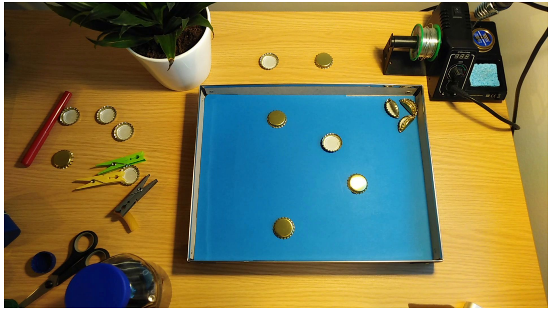
Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bottlecap challenge

Manoj Kolpe Lingappa, 9038585

Introduction

- Bottlecap detection is a task to detect, recognize and localize the bottlecap.
- Computer vision technique to solve the object detection problem.
- Video dataset containing the bottlecaps in one or more frames.
- Objects class in the video:
  - Bottlecap faceup
  - Bottlecap facedown
  - Bottlecap deformed
  - Distractors
    - Coins, nuts, different class of bottlecaps, clips,...etc



Motivation

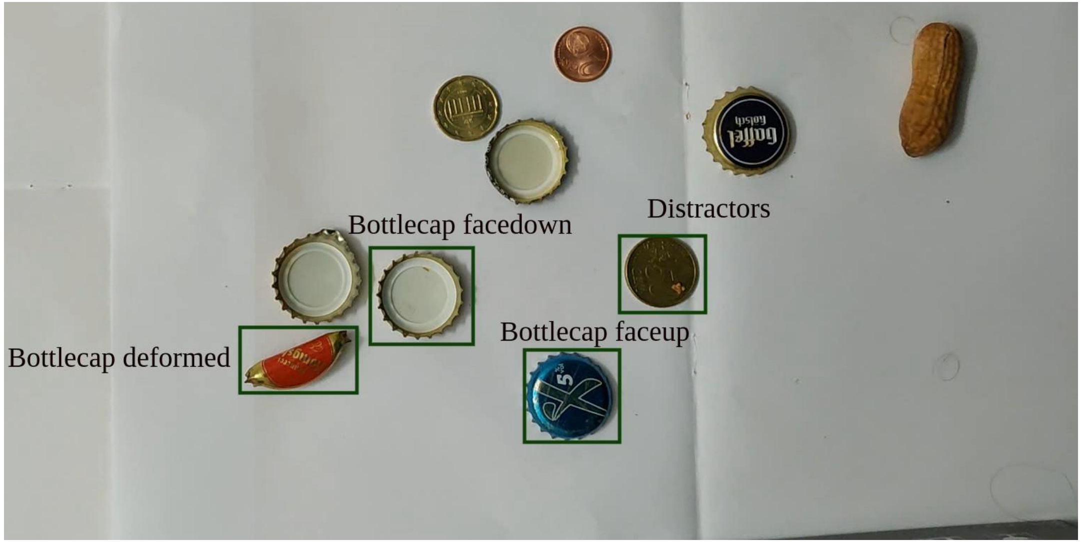
- Want to recognize a known object from unknown viewpoint

Bottlecap faceup

Bottlecap deformed

Bottlecap facedown

Bottlecap distractors



Detection of object class

Literature review

- Traditional computer vision approaches
  - Canny edge detection, Harris corner detector. [1]
  - SIFT, SURF, BRIEF, etc [2]
- Sometimes deep learning approach is overkill for certain task.
- Deep learning based object detection methods:
  - R-CNN [3]
  - Single-shot detector (SSD)[4]
  - YOLO[5]
- Transfer learning using YOLOv3.
- Capable of real-time object detection.
- Hybrid techniques that combine classical and modern approach.

Problem decomposition

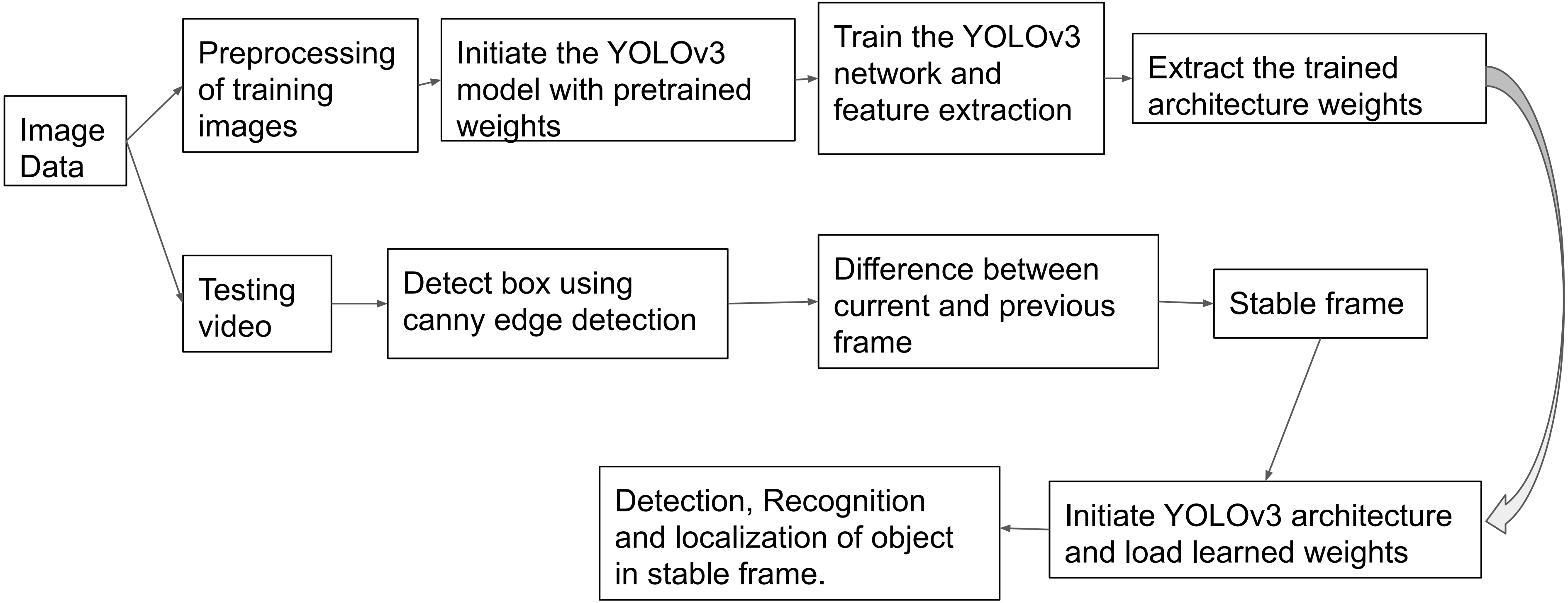
Training phase

Training data source → Preprocessing of data → Define a pretrained model → Train the model → Validate the model → Save the model

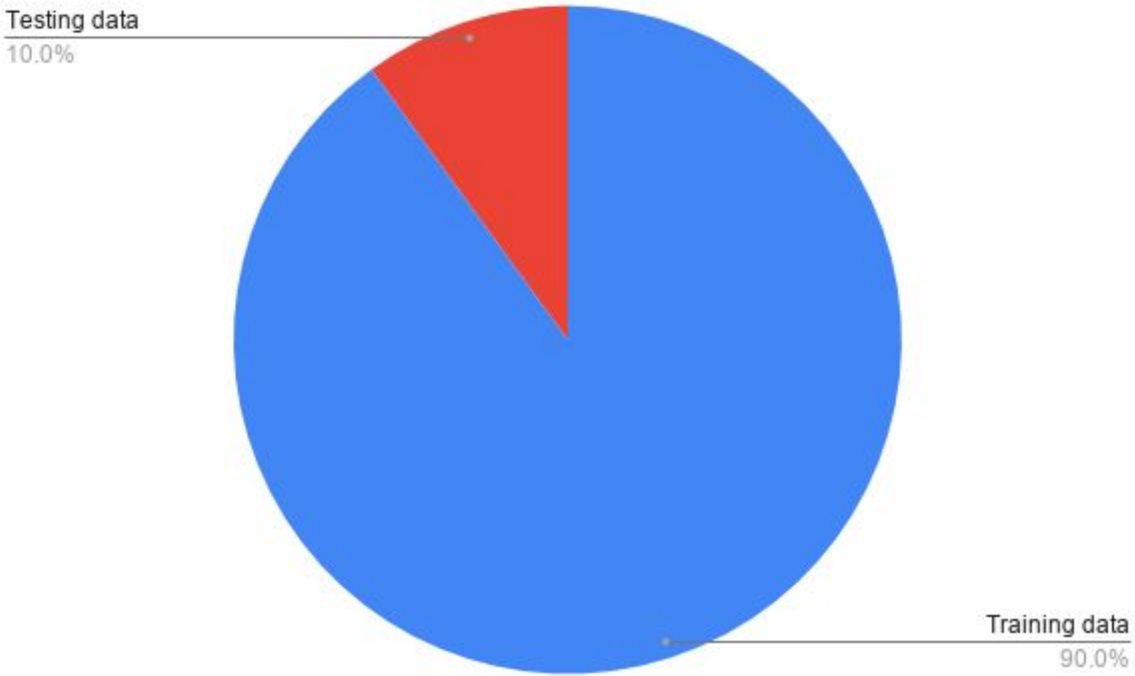
Testing phase

Testing data source → Preprocessing of data → Extraction of stable frame → Load the trained model → Detect, recognize and localize the target object class

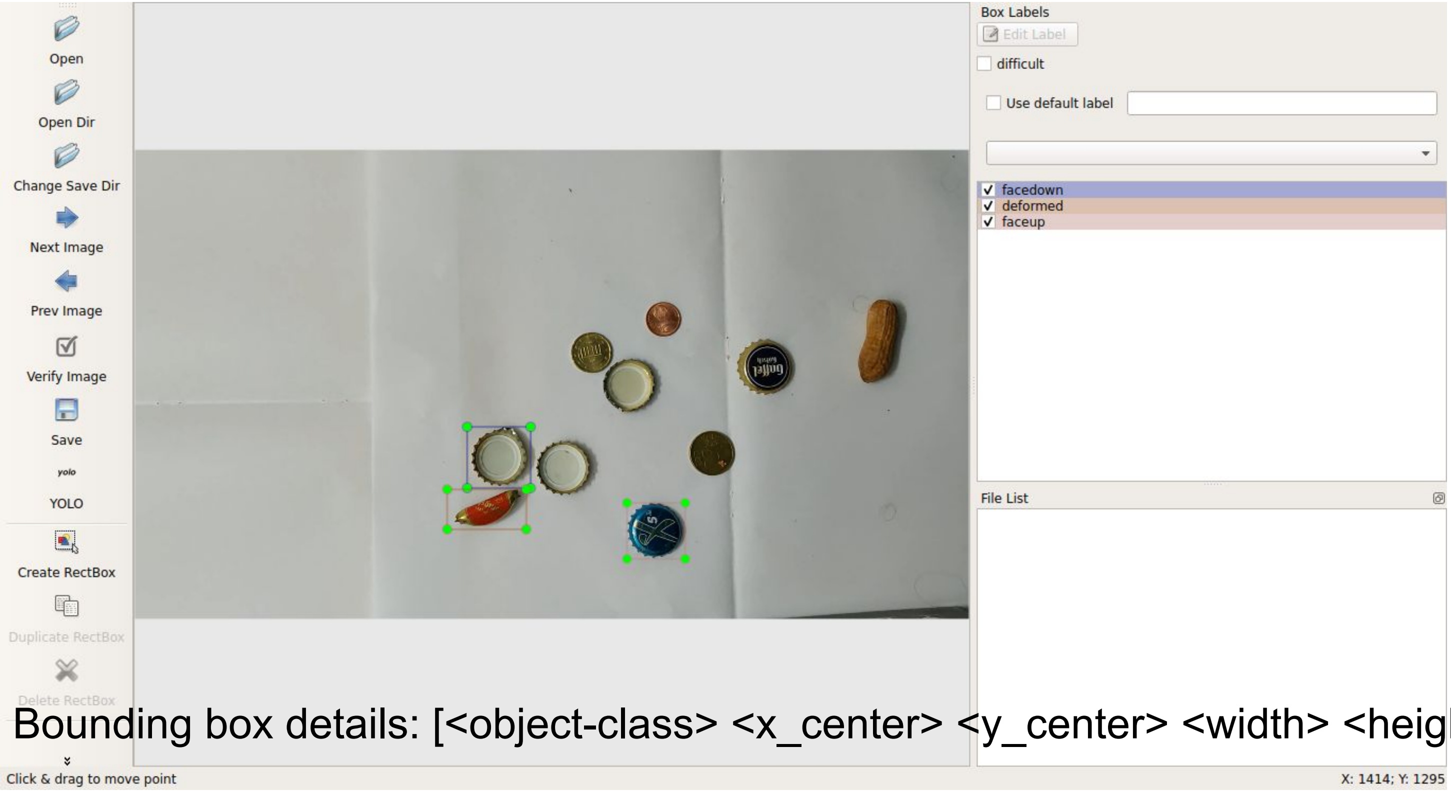
Plan of attack



Data



Pre-Processing of training data



- Bounding box details: [<object-class> <x_center> <y_center> <width> <height>]

YOLOv3

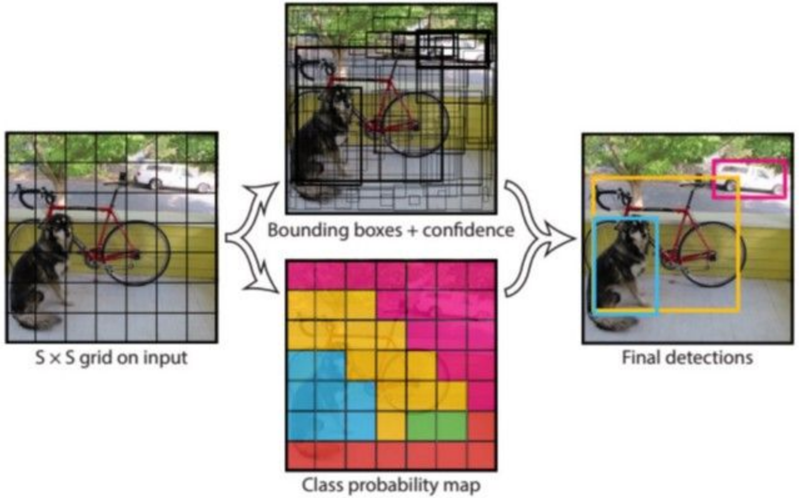
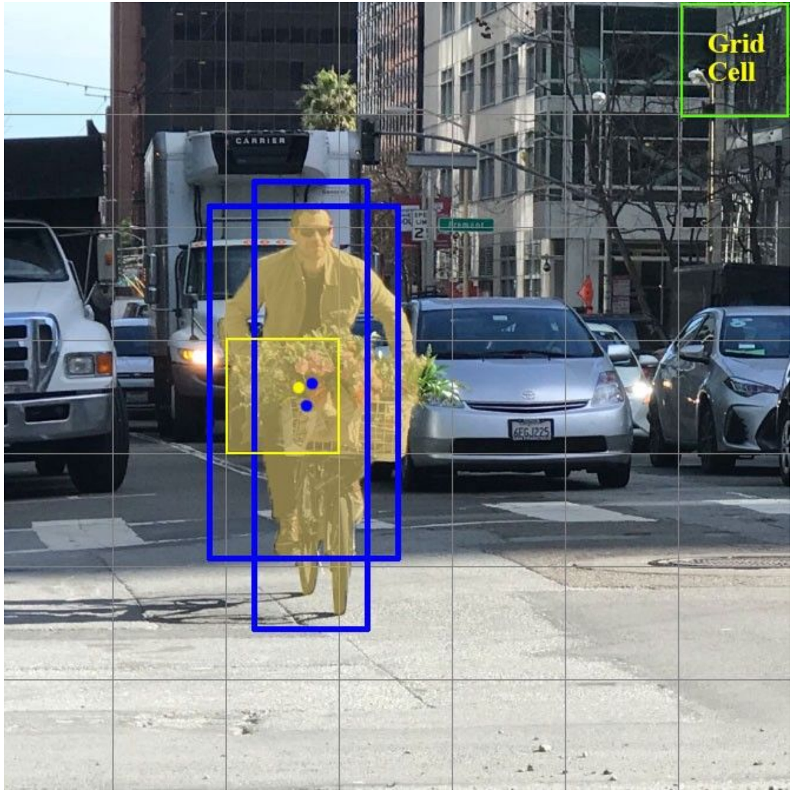


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

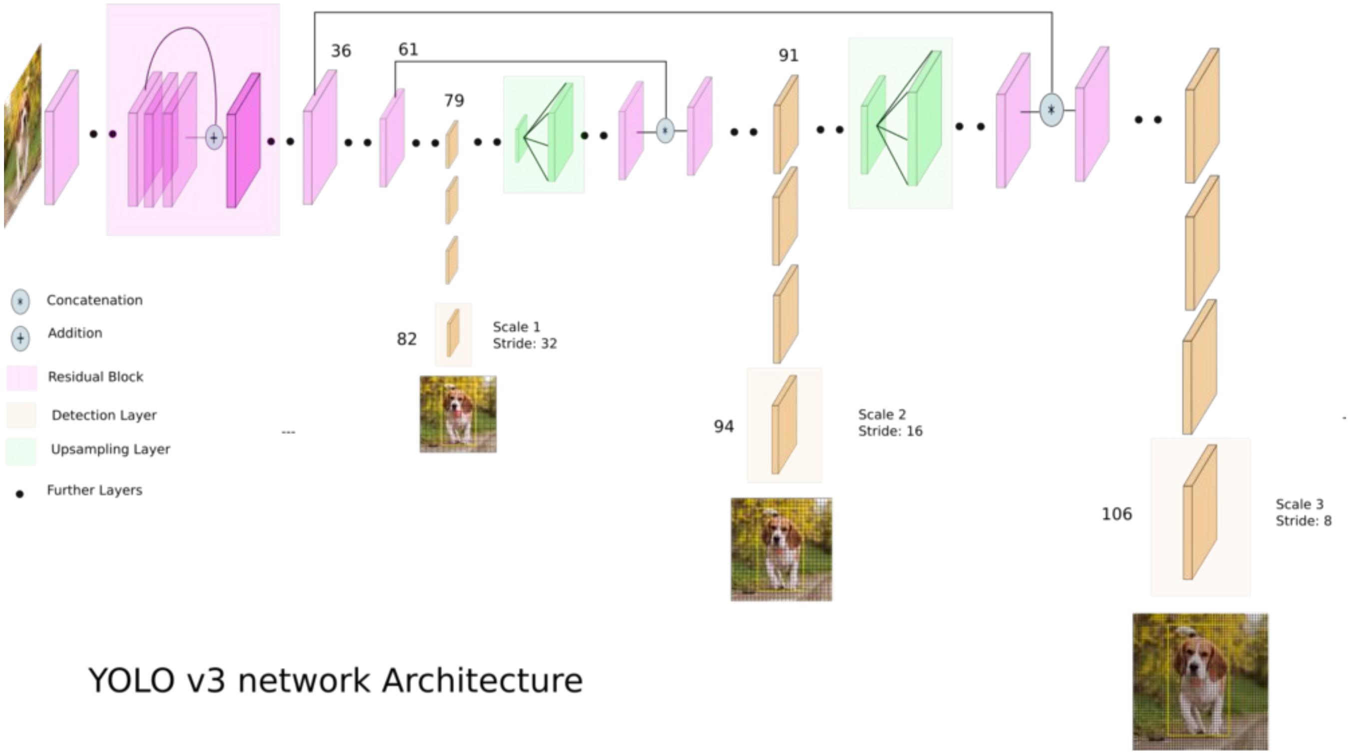
Courtesy of [6]



Courtesy of [7]

Bottlecap challenge - Manoj

YOLOv3

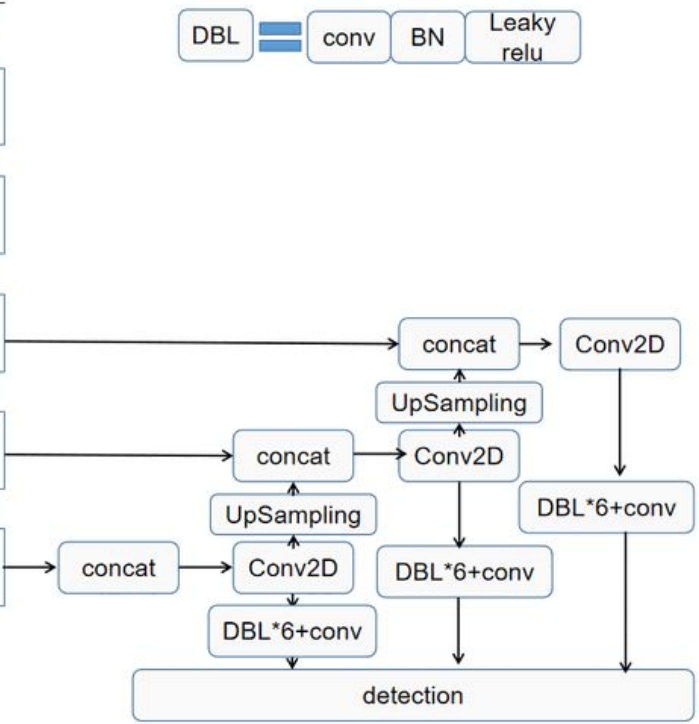


YOLO v3 network Architecture

Courtesy of [8]

YOLOv3

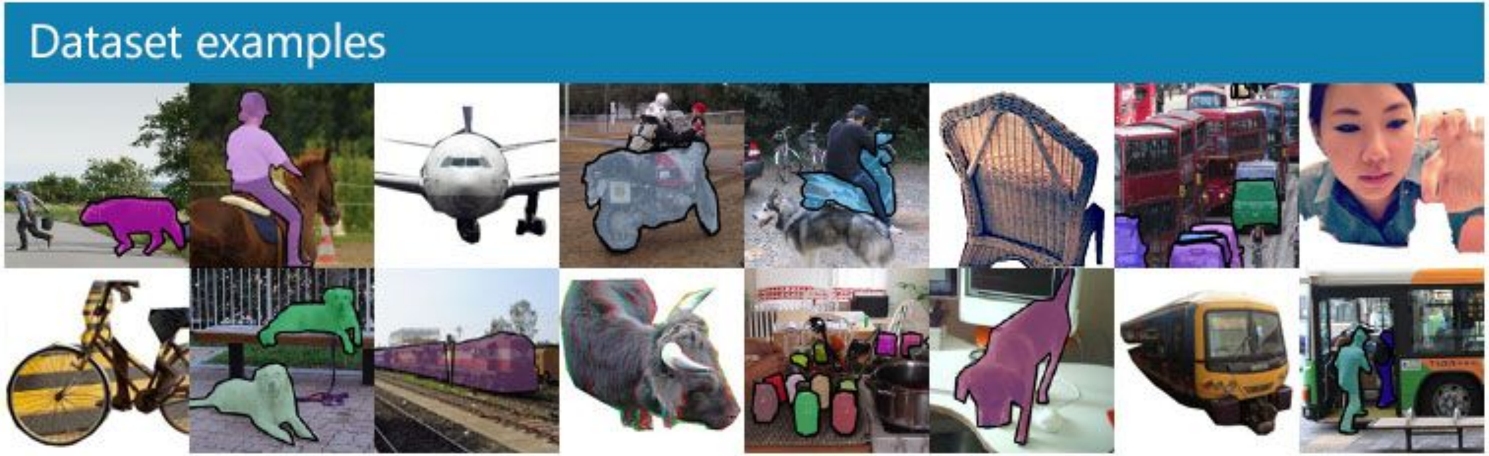
	type	filters	size	output
	convolutional	32	3x3	416x416
	convolutional	64	3x3/2	208x208
1x	convolutional	32	1x1	
	convolutional	64	3x3	
	convolutional			208x208
	convolutional	128	3x3/2	104x104
2x	convolutional	64	1x1	
	convolutional	128	3x3	
	convolutional			104x104
	convolutional	256	3x3/2	52 x 52
8x	convolutional	128	1x1	
	convolutional	256	3x3	
	convolutional			52 x 52
	convolutional	512	3x3/2	26 x 26
8x	convolutional	256	1x1	
	convolutional	512	3x3	
	convolutional			26 x 26
	convolutional	1024	3x3/2	13 x 13
4x	convolutional	512	1x1	
	convolutional	1024	3x3	
	convolutional			13 x 13



Courtesy of [9]

Pre-trained YOLOv3 with COCO dataset

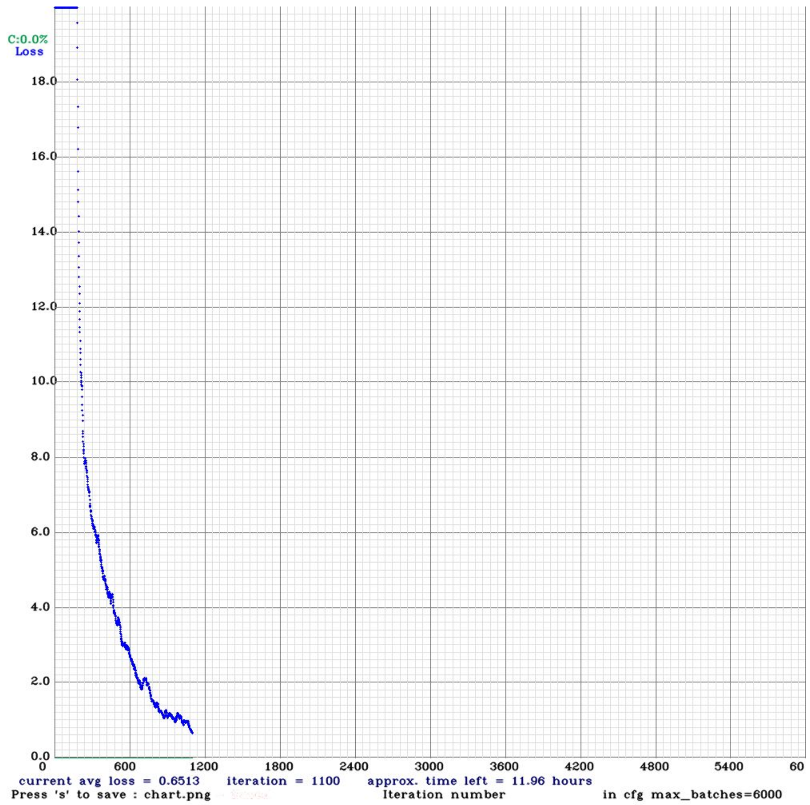
- COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features.



Courtesy of [10]

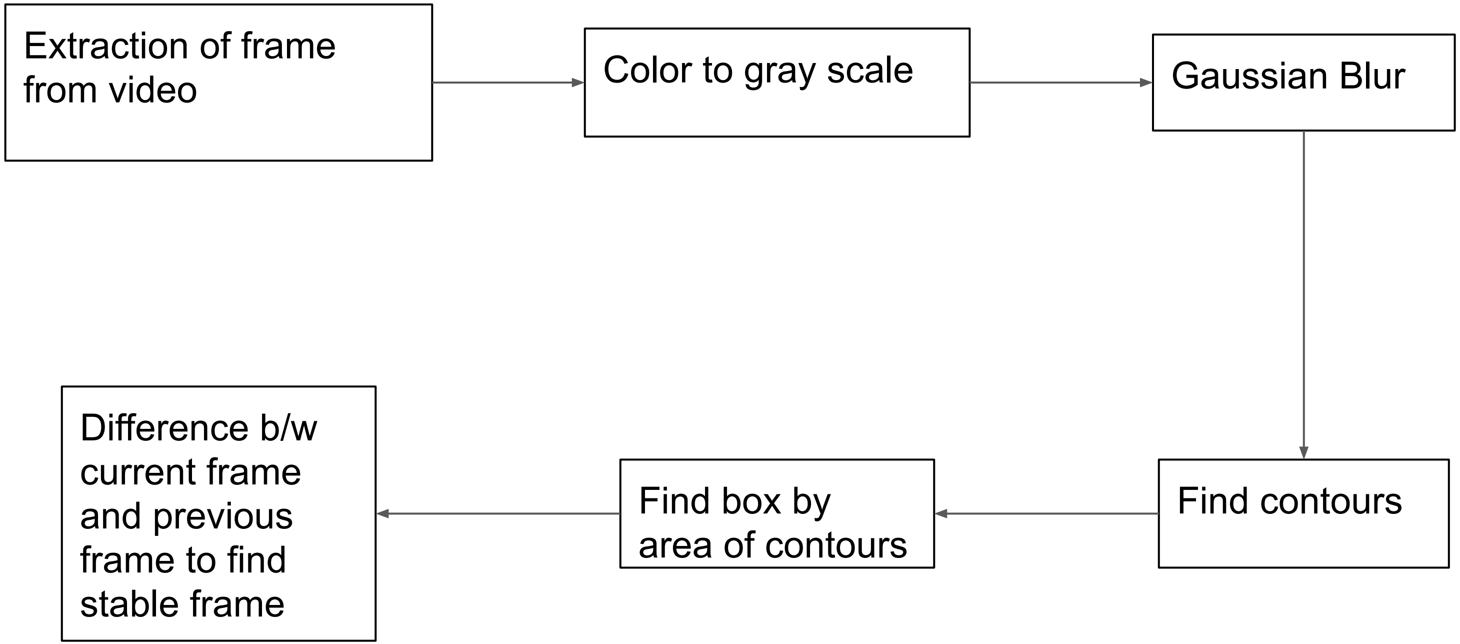
Courtesy of [9]

Training of YOLOv3

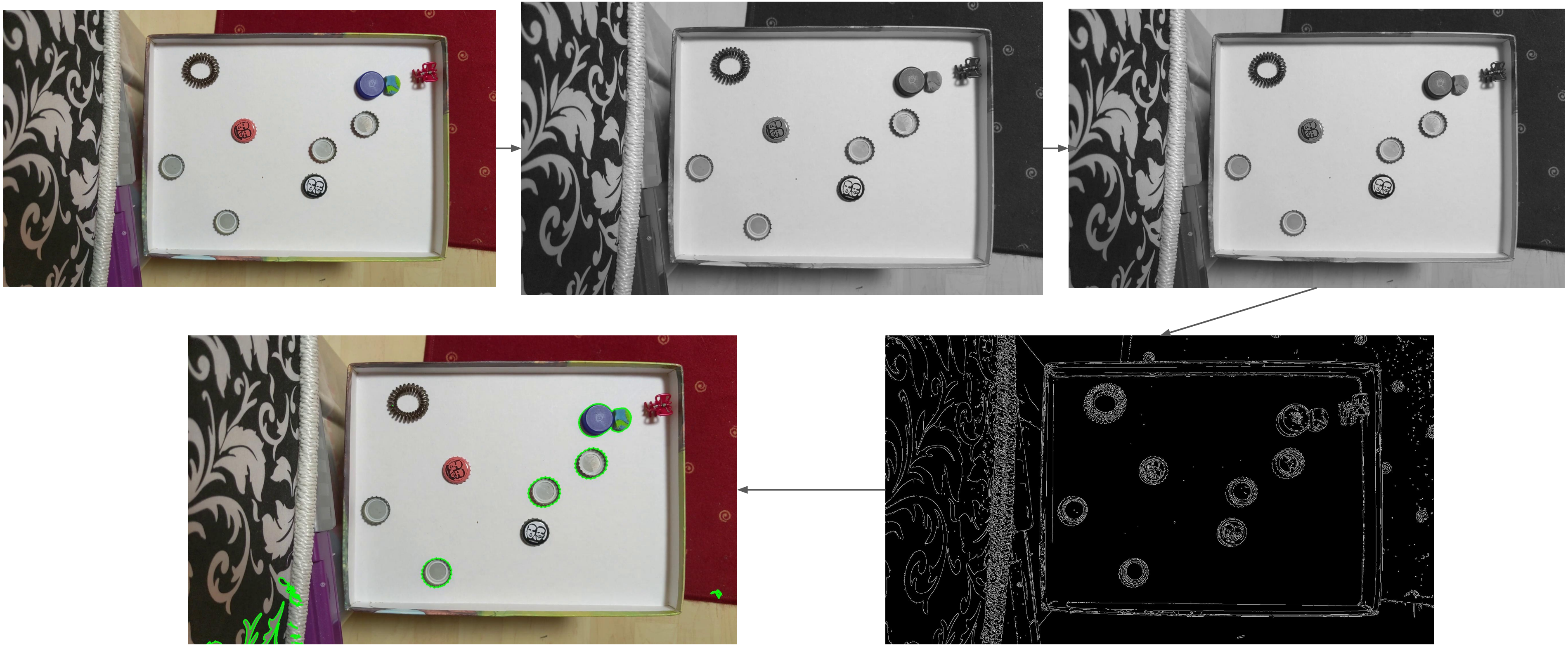


Learning rate	0.001
Batch	64
Optimizer	Adam
Epoch	1000

Extraction of stable frame



Extraction of stable frame



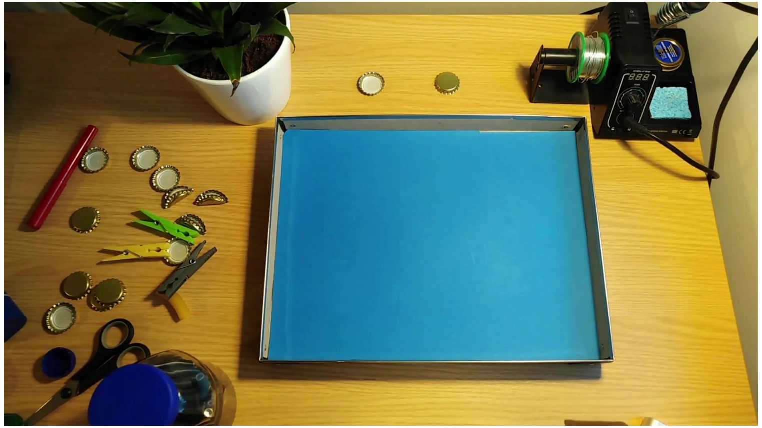
Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

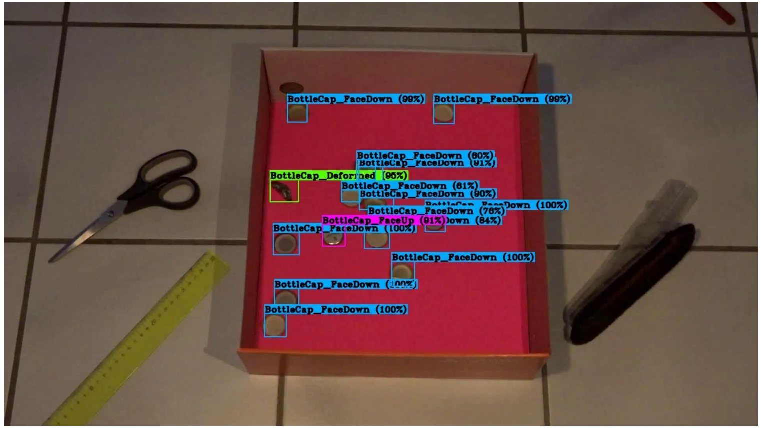
Bottlecap challenge - Manoj

Extraction of stable frame

- Stable frame extraction using canny edge detection is failed due to not proper detection of the contours.
- Trick to find stable frame with assumptions.
- Assumption:
    - YOLOv3 detect object only inside the box
    - YOLOv3 detect object with greater than 90% confidence rate.
- Extract frame every 20th frame of the video.
- Compute the object in the frame.
- Find the frame that have maximum number of detected object.
- Frame with the maximum number of frame is a stable frame.

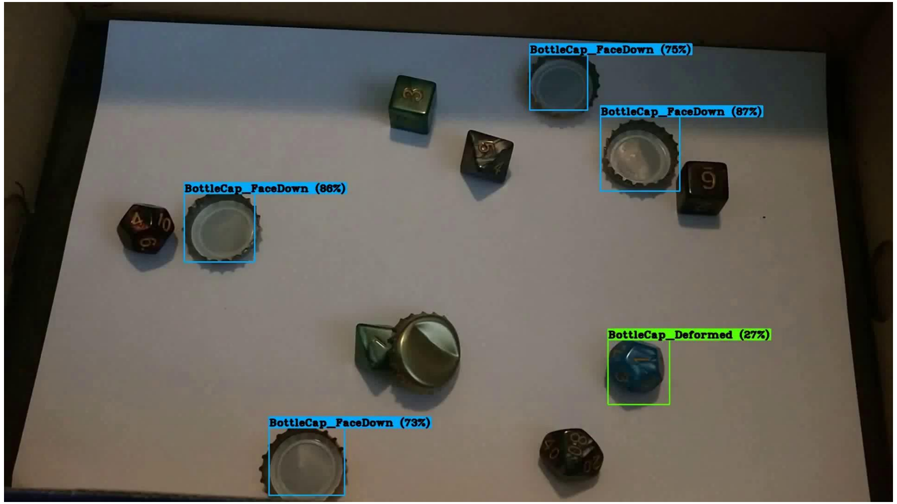
Final results





- Test data are fed into the trained network for evaluation purpose.
- Some distractors are classified as main class.
- Train the model with more data and apply data augmentation on images for better generalizable capability.

Final results



- The YOLOv3 is able to detect, recognize and localize the object class from the testing video.
- The deep neural network architecture is recognizing main object class with greater than 90% confidence score.

Final results

Precision = True positive / (True positive + False positive)

Recall = True positive / (True positive + False negative)

F1-score = 2x(Precision*Recall)/(Precision +Recall)

Final results

Predicted Label					
	Bottlecap_faceup	28	0	1	3
	Bottlecap_facedown	2	36	2	1
	Bottlecap_deformed	1	1	12	3
	Distractors	0	0	0	0
		Bottlecap_faceup	Bottlecap_facedown	Bottlecap_deformed	Distractors
		True Label			

	Precision	Recall	F1-score	Accuracy
Bottlecap_faceup	0.88	0.90	0.89	0.80
Bottlecap_facedown	0.88	0.97	0.92	0.86
Bottlecap_deformed	0.71	0.80	0.75	0.60

References

1. Guiming, S., & Jidong, S. (2018, August). Multi-Scale Harris Corner Detection Algorithm Based on Canny Edge-Detection. In 2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET) (pp. 305-309). IEEE.
2. O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2019, April). Deep learning vs. traditional computer vision. In Science and Information Conference (pp. 128-144). Springer, Cham.
3. Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. International journal of computer vision, 128(2), 261-318.
4. Han, J., Zhang, D., Cheng, G., Liu, N., & Xu, D. (2018). Advanced deep-learning techniques for salient and category-specific object detection: a survey. IEEE Signal Processing Magazine, 35(1), 84-100.
5. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
6. https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006

References

7. https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088
8. https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b
9. Ding, X., & Yang, R. (2019, October). Vehicle and parking space detection based on improved yolo network model. In Journal of Physics: Conference Series (Vol. 1325, No. 1, p. 012084). IOP Publishing.
10. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

Thank You