Unmanned Aerial Vehicle and Artificial Intelligence for Thermal Target Detection in Search and Rescue Applications

Joseph McGee ¹, Sajith J Mathew ², Felipe Gonzalez ³

Abstract- Recent developments in unmanned aerial vehicles (UAV), artificial intelligence and miniaturized thermal imaging systems represent a new opportunity for search and rescue (SAR) experts to survey relatively large areas. The system presented in this paper includes thermal image acquisition, the video processing pipeline that performs object detection and classification of people in need of SAR in outdoor environments. The system is tested on thermal video data from ground based and test flight footage and is found to be able to detect all the target people located in the surveyed area. The training dataset is a combination of gathered data and internet sourced data. The initial data procurement utilised online academic thermal databases and also the simulation of the UAV mounted camera environment. Ground based data was collected at Kangaroo Point cliffs in Brisbane, Australia, giving an approximate elevation of 26m. Airborne datasets were collected at South Bribie Island in Queensland, Australia, at a range of heights and vegetation density. These datasets where collected at different times of the day, allowing for a range of contrast levels between background and intended target. Once all data was collected, individual frames where extracted from each image and augmentation and annotation was completed. The images were gaussian blurred, lightened and darkened, once all annotation was completed. A total of 2751 original images were annotated, with the augmented dataset comprising of 10380 images. The YOLOV3 algorithm was selected as the neural network (NN) to be used for this experiment throughout training and classification. The ‘Experiencor’ GitHub pipeline was also used throughout the entirety of this project for data output and analysis purposes. The algorithm training was implemented on the Queensland University of Technology (QUT) High Performance Computing (HPC) cluster. Two main models were trained using different hyperparameters for comparison purposes. The first model had a loss of 3.81, AP of 98.6 after ~88 hours of training, with model two having a loss of 4.73, AP of 97.7 after ~184 hours of training.

The comparison shows the importance of the chosen parameters when training detection algorithms of this nature, as minor changes can be the difference between an efficiently trained model and an inefficient failed training attempt. The prediction testing was completed on the test data sets that were not included in the training of the two models. This was done to remove any bias from the system, although it is noted that due to the shared environments of two of the test sets with the training sets, a small amount of bias will exist. Predictions made on the two data sets sharing environmental conditions with the training data, showed good prediction results for both trained models, with very limited false positive and missed detections. The system is flexible in that the user can readily define the types of objects to classify and the object characteristics that should be considered during training.

I. INTRODUCTION

The design and application of UAVs has seen an unprecedented growth over the past few years [1-4]. Recent developments in unmanned aerial vehicles (UAV), artificial intelligence and miniaturised thermal imaging systems represent a new opportunity for SAR experts to survey relatively large areas. The system presented in this paper includes thermal image acquisition as well as a video processing pipeline to perform object detection and classification including people [5]. This paper presents the development of a deep learning algorithm that can be utilised with thermal video and UAV technology. All data was captured using the FLIR Tau 2 640 with TEAX pack and a Phantom 3 Pro. The developed system and algorithm allow for the exploration of the limitations of the detection method, in order to produce recommendations for Search and Rescue (SAR) personnel. These recommendations will allow SAR personnel to make evidence backed decisions, in regard to asset deployment and the dissemination of public information, aiding with SAR scenarios. These recommendations will range from the effectiveness of the detection method in different environments to information to aid in the development of ‘actions on’ procedures for the public [6] [7].
II. LITERATURE REVIEW

The 2017 Search and Rescue Council Meeting highlight the events from Australia in the 2016-17 financial year [8]. The minutes highlight the total number of recorded incidents at approximately 21000, covering all SAR events including maritime and vehicle assistance callouts [8]. This high number of yearly incidents highlight the need for optimal resource management, to allow SAR organisations to operate efficiently. The use of Remotely Piloted Aircraft Systems (RPAS) for SAR purposes still remains in its infancy and the platforms represented are remotely piloted, with no mention of autonomous systems throughout the minutes [8]. The council mentions that RPAS will play an important role in the future in SAR operations, particularly those that may cause a danger to SAR personnel [8]. Although most SAR operations are performed under the maximum RPAS altitude limitations, safety issues are a concern with potential near misses occurring with other aircraft in the area of operation. The Australian Maritime Safety Authority (AMSA) is an active member of CASA’s Unmanned Aircraft Standards Sub-Committee (UASSC) and liaise directly with CASA on regulatory issues pertaining to the operation of UAVs in the SAR environment [8]. This UASSC involvement shows that UAVs are beginning to be seen as a resource and possible asset to SAR organisations, with the implementation and safety concerns being actively discussed.

UAVs are being increasingly utilised in environments that pose dangers to human SAR personnel. In addition to these dangerous environments, UAVs are being utilised in order to increase the total amount of searchable area in a shorter amount of time, compared to traditional search methods [5]. Yunus et al. [5] compared a traditional on foot mountain rescue method called ‘Classical Line Search Technique’ (CLT) and a Drone and Snowmobile Technique (DST), in order to locate a simulated unconscious person. CLT involved five experienced mountain rescue personnel to search in a line, with 10 metre spacing between each member for a mannequin located somewhere on a snow-capped mountain [5].

The DST method utilised a DJI Phantom 3, UAV operator, snow-mobile operator and a rescuer monitoring a live feed from the UAV. The snowmobile began moving once the UAV located the target mannequin. As the DST utilises a vehicle and transportation times are expected to reduce the over search time, it can be seen that the use of a UAV dramatically decreases the search time, while increasing the overall search area. The weather was clear and ideal for UAV operation during the experiment, although Yunus et al. [5] explain that snowy conditions would stop the UAV method from being conducted.

Pólka, Ptak & Kuziora [6] conducted a survey on SAR personnel including members from the state fire service, volunteers, police, railway services and army, in regard to the technical performance they desired in a UAV platform. The outcomes of the survey show that the optimal accuracy required in an identification system would be less than 5 metres, have a flight duration to allow the search of a 10 000 square metre area and be able to operate in environments ranging from negative 20 degrees Celsius to 47.5 degree Celsius. The overall accuracy of the target detection is a compound problem involving target detection and geolocation. Ma et al. [9] explain that the large thermal signature of a human being allows for thermal detection to be a viable method for target identification and tracking purposes. The paper explores a thermal detection method, utilising low resolution imagery in order to track and identify pedestrians. The method used is a two staged ‘blob’ extraction method, where a regional gradient feature and filtering is used to extract the Regions of Interest (ROI) target blobs, followed by the use of a support vector machine (SVM) for classification.

Andrea et al. [10] explore the effectiveness of a particular trained ANN algorithm and the operational conditions of a Phantom 3 UAV [10]. The method utilises a number of positive and negative images which are then mirrored via software methods, in order to increase the size of the total training set by a factor of two [10]. The thermal images are then segmented into binary images and the algorithm is trained on the shape of the human shaped ‘blobs’ caused by the silhouettes of people [10]. This method of target detection was found to be optimal between 15 and 17 metres in height, with a UAV speed of 2 metres per second. The experiment conducted by Andrea et al. [10] also examines a geolocation method for the identified target, utilising the UAVs inbuilt GPS system. This method utilised the focal length and height of the UAV to gain an estimation of the target location, allowing for the calculation of the ground distance between the UAV position and the target. The UAVs speed was then used to estimate the time until the target was directly below the aircraft, which allowed a latitude and longitude to be recorded from the UAV, locating the target [10].

In this work we use Deep learning methods and specifically, YOLOv3 detection method resizes an input image, before using a convolutional NN (CNN) called Darknet-53 on an input image once. Bounding boxes are then generated and class probabilities for each box are produced [11]. The CNN only requires using the image once, creating bounding boxes and predictions simultaneously, making YOLOv3 an extremely fast detection method [11]. YOLOv3 breaks an image up into an S x S grid, before the CNN generates the bounding boxes and class predictions [11]. Each bounding box contains five predictions (x, y, w, h, and confidence), where (x, y) are the coordinates of the bounding box centre in reference to the grid, (w, h) are in reference to the entire input image, and confidence is the intersection over union (IoU) of the predicted box in relation to the ground truth [11] of the predicted box in relation to the ground truth [11].

YOLOv3 utilises independent logistic classifiers for multiclass classification in order to allow for more than one class to lie within the bounding box prediction, e.g. person, woman, etc. [11]. This prediction method overcomes the issue of previous YOLO versions that had difficulty in
identifying small objects in an image, although the prediction performance for medium to large objects has deteriorated compared to the previous versions [11]. This Darknet-53 NN is a combination of Darknet-19 (used in YOLOv2) and a residual network, making it more powerful than Darknet-19 and more efficient than other NNs like ResNet-101 and ResNet-152 [11].

III. MATERIALS AND METHODS

Our approach to generate a predictive detection monitoring system employs multiple stages [12] (Figure 1).

The first is the system development, followed by ground and thermal imagery acquisition and finally, the deep learning process that included data labelling, data augmentation, algorithm development and testing [13].

- Deep learning method selection.
- FLIR TEAX thermal camera selection.
- 3D modelling and printing of aircraft and potential mounting hardware.
- Hardware bench testing.

- Online database imagery.
- Kangaroo Point ground-based data gathering.
- UAV mounting Bribie island data gathering.

- Image extraction and selection.
- Labelling and augmentation.
- YOLOV3 training.
- YOLOV3 testing.

Figure 1. Research approach

A. System Architecture

The system used in this experiment can be divided into airborne and ground segments as presented in Figure 2. The system was developed around the FLIR Tau 2 640 with TEAX pack and a Phantom 3 Pro aerial platform however, other cameras and UAV and UAWS can be used. A mounting bracket was developed for the Phantom using CAD software and a 3D printer. Figure 3 below shows the bracket design and the UAV.

The battery and sensor where mounted on opposing sides of the UAV in an attempt to minimise the alteration of the centre of gravity.
B. Data Collection, Labelling and Augmentation
Initial data was gathered on the 15th of May 2019, this data set was a mix of ground based FLIR images, collected at Kangaroo Point Cliffs, QLD, Australia (Figure 4) in Brisbane, and an assortment of thermal data sourced online from a thermal database, hosted by the Ohio State University [14] (Figure 5).

The third environment that the data was collected was Red Beach, South Bribie Island, QLD, Australia. This location allowed for varying canopy cover throughout multiple environments including beach and forest. This data was collected over two separate days utilizing the UAV and FLIR. The first collection of the Bribie Island data was completed on the 19th August 2019 at 0730, with an ambient temperature of approximately 16 degrees Celsius. The data was gathered utilizing 20, 30 and 40 meters Above Ground Level (AGL), with the thermal target conduction walking, running, crouching and sitting (Figure 6).

The second collection of data was completed on the 3rd September 2019 at 1430, with an ambient temperature of 23 degrees Celsius. The data was gathered using the same technique as the first day, although the hotter environment reduced the contrast between the background hotspots and the target, and multiple people (Figure 7), including a large dog for noise purposes, were present in the data set.

A total of 3400 images were extracted from the data gathered at South Bribie Island. Labelling of these images was completed through the use of the software ‘LabelImg’ [15]. During the labelling process, great care was needed to select the pictures that could be identified as a human being. Due to the large amount of time required to complete such task, care had to be taken in order to select targets in the images that were clearly identifiable as a person. The person labelling the data had to take into account that each image had to be treated as an isolated frame and assumptions could not be made based on previous frames [10]. Figure 8 below shows the labelling process with the dog (red circle), not included. After all un-acceptable images were removed from the dataset, the total un-augmented dataset was comprised of 2751 images.

The data and annotations where then augmented by Gaussian blurring, darkening, lightening and some mirroring. Once the dataset was augmented, a total of 10380 images existed in the training set. Figure 9 below shows the augmented and labelled data.

C. Algorithm training
The deep learning method selected for the object detection is the YOLOV3 algorithm. The ‘Experiencer’ GitHub YOLOV3 repository [16] was utilised for the training of the object detection method. All training was conducted on the QUT HPC cluster via remote access.
The training was split into two different sets, utilising two separate sets of parameters to allow for comparative purposes. Set 1 utilised the following parameters:

- An initial learning rate of 0.0001
- Reduction factor of 0.01
- Loss threshold of 0.01
- Epoch threshold of 5
- Minimum learning rate of 0

Set 2 utilised the following parameters:

- An initial learning rate of 0.01
- Reduction factor of 0.1
- Loss threshold of 0.01
- Epoch threshold of 5
- Minimum learning rate of 0

The training continues with the initial learning rate until the loss had not reduced by the loss threshold over the epoch threshold, then it multiplies the learning rate by the reduction factor. The training script sourced from the repository [16] contained and early stopping function. This function terminates the training early if the loss does not reduce by a set threshold over a set number of epochs. This was modified so that positive reduction of the loss was 0.0001 and the set number of epochs was set to 100. This removed the issue of early stopping from the training. The first row in each table is the initial training information, with each row after, loading the previous trained weights to continue training.

Table II outlines the training of set 1:

### TABLE II. SET 1 TRAINING TABLE

| Image Number | Initial learning rate | Epochs | Loss | mAP (%) | Training Time (HH:MM) | GPU | Mem (GB) | CPUs | Batch Size |
|--------------|-----------------------|--------|------|---------|------------------------|-----|----------|------|-----------|
| 10380        | 0.0001                | 23     | 96.92| 14:19   | 14:19                  | P100| 40       | 12   | 10        |
| 10380        | 0.0001                | 20     | 97.66| 12:25   | 12:25                  | P100| 40       | 12   | 10        |
| 10380        | 0.0001                | 100    | 98.6 | 61:12   | 61:12                  | P100| 40       | 12   | 10        |
| Total        |                       | 143    | 98.6 | 87:56   |                        |     |          |      |           |

Table III outlines the training of set 2:

### TABLE III. SET 2 TRAINING TABLE

| Image Number | Initial learning rate | Epochs | Loss | mAP (%) | Training Time (HH:MM) | GPU | Mem (GB) | CPUs | Batch Size |
|--------------|-----------------------|--------|------|---------|------------------------|-----|----------|------|-----------|
| 10380        | 0.01                  | 50     | 6.676| 30.56   | 30.56                  | P100| 40       | 12   | 10        |
| 10380        | 0.01                  | 250    | 4.73 | 153     | 153                    | P100| 40       | 12   | 10        |
| Total        |                       | 300    | 4.73 | 183:56  |                        |     |          |      |            |

IV. RESULTS

A. Hardware Testing

Table IV outlines the bench testing completed:

### TABLE IV. BENCH TEST OUTCOMES

| Test | Pass | Fail |
|------|------|------|
| FLIR is secured in the bracket without excessive movement | X | |
| FLIR and bracket do not hinder DJI gimble movement | X | |
| FLIR and bracket do not obscure DJI Camera view >50% (Figure 3) | X | |
| FLIR bracket and battery bracket secure firmly to airframe | X | |
| All cables are secure and clear of propellers | X | |
| Propellers are free of obstacles | X | |
| CG does not sit to one side excessively | X | |
| FLIR lanyard attached | X | |

A licensed REPL pilot from Aspect UAV Imaging confirmed that the aircraft and attachments where flight worthy before each flight. The aircraft was placed into a low hover of approximately 8m, where the control of the UAV was manipulated to determine how it reacted while carrying the FLIR and battery.

B. Initial Bench Testing

Table V was used to ensure that all training is being conducted efficiently and without error. The loss and average precision metrics are also examined, in order to gauge the effectiveness of the training and determine if further training is required.

### TABLE V. HPC TEST TABLE

| Test | Pass | Fail | Notes |
|------|------|------|-------|
| Weights file is being updated during training | X | This shows that the loss is reducing, and the algorithm is saving new updated weights |
| Average precision is >90% | X | This is examined at end of training |
Training completes | X | Without error or early stopping
Loss is < 1 | X | More training needed and parameter tuning

Figure 10 produced from the training data, shows the loss versus epochs for both of the training sets. The learning rate was included on the graph to show how the learning rate affects the reduction of loss throughout the training.

Figure 10. Loss graph for set 1

It can be seen in Figure 10, that the initial learning rate of 1e-2 caused a slow reduction in loss for the first ~28 epochs. The reduction in the learning rate to 1e-6 then caused a significant drop in loss until ~42 epochs before levelling out. Reductions in the learning rate here had minor effects on the rest of the training, before the loss reduction had zero effect on the loss at all (~52nd to 100th epoch). As the reduction factor here was 1e-2, the learning rate kept reducing and had no effect on the loss. It appears that for the training of set 1, the most effective range for the learning rate is between 1e-2 and 1e-6.

Set 2 training (Figure 11) has a similar trend, with reduction of the loss being achieved over a longer time frame. The initial learning rate for set 2 was 1e-2, this reduced the loss gradually until ~40 epochs. When the learning rate reduced to 1e-3 at ~40 epochs, the loss reduced quickly until ~55 epochs. The learning rate reduction had little effect on the rest of the training. For set 2, the most effective loss reduction had taken place with a learning rate between 1e-2 and 1e-4.

Figure 11. Loss graph for set 2

C. Prediction Testing
Testing was completed on two separate test videos that were isolated from the training data:
- Bribie Island – 40 meters above ground, extremely poor (vibrations) video quality, late afternoon, with poor contrast between background and target
- Kangaroo Point Cliffs – 26 meters above ground, good video quality, night-time, with excellent contrast between target and ground

Each video will be examined with both trained sets of weights, with the prediction results being printed to CSV file. The CSV data will then be written to each frame of the video and side by side comparisons will be done between sets. Only the significant differences will be reported as there are thousands of frames across all three videos.

V. Prediction Results
Table 5 shows the results from the Kangaroo Point test video. The prediction in this video is extremely high, with a prediction made correctly in the majority of frames. No false predictions were made in this test. This outcome was expected as this video is filmed from the exact same location and time, while containing the same target as the training data. This video was used as a baseline test throughout training to ensure that the prediction function was working as expected.

Set 1 for Bribie Island outperforms set 2 during the prediction. There are clearly more true positive predictions with Set 1. Set 1 also contains a number of false positives and missed predictions, although this is expected due to the high loss. The video shown in the frames above was extremely difficult to watch, due to the vibration. During the output prediction video there are several parts that it is impossible to make out the target with the human eye. The predictions flashing on screen appear to be errors, until the data can be examined frame by frame. Set 1 and set 2 clearly outperform human observation in this test.
### TABLE V. BRIBIE ISLAND RESULTS

| Frame | Set 1 | Set 2 |
|-------|-------|-------|
| 52 | one TP, set 2 – one double TP | |
| 521 | one double prediction, set 2 – one MP | |
| 708 | one MP, set 2 – one TP | |
| 720 | one TP, set 2 – one MP | |
| 1116 | one TP, set 2 – one TP | |
| 1313 | one MP and one FP, set 2 – two MP | |
| 907 | one MP, one TP and one FP, set 2 – two double TP | |

### TABLE V. KANGAROO POINT RESULTS

| Frame | Set 1 | Set 2 |
|-------|-------|-------|
| 134 | one TP, set 2 – one MP | |
| 224 | one TP, set 2 – one MP | |
| 224 | two TP, set 2 – one MP | |
| 304 | two double TP, set 2 – two double TP | |
| 393 | one MP, set 2 – one TP | |
| 425 | one MP and one FP, set 2 – two MP | |
| 907 | one MP, one TP and one FP, set 2 – two double TP | |
VI. CONCLUSIONS AND RECOMMENDATIONS

This paper presented a system and two models for the detection of personnel during a SAR mission. Overall, both sets of models made acceptable predictions on the Bribie Island and Kangaroo point test videos. The first model had a loss of 3.81, AP of 98.6 after ~88 hours of training, with model two having a loss of 4.73, AP of 97.7 after ~184 hours of training. The comparison shows the importance of the chosen parameters when training detection algorithms of this nature, as minor changes can be the difference between an efficiently trained model and an inefficient failed training attempt. The most interesting test outcome was the effectiveness of both models on the Bribie Island test video. Watching this video in real time does not allow for many correct observations to be made from a human observer. Both models are able to make more positive predictions than missed and false positive predictions combined. Some of the frames by themselves are hard for a human observer to find the target, due to all of the hotspots found in the image. This is mainly due to the fact that the algorithm is breaking down the video frame by frame and running the prediction on each individual frame.

Optimal conditions are during the night, or when the background temperature is significantly cooler than the target. As partially obscured targets were specifically used in the training images, the predictions handle the introduction of obstacles quite well. The most difficult part of this process is deciding what should be used as training and what should not be used. During the test video at Bribie Island, the poor contrast between target and background due to heat, causes some false positive predictions on random hotspots throughout the video.

All videos contain what was labelled as a ‘Double Prediction’. The hypothesis behind these predictions is that this is occurring due to the high level of loss and the nature of the training images [10]. There are many images that were included during training that only half of the person was observable due to obstacles. Thus, double predictions were occurring during the prediction phase of the testing. Further examination of the model would be required in order to confirm this hypothesis. In addition to this, throughout the testing tables some of the prediction bounding boxes appear to me smaller than the target. This may be the result of the double predictions, as well as the rounding to two decimal places of the scaled bounding box coordinates for the CSV file.

The following recommendations have been made in order to further the project. These recommendations would allow for the optimisation of the training and project, to achieve the goal of being utilised by SAR teams.

A. Minimum Learning Rate

The training completed for this project utilised a reducing learning rate function, with the minimum being set to zero. As seen in Figure 10 and 11, the learning rate reducing below 1e-6 had little to no impact on the reduction of the loss. Limiting this to 1e-8 and changing the reduction factor from 0.1 to between 0.5 and 0.9, would allow for smaller decrements in the learning rate. This would allow for a more thorough examination of the loss vs epochs across effective learning rate values.

B. Early stopping

The early stopping functionality was removed due the que times on the HPC. This could be re-introduced to the training if further implementation was completed on a local computer. This would allow for the monitoring of effective changes in the learning rate.

C. Data and Training:

Utilising more training data, that includes different environments and angles would increase the environmental robustness of the models. Once tuning to the learning rate has been optimised, more training would allow for the loss to be reduced to less than one.

In addition to more training data, the rotational augmentations could be re-introduced into the dataset. This would require debugging of the current augmentation script in order to ‘catch’ the cases that are causing the anchor errors.

D. Airframe and Payload

The DJI Phantom used for data gathering was operating below optimal levels for the first lot of data gathered at Bribie Island. Future data gathering should utilise a Phantom with specific thermal gimbal in place of the default RGB gimbal and camera. If this is not possible, then another airframe must be utilised to carry the FLIR TEAX. As mentioned above, without a gimbal compensating for the vibration and movement, some of the gathered data can go to waste.

REFERENCES

[1] D. S. Lee, L. F. Gonzalez, K. Srinivas, D. J. Auld, and K. C. Wong, “Aerodynamic/RCS shape optimisation of unmanned aerial vehicles using hierarchical asynchronous parallel evolutionary algorithms”, in 24th AIAA Applied Aerodynamics Conference, Fluid Dynamics and Co-located Conferences, 2006, pp. 1358–1378, doi: 10.2514/6.2006-3331.
[2] J. Sandino, G. Pegg, F. Gonzalez and G. Smith, "Aerial mapping of forests affected by pathogens using UAVs hyperspectral sensors and artificial intelligence", Sensors, vol. 18, no. 4, pp. 944, 2018.
[3] J. J. Mitchell, N. F. Glenn, M. O. Anderson, R. C. Hruska, A. Halford, C. Baun, et al., ”Unmanned aerial vehicle (UAV) hyperspectral remote sensing for dryland vegetation monitoring”, Proc. Hyperspectral Image and Signal Processing, pp. 1-10, 2014.
[4] S. Ward, J. Hensler, B. Alsalam, L.F. Gonzalez. “Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision. In Proceedings of the 2016 IEEE Aerospace Conference, Big Sky, MT, USA, 5–12 March 2016.
[5] Y. Karaca, M. Cicek, O. Tatli, A. Sahin, S. Pasli, M. Beser and S. Turedi, "The potential use of unmanned aircraft systems (drones) in mountain search and rescue operations", The American Journal of Emergency Medicine, vol. 36, no. 4, pp. 583-588, 2018.
[6] M. Półka, S. Ptak and Ł. Kuziora, "The Use of UAV’s for Search and Rescue Operations", Procedia Engineering, vol. 192, pp. 748-752, 2017.

[7] V. Racanell, S. Bartolozzi, E. Guerri and F. Guerri, "Use of remotely piloted aerial systems (R.P.A.S.) for wildlife monitoring", in 17th International Scientific Conference Engineering for Rural Development, Latvia, 2018, pp. 1611-1617.

[8] National Search and Rescue Council, "Forty-first meeting of the Australian National Search and Rescue Council", National Search and Rescue Council, Canberra, 2017.

[9] Y. Ma, X. Wu, G. Yu, Y. Xu and Y. Wang, "Pedestrian Detection and Tracking from Low-Resolution Unmanned Aerial Vehicle Thermal Imagery", Sensors, vol. 16, no. 4, p. 446, 2016.

[10] C. Andrea, J. Byron, P. Jorge, T. Inti and W. Aguilar, "Geolocation and counting of people with aerial thermal imaging for rescue purposes", in Augmented Reality, Virtual Reality, and Computer Graphics - 5th International Conference, Otranto, 2018, pp. 171-182.

[11] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", arXiv.org, 2016. [Online]. Available: https://arxiv.org/abs/1506.02640.

[12] J. Chaki and N. Dey, A beginner's guide to image pre-processing techniques. Kolkata: CRC Press, pp. 4-10,27-28.

[13] S. Khan, H. Rahmani, S. Shah and M. Bennamoun, A guide to convolutional neural networks for computer vision. Morgan & Claypool Publishers, p. 74.

[14] "OTCBVS", Vcipl-okstate.org, 2019. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/.

[15] "tzutalin/labelImg", GitHub, 2019. [Online]. Available: https://github.com/tzutalin/labelImg

[16] "experiencor/keras-yolo3", GitHub, 2019. [Online]. Available: https://github.com/experiencor/keras-yolo