Deep Neural Network based Automatic Litter Detection in Desert Areas Using Unmanned Aerial Vehicle Imagery

1st Guoxu Wang
Department of Computer Science
Loughborough University
Loughborough, UK
G.Wang@lboro.ac.uk

2nd Andrew Leonce
College of Technological Innovation
Zayed University
United Arab Emirates
Andrew.Leonce@zu.ac.ae

3rd Hakim Hacid
Technology Innovation Institute
Masdar city, Abu Dhabi, UAE
Hakim.Hacid@tii.ae

4th E.A. Edirisinghe
School of Computing & Mathematics
Keele University
Keele, UK
E.Edirisinghe@keele.ac.uk

2023 International Symposium on Networks, Computers and Communications (ISNCC) | 979-8-3503-3559-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISNCC58260.2023.10323960

***Abstract*—The United Arab Emirates (UAE) values its relationship with the desert, considering it a crucial part of its heritage and culture. However, the desert faces environmental challenges due to the improper disposal of garbage by visitors and the dumping of waste, as some perceive the desert as an empty wasteland. The rise in tourism exacerbates the problem, as litter negatively impacts the desert's ecology, wildlife, and natural habitats. Traditional litter collection methods involving human patrols are inadequate for the vast desert terrain. Drones equipped with high-resolution cameras offer a potential solution by conducting aerial surveys quickly and efficiently. However, the manual review of drone footage to detect litter is time-consuming. This paper explores the use of deep neural network architectures, such as Faster R-CNN, SSD, and YOLO, to develop litter detection models. These models focus on distinguishing litter from other man-made objects. The training dataset consists of thousands of samples, and the models are evaluated based on their performance in detecting and locating litter in drone images captured at different altitudes and environmental conditions. The evaluation includes objective and subjective analyses. The research aims to alleviate the practical challenges of litter detection in the desert by automating the process through computer vision-based object detection methods.**

***Index Terms*—Convolutional Neural Networks, Litter Detection, Object Detection, YOLO-V5, Drone Imagery, Desert Ecology, Environment Protection, Desert Environment**

I. Introduction

The global expanse of desert land measures approximately 47 million square kilometres, with deserts found extensively across the African, Asian, Australian, American, and European continents [1]. Prominent deserts include the Sahara in North Africa, the Rub Khali in Saudi Arabia, and the Taklamakan in China [2]. Within the United Arab Emirates, the Dubai Desert Conservation Reserve (DDCR) safeguards an area spanning roughly 250 square kilometres, serving as a protected habitat for numerous endangered wildlife species and flora [3]. The DDCR actively engages in diverse desert conservation initiatives encompassing wildlife breeding, controlled release programs, and preservation of endangered tree species. Each year, the DDCR attracts a multitude of tourists, scientists, and researchers. Nevertheless, frequent human presence in nature reserves such as the DDCR often results in the significant accumulation of litter, posing a detrimental threat to the natural environment, flora, and fauna.

Current practices for litter removal in desert regions primarily involve deploying teams of litter pickers, comprising workers and volunteers, to high-probability litter areas, particularly popular recreational sites. These groups undertake the laborious task of manually searching for and collecting litter, necessitating navigation across challenging terrains either on foot or in vehicles. However, this approach proves arduous, as it often entails covering extensive areas where litter may ultimately be absent. Consequently, valuable time, effort, and resources are wasted, while individuals face potential risks. In recent times, the prevalence of low-cost drones has led to their utilization for terrain surveillance. By flying at significant altitudes, drones offer the ability to swiftly cover expansive ground areas. The subsequent images or videos captured by drones are manually inspected by humans to identify litter presence and mark corresponding locations. Nevertheless, this manual review process is also burdensome, as operators of the video surveillance system must subjectively analyse all drone footage. Once human observations are made, litter pickers can

be directed to marked areas, potentially via GPS coordinates, for eventual collection.

To address these challenges, computer vision-based approaches can be employed to automatically detect and locate litter in drone footage. Traditional machine learning methods require the identification of distinctive litter features through manual feature engineering, followed by the use of feature-based object classifiers to differentiate litter from other objects. However, a significant challenge lies in defining the most effective features that enable accurate differentiation from other objects. Previous attempts in the literature to employ machine learning for litter detection have yielded limited success, with recorded accuracy levels remaining relatively low [4]–[6].

In recent years, deep learning, a subset of Artificial Intelligence (AI), has emerged as the dominant technology for addressing object detection and classification challenges in computer vision, surpassing traditional machine learning approaches [7]. The key advantage of deep learning lies in its ability to automate feature engineering, unlike classical machine learning methods that require manual feature engineering [8]. In the realm of 2D image processing applications, Convolutional Neural Networks (CNNs) are popular deep neural networks employed. They can be categorised into two main groups. The first category is Region-based CNN (R-CNN), which includes variations such as Fast R-CNN and Faster R-CNN [9]–[11]. These networks operate in two stages: initially generating region proposals and subsequently performing classification and regression on these proposals. The second category encompasses one-stage CNN architectures like You Only Look Once (YOLO) and single shot detector (SSD) [12], [13]. These architectures employ a single CNN network to directly predict the categories and locations of different targets. YOLO has evolved over the past five years, progressing from YOLO-V1, v2, v3 to the v5. Our recent research and related work [14]–[18] suggest that YOLO-V5, the more established variant, outperforms earlier versions of YOLO, SSD, and R-CNN architectures for most object detection and classification tasks.

In this paper, we extensively investigate the utilization of YOLO-V5 and its four variants representing deep neural network architectures of varying depth: S (Small-shallowest), M (Medium), L (Large), and X (Extra Large). However, we do not investigate the use of YOLO-V7 and YOLO-V8 in this study as they were released during the concluding stage of this research. Our approach focuses on single-class detection, specifically distinguishing litter objects from the background of the scene, without considering any other types of objects. To thoroughly evaluate and analyse the effectiveness of this litter detection approach, we conduct experiments in two distinct environments: remote areas primarily composed of natural objects in the background, and sub-urban areas where the background includes man-made objects like camp sites.

To enhance the clarity of presentation, this paper is structured into five sub-sections. Section 1 introduces the application and research context of the study. In Section 2, we present the YOLO-V5 network architecture and define the objective metrics used to measure and compare the performance of the trained models. Section 3 outlines the research methodology, including the experimental design details for the two approaches adopted for litter detection. Additionally, we provide information on dataset preparation, data labelling, training, and the testing procedure employed to evaluate the performance of the Deep Neural Network (DNN) model. Section 4 offers a comprehensive analysis of the litter detection results, along with detailed insights into the performance of the trained model under the chosen litter detection approach. Finally, in Section 5, we conclude the paper, provide suggestions for future work, and propose potential improvements for the established DNN model.

II. Background

A. Quantitative Performance Evaluation Metrics

In our study, we employ various metrics to objectively analyse the performance of the four object detection models obtained through training the sub-versions of YOLO-V5. These metrics serve as quantitative measures for evaluating the effectiveness of the models. Additionally, in section 4, we present the experimental results by utilizing these metrics and conduct a subjective performance comparison to provide further insights.

1) Intersection over Union (IoU): Intersection over Union (IoU) measures the overlap between the predicted bounding box and the ground truth bounding box of an object. To calculate the IoU, the area of intersection (the overlapping region) between the predicted bounding box and the ground truth bounding box is divided by the area of union (the combined region) between the two bounding boxes. The formula for IoU is shown in Figure 1.

The IoU value ranges from 0 to 1, where a value of 1 indicates a perfect overlap between the predicted and ground truth bounding boxes, while a value of 0 indicates no overlap at all. IoU is used as a measure of how well the object detection model is able to accurately localize the objects in an image. It is often used as a criterion for evaluating the performance of object detection algorithms and determining the accuracy of bounding box predictions. A higher IoU score indicates a better detection performance. In all our experiments we have used an overlap of 0.5 as default.

2) Mean Average Precision (mAP): Mean Average Precision (mAP) is a measure of the overall performance of an object detection model in terms of both precision and recall. Precision refers to the accuracy of the predicted bounding boxes, while recall refers to the ability of the model to detect all instances of a given object class. mAP takes into account the precision and recall values at different levels of confidence thresholds or IoU thresholds. The calculation of mAP involves the following steps:

1. For each class in the dataset, the precision and recall values are computed by comparing the predicted bounding boxes with the ground truth bounding boxes.

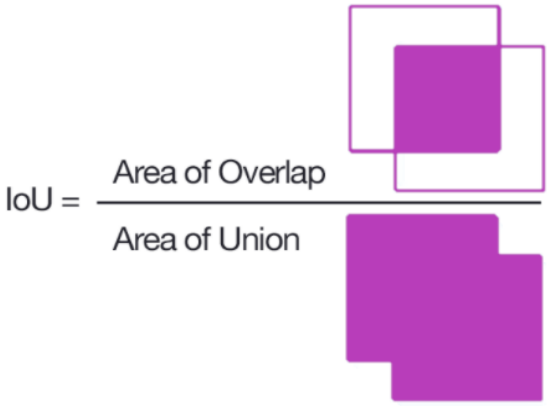


Fig. 1. Intersection of Union (IoU)

2. A precision-recall curve is generated by plotting the precision values against the corresponding recall values at different confidence thresholds or IoU thresholds.
3. The average precision (AP) is calculated for each class by computing the area under the precision-recall curve.
4. Finally, the mAP is computed by taking the mean of the AP values across all classes. mAP provides a comprehensive evaluation of the object detection model's performance by considering both precision and recall across different object classes. It is a widely used metric for comparing and benchmarking different object detection algorithms or models. Higher mAP values indicate better overall performance in terms of both localization accuracy and object detection capabilities. In this study, we set a threshold of 0.5.

III. Methodology

The proposed approach to litter detection is based on a generic workflow, as depicted in Figure 2. This workflow consists of two main stages: the training stage (comprising training and validation) and the testing stage. Prior to training, the collected dataset must undergo preparation to ensure its suitability for training, and the objects of interest need to be accurately labelled for the purposes of training, validation, and testing. Subsequently, the Deep Neural Network is trained using the prepared dataset. Upon completion of the training process, a model is generated, which can serve as an object detector/classifier. During the testing stage, the model is applied to a set of test images, where it identifies and classifies various object types present in each image. In the case of a single object detector model, the trained model specifically focuses on detecting one type of object. The specific details and procedures pertaining to each phase, including data preparation, labelling, training, validation, and testing, are described in the subsequent sub-sections.

A. Data Preparation

We conduct a comprehensive investigation into a specialized approach for litter detection, focusing exclusively on a single object class: litter objects. The performance and effectiveness of this approach in accurately detecting litter in both natural and sub-urban environments are thoroughly examined and presented in Section 4. The results showcase the strengths and capabilities of our approach in successfully identifying and detecting litter objects in these specific environmental settings.

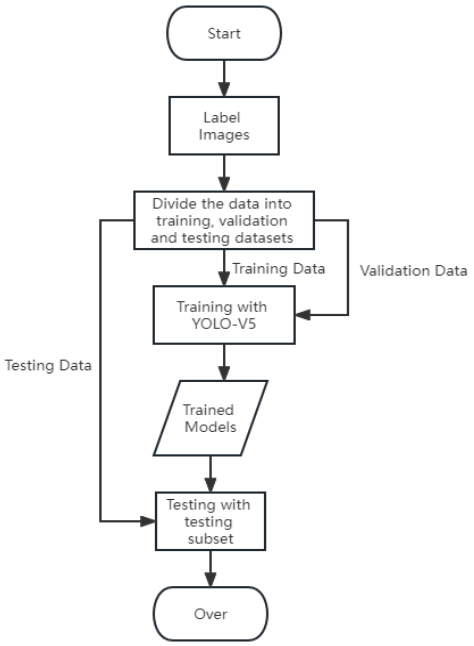


Fig. 2. The workflow of the proposed method

For the single-class litter detection approach, a dataset of 913 images was gathered from drone flights conducted at various altitudes within the nature/natural areas of the DDCR. These images were randomly selected and contained over 8000 instances of litter objects. The chosen dataset predominantly consisted of natural habitat, devoid of substantial human-made objects in the background. To facilitate training, validation, and testing, the dataset was partitioned into three subsets: training (60%), validation (20%), and testing (20%). Notably, due to variations in the number of litter objects present in each image, the allocation of images across the subsets differed. Specifically, the training, validation, and test subsets comprised 512, 236, and 165 images, respectively.

The authors collected all the drone imagery within the nature reserve areas of the Dubai Desert Conservation Reserve (DDCR) using DJI Phantom 4 and DJI Mavic 2 Pro drones. The data collection process involved flying the drones at different heights/altitudes, speeds, and with varying camera angles to capture comprehensive aerial views.

TABLE I
Number of Labelled Litters in Each Data Subset

Dataset	Number of labelled litters
Training	5000
Validation	1600
Testing	1600

B. Data Labelling

The drone captured image set comprised images taken in different areas of the DDCR demarcated land, including nature and camp sites. These images were captured at various altitudes, camera angles, and different times of the day, specifically during daytime. The density or sparsity of litter within each image varied across samples.

The nature-site images consisted solely of natural objects and litter, with very few instances of man-made, non-litter objects. On the other hand, the camp-site images contained both litter and man-made objects/structures, making it challenging to differentiate litter without considering the image context. A subset of images was randomly selected from the drone captured data for training, validation, and testing purposes, as shown in Table 1.

To label the images, the authors utilised the labelImg tool [19], which involved placing rectangles around identifiable objects. In the 913 images captured in the natural environment, all litter objects were labelled as 'litter', resulting in a total of 8200 labelled litter objects. Among these, 5000 objects were labelled for training, 1600 for validation, and 1600 for testing subsets of images. It is important to note that the litter objects encompassed various types such as glass bottles, plastic bottles, cardboard boxes, plastic bags, etc. However, for the purpose of detection, all litter objects were labelled under a single object class, 'litter', without attempting sub-classification.

The background surrounding the litter objects exhibited variation, predominantly consisting of sand but occasionally including bushes, shrubs, or rarely man-made structures. During the labelling process, the rectangles were drawn tightly around each litter sample, aiming to encompass the litter object's shape while potentially including a portion of the background, which could be desert sand or parts of other objects present in the background. In cases where a litter object was partially occluded by other objects, the non-occluded shape of the litter was inferred, and the rectangle was drawn to include the potentially covered area of the litter object's body. Additionally, efforts were made to capture isolated litter objects with clear shape, boundary, texture, and colour. When labelling litter objects with shadows, minimizing the inclusion of the shadow within the rectangle was considered. All the aforementioned labelling criteria were adopted to enable the Deep Neural Network (DNN) to learn and recognize litter objects under various conditions, including variations in illumination, occlusion, size, and clarity.

C. Training & Validation

The experiments were conducted on a computer system consisting of an Intel Core i7-6850k CPU, NVIDIA GeForce GTX-1080ti GPU, 32 GB of RAM, and running the Windows 10 operating system. In order to evaluate the effectiveness of popular Deep Neural Network architectures in object detection, we adopt the proposed Single-Class Object Detector approach for litter detection. The training data subset, which contains labelled information specifically related to objects belonging to the 'litter' class, is utilized to train the SSD, Faster R-CNN, and four sub-versions of the YOLO-V5 Convolutional Neural Networks (CNNs). Furthermore, the labelled litter samples from the validation data subset are employed to fine-tune the CNN architectures and optimize their performance during the training process. This involves determining the optimal values of the network's hyperparameters to enhance its detection capabilities.

D. Testing

Upon completion of the training process, the trained CNN model is employed as a litter/object detector to identify litter within the test data subsets. Each test image, initially sized at 3840×2160 pixels, is automatically adjusted by the network to 608×608 pixels while maintaining the original aspect ratio. The processed image is then passed through the network to extract features, generating feature maps of three different sizes: 76×76, 38×38, and 19×19 pixels, respectively, for prediction. The 76×76 feature map is utilized for detecting small-sized litter, while the 38×38 and 19×19 feature maps are used for detecting medium and large-sized litter, respectively.

At each prediction scale, the model predicts three bounding boxes using anchor box scales that were learned automatically during the training process. Each bounding box consists of coordinates, width, height, a prediction confidence score, and class probability. The confidence score reflects the level of confidence in the bounding box containing the object. Bounding boxes with low or zero confidence scores indicate the absence of objects. By applying a threshold, the model can filter out bounding boxes with low confidence, retaining only those that contain objects of interest.

It should be noted that when dealing with larger input image sizes, it may be necessary to divide the input images into smaller tiles (at least 3840×2160 pixels, as utilized in this study) to ensure the accuracy of object detection is not compromised due to the automatic re-scaling to the 608×608 pixel image size before object detection is performed.

IV. Experimental Results and Discussion

In this study, we present the experimental results and a comprehensive performance analysis of all CNN models developed. The evaluation of these models is carried out using a combination of quantitative and qualitative approaches. The testing phase involves a dedicated test image set, which has been reserved exclusively and not utilized in any previous training or validation processes. This test set comprises drone images captured within both the nature areas and the campsites of the DDCR.

The effectiveness of the four object detection models, generated through the training of the four sub-versions of YOLO-V5, is thoroughly examined. Additionally, we compare these models with others created by training widely adopted Deep Neural Networks, specifically SSD (Single Shot Detector) and Faster R-CNN (Faster Region-based Convolutional Neural

Networks). All six networks undergo training, validation, and testing on identical datasets, as outlined in Table 1.

A. Quantitative Performance Comparison

The comparison of quantitative results is presented in Table 2. Our analysis shows that SSD has the fastest training convergent time, but requires significant time for testing (i.e. high deployment time / resources), and its recall, precision, and mAP@0.5IoU values remain substantially lower than those of the four YOLO-V5 models. Faster R-CNN, on the other hand, takes the longest time for the convergence of training but has relatively good testing speeds / deployment costs as compared to the YOLO-V5 based models. However, the accuracy, precision, and mAP@0.5IoU values of the Faster R-CNN model are inferior to those of the four YOLO-V5 models. Comparing the performance of the four YOLO-V5 models, we observed that as the complexity/depth of the architecture increases, more time is required for training and testing, with YOLO-V5l and x taking significantly more time than sub-versions s and m. YOLO-V5l achieved the highest mean average precision (71.5%) in litter detection as compared to other models, as presented in Table 2. Given the above observations, considering the objective performance metrics, we recommend the use of YOLO-V5l for litter detection.

TABLE II
PERFORMANCE COMPARISON OF DNN BASED OBJECT DETECTION MODELS

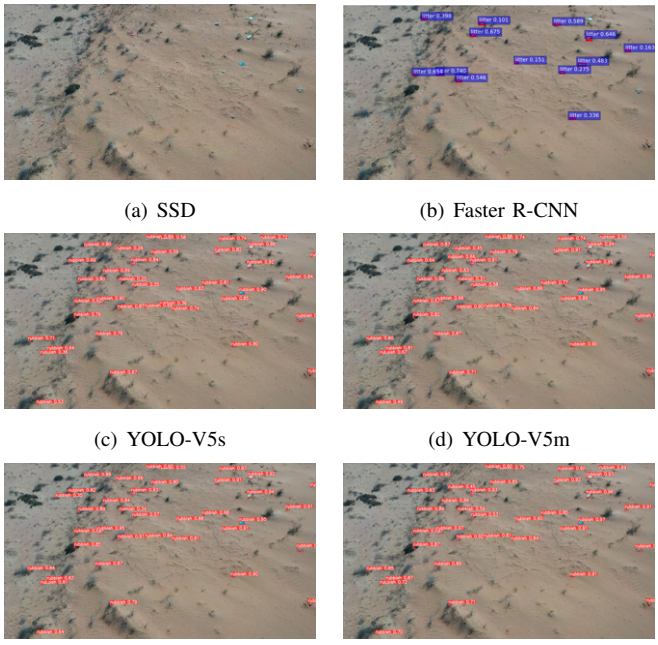
Model	Training Hours	Recall	Precision	mAP
SSD	**10.5**	**0.30**	**0.21**	**15.9%**
Faster-RCNN	**15.6**	**0.28**	**0.14**	**20.2%**
YOLO-V5s	**14.6**	**0.61**	**0.76**	**65.3%**
YOLO-V5m	**14.8**	**0.62**	**0.81**	**69.5%**
YOLO-V5l	**15.1**	**0.65**	**0.76**	**71.5%**
YOLO-V5x	**15.2**	**0.65**	**0.79**	**71.3%**

In summary, based on the objective performance values presented and discussed above, it can be concluded that the object detector models generated from YOLO-V5 and its sub versions are far superior in performance as compared to models generated by other popular CNNs such as SSD and Faster-RCNN. The deeper the architecture, the objective performance improves when comparing the different sub-versions of YOLO-V5.

B. Visual Performance Comparison

As indicated in Table 2, all four versions of YOLO-V5 exhibit the ability to detect various types of litter in the images, achieving precision levels of over 76%. The exceptional precision values, along with strong recall rates, further validate the successful performance of all YOLO-V5 models. This objective assessment aligns consistently with the subjective evaluation presented in Figure 3, which showcases the models' proficient performance.

In contrast, both SSD and Faster R-CNN demonstrate subpar performance. These models exhibit deficiencies in accurately detecting numerous litter objects, as clearly depicted in Figure 3. The lacklustre performance of SSD and Faster R-CNN models is further highlighted by their lower precision and recall values, as evidenced in Table 2.



(a) SSD (b) Faster R-CNN (c) YOLO-V5s (d) YOLO-V5m (e) YOLO-V5l (f) YOLO-V5x

Fig. 3. The visual performance comparison of litter detector models derived from DNN architectures, (a) SSD, (b) Faster-RCNN, (c) YOLO-V5s, (d) YOLO-V5m, (e) YOLO-V5l, and (f) YOLO-V5x

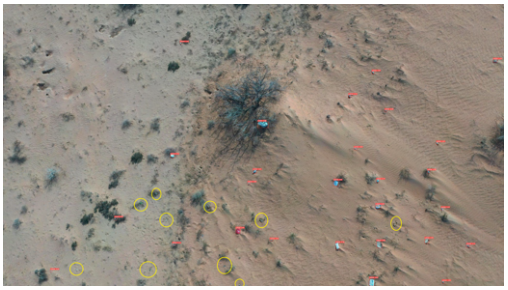
It is important to note that the relative performance of the models remains consistent across all test images utilized in the evaluation. Considering this consistent pattern and the comprehensive results provided in Table 2, future performance comparisons exclude the SSD and Faster R-CNN based models.

Figure 4 presents a visual assessment of litter detection by employing the four sub-versions of YOLO-V5 models. While all four models successfully detect all litter objects, it should be noted that the YOLO-V5s and YOLO-V5m models missed more litters than l and x did.

Additional experimental results, comparing the performance of litter detection models generated using the four sub-versions of YOLO-V5, are presented in Figures 5 and 6. In these figures, the 'yellow circles' represent instances where the models failed to detect litter objects, as compared to human observation. In Figure 6, all of the 4 models detected all the litters in the testing image.

The visual comparison results presented above demonstrate the effectiveness of each of the four sub-versions of YOLO-V5 in litter detection. When the UAV's flight height is relatively low or the litter size is large, all four models exhibit excellent detection performance (as seen in Figure 6). Regardless of the litter subtype (e.g., bottles, paper, bags, boxes, etc.), litter objects are consistently detected, particularly when common objects used in training (e.g., plastic bottles) and well-defined objects with distinct shapes are involved, especially during

(a) YOLO-V5s

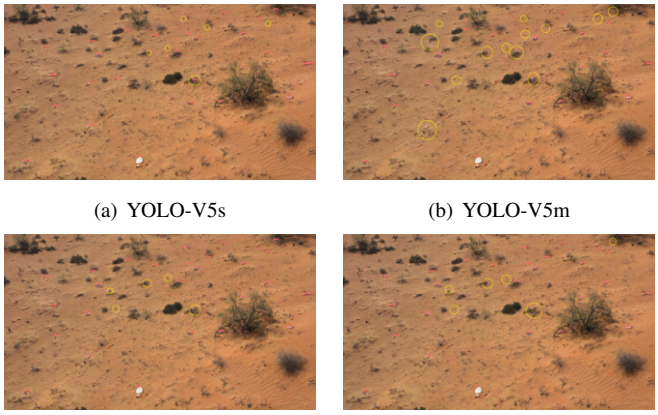


(b) YOLO-V5m

(c) YOLO-V5l

(d) YOLO-V5x

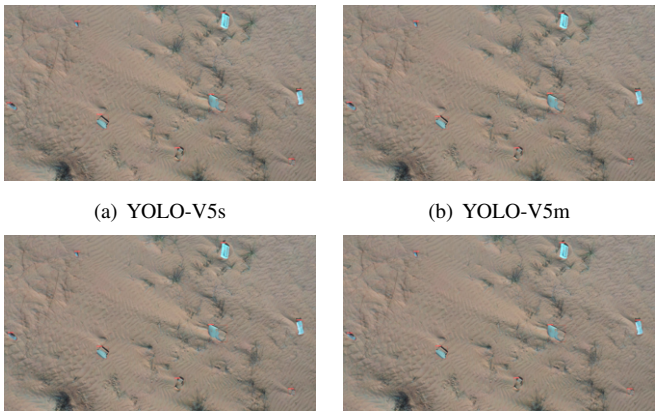
Fig. 4. The results of litter detection in drone imagery using the YOLO-V5s, YOLO-V5m, YOLO-V5l, and YOLO-V5x



(a) YOLO-V5s (b) YOLO-V5m

(c) YOLO-V5l (d) YOLO-V5x

Fig. 5. The results of litter detection in drone imagery using the YOLO-V5s, YOLO-V5m, YOLO-V5l, and YOLO-V5x



(a) YOLO-V5s (b) YOLO-V5m

(c) YOLO-V5l (d) YOLO-V5x

Fig. 6. The results of litter detection in drone imagery using the YOLO-V5s, YOLO-V5m, YOLO-V5l, and YOLO-V5x

high-altitude drone flights.

Figure 7 showcases an image example that highlights the impact of employing the YOLO-V5x-based litter detection model. The image, captured from a drone at a significant altitude, contains several very small litter objects on the top right side, which are not detected due to their size. The YOLO-V5x model outperforms in detecting the smallest litter objects, and our analysis suggests that further improvements can be achieved by augmenting the training data with a larger number of very small litter objects. However, it is worth noting that shallower YOLO-V5 models, such as 's', may result in false positives, wherein very small non-litter objects are detected as litter. To accurately differentiate between very small litter and non-litter objects, the model's depth should be sufficient to perform detailed feature analysis. Additionally, in tested example images, YOLO-V5l demonstrates slightly superior performance compared to YOLO-V5x. This can be attributed to the need for a larger dataset to train the deeper YOLO-V5x network, enabling more accurate detection.

C. Discussion – Single-Class Litter Detector Performance

According to the review of literature we conducted, this is the first attempt carried out in literature to identify litter in drone captured footage using the latest advancements in Deep Neural Network architectures. We have shown that single-class litter detection models based on YOLO-V5 sub-versions, result in mean average precision values of above 71.5% and precision values over 76%. This is a promising step toward real-time detection of litter using aerial image data, despite the small sample size of training images employed in the training of the DNN architectures, in the proposed research. Further

Fig. 7. Detecting very small objects of litter using YOLO-V5x based model

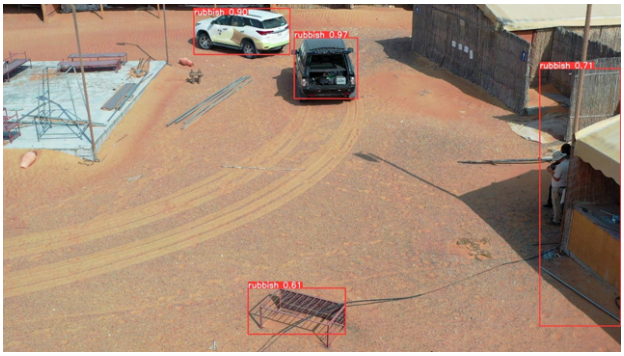


Fig. 8. Test results of single-class litter detection models in desert camp-sites

Fig. 9. Test results of single-class litter detection models in desert camp-sites

investigation of the resulting models' training times indicated higher training times of the models generated from the more complex and deeper networks (i.e., YOLO-V5, x and l sub-versions) were mainly due to the limitations of the processing power of the computational hardware used in this research. The training and testing times could be considerably reduced by using faster computer hardware.

In the experiments conducted we only investigated a single-class litter detector in which all objects of litter, e.g., regardless of whether they are bottles, cans, paper, boxes, or any other common objects of litter typically left behind after human consumption, was labelled as a single type of object, 'litter'. Our detailed investigations revealed that it is important still to have a sub-class balance, i.e., similar number of different types of litter objects being used in training, despite the fact that all litter types are classed as one type, for testing accuracy for all sub-types to be similar. For example, in our training data the least amount of litter was 'drink cans' and such objects had the highest chances of being missed or mis-classified. Therefore, within the training process, we attempted to balance the amount of different sub-types of litter as the available sample data on some sub-types of litter (such cans) was relatively scarce, e.g., drink cans. Our investigations revealed that the models' performance will be better if we collected sufficient data for all sub-types of litter, e.g., 2000 samples of each sub-type.

As recommendations for the future improvement of performance of models, in particular that of the models created from the deeper DNN architectures such as YOLO-V5 x and l, it is recommended that more UAV data to be captured at very different altitudes of drone flights, i.e., capturing at altitudes resulting in visibly distinct litter types in addition to capturing images at altitudes where the litter types are indistinguishable. Such data, when used in training results in better generalization of models in litter detection. Despite the limitations of data available for training, we have demonstrated the superior capability of the models based on the most popular DNN architectures in litter detection.

In the investigation conducted in this section and the results presented, the use of training and test data was limited to desert regions of natural habitats, where litter had been left behind by visitors to the DDCR desert conservation areas. The scenes/images rarely consisted of any human-made or non-natural objects. This led to our curiosity to determine the potential capability of the developed models to detect litter in areas that is usually occupied by humans, such as camp sites. In this section we present the results of our investigation and a resulting alteration to our litter detection approach to enable more accurate detection of litter in such regions.

We only use the best performing litter detection based on the YOLO-V5L sub-version (71.5% mAP value) for single-class litter detection in camp sites of the DDCR desert regions. Representative sample of results that are illustrated in Figures 8 and 9, clearly show that the single-class litter detection model detects many human-made objects (such as cars and freezers), as litter. Not all human-made objects are detected as litter. Most false positives are of a distinct colour (not relevant to typical colours present in natural desert regions) or shape (e.g., those with straight line edges). Although one could argue that the scene in Figure 8 has some objects that can be defined as litter, the scene in image 10 does not have any objects that can be defined as litter. It is noted that human-made objects that is left around in a non-organized manner, in a locality that they are no generally expected, could be contextually defined as litter, typically. An attempt to detect individual objects of litter, without considering the context in which they appear within a scene, has limitations. However, our current research efforts are limited to identifying

litter based on single object detection. The fact that 'litter' objects are not natural objects and are human-made, make the challenge further complex. Therefore, it is important that we differentiate litter objects that a human would define as litter (based on a visual context analysis) from man-made, but non-litter objects (human visual judgment based on context). It is noted that both object types are human-made and the single class litter detection approach we adopted in section 4 will therefore fail to differentiate the two types of objects. Further it is noted that as we trained our single-class litter detector only on natural images, the background of labelled litter objects used in training and validation, only consists of natural objects. Therefore, when we apply the resulting object detectors on campsites any non-natural or man-made objects, which we might not define as litter, is still more likely to be classified as litter, than being classified as a part of the scene background. This is a further reason behind the failure of the single-class litter detector, when applied in camp-site images. Nevertheless, we demonstrated that the single-class approach to litter detection is an ideal and simple solution to litter detection in nature reserves such as most of the land area managed/conserved by the DDCR.

V. Conclusion

In this study, we suggest employing the most recent YOLO-Version to find litter in UAV photographs taken over the United Arab Emirates (YOLO-V5). All YOLO-V5 networks have a detection precision of over 76%, according to the findings of the experiments described in this article. The performance of the YOLO-V5l in terms of litter detection is the best, with a rate of over 71.5%(mAP@0.5IoU). When images are blurry, overlapping, or obscured by different backgrounds, YOLO-V5x outperforms the other YOLO-V5 networks at spotting many litters. Over 5000 litter samples were used during training, and over 913 photographs total were used in this study. The detection performance will enhance even more if this number is raised to 10,000 or higher since the models will be able to generalise previously unobserved data. Furthermore, this approach offers new inspiration and challenges for object classification with strong subjectivity, and has the potential to enhance the accuracy and efficiency of object detection in a variety of real-world applications. The results of this study, despite the limits of the data set, open the door for the real-time detection of litter via aerial surveillance, aiding in the preservation of the environment and desert in the UAE.

Acknowledgment

This grant was funded by a RIF grant, code - R19097, awarded by Zayed University, UAE. The authors would like to thank the DDCR, Dubai, UAE for permission to capture data within their nature reserve areas.

References

[1] G. Sharma, "A review on the studies on faunal diversity, status, threats and conservation of thar desert or great indian desert ecosystem," in *Biological Forum–An International Journal*, vol. 5, no. 2. Citeseer, 2013, pp. 81–90.

[2] H. Yang, X. Zhang, F. Zhao, J. Wang, P. Shi, and L. Liu, "Mapping sand-dust storm risk of the world," *World Atlas of Natural Disaster Risk*, pp. 115–150, 2015.

[3] D. Gallacher and J. Hill, "Status of prosopis cineraria (ghaf) tree clusters in the dubai desert conservation reserve," *Tribulus*, vol. 15, no. 2, 2005.

[4] S. Majchrowska, A. Mikołajczyk, M. Ferlin, Z. Klawikowska, M. A. Plantykow, A. Kwasigroch, and K. Majek, "Deep learning-based waste detection in natural and urban environments," *Waste Management*, vol. 138, pp. 274–284, 2022.

[5] M. Wolf, K. van den Berg, S. P. Garaba, N. Gnann, K. Sattler, F. Stahl, and O. Zielinski, "Machine learning for aquatic plastic litter detection, classification and quantification (aplastic-q)," *Environmental Research Letters*, vol. 15, no. 11, p. 114042, 2020.

[6] M. Córdova, A. Pinto, C. C. Hellevik, S. A.-A. Alaliyat, I. A. Hameed, H. Pedrini, and R. d. S. Torres, "Litter detection with deep learning: A comparative study," *Sensors*, vol. 22, no. 2, p. 548, 2022.

[7] V. Kakani, V. H. Nguyen, B. P. Kumar, H. Kim, and V. R. Pasupuleti, "A critical review on computer vision and artificial intelligence in food industry," *Journal of Agriculture and Food Research*, vol. 2, p. 100033, 2020.

[8] K. Faust, S. Bala, R. Van Ommeren, A. Portante, R. Al Qawahmed, U. Djuric, and P. Diamandis, "Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning," *Nature Machine Intelligence*, vol. 1, no. 7, pp. 316–321, 2019.

[9] P. Bharati and A. Pramanik, "Deep learning techniques—r-cnn to mask r-cnn: a survey," *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, pp. 657–668, 2020.

[10] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[14] T. Jintasuttisak, E. Edirisinghe, and A. Elbattay, "Deep neural network based date palm tree detection in drone imagery," *Computers and Electronics in Agriculture*, vol. 192, p. 106560, 2022.

[15] S. Bilik, L. Kratochvila, A. Ligocki, O. Bostik, T. Zemcik, M. Hybl, K. Horak, and L. Zalud, "Visual diagnosis of the varroa destructor parasitic mite in honeybees using object detector techniques," *Sensors*, vol. 21, no. 8, p. 2764, 2021.

[16] Y. Fang, X. Guo, K. Chen, Z. Zhou, and Q. Ye, "Accurate and automated detection of surface knots on sawn timbers using yolo-v5 model." *BioResources*, vol. 16, no. 3, 2021.

[17] L. Wang and W. Q. Yan, "Tree leaves detection based on deep learning," in *Geometry and Vision: First International Symposium, ISGV 2021, Auckland, New Zealand, January 28-29, 2021, Revised Selected Papers 1*. Springer, 2021, pp. 26–38.

[18] J. Zhao, X. Zhang, J. Yan, X. Qiu, X. Yao, Y. Tian, Y. Zhu, and W. Cao, "A wheat spike detection method in uav images based on improved yolov5," *Remote Sensing*, vol. 13, no. 16, p. 3095, 2021.

[19] A. Yakovlev and O. Lisovychenko, "An approach for image annotation automatization for artificial intelligence models learning," , vol. 1, no. 36, pp. 32–40, 2020.