Article

Pest animal's Detection, and Habitat Identification in Low-resolution Airborne Thermal Imagery

Anwaar Ulhaq [1,†,‡,*], **Peter Adam** [2,‡] **Tarnya Cox** [3,‡], **Asim Khan** [4,‡], **Tom Low** [5,‡] **and Manoranjan Paul** [1]

[1] School of Computing and Mathematics, Charles Sturt University, NSW, Australia;
[2] Department of Primary Industries and Regional Development, WA, Australia; peter.adams@dpird.wa.gov.au
[3] Department of Primary Industries , NSW, Australia; tarnya.cox@dpi.nsw.gov.au
[4] Victoria University, Melbourne, VIC, Australia; asim.khan@vu.edu.au
[5] Tomcat Technologies, Australia; tom@kargow.com
* Correspondence: aulhaq@csu.edu.a

Version September 21, 2020 submitted to Preprints

Abstract: Invasive species are significant threats to global agriculture and food security being the major causes of crop loss. An operative biosecurity policy requires full automation of detection and habitat identification of the potential pests and pathogens. Unmanned Aerial Vehicles (UAVs) mounted thermal imaging cameras can observe and detect pest animals and their habitats, and estimate their population size around the clock. However, their effectiveness becomes limited due to manual detection of cryptic species in hours of captured flight videos, failure in habitat disclosure and the requirement of expensive high-resolution cameras. Therefore, the cost and efficiency trade-off often restricts the use of these systems. In this paper, we present an invasive animal species detection system that uses cost-effectiveness of consumer-level cameras while harnessing the power of transfer learning and an optimised small object detection algorithm. Our proposed optimised object detection algorithm named Optimised YOLO (OYOLO) enhances YOLO (You Only Look Once) [27] by improving its training and structure for remote detection of elusive targets. Our system, trained on the massive data collected from New South Wales and Western Australia, can detect invasive species (rabbits, Kangaroos and pigs) in real-time with a higher probability of detection (85–100 %), compared to the manual detection. This work will enhance the visual analysis of pest species while performing well on low, medium and high-resolution thermal imagery, and equally accessible to all stakeholders and end-users in Australia via a public cloud.

Keywords: invasive species; thermal imaging; habitat identification; deep learning

1. Introduction

A pest animal is defined as any animal that has or has the potential to have an adverse economic, environmental or social/cultural impact [8]. Pest animals have adverse effects on Australian agri-ecosystem as they cause significant crop damage, competing with native species for pasture or causing soil erosion, and acting as reservoirs for diseases. Invasive population expansion of European rabbit (Oryctolagus cuniculus) and Feral pig (Sus scrofa) are among the more pervasive invasions to crops and agricultural lands. The kangaroo (Macropodidae) is a symbol of Australia and not considered a pest [13]. However, they are involved in more than eighty per cent of the 20,000-plus vehicle-animal collisions reported each year [14]. Rabbit-proof fencing, ground and helicopter culling of pigs and Kangaroos, pest trapping and poisoning, are some of the control strategies in place. However, the pest population is rapidly spreading despite substantial investment in control.

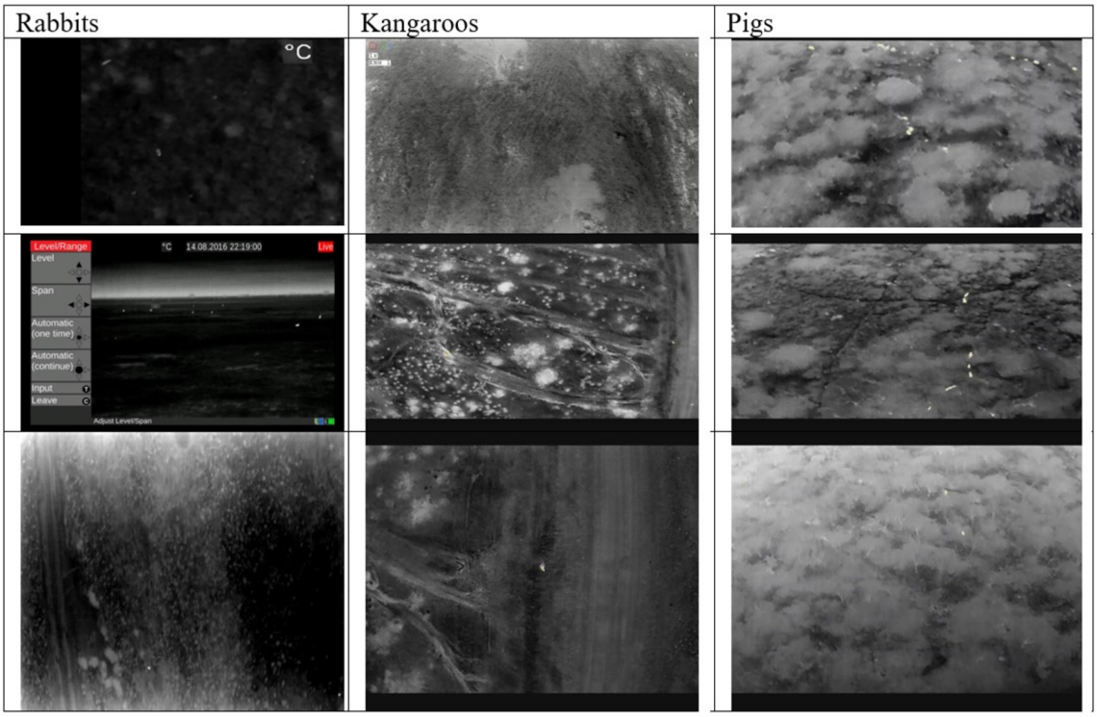


Figure 1. An exemplar shots of dataset shots of invasive animals (rabbits, pigs and Kangaroos) captured in airborne thermal imagery that show visual challenges for object recognition from a distance These videos are captured using helicopter-based surveys of vast farmlands in states of Western Australia (WA) and New South Wales (NSW), Australia.

Recent advancements in drone and imaging technologies have enabled non-invasive monitoring of pest animals [4,10,18,32]. However, manual detection of pest animals, habitat identification and estimation of pest population size is cumbersome as it requires frame by frame analysis of hours of video data. Some automated approaches are proposed in recent years [2,6,19,25,31]. However, they often lack usability due to low accuracy, ineffectiveness against occlusion, limitations of the visible spectrum and low detection speed. It requires a need for an intelligent, real-time, fully automated and around the clock monitoring system that is not only limited to animal detection but also their habitat identification. Thermal imagery can provide crucial information about animal habitats that look more active and warm in thermal heat maps.

However, such approaches work better for large mammals as the drastic changes in the temperature gradient between mammals, and their cold background can distinguish their thermal signatures, facilitating their detection and count. In the pest animal remote sensing scenario, both small and elusive signatures are available that decrease the accuracy of computer vision algorithms. Figure 1 shows some dataset shots of invasive animals (rabbits, pigs and Kangaroos) captured in airborne thermal imagery that show visual challenges for object recognition from a distance. As object size becomes very small, even manual identification and tagging of correct thermal signatures is problematic.

Deep learning has revolutionalised the field of object detection, and various deep object detection approaches exist in the literature. Some of the notable techniques include RCNN [12], Fast-RCNN [11], Faster-RCNN[29], Mask-RCNN [16], FPN [20], SSD [23] and YOLO [27]. RCNN family of detectors are two-stage detectors based on the concept of region proposals requiring considerable processing time and unsuitable for fast and real-time object detection. SSD (single-shot multibox detector) and YOLO (You Only Look Once) are one stage or one-shot detectors. SSD is very slow for detection tasks due to the sliding window approach while YOLO outperforms these approaches in both accuracy and processing time. As YOLO initially is trained on MS COCO dataset [21], its performance suffers if objects are tiny and receptive field is limited. YOLOv3 [28] uses DarkNet-53 for feature extraction and introduces the Feature Pyramid to detect small objects at different scales. FPN predicts small-scale objects in the shallower layers with low semantic information which might not be sufficient to classify small objects.

Similarly, due to striding and pooling, the small-scale objects disappear in the deep convolution layers. Therefore, the removal of pooling and striding can improve YOLO to detect smaller objects. Meanwhile, YOLOv4 [3] presents new findings. However, its scope is to increase the overall speed and accuracy on MS COCO dataset using a different bag of features and bot to increase small object detection in thermal imaging. In this work, we address the above weaknesses by introducing an optimised version of YOLO for small object detection. We named it Optimised YOLO. It enables us to propose a real-time pest animal detection with improved accuracy on imagery captured from consumer-level thermal cameras counted on an unmanned aerial vehicle.

We claim the following contributions in this paper:

- We introduce a real-time pest animal detection with improved accuracy and speed using deep learning-based small object detection approach.
- We optimise traditional YOLO by improved model training and structure optimisation for detecting smaller objects.
- We validate our approach on an extensive thermal video data set collected by the Department of Primary Industries, NSW, Australia. This dataset is very challenging due to low resolution, the small size of pets like rabbits and elusive signatures of similar thermals signatures of pigs and Kangaroos.

We have organised the paper as follow: Section 2 describes the related work. Section 3 provides a detailed description of our methodology. Section 4 illustrates our finding with the help of experimental results. Section 5 presents the discussion and future work directions, followed by concluding remarks and references.

2. Related Work

One of the traditional approaches to animal detection and activity monitoring is the use of camera traps. They have been used to investigate 13 broad areas of wildlife monitoring in Australia over the last twenty-four years [24**?**]: However, the field of view and coverage of camera traps is limited, and it has not proved to be a reliable tool to monitor cryptic pest animals and their activities[33]. An alternative way to airborne monitoring through unmanned aerial vehicles (UAVs) and helicopters [1,4,9,10,18,32].

The recent revolution in the field of deep learning [30] has enabled scientists to automate various vision-based problems. Early use of deep learning for automated animal classification involved sufficient pre-processing and limited recognition accuracy [5,7]. Domain adaptation and transfer learning [19] can increase detection accuracy across different domain and tasks.

Animal detection work can also be categories can object detection problem as there is more interest in object recognition and location detection than in simple classification. Various object detection methodologies can be used that include RCNN [12], Fast-RCNN [11], Faster-RCNN[29], Mask-RCNN [16], FPN [20], SSD [23] and YOLO [27]. However, all these approaches use high-resolution data for training and object scales are generally larger and clear. Therefore, their performance decreases for small animal detection from UAV, especially in low-resolution thermal video sequences resulting in low accuracy, slow detection or overfitting.

Our work is related to YOLO YOLO [27] and its improved versions [3,28]. Some recent work on small object detection from a distance is related to our work. An improved version of YOLO for UAV called UAV-YOLO [22] tried to improve small object detection through YOLO. It has included a few more convolution layers and shortcut connections to improve the model. However, the basic limitations of subsampling remain unaddressed. In this work, we addressed the major weakness of convolution operation and aggressive subsampling and proposed an optimised YOLO.

3. Materials and Methodology

In this paper, we used the Convolutional neural network (CNN)-based object detection method for pest animal detection in thermal imaging. Data collection. In order to perform this study, we first established the Australian pest animal database that was collected by two different teams. One team from the department

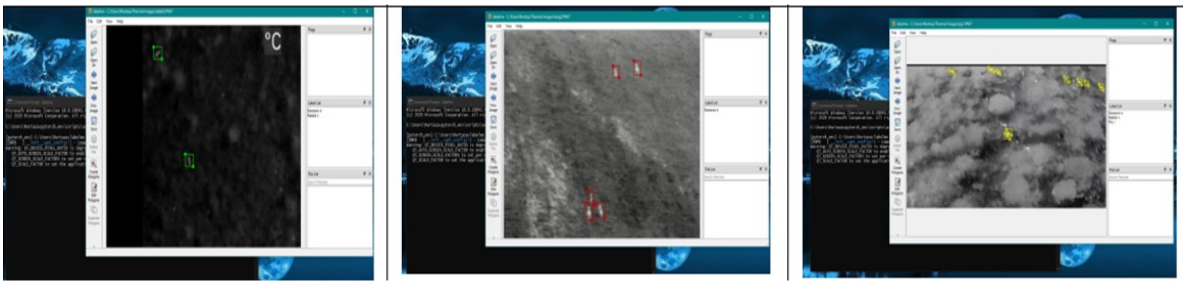


Figure 2. Data annotation samples for different pest species in our pest dataset.

of primary industry, NSW was responsible for the collection of rabbit movement and warren footage using helicopter-based surveys. The other team at Department o primary industry and regional development, Western Australia was responsible for data collection related to wild pigs and Kangaroos using drones.

3.1. Data collection and Annotation

In order to perform this study, we first established the Australian pest animal database that was collected by two different teams. One team from the department of primary industry, NSW was responsible for the collection of rabbit movement and warren footage using helicopter-based surveys. The other team at Department o primary industry and regional development, Western Australia was responsible for data collection related to wild pigs and Kangaroos using drones.

From the thermal footage, we extracted frames to prepare training dataset. As the video framerate 60fps, so thus we had a huge number of frames. However, the majority of frames has no evidence of the presence of any invasive animals; we used only those frames that had confirmed the presence of targetted pest animals.

We manually labelled the dataset. We used python based library Labelme, which is a graphical image annotation tool inspired by http://labelme.csail.mit.edu. We also observed that target objects were very small in some of the frames that had been collected from a high altitude. Similarly, some of the targets were obscure, and even manual classification of their thermal signatures was challenging. We had to magnify such frames/images to label them accurately. Some sample shots of the manual annotation of our thermal dataset are shown in Figure 2.

Table 1. Dataset used for training purpose

Class Name	Labeled	Total Images
Rabbit	Rabbit	1246
Kangaroos	Kangaroo	4211
Pigs	Pig	6000

3.2. Pest Animal Annotation, Model Training and Detection

Our footage library was extensive, so we divided it into three groups of datasets: training dataset, evaluation dataset and testing dataset. We annotated interested target pest animals in our training dataset using Python-based annotation tool. We then trained our proposed OYOLO (Optimised You Look Only Once) model on the trained dataset. The detailed description of OYOLO is provided in coming subsections.

During data collection, our crew took both far and nearer footage of animals by different camera zoom. Therefore, to optimise the performance of YOLOv3 for small object detection, we divided our dataset into two categories named “zoom in”, “zoom-out” groups by taking the distance and receptive field in consideration. We also used data augmentation to balance their sizes. k-means [15] is then used to cluster different numbers of anchor boxes to find the optimised number and size for better results. Finally, the model is retrained by zoom out-category data. On the other hand, the backbone structure of YOLO3 is improved to improve performance.

A brief introduction to YOLO:

YOLOv3 is a more established one-shot detector that is an incremental model of the former YOLO[26] and YOLO9000 [27] and deals object detection as a regression problem.

YOLOv3 backbone known as Darknet 53 includes 53 convolution layers and Resnet [17] short cut connections. In Ithe prediction stage, it uses FPN (Feature Pyramid Network) that uses three scale feature maps, where small feature maps provide semantic information, and large feature maps provide finer-grained information. YOLOv3 uses independent logistic classifiers rather than softmax with binary cross-entropy loss for the class predictions in the training stage. FPN uses three scales of detection with different receptive fields, where the32-fold downsampling is suitable for large objects, the 16-fold for middle size objects, and the 8-fold for small size objects.

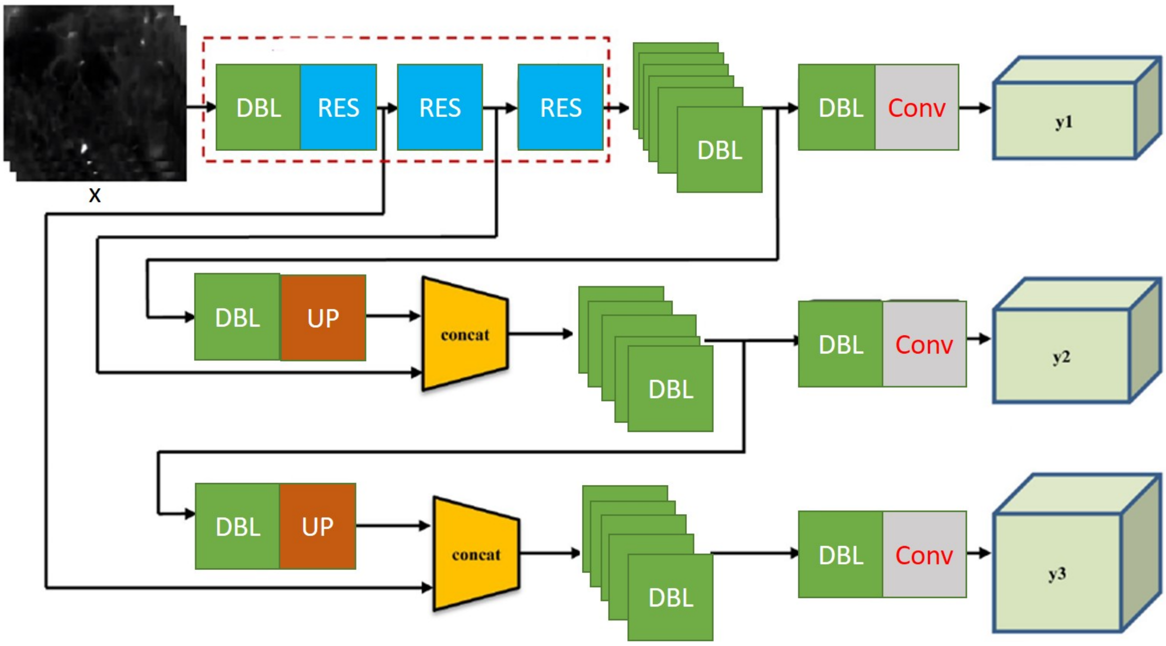


Figure 3. YOLOv3 architecture with input image size 416 x 416 and 3 types of feature map (13 x 13 x 3, 26 x 26 x 3 and 52 x 52 x 3) as output; (B) DBL is Darknet conv2D BN Leaky composed of one convolution layer, one batch normalisation layer and one leaky relu layer.; ResUnit includes two "DBL" structures followed by one "add" layer leads to the residual-like unit, "ResBlock" has several "ResUnit" with one zero-padding layer and "DBL" structure forward generates a residual-like block, "ResBlock." is the module element of Darknet–53.

An architectural diagram of YOLOv3 is shown in Figure 3 that takes an input image size 416 x 416 and 3 types of feature map (13 x 13 x 3, 26 x 26 x 3 and 52 x 52 x 3) as output; (B) DBL is Darknet conv2D BN Leaky composed of one convolution layer, one batch normalisation layer and one leaky relu layer.; ResUnit includes two "DBL" structures followed by one "add" layer leads to the residual-like unit, "ResBlock" has several "ResUnit" with one zero-padding layer and "DBL" structure forward generates a residual-like block, "ResBlock." is the module element of Darknet 53. This architecture is shown in Figure 2.

OYOLO: The optimised YOLO:

One of the problems with traditional CNN networks is their inability to handle low resolution and receptive field ar both pooling and striding may cause loss of small targets. The semantic information about the small objects will vanish or weaken with a decreased spatial resolution of feature maps in subsequent layers. Low semantic information nay is not enough to recognise the small object category in thermal images.

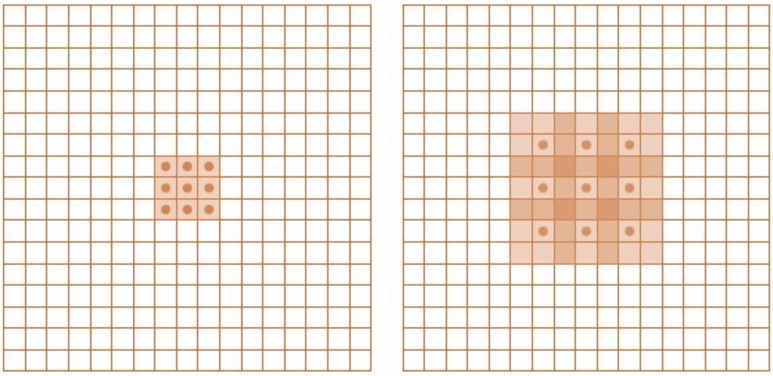


Figure 4. Dilated Convolutions: Figure 3(A) on left shows F2 that is generated from F1 by a 1-dilated convolution; each element in F2 has a receptive field of 3×3. Figure 3(B) on right shows F3 that is generated from F2 by a 2-dilated convolution; each element in F3 has a receptive field of 7×7.

147 A region of the input on which a pixel value in the output depends on is called the receptive field. CNN's 148 progressive reduce resolution and removing subsampling can help, but it reduces the receptive field. Dilated 149 convolutions [34] can increase the resolution of the output feature maps without harming the receptive field of 150 individual neurons. Dilated convolution is also called as "convolution with a dilated filter" as it is a similar filter 151 that is used for wavelet transformation. This concept is explained in Figure 4.

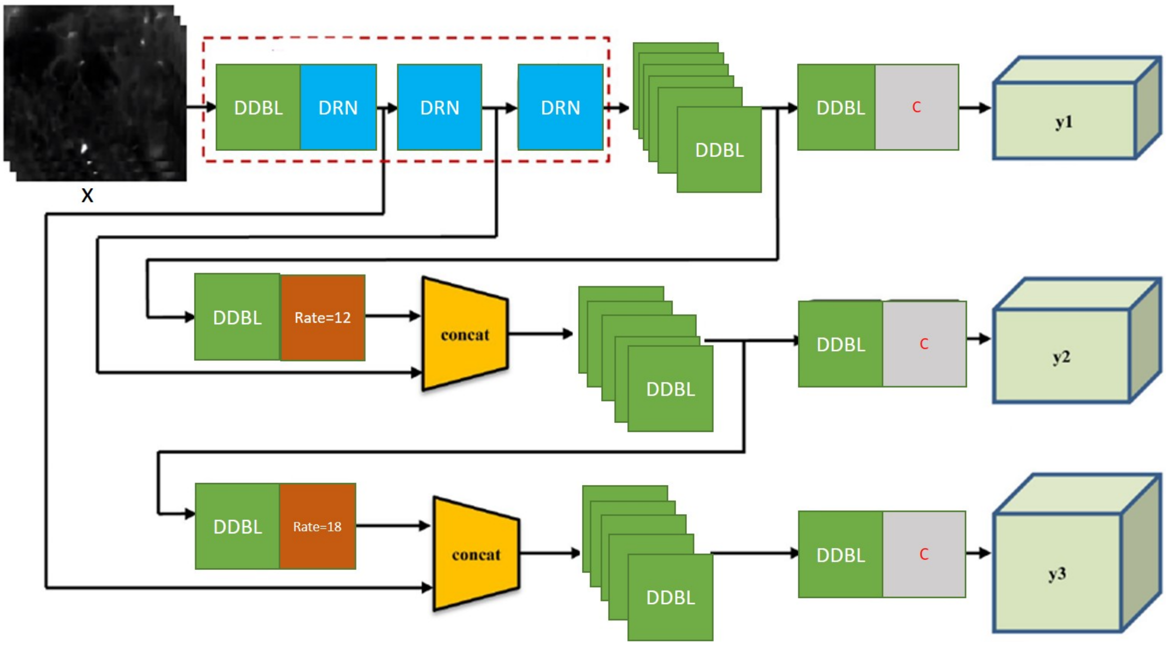


Figure 5. OYOLO architecture with input image size 416x 416 and 3 types of feature map (13 x 13 x 3, 26 x 26 x 3 and 52 x 52 x 3) as output; (B) DDBL is Darknet dilated conv2D BN Leaky composed of one convolution layer, one batch normalisation layer and one leaky relu layer.; DRN (Dilated Residual Network) provides residual like connection with dilated convolutions. Similarly, for multiscale spatial pooling, we use different dilation rates and replace upsampling with dilation filtering.

152 Let $F : Z^2 \rightarrow R$ is a discrete function, $\Phi_n = \lceil -n, n \rceil^2$ and let $f = \Phi_n \rightarrow R$ is another discrete, the convolution 153 operator $*$ can be defined as :

$$(F * f)(x) = \sum_{s+t=x} F(s) f(t) \tag{1}$$

154 Let d be a dilation factor and let $*_d$ be defined as:

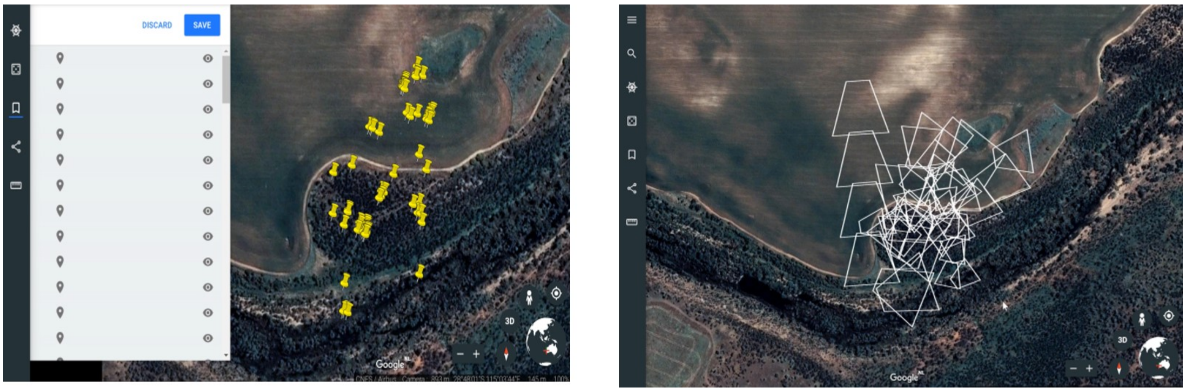


Figure 6. (A) The geo-tagging of detected pest animals from drone data that points out their detection location and (B) visualisation of pest movements that display the area of their activity.

$$(F *_d f)(x) = \sum_{s+*_d t=x} F(s)f(t) \tag{2}$$

where $*_d$ is a dilated convolution or a d-dilated convolution. The tradition CNN convolution is simply the 1-dilated convolution. Dilated convolution supports an exponential expansion of the receptive field without loss of resolution—figure 3 illustrated outcome of dilated convolution. Figure 3(A) shows F2 that is generated from F1 by a 1-dilated convolution; each element in F2 has a receptive field of 3×3. Figure 3(B) shows F3 that is generated from F2 by a 2-dilated convolution; each element in F3 has a receptive field of 7×7.

Therefore, to increase the receptive field of YOLO to handle small objects, we integrated dilated convolutions in its architecture. For this purpose, we replaced DDL block with DDDL block that uses dilated convolution followed by batch normalisation and leaky Relu. RES block is replaced with DRN (Dilated Residual Network) [35]. Similarly, for multiscale spatial pooling, we use different dilation rates and replace upsampling with dilation filtering. Finally, semantic information from three scales is concatenated to detect objects and their categories. Optimised YOLO architecture is shown in Figure 5.

3.3. Geo-tagging and Visualizing of Detected Targets:

Finally, geo-tagging of detected pest animals is done by embedding Google maps platform done flight GPS data for locating and visualising targets in real-time. It provides precise tracking of target locations and visualisation of their movement within the surrounding. Such information is key to monitor pest animal movement patterns and valuable insight about their activities. Figure 6 illustrates the process of geo-tagging of detected pest animals from drone data and also provides visualisation of their movements during the time of flight.

3.4. Results

In this section, we describe the description of our image dataset, system parameters, list of experiments and their results. We would also discuss our experimental results and future work directions.

Two different teams collected our thermal image dataset. One team from the department of primary industry, NSW was responsible for the collection of rabbit movement and warren footage using helicopter-based surveys. The other team at Department o primary industry and regional development, Western Australia was responsible for data collection related to wild pigs and Kangaroos using drones.

Platform specifications: One of our team used DJI Matrice 210 drone with DJI Zenmuse XT (thermal) cameras of 640 x 512 resolution, 9 Hz collected footage at 60 fps. This flying time (with each flight 5-6 hours) is all during night footage was collected. Each 12-15 min length file is approximately 250 MB. This survey included rabbits, pigs and Kangaroosand flight area covered farmlands in Western Australia. At the same time, the other team collected footage of rabbits and their warrens using a helicopter-based thermal imaging platform for farmlands in New South Wales, Australia. Our helicopter crew used (DJI M600 and old footage from DJI

S1000 and DJI Inspire 2) Camera is Vayu HD by Sierra Olympic with a flight time of 3-5 hours for drone and, 10-15 for helicopter: The size of the collected video was around 3 TB.

Table 2. System Specifications for Training/Testing

System Hardware / Software (Operating System)	Specifications
RAM	64 GB RAM
CPU	Intel 9th Gen i9 9900K
GPU(s)	2x NVIDIA RTX 2080 Ti 11 GB VRAM
Operating System	Windows 10 Professional and Ubuntu 18.04

The whole dataset was divided into training and validation as 85% and 15% respectively, as shown in the table 7 to get the optimised results and overcome the issue of over-fitting. For our deep model, we carried out the training process, experiments both on Windows and Ubuntu operating systems. We used deep learning framework PyTorch and related Python libraries for system training and testing. Training and testing were performed on both windows and ubuntu operating systems workstation. It had intel ninth gen i9 CPU, i.e. 9900k, 64 GB RAM and Nvidia dual RTX 2080 Ti 11 GB VRAM GPUs. Table 2. shows the system specifications.

Table 3. Data Split for Testing / Training & Accuracy Obtained

Dataset (Train/Test) Split in %	Accuracy [%]				
	10 Epochs	20 Epochs	30 Epochs	40 Epochs	50 Epochs
85 – 15	92.31	95.84	96.86	97.39	**98.38**

We first tried to establish the baseline by training a YOLOv3 based detection, For this purpose, we used the size of input frames as an integer multiple of 32 (416 x 416), with a total of 5 steps for downsampling operation leading to the largest stride size of 32. As this version used multi-scale analysis, y1,y2 and y3 lead to three different sizes of feature maps. Information for detection of final bounding boxes comes from the combination of all three scales. Please see Figure 4 for details. For training, we fine-tune a pre-trained YOLOv3 model, with a mini-batch size of 32, 10,500 batches, and subdivisions of 15 on 1 GPU, a momentum of 0.8 and a weight decay of 0.0004. We adopt the multistep learning rate with a base learning rate of 0.0001, and the learning rate scales of [0.1,0.1].

We then designed OYOLO by replacing convolutions by dilated version, For this purpose, we used the size of input frames as an integer multiple of 32 (416 x 416), without downsampling operation and introduced dilation rates of 6, 12 and 18 at different levels. The rest of the design remains the same. Information for detection of final bounding boxes comes from the combination of all three scales. However, the original model size remains the same as of YOLOv3. We used similar training specifications for our baseline model. Please see Figure 6 for details.

Training and validation loss and accuracy was calculated for our training and validation set—figure 7 displays our training and validation loss. In Figure 7 (a) (b), the accuracy and loss for both training and testing/validating are presented for each epoch. These graphs were generated for the data split of 85% – 15%. The accuracy graph visually shows that accuracy for both training and testing increases gradually and then tends to converge on a specific point. It also shows that after 40 epochs, the change in accuracy reduces as the validation accuracy appears to be equivalent to training accuracy. Similarly, the right graph shows how the loss starts decreasing gradually as the model learns on a given dataset. The loss of validation data becomes stable after 43 epochs and thus tends towards a specific value.

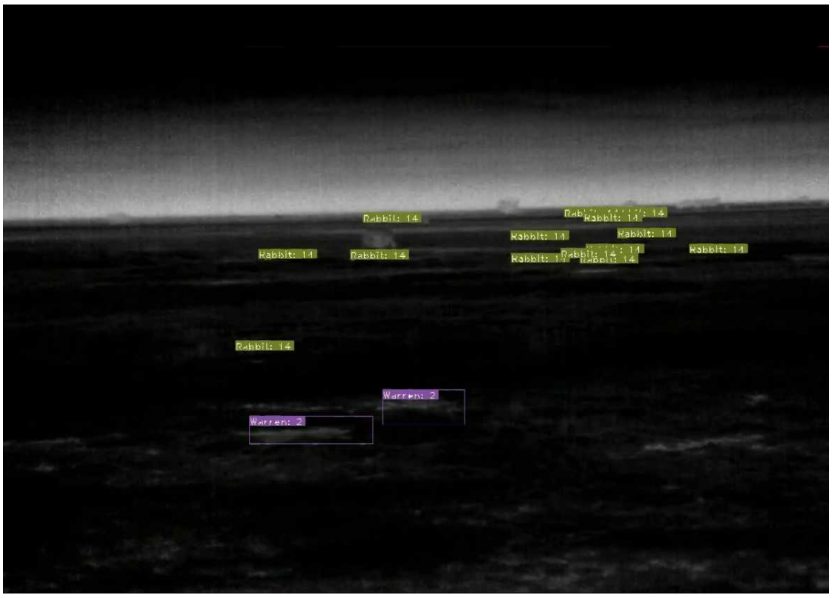
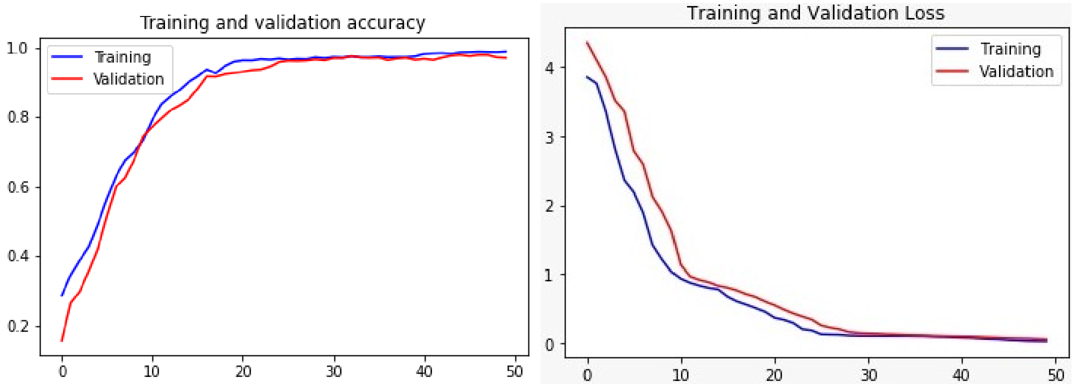


Figure 8. Identification of rabbits and their warren is shown with their respective labels found by our model. Yellow colour labels belong to rabbits, while purple colour labels are their warrens.



(a) Training and Validation Accuracy

(b) Training and Validation Loss

Figure 7. Training and Validation Plots

We tested our approach on the data that was not part of our training or validation set. We first detected all bounding boxes and used them for counting the number of detected pest animals. To remove double counting, we sustained our count till 10th frame. This value was found empirically based on manual inspection of frames and detected animals. For verification purposes, we counted ground truth detections and compared with the automated population count of pest animals. This process also verified our detection results accuracy. Some of the sample detection results are shown in Figure 10. Detected labels and their sizes are intentionally made small to show small bounding boxes. Similarly, we also trained our model for automated identification of pst habitats like rabbit warrens as such places are generally visible in thermal images due to enhanced animal activity in these regions and their underground presence. Figure 8 Illustrates the detection of rabbits and their warrens.

During the testing phase, we achieved an average accuracy of 98.33% for OYOLO compared to 92.33% accuracy for baseline YOLO model. For Pig class (accuracy = 97.34%, recall = 96.89%, precision = 96.37% and f1-score = 96.35%) , Rabbit class (accuracy = 98.17%, recall = 96.70%, precision = 96.48% and f1-score = 97.48%) and Kangaroo class (accuracy = 99.48%, recall = 96.96%, precision = 97.30% and f1-score = 98.60%). Figure9 visualize the above results. For warren detection, we achieved (accuracy = 93.34%, recall = 96.89%, precision = 96.37% and f1-score = 96.3%).

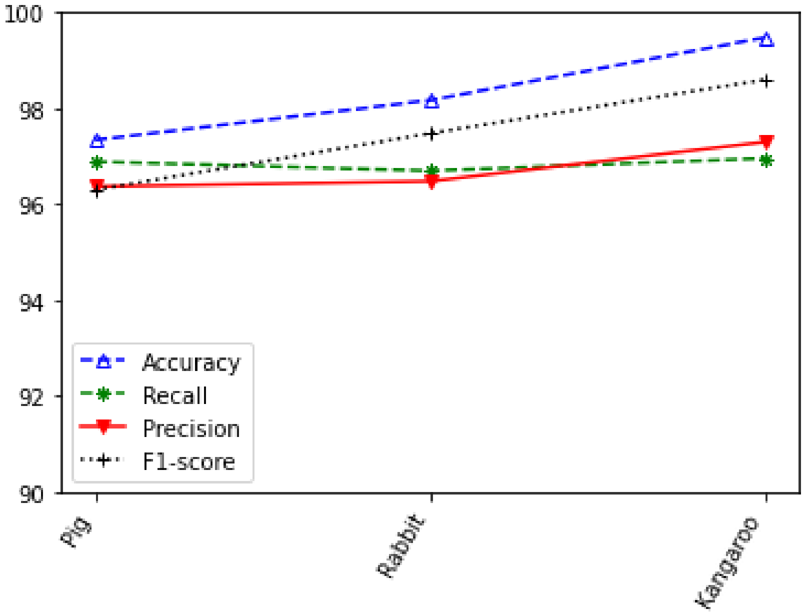


Figure 9. System performance metrics for our proposed system; different metrics are shown for each class of pest animals.

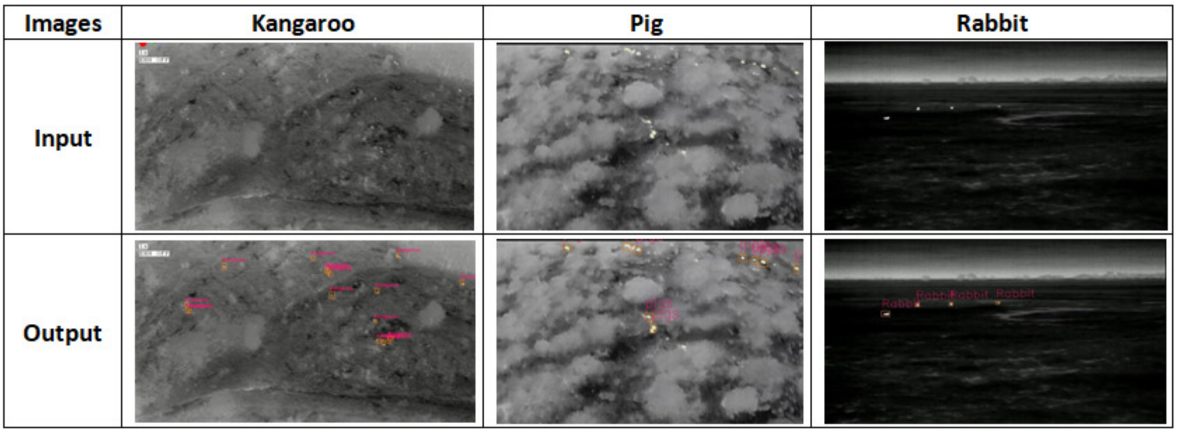


Figure 10. Sample Results: the first row includes input images while the second row shows respective output images.Both bounding boxes and labels are shown.

Our system enabled real-time detection of small pest animals in low-resolution video sequences. However, there are still some weaknesses in our system that we intend to improve in our future work. Our current model has only trained fr three classes of invasive pest animals including pigs, rabbits and Kangaroos. We want to extend it to include several species of pest animals in our future work. Another aspect that needs improvement is the removal of double counts as in some instances; the same animal is being counted twice. As accurate count is not claimed in this paper, we intend to develop some robust strategy to manage this problem in our future work.

4. Conclusion

In this paper, we proposed a robust and real-time detection system for identification of the potential pests and their and habitat from aerial thermal imaging data. Our dataset had several challenges as the size of target animals was not only cryptic and small, but the resolution of our cameras was also low. The aim of the project was to develop a robust system for identification of pest animals in consumer-level cameras. For this purpose, we optimised the object detection algorithm named YOLO (You Only Look Once) [27] by improving its training and structure for remote detection of elusive small targets. Our system, trained on the massive data collected from

New South Wales and Western Australia, can detect invasive species (rabbits, Kangaroos and pigs) in real-time with a much higher probability of detection compared to the manual detection. This work will facilitate farmers to monitor activities of pest animals in their farmlands.

5. Acknowledgement

This research work is supported by a research grant number 0000102841 by mutually funded by the department of primary industries, New South Wales and department of primary industries and regional development, Western Australia.

References

1. Bajiou Mroczkowska, Daniel. 2018. Development of a system for detection, control and prevention of locust pests using uav platforms. B.S. thesis, Universitat Politècnica de Catalunya.
2. Berg, Amanda, Joakim Johnander, Flavie Durand de Gevigney, Jorgen Ahlberg, and Michael Felsberg. 2019. Semi-automatic annotation of objects in visual-thermal video. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 0–0.
3. Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
4. Burke, Claire, Maisie Rashman, Serge Wich, Andy Symons, Cobus Theron, and Steve Longmore. 2019. Optimizing observing strategies for monitoring animals using drone-mounted thermal infrared cameras. *International Journal of Remote Sensing 40*(2), 439–467.
5. Chen, Guobin, Tony X Han, Zhihai He, Roland Kays, and Tavis Forrester. 2014. Deep convolutional neural network based species recognition for wild animal monitoring. In *2014 IEEE international conference on image processing (ICIP)*, pp. 858–862. IEEE.
6. Corcoran, Evangeline, Simon Denman, Jon Hanger, Bree Wilson, and Grant Hamilton. 2019. Automated detection of koalas using low-level aerial surveillance and machine learning. *Scientific reports 9*(1), 1–9.
7. Falzon, Greg, Christopher Lawson, Ka-Wai Cheung, Karl Vernes, Guy A Ballard, Peter JS Fleming, Alistair S Glen, Heath Milne, Atalya Mather-Zardain, and Paul D Meek. 2020. Classifyme: a field-scouting software for the identification of wildlife in camera trap images. *Animals 10*(1), 58.
8. Fleming, Peter JS, Guy Ballard, Nick CH Reid, and John P Tracey. 2018. Invasive species and their impacts on agri-ecosystems: issues and solutions for restoring ecosystem processes. *The Rangeland Journal 39*(6), 523–535.
9. Gentle, Matthew, Neal Finch, James Speed, and Anthony Pople. 2018. A comparison of unmanned aerial vehicles (drones) and manned helicopters for monitoring macropod populations. *Wildlife Research 45*(7), 586–594.
10. Georgieva, M, G Georgiev, P Mirchev, E Filipova, et al. 2019. Monitoring on appearance and spread of harmful invasive pathogens and pests in belasitsa mountain. In *X International Agriculture Symposium, Agrosym 2019, Jahorina, Bosnia and Herzegovina, 3-6 October 2019. Proceedings*, pp. 1887–1892. University of East Sarajevo, Faculty of Agriculture.
11. Girshick, Ross. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.
12. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.
13. GORDON, C. 2001. Sustainable use and pest control in conservation: kangaroos as a case study. *Conservation of exploited species 6*, 403.
14. Green-Barber, Jai M and Julie M Old. 2019. What influences road mortality rates of eastern grey kangaroos in a semi-rural area? *BMC Zoology 4*(1), 1–10.
15. Guo, Wei, Weihong Li, Weiguo Gong, and Jinkai Cui. 2020. Extended feature pyramid network with adaptive scale training strategy and anchors for object detection in aerial images. *Remote Sensing 12*(5), 784.
16. He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.
17. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

doi:10.20944/preprints202009.0480.v2

18. Jepsen, Emma M, André Ganswindt, Celiwe A Ngcamphalala, Amanda R Bourne, Amanda R Ridley, and Andrew E McKechnie. 2019. Non-invasive monitoring of physiological stress in an afrotropical arid-zone passerine bird, the southern pied babbler. *General and Comparative Endocrinology 276*, 60–68.
19. Kellenberger, Benjamin, Diego Marcos, Sylvain Lobry, and Devis Tuia. 2019. Half a percent of labels is enough: Efficient animal detection in uav imagery using deep cnns and active learning. *IEEE Transactions on Geoscience and Remote Sensing 57*(12), 9524–9533.
20. Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125.
21. Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer.
22. Liu, Mingjie, Xianhao Wang, Anjian Zhou, Xiuyuan Fu, Yiwei Ma, and Changhao Piao. 2020. Uav-yolo: Small object detection on unmanned aerial vehicle perspective. *Sensors 20*(8), 2238.
23. Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer.
24. Meek, Paul D, Guy-Anthony Ballard, Karl Vernes, and Peter JS Fleming. 2015. The history of wildlife camera trapping as a survey tool in australia. *Australian Mammalogy 37*(1), 1–12.
25. Meena, Divya and L Agilandeeswari. 2020. Invariant features-based fuzzy inference system for animal detection and recognition using thermal images. *International Journal of Fuzzy Systems*, 1–12.
26. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.
27. Redmon, Joseph and Ali Farhadi. 2017. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271.
28. Redmon, Joseph and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
29. Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99.
30. Sejnowski, Terrence J. 2018. *The deep learning revolution*. Mit Press.
31. Shepley, Andrew Jason, Greg Falzon, Paul Meek, and Paul Kwan. 2020. Location invariant animal recognition using mixed source datasets and deep learning. *bioRxiv*.
32. van Hespen, Rosanna, Cindy E Hauser, Joe Benshemesh, Libby Rumpff, and José J Lahoz Monfort. 2019. Designing a camera trap monitoring program to measure efficacy of invasive predator management. *Wildlife Research 46*(2), 154–164.
33. West, Peter and Glen Saunders. 2003. Pest animal survey 2002. *An analysis of pest animal distribution and abundance across NSW and the ACT. NSW Agriculture, Orange*.
34. Yu, Fisher and Vladlen Koltun. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.
35. Yu, Fisher, Vladlen Koltun, and Thomas Funkhouser. 2017. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 472–480.