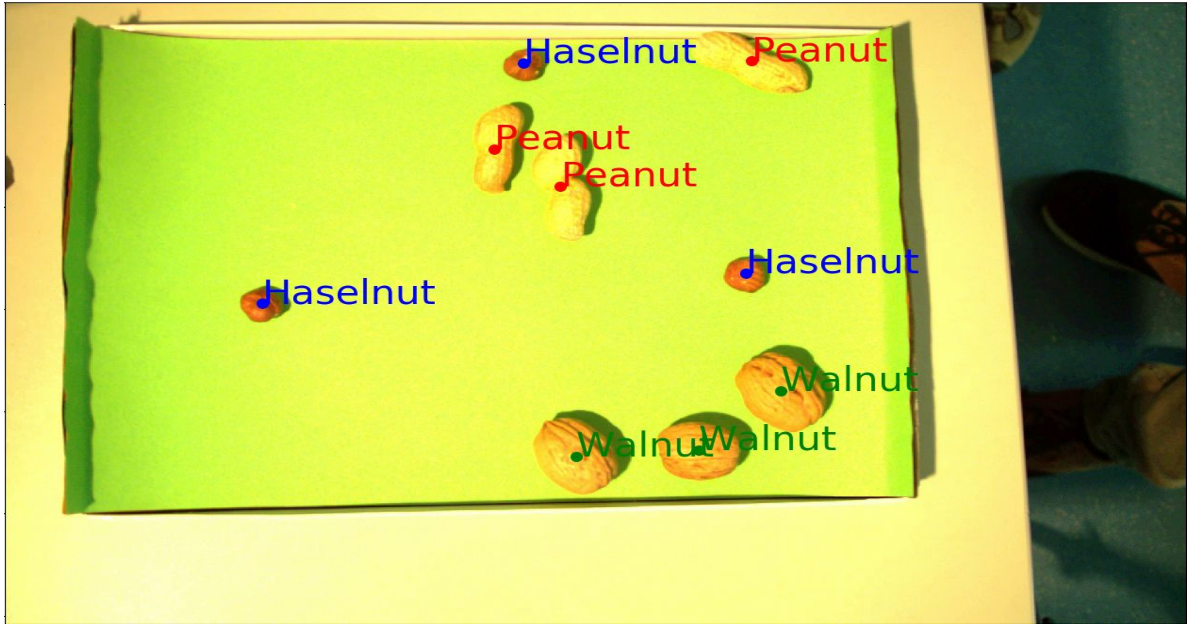

Figure 5 : Output frame
(Proper Illumination condition)

- Red : Peanut
  - Actual : **3**
  - Detected : **3**
- Green : Walnut
  - Actual : **3**
  - Detected : **3**
- Blue : Haselnut
  - Actual : **3**
  - Detected : **3**

Results 3 of 6:

```
34 [849.8522658348083, 119.4853925704956] Haselnut
34 [1271.525165578613, 760.4579963684082] Walnut
34 [935.1773064136505, 888.0472030639648] Walnut
34 [1136.6842765808105, 876.1532135009766] Walnut
34 [1214.3751859664917, 530.9792022705078] Haselnut
34 [421.6336944103241, 589.111346244812] Haselnut
34 [801.9133710861206, 287.1815164733887] Peanut
34 [1223.410813331604, 115.0017379522336] Peanut
34 [910.9826664924622, 360.8621082305908] Peanut
```

Figure 6 : Stationary frame text output
(Proper Illumination condition)
Origin : Top Left corner

Results 4 of 6:

Figure 7 : Stationary frame
(Improper Illumination condition)

Results 2 of 6:

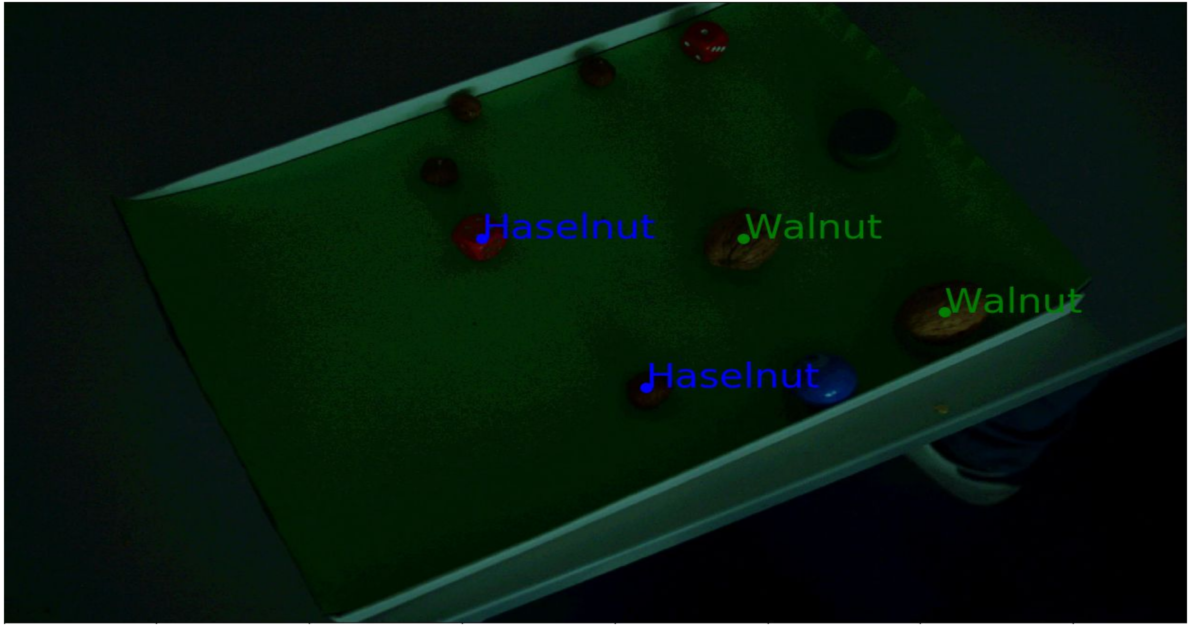

Figure 8 : Output frame (Improper Illumination condition)

- Red : Peanut
  - Actual : **0**
  - Detected : **0**
- Green : Walnut
  - Actual : **2**
  - Detected : **2**
- Blue : Haselnut
  - Actual : **4**
  - Detected : **2** (1 incorrect detection)

- Incorrectly detected red die as Haselnut

Results 3 of 6:

```
54 [782.7833132743835, 463.7918815612793] Haselnut
54 [1211.8257732391357, 463.4927234649658] Walnut
54 [1539.0790672302246, 608.1815423965454] Walnut
54 [1051.7555875778198, 755.7490196228027] Haselnut
```

Figure 9 : Stationary frame text output
(Improper Illumination condition)
Origin : Top Left corner

Failures, Learnings and Scope-

- Approaches such as-
    - Shape detection
    - Template matching
    - Color identification

  were explored before switching to deep learning

- Owing to drastic variations in data, the methods were failing to generalize
- Static frame not always correctly isolated
- Spurious detections need to be addressed
- Bigger network can be used in place of lighter SSD
- More data can be used for training

References

[1] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In *European conference on computer vision*, pp. 21-37. Springer, Cham, 2016.

[2] https://docs.opencv.org/3.4/d1/dc5/tutorial_background_subtraction.html

[3] https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_template_matching/py_template_matching.html

[4] https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html