Proceedings of 13th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2023 Month: April Volume: 13

YOLO Based Abandoned Garbage Detection From Video Stream

Ajit Gauli [a], Smita Adhikari [b], Bamdev Bhandari [c], Bibek Thapa [d]

[a, b, c, d] *Department of Electronics and Computer Engineering, Pashchimanchal Campus, IOE, Tribhuvan University, Nepal*
[a] ajitgauli111@gmail.com, [b] adsmita1@gmail.com, [c] bamdevb09@wrc.edu.np, [d] bibekthapa619@gmail.com

Abstract
Abandoned garbage detection and classification mechanism through the analysis of video frames that can be useful for smart waste management. Implemeting deep learning techniques has been a growing topic of research for waste detection and classification due to quick advancement in computational capabilities. The main objective of this thesis is to create an object detection model that can identify abandoned waste by analyzing video streams, and subsequently categorize the detected waste to six different classes of garbage blob, trash, dustbin, solid waste, garbage bag and organic waste. This paper analyses the state-of-art method for waste detection and classification in custom dataset, in which categories of waste and rules of annotation are specified.The custom dataset consists of significant number of images that were collected from different places of Pokhara. The experimental result shows that among different existing single stage state-of-art classifiers, pre-trained YOLO-v7 model with Extended-ELAN as backbone architecture on COCO dataset outperforms other models. The dataset given here might be applied by other researchers as challenging data or further training data for an innovative garbage detection and classification system. The findings indicate that the suggested methodology may significantly improve efficiency for waste management in smart cities.

Keywords
CNN, ELAN, Garbage Classification, Garbage detection, YOLOv7-tiny model

1. Introduction

Cities are starting to develop into "smart cities" nowadays. Smart waste management is an integral part of any smart city. Excessive waste are generated due to overpopulation and rapid urbanization. Various scenarios may arise that can cause abudant availabilty of garbage in our environment:

- There are individuals who do not bother to transport the waste to the appropriate locations, abandoning the waste in inappropriate area point of view.
- There are illegal landfills that often arise even when the specialized entities no longer have space to store the waste.
- Many times during the daily collection of waste by specialized entities, parts of the city to be cleaned up are forgotten
- There are situations where waste is deliberately left on site because there are not enough workers or means like dustbin available.

Beyond causing the degradation of the area, abandoned garbage can cause pollution and have a negative impact on the quality of life of residents in those areas. So, an effective and efficient waste management system would be of great societal benefit. CNN-based approaches have become mainstream for object detection. They can be divided into two classes: the network of two stages and the network of one stage.The crucial distinction among them is the proposal for region. Single stage detectors extract one time function and proposes final regression layer area like You Only Look Once YOLO[1] and Single Shot Detector SSD[2]. There is accuracy precision trade off between single stage and double stage detector. The single stage detector is fast but less precise, while the two-stage detector is more precise but slow. Existing state-of-art garbage detection method are not so robust and fast enough for detection in video streams. As object detection in video streams requires high speed detection, the targeted problem of the research is formulated as garbage detection and classification using YOLOv7 pre-trained model, which is a state-of-art single stage object detection algorithm. Using the transfer learning approach helps to improve efficiency of advanced state-of-the-art techniques trained on generalized dataset but transfer learning method can be used with a new dataset for training, to increase the scene specific precision with a pre-trained model. But this requires preparation of new data to train and test the model. Data annotation, itself is a tedious job which requires a huge effort and time.

This research contributes a new custom dataset collected in Pokhara city at different scenes (sunny day, rainy day, foggy day, fine day). Garbage are classified in 6 main categories: Garbage blob, Trash, Solid waste, Organic waste, dustbin and garbage bags. In addition, the performance of representing state-of-art models on this new dataset are compared. In this study, for comparing the performance of the model, time taken by the model to make inference for a frame is considered. This is essential for determining the feasibility of model to make real-time detection. The metrics considered for this comparison is mean Average precision(mAP). A score is returned by mAP by contrasting the bounding box of ground truth with the box detected. The higher the score, the more accurate the model is in its detection. This research illustrates that YOLO-v7 model performs well for the garbage detection and classification task for the collected custom dataset.

2. Objective

The objective of this research is to design a YOLO based model, that can detect and classify abandoned garbage in video streams.

3. Related Works

With the increased interaction in the fields of computer graphics and computer vision, a major shift came about at the end of the 1990s. This included rendering based on images, image morphing, interpolation of views, panoramic image stitching and early rendering of light fields[3]. Since the emergence of the very first research on object detection, image classification systems have become a prospective high valued research field. Garbage detection and classification has always been a focused area for intelligent waste management, which can be done with greater efficiency from video surveillance as compared to other methods.

Automated garbage detection in video streams is an important task for maintaining clean environments.The technique of locating and identifying objects in a video or picture stream is known as object detection. Many methods, including Haar cascades, HOG, and CNNs, have been employed for garbage detection utilizing object detection. By training the classifier on both positive and negative images, researchers have employed Haar cascades to detect trash in trash cans. In order to train the classifier, researchers have also used HOG for garbage identification. These features include color, texture, and shape. CNNs have also been used to detect rubbish, with the model trained on both garbage- and non-garbage-containing pictures. Researchers have used deep learning for garbage detection by training a neural network on garbage and non-garbage images. The neural network can then identify garbage in a video stream.

In [4] Improved FMA-YOLOv5 Based Detection Algorithm for Floating Debris in Waterway was proposed. It uses Cross Stage Partilal Network(CSPN) as backbone, Feature Pyramid Network(FPN) as Neck and FCN as head for localization and classification of floating water debris. This architecture was trained on the custom dataset and it reached an accuracy of 78.19% mAP for the task of detection of garbage.

In [5] a network called GarbNet was proposed. This architecture was trained on the GINI Dataset and it reached an accuracy of 87.69% for the task of the detection of garbage. However, GarbNet produced wrong predictions when in an image are detected objects similar to waste or when they are in the distance.

Ying Wang and Xu Zhang [6] developed a Faster R-CNN based WasteNet: open source framework with region proposal network and ResNet pre-trained on Dataset COCO [7]. They reached an accuracy of 89%. However, the model produced false positives when in an image there were also other objects in addition to waste.

In [8] a Faster R-CNN model pre-trained on PASCAL VOC dataset was used, which allows to recognize landfill, recycling, and paper. The work presented reached a mAP of 68.30%. However, the authors created a dataset of images with white background where individual waste is artificially combined, and this does not reflect real life scenarios. The biggest problem with the R-CNN family of network is their speed, they are incredibly slow, obtaining only 5 FPS on GPU.

In [9] YOLOv2 is used. The architecture of the proposed neural network used MobileNet as feature extractor, and it was pre-trained on ImageNet [10]. Model reached an accuracy of 89.71% was reported, but there was only one class in the dataset and they did not test on negative examples.

4. Methodology

Focusing on the test speed and precision, this paper compares the efficiency of the leading state-of-the-art single stage classifier on our custom dataset. We divide the work in two major sub-task: Dataset Preparation and Model Analysis.

4.1 Dataset Preparation

4.1.1 Dataset

Several studies infer that natural pictures can yield strong identification results based on deep learning algorithm. To improve the state-of-the-art methods for specific task object identification and classification using CNN networks, a number of datasets like ImageNet and PASCAL VOC were already developed. Building precise and reliable machine learning models requires access to a sizable and varied dataset. It lets the model to gain knowledge from multiple data inputs, generalize that knowledge, and enhance its functionality. We present a novel custom garbage data-set to develop garbage detection model. The colleceted dataset can also support for the further research in the field of garbage detection and classification.

4.1.2 Data collection

Research is based on the images captured at various places in pokhara city. Most of the images were captured at roadsides, bare land and landfills. Collected dataset also contains clipped video having at-least a instance of a waste in the frame. Further google images[11], coco dataset[7] were also added to dataset seeking for the robustness of model. Out of all collected dataset 2014 images were selected for augmentation. Augmentation tasks like fliping, translation, rotation, zooming and brightness adjustment were performed in order to increase model's performance and also for reducing class imbalance problem.

Figure 1: Some collected garbage images

4.1.3 Garbage Annotations

In the (labelImg)[12], all related annotation tasks are performed. . Among varieties of annotation tools like LabelImg, VGG annotator tool, COCO annotator tool we selected labelImg for its robustness in data annotation and deployment. labelImg is a standalone annotation tool with advanced annotation functionalities. The User Interface (UI) of the labelImg is shown in Figure 1.

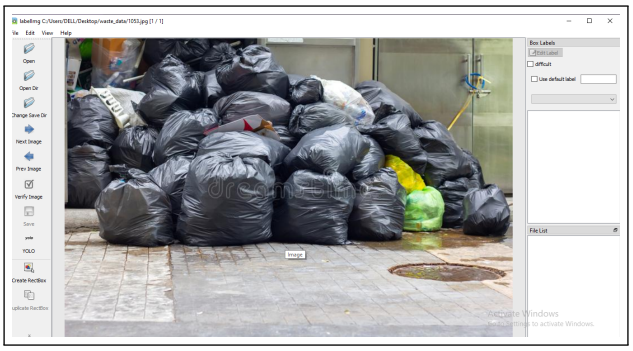

Figure 2: labelImg UI

Garbages were classified into six categories: Garbage blob, Trash, Solid waste, Organic waste, Garbage bag and Dustbin.

The performance of the data set relies on the annotation rule used, which are:

1. Small Target: The farther the object gets, the less waste class features are acquired. Small items are also carefully labeled with potential compressed bounding boxes from the very far point of view.

2. Special Samples : A few special samples with category ambiguity like trees, people, vehicles are kept in the data set but are not labelled as any of the Classes used in this data set.

5769 instances were annotated across 2084 images that contained objects of the six custom classes. Garbage images were annotated into six categories: Garbage blob, Trash, Solid waste, Organic waste, Garbage bag and Dustbin. Instances distribution in dataset is shown in figure below:

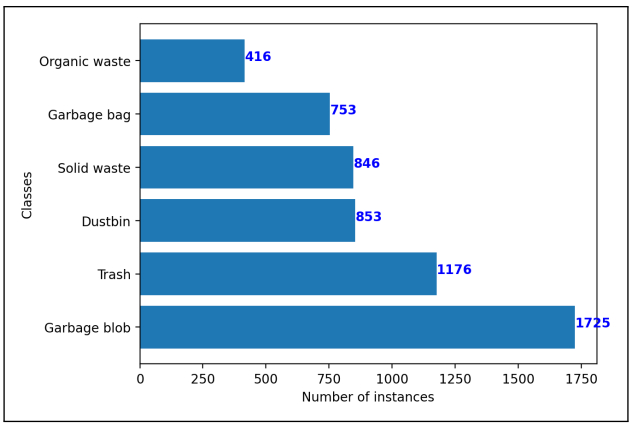

Figure 3: Classes distribution of dataset

4.1.4 Dataset Splitting into Training and Validation Set

A dataset typically has to be split into training, validation, and test sets. We chose to split the dataset into a training set and a validation set because we did not collect a significant number of photographs for our research. Images that can represent several items of interest make up the instances in the dataset used to train an object detection system in the application domain. Then, rather than taking into account the number of photographs that were accessible, we chose to divide it by the number of instances for each class. However, because to the reliance on the distribution of the number of objects per class in the various photos, this cannot be done precisely. Therefore, there might not be a combination that has exactly 80% training data and 20% validation data. A Python script has been written to get a good dataset split into training and validation set. Classes distribution of training and validation set after split are shown in figures below:

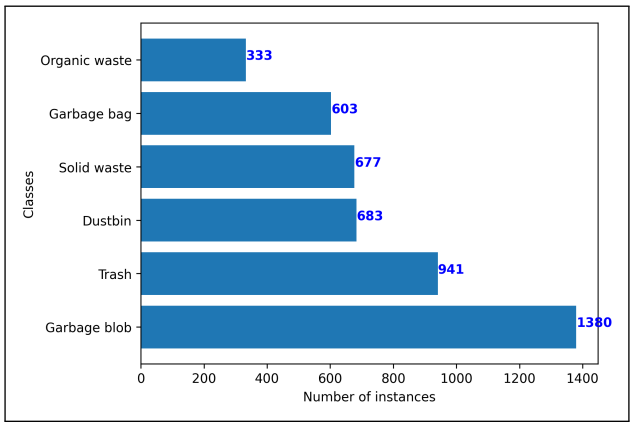

Figure 4: Classes distribution of training dataset

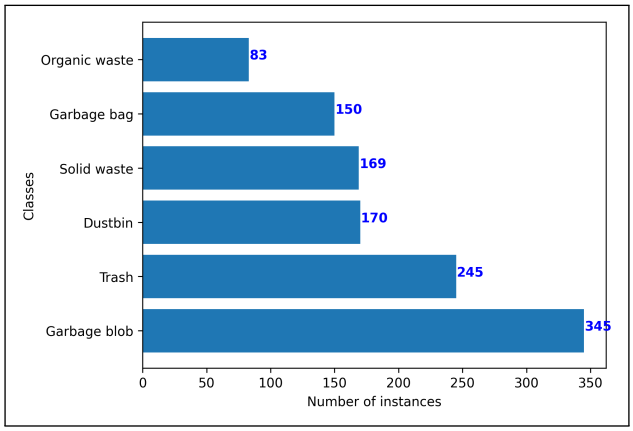

Figure 5: Classes distribution of validation dataset

4.2 Model architecture

In the recent trend, YOLOv7[13], among the single stage classifier are widely used. Recent research results performed on this field suggests YOLOv7 generates inference in real-time which is desirable to apply in some online system like garbage detection. YOLOv7 model has also been used in challenging weather conditions showing comparative higher efficiency in drastic weather condition. YOLOv7 architecture design is based on ELAN (efficient layer aggregation network) as backbone of architecture that is responsible for feature extraction. ELAN

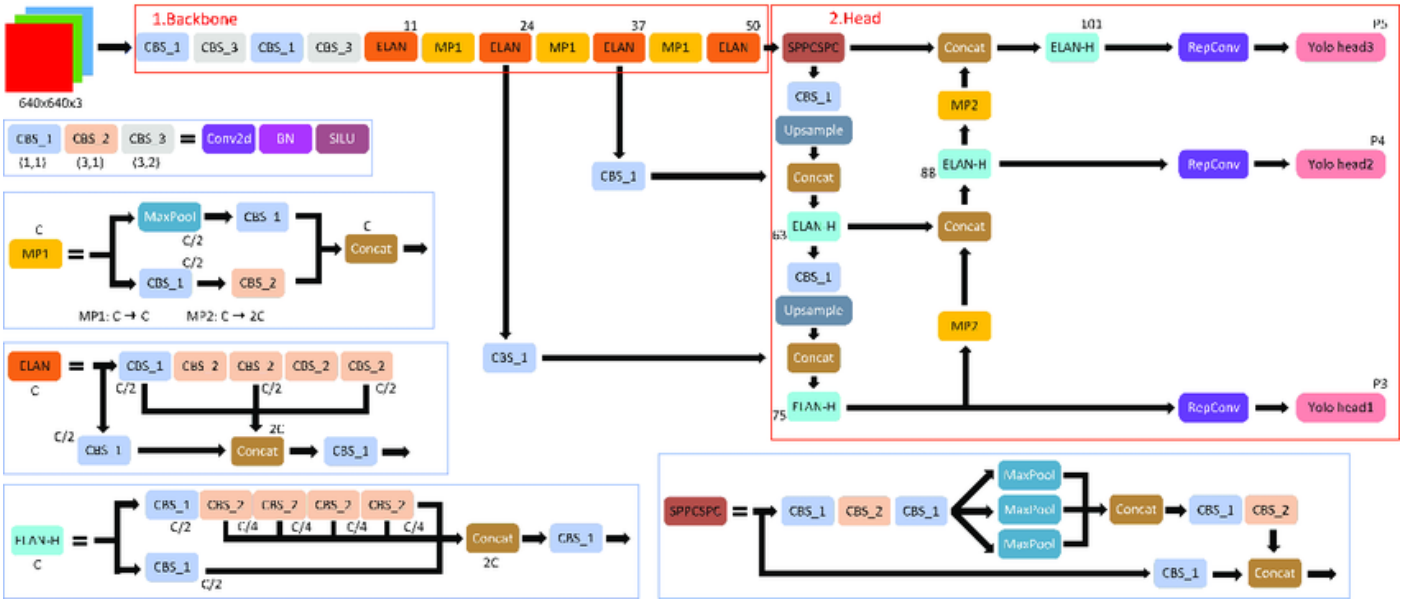

Figure 6: YOLOv7 model architecture.

considers designing an efficient network by controlling the shortest and the longest gradient path, so that deeper networks can converge and learn effectively. Neck consists of blocks of spatial pyramids networks and cross stage partial networks that combinedly work for feature aggregation. Repconv (Repeated convolution) layer is placed before the head of YOLOv7 network allows it to extract more meaningful features from the feature maps produced by the backbone network, which in turn helps to increase the accuracy of the object detection and classification tasks performed by the model. The head is responsible for production of predicted outputs of model. To improve training and performance, YOLOv7 uses a technique called Deep Supervision, which allows for multiple heads instead of just one. The primary head that is responsible for the final output is called the lead head, and the additional head(s) used to help with training in intermediate layers are referred to as auxiliary head(s). In general, YOLOv7 provides a faster and stronger network architecture that provides a more effective feature integration method, more accurate object detection performance, a more robust loss function, and an increased label assignment and model training efficiency.

4.3 Evaluation protocol

We compare mean average precision(mAP), at detection IOU threshold 0.5, to find the model's efficiency as used in several major detection benchmarks[14], [15]. This evaluation protocol is performed for the key frame at 640*640 resolution. The confidence threshold for detecting the garbage is set to 0.5 as an standard threshold metrics to perform the inference.The inference speed of the models are also compared on varying resolutions of the image.

5. Results

For preparing our data-set, we selected the samples in accordance to the sample balance principle to solve the problem of not having comparative ratio between instances of all categories or physical environment. However, the number of instances of garbage blob is significantly higher in our collected data-set which demonstrates a sub-sampled characteristic of today's real world's abundant presence of garbage.

S.N.	Batch Size	Epochs	mAP at IOU 0.5
1	4	75	73.20%
2	4	100	76.40%
3	8	200	68.60%
4	16	200	63.60%
5	16	300	65.70%

Figure 7: Experimental Result Summary of Yolo model for garbage Detection

The IOU threshold value was kept 0.5. Observing the figure 7 above, it can be inferred that, the training time can be shortened by increasing the batch size because the training model can handle more photos concurrently. However, if the hardware is inadequate, a bigger batch size necessitates additional memory, which could result in slower training times. A smaller batch size can improve generalization performance since it makes the model learn more carefully from each image. This may aid in preventing over-fitting, which occurs when a model memorizes training data rather than discovering underlying patterns. On the other hand, if the model is unable to generalize adequately to new images, a higher batch size may result in worse generalization performance. It was also observed that increasing number of epochs can increase generalization performance of the model.

Confusion Matrix, PR-curve and F1-curve for developed YOLOv7 tiny version by fine-tuning with our data-set taking batch-size of 4 and was iterated for 100 epochs, are shown below:

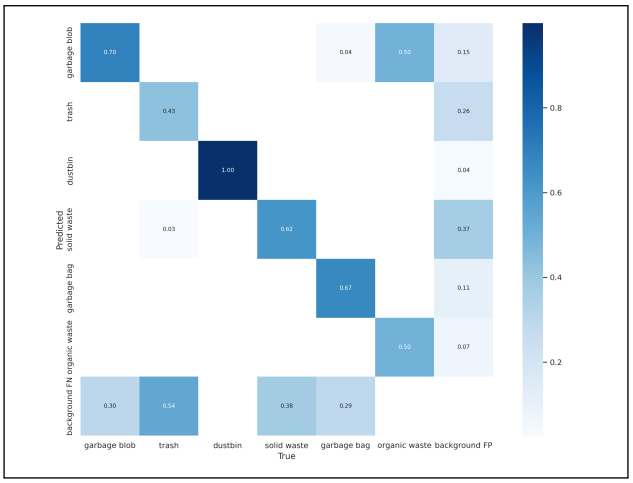

Figure 8: Confusion Matrix

Figure 8 above is confusion matrix for YOLOv7 tiny model with batch-size 4 and for 100 epochs. Row represents ground truth value and column represents predicted value of the model. Here for instance garbage blob 70% of predicted values were true positive. Although 30% of garbage blob were not detected and ignored as background. 4% of garbage bag were falsely predicted as garbage blob. 50% of organic waste were falsely detected as garbage blob and 15% background parts were also falsely detected as garbage blob.

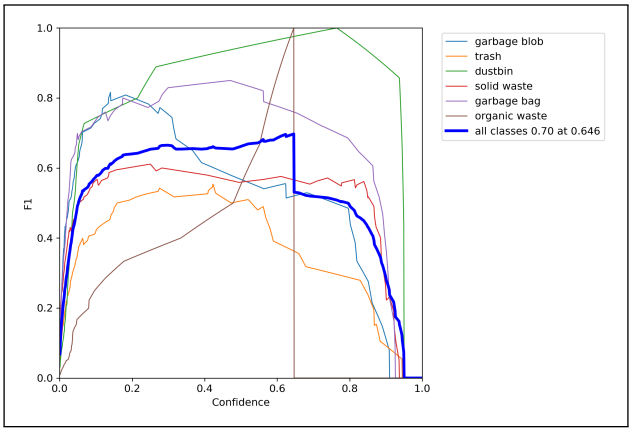

Figure 9: F1-curve

From figure 9, it can be observed that the developed model which had reached F1 score of 70% at confidence 0.646, which is a good score for object detection task.This indicates that the model is able of identifying 70% of the related instances (i.e., true positives) while minimizing the number of false positives and false negatives. F1-score for dustbin has reached to 1 and has vertically decreased because the model became conservative so it stopped predicting positive for the class organic waste and dustbin with confidence greater than 0.65.

PR-curve in figure 10 gives mean average precision of 76.4% at the confidence of 0.5. It means the model correctly identifies 76.4% of the objects in the dataset with an IoU threshold of 0.5 or higher.

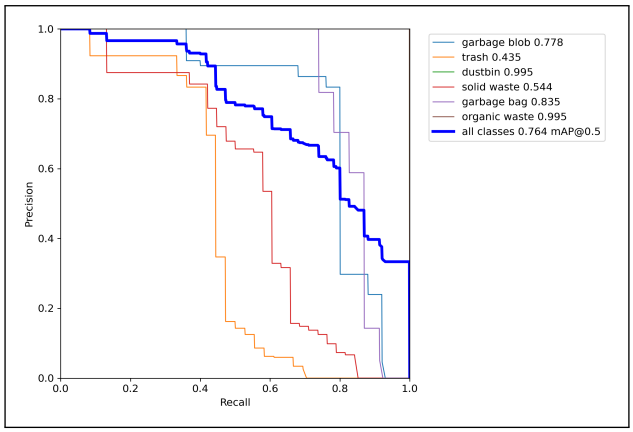

Figure 10: PR-curve

The model was tested for some random images which showed good results.

Figure 11: Garbage detection model output for random test image

The model is trained tested in Tesla T4 GPU on our collected data-set.The inference time taken by the YOLOv7 model seems to have improvement in speed and accuracy on the test frames. This was also inferred that YOLOv7 model can be used without altering the frame resolution, since the performance is not degraded haphazardly with varying resolution.

The model was then tested for its performance on video stream which was captured on streets between Paschimanchal campus gate to Lamchaur chowk. The inference speed was fast (with average of 41-68 FPS on Tesla T4 GPU) with improved accuracy. This shows a substantial evidence for the application of YOLOv7 model in implementing real time garbage detection.

6. Disccussion and Analysis

The developed YOLOv7 model reached accuracy of 76.4% mAP, which is greater than previously developed model which was upto 68%mAP that was developed with YOLOv5 model architecture.The inference (detection) speed on video was 41-68 FPS, which is also improved than previously developed waste detection model that had average inference speed of 25FPS.

7. Conclusion

In this work, we formulate a new custom data-set of abandoned garbage. Our data may be used in other experiments as an alternative data base or a difficult research collection, considering the sophistication of the proposed data-set. From the above experiment, we conclude that YOLO V7 model architecture is a best choice among other object detection models for abandoned waste detection. As this model gave fast inference speed and good accuracy, it is suitable for real time garbage detection in video streams.

As a future work, the main goal will be to increase the performance of YOLO garbage detection model by increasing the number of images in the data-set, because there is an imbalance between the number of images and the number of objects. Morever, we want to increase the number of classes to be recognized. We will use YOLOv7 architecture for training, on our custom data-set, with some hyper-parameter tuning to design a novel model for detecting abandoned garbage present in our environment with better accuracy.

References

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. 2016.

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

[3] Ahmad Arib Alfarisy, Quan Chen, and Minyi Guo. Deep learning based classification for paddy pests & diseases recognition. 2018.

[4] Qiannan Jin Feng Lin, Tian Hou and Aiju You. Improved yolo based detection algorithm for floating debris in waterway. 2021.

[5] Gaurav Mittal, Kaushal B Yagnik, Mohit Garg, and Narayanan C Krishnan. Spotgarbage: smartphone app to detect garbage using deep learning. 2016.

[6] G. Thung and M. Yang. Classification of trash for recyclability status. 2016.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft ´ coco: Common objects in context. 2016.

[8] The growing global landfill crisis. 2020.

[9] Ying Liu, Zhishan Ge, Guoyun LV, and Shikai Wang. Research on automatic garbage detection system based on deep learning and narrowband internet of things. 2018.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei Fei. Imagenet: A largescale hierarchical image database. 2009.

[11] Google "search engine", 2001.

[12] TzuTa Lin. Labelimg. `https://github.com/tzutalin/labelImg`, 2015.

[13] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan, and Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2022.

[14] Longyin Wen, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu. Ua-detrac: A new benchmark and protocol for multi-object detection and tracking. 2020.

[15] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.