A COMPARISON OF DEEP LEARNING METHODS FOR OBJECT IDENTIFICATION IN AUTONOMOUS OUTDOOR CLEANING ROBOT

Guided by,
Dr. Sreeni K G
Dept. Of Electronics and Comm.
CET, Trivandrum

Presented by,
Mohammed Raheem P
Robotics and Automation
Reg No: TVE16ECRA15

CONTENTS

- Introduction
- Literature Review
- Objectives
- Proposed Methods
- Experimental Results
- Contribution of Thesis
- Publication based on Thesis
- Problems faced during works
- Conclusions
- References

INTRODUCTION

- Visual system is introduced for proposed cleaning vehicle to optimize the power consumption during cleaning tasks.
- Object identification and classification can be more accurately done with the help of deep learning frameworks.
- Keras-Tensorflow based approach and YOLO mark object detection methods are compared in this work.

LITERATURE REVIEW

Paper	Methodology	Advantages	Disadvantages
"Convolutional neural networks for image classification." (International Conference on Advanced Systems and Electric Technologies (IC_ASET). IEEE Jmour, Nadia, Sehla Zayen, and Afef Abdelkrim. (2018.)	• Fine tuning technique. • Transfer learning. • Imagenet.	• Very effective. • Accurate	• Long time for training. • Need more training images.
"Implementation of deep-learning based image classification on single board computer." (Electronics and Smart Devices (ISESD) International Symposium on. IEEE) Shiddieqy, Hasbi Ash, Farkhad Ihsan Hariadi, and Trio Adiono(2017)	• Used tflearn • 5 hidden layers used.	• Model deploy and running in three platform CPU, GPU and Single Board Computer. • Improve accuracy with increase size of network.	• Not used deep network for real time application using simple computers.

LITERATURE REVIEW (C0NTD...)

Paper	Methodology	Advantages	Disadvantages
"You only look once: Unified, real-time object detection." (Proceedings of the IEEE conference on computer vision and pattern recognition.) 2016.	YOLO marking methods.	• Easy to construct • Fast • 45 FPS	• Struggles with small objects. • Each grid cells only predicts two objects with one class.
"YOLO9000: better, faster, stronger" Redmon, Joseph, and Ali Farhadi. (2017).	YOLO marking methods.	• Faster • Better	• Only for high classification over 9000 classes. Not use for few classes.

OBJECTIVES

- Classification of waste materials and obstacles from real time video.
- Marking and locating the objects.
- Comparative study between Keras-Tensorflow method and Yolo mark object detection method.

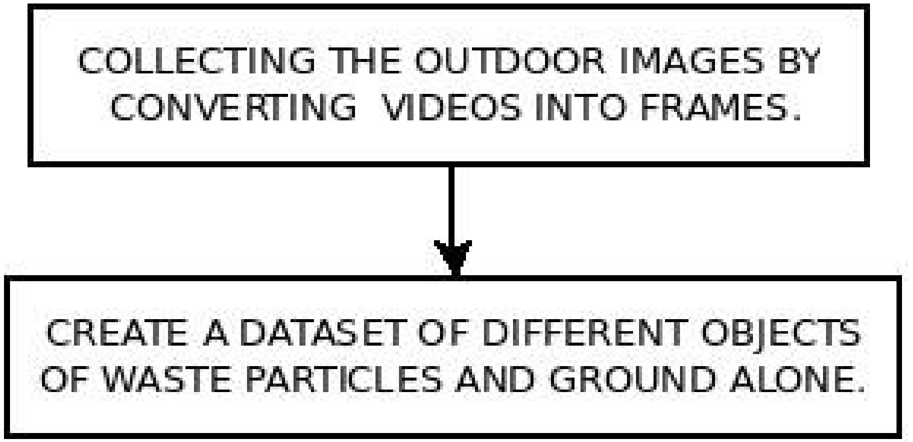
PROPOSED METHODS

(1) **Using the Keras-Tensorflow based classification**

- Keras is used to model the architecture. This can be model in the top of Tensorflow backend.
- Convolutional neural network is used in this work.
- Image dataset of waste particles and ground alone are created for the training process.

PROPOSED METHODS (CONTD...)

- Input images are collected from different environments such as college and hostel campuses, roads etc.
- Plastic covers, leafs, papers etc are taught as waste particles using the algorithm.



PROPOSED METHODS (CONTD...)

- 5 hidden layers are used in the proposed method to classify objects and grounds.
- Format of layers:

CONV – POOL – CONV – POOL – CONV – POOL – CONV – POOL – CONV – POOL – FC -FC

CONV = Convolution + ReLu

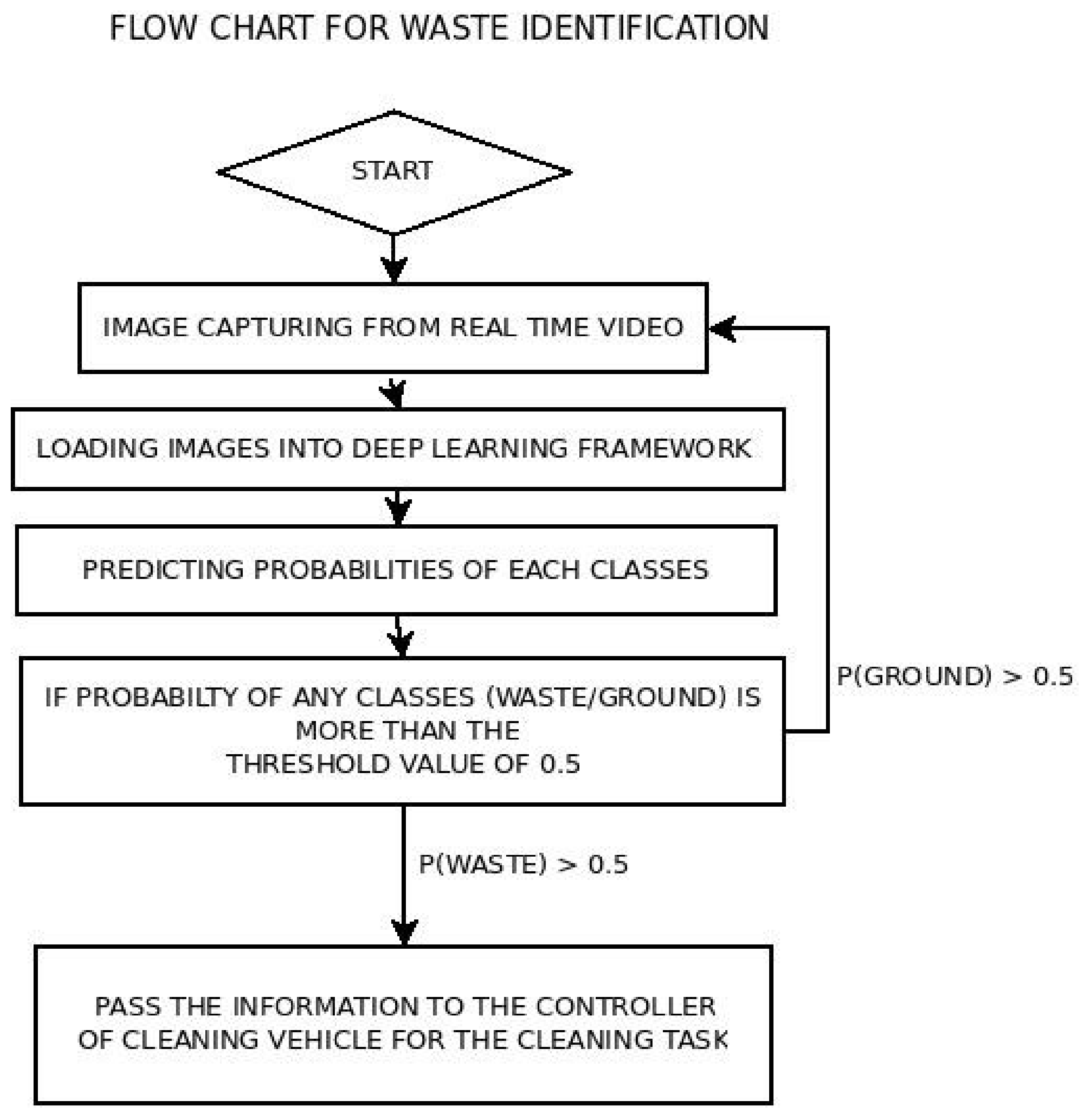
CONV + POOL = HIDDEN LAYER

PROPOSED METHODS (CONTD...)

Image showing the class of ground alone.(15000 similar images is used).

Image showing the class of waste particle. (15000 similar images is used).

PROPOSED METHODS (CONTD...)



PROPOSED METHODS (CONTD...)

- Advantages of keras model :

(1) Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).

(2) Supports both convolutional networks and recurrent networks, as well as combinations of the two.

(3) Runs seamlessly on CPU and GPU

PROPOSED METHODS (CONTD...)

(2) **Using YOLO object detection method**

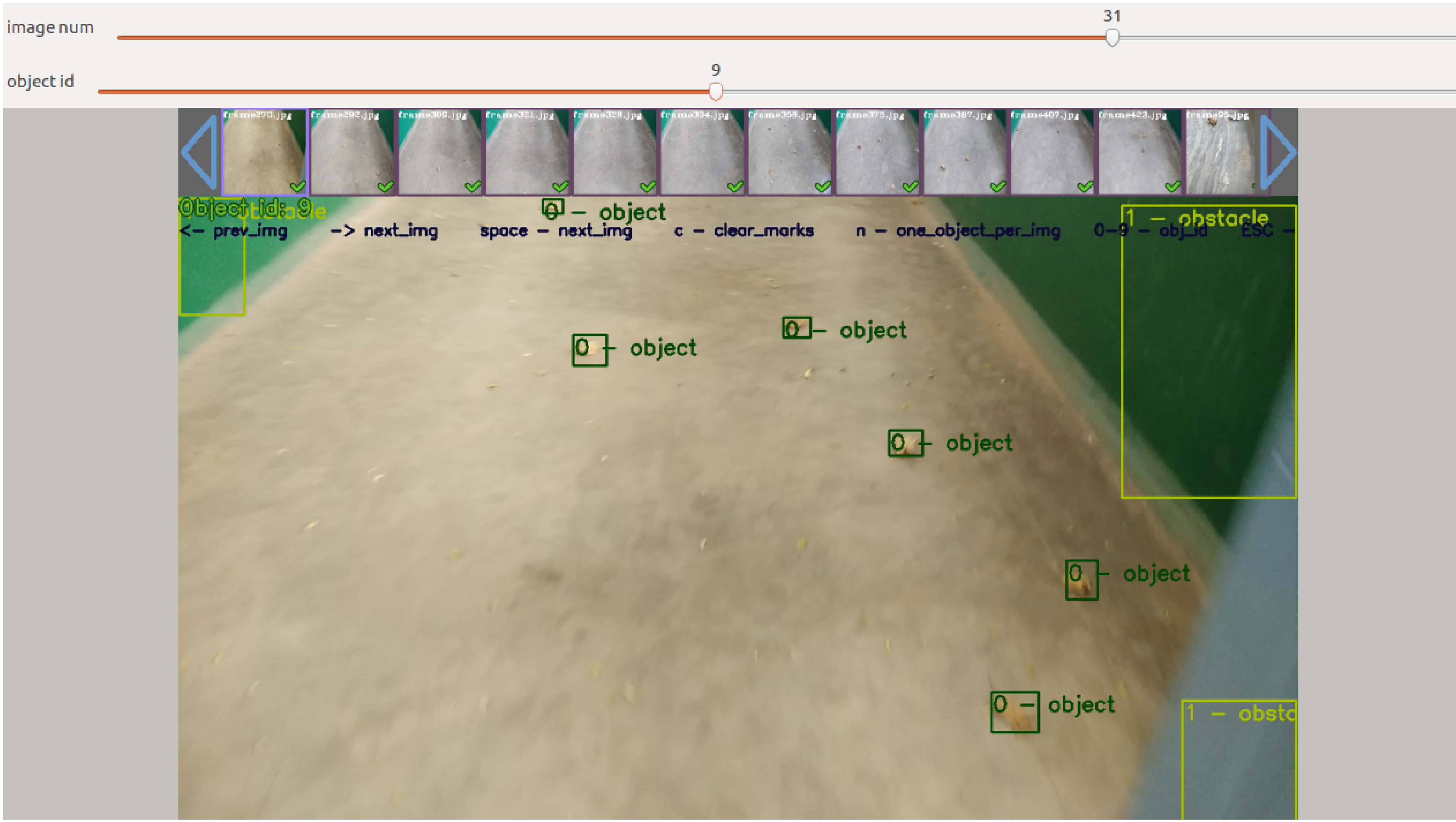
- YOLO Object Mark :

A) Training is carried out by marking objects and obstacles on every images. After,

B) A text file that contains object locations on the every images is created (Annotation).

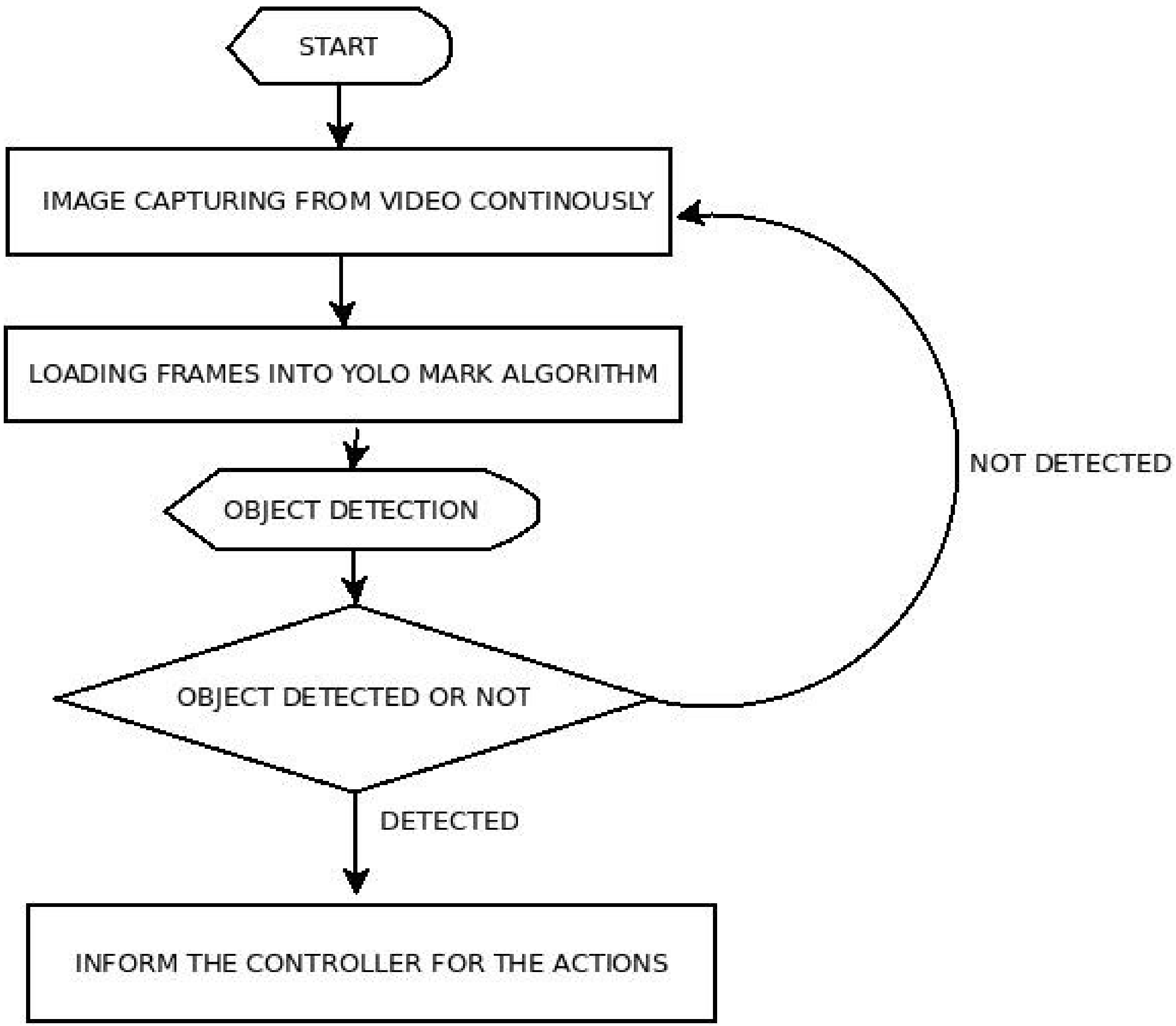
C) Marked objects and text files are moved to Darknet for training purposes.

PROPOSED METHODS (CONTD...)



YOLO Marking Method

PROPOSED METHODS (CONTD...)



Algorithm to detect objects from images

PROPOSED METHODS (CONTD...)

- Training Using Darknet

- ✓ Used to understand the objects and obstacles from images using Convolutional Neural Networks.
- ✓ 31 hidden layer is used in this work (24 CONV + 7 POOL).
- ✓ The annotation file and images from yolo mark detection for the training process are obtained.

PROPOSED METHODS (CONTD...)

layer		filters	size	input		output
0	conv	32	3 x 3 / 1	416 x 416 x 3	->	416 x 416 x 32
1	max		2 x 2 / 2	416 x 416 x 32	->	208 x 208 x 32
2	conv	64	3 x 3 / 1	208 x 208 x 32	->	208 x 208 x 64
3	max		2 x 2 / 2	208 x 208 x 64	->	104 x 104 x 64
4	conv	128	3 x 3 / 1	104 x 104 x 64	->	104 x 104 x 128
5	conv	64	1 x 1 / 1	104 x 104 x 128	->	104 x 104 x 64
6	conv	128	3 x 3 / 1	104 x 104 x 64	->	104 x 104 x 128
7	max		2 x 2 / 2	104 x 104 x 128	->	52 x 52 x 128
8	conv	256	3 x 3 / 1	52 x 52 x 128	->	52 x 52 x 256
9	conv	128	1 x 1 / 1	52 x 52 x 256	->	52 x 52 x 128
10	conv	256	3 x 3 / 1	52 x 52 x 128	->	52 x 52 x 256
11	max		2 x 2 / 2	52 x 52 x 256	->	26 x 26 x 256
12	conv	512	3 x 3 / 1	26 x 26 x 256	->	26 x 26 x 512
13	conv	256	1 x 1 / 1	26 x 26 x 512	->	26 x 26 x 256
14	conv	512	3 x 3 / 1	26 x 26 x 256	->	26 x 26 x 512
15	conv	256	1 x 1 / 1	26 x 26 x 512	->	26 x 26 x 256
16	conv	512	3 x 3 / 1	26 x 26 x 256	->	26 x 26 x 512
17	max		2 x 2 / 2	26 x 26 x 512	->	13 x 13 x 512
18	conv	1024	3 x 3 / 1	13 x 13 x 512	->	13 x 13 x1024
19	conv	512	1 x 1 / 1	13 x 13 x1024	->	13 x 13 x 512
20	conv	1024	3 x 3 / 1	13 x 13 x 512	->	13 x 13 x1024
21	conv	512	1 x 1 / 1	13 x 13 x1024	->	13 x 13 x 512
22	conv	1024	3 x 3 / 1	13 x 13 x 512	->	13 x 13 x1024
23	conv	1024	3 x 3 / 1	13 x 13 x1024	->	13 x 13 x1024
24	conv	1024	3 x 3 / 1	13 x 13 x1024	->	13 x 13 x1024
25	route	16				
26	conv	64	1 x 1 / 1	26 x 26 x 512	->	26 x 26 x 64
27	reorg		/ 2	26 x 26 x 64	->	13 x 13 x 256
28	route	27 24				
29	conv	1024	3 x 3 / 1	13 x 13 x1280	->	13 x 13 x1024
30	conv	125	1 x 1 / 1	13 x 13 x1024	->	13 x 13 x 125
31	detection					

Training process

PROPOSED METHODS (CONTD...)

- Epoch vs Batch size vs Iterations

➔ Epoch : One Epoch is said to occur when an entire dataset is passed forward and backward through the neural network only once.

➔ Batch size : Total number of training examples present in a single batch.

➔ Iterations : Iterations is the number of batches needed to complete one epoch.

PROPOSED METHODS (CONTD...)

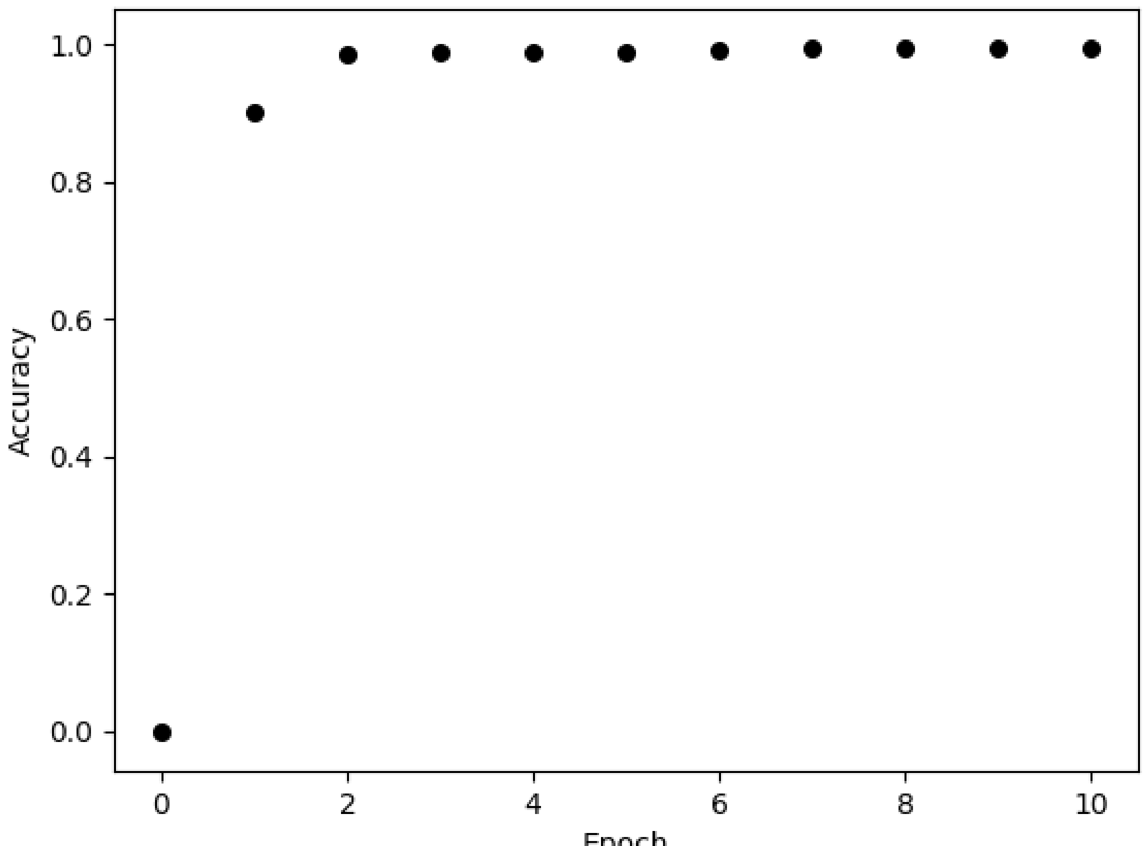
COMPARISON BETWEEN TWO DIFFERENT DEEP LEARNING FRAMEWORKS

KERAS-TENSORFLOW	YOLO MARK DETECTION
• Keras is used for the modelling the network and the tensorflow used for the executing the neural network. • Easy to construct the model • Need more input images to create dataset • Getting outputs in the form of probabilities	• Graphical based learning with deep neural network. (Bounding box is used) • Difficult to mark the objects on the images. • Need less number of images to learn the inputs as compared to the keras based model • Output is getting bounding box of every classes.

EXPERIMENTAL RESULTS

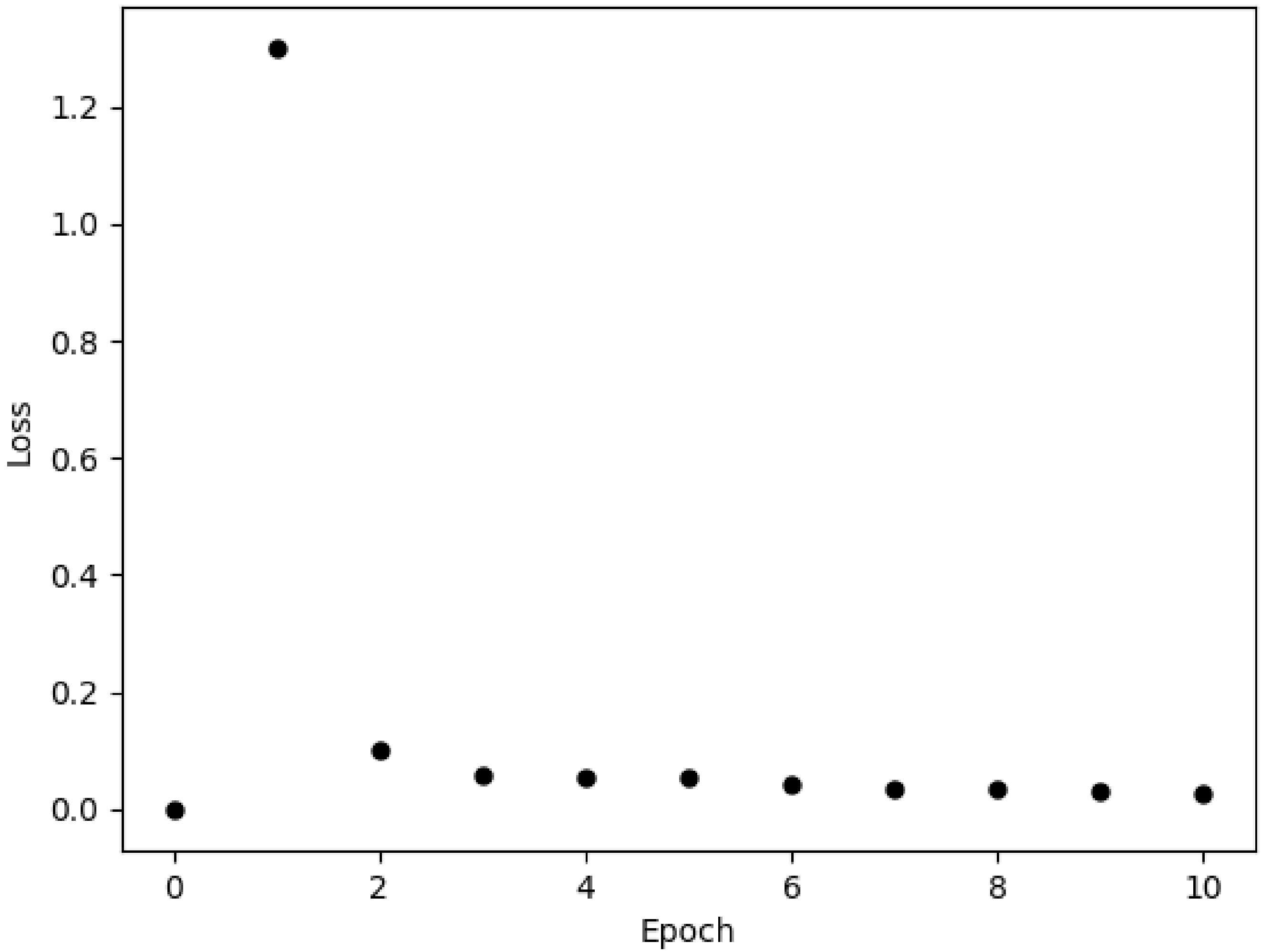
(1)Keras model (during training process)

a) Graph between accuracy and epoch



EXPERIMENTAL RESULTS (CONTD...)

b) Graph between Loss vs epoch

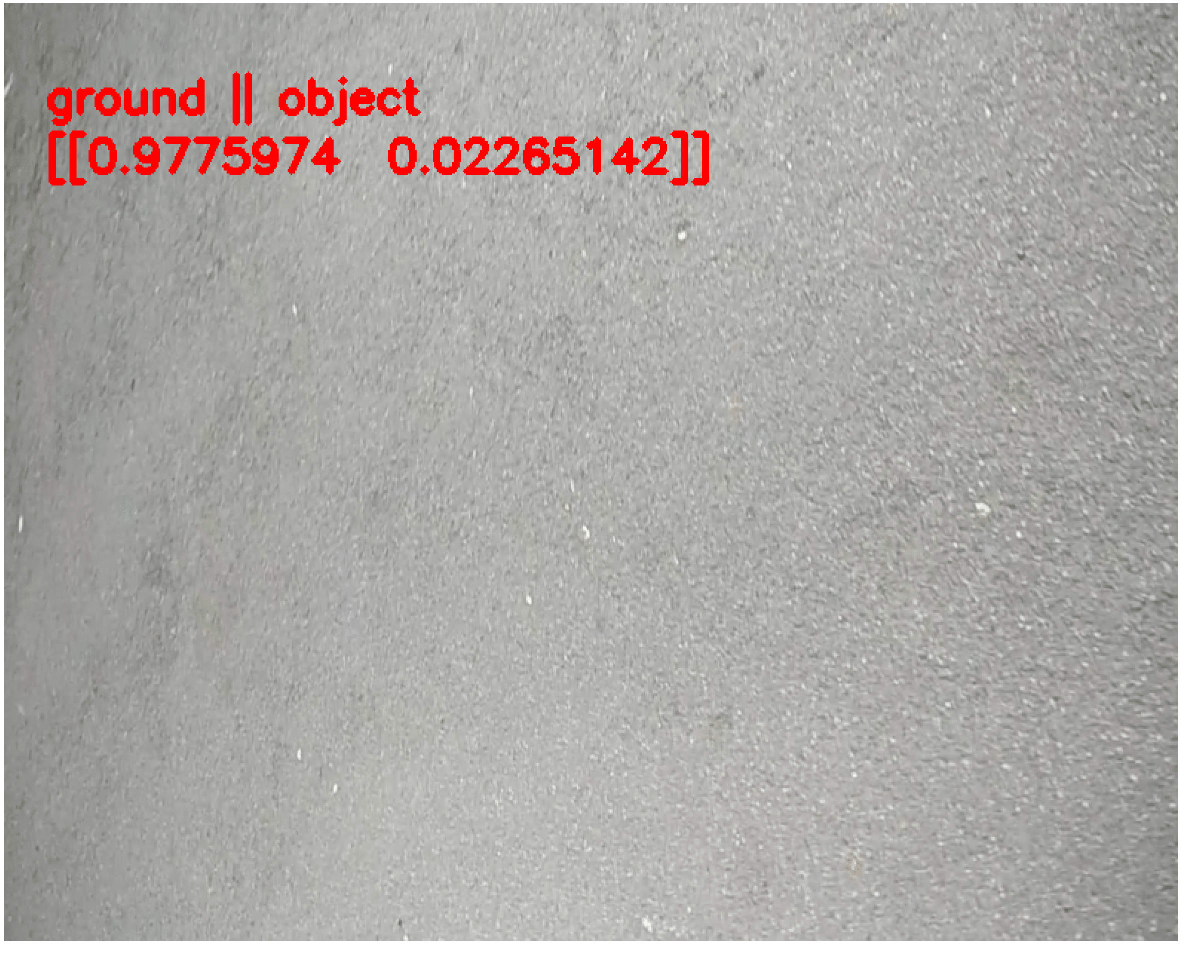


EXPERIMENTAL RESULTS (CONTD...)

- Predicted image for an input frame.
- P(object) > 0.5
- P(object)>P(waste)

EXPERIMENTAL RESULTS (CONTD...)

- For a ground alone input image
- P(ground) > 0.5
- P(ground)>P(object)

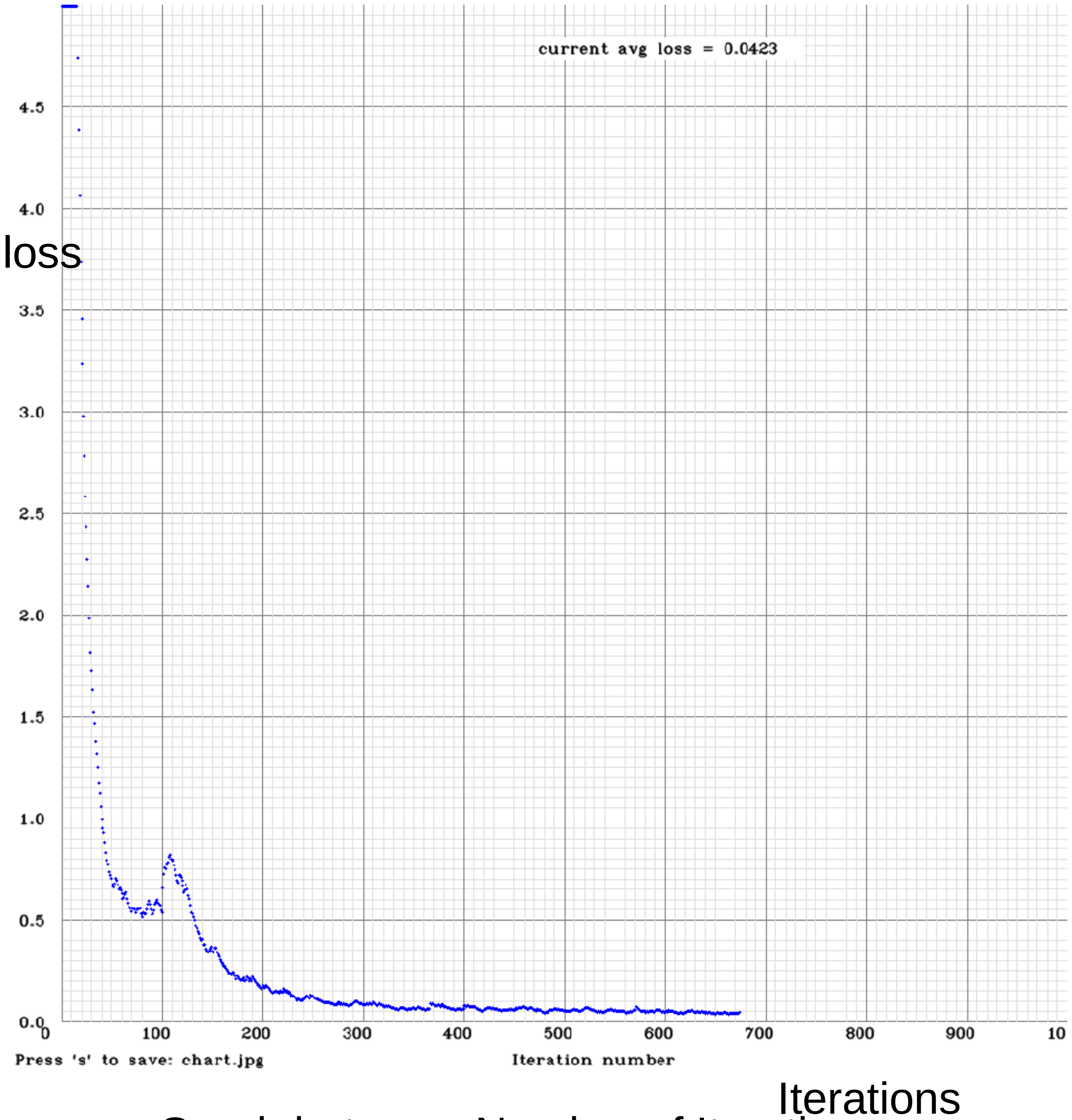


EXPERIMENTAL RESULTS (CONTD...)

(2) For the YOLO mark detection

<u>Losses during Training</u>

- Graph shows that loss is gradually decreasing with increasing number of Iterations.
- Number of Iterations =1000
- Number of images used for training = 1000
- Average loss = 0.0423

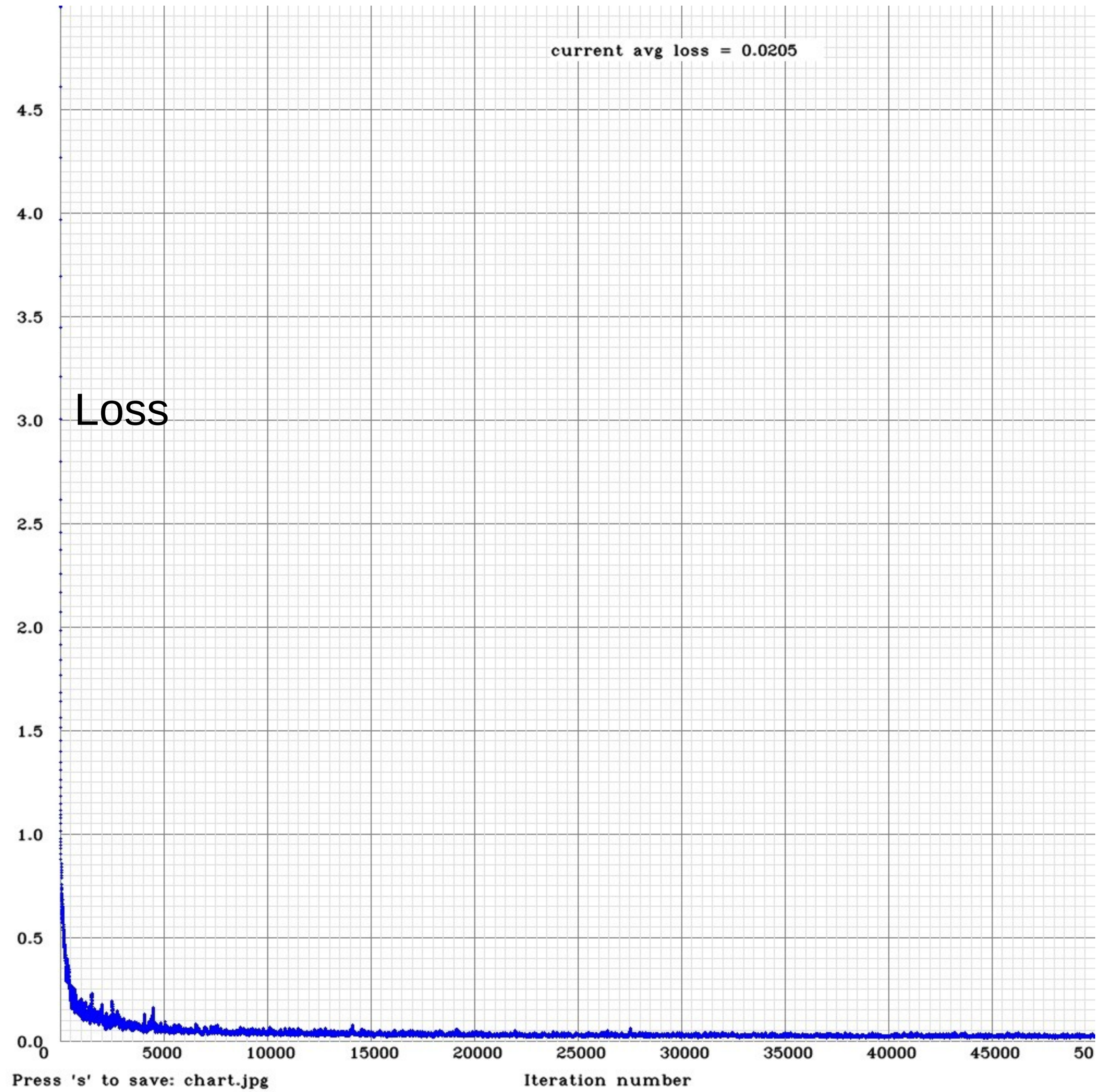


Graph between Number of Iterations and Losses.

EXPERIMENTAL RESULTS (CONTD...)

<u>Losses for different Iterations</u>

- Number of Iterations =50000
- Number of images used for training = 1000
- Average loss = 0.0205



Iterations

Graph between Number of Iterations and Losses.

EXPERIMENTAL RESULTS (CONTD...)

Test data with same environment condition of train data

Test Input Image

EXPERIMENTAL RESULTS (CONTD...)

Output Image

EXPERIMENTAL RESULTS (CONTD...)

Test data with different environment condition of train data

Fig 3 : Test Input Image

EXPERIMENTAL RESULTS (CONTD...)

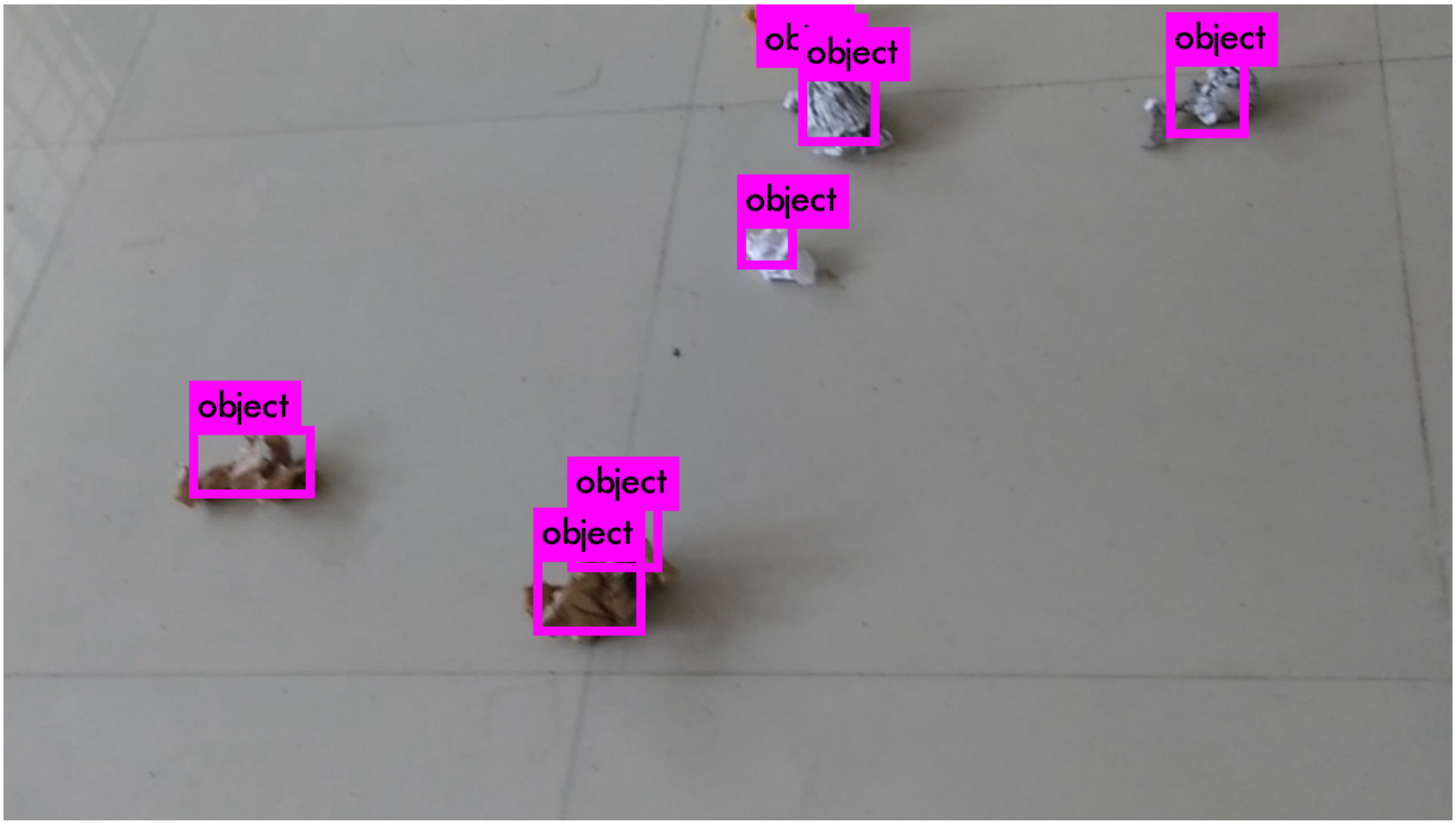


Fig 4 : Output Image

EXPERIMENTAL RESULTS (CONTD...)

Detection Accuracy Comparison

Input Data Environment	Total Detected	Wrong Detection	Total Input Images used	Accuracy
Same Environment (seen)	19	1	20	95 %
Different Environment (unseen)	17	3	20	85 %

Table 1 : Inputs and Accuracies

EXPERIMENTAL RESULTS (CONTD...)

OUTPUT COMPARISON BETWEEN DIFFERENT DEEP LEARNING FRAMEWORKS

KERAS-TENSORFLOW	YOLO MARK DETECTION
1) Less accuracy as compared with the YOLO mark detection method 2) Need more images for training to get more accuracy. 3) No any idea about location of objects without using depth sensors 4) Probability based output prediction	1) High accuracy 2) Less number of input images needed as compared with the keras model to get good accuracy 3) It can locate the objects and getting depth of objects 4) Bounding box based output

CONTRIBUTION OF THESIS

- Real time object detection within small duration.
- Adding different environments into inputs to get more accurate results.
- Avoid the obstacles easily.
- Can detect even detect small objects.
- Comparison of different Deep learning frameworks used for classifying objects.

PUBLICATION BASED ON THESIS

- Submitted to 11th Indian conference on Computer Vision, Graphics and Image Processing (ICVGIP) IIIT Hyderabad.

PROBLEMS FACED DURING THE WORK

- Training of the dataset take more time.
- YOLO based marking of input images is very difficult.
- Collection of images from different environment was a difficult task.
- In certain conditions, stones were also detected as waste particles (Which can be avoided by putting more similar input images to learning the dataset).

CONCLUSIONS

- All objects were identified for unseen and seen environments.
- Compared the different models.
- Obtained good accuracy for the real time detection problem.
- Detection within less time.

REFERENCES

1. Jmour, Nadia, Sehla Zayen, and Afef Abdelkrim. "Convolutional neural networks for image classification." 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET). IEEE, 2018.

2. Shiddieqy, Hasbi Ash, Farkhad Ihsan Hariadi, and Trio Adiono. "Implementation of deep-learning based image classification on single board computer." Electronics and Smart Devices (ISESD), 2017 International Symposium on. IEEE, 2017.

3. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

4. Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." arXiv preprint (2017).

Thank You