An Investigation Into Using Neural Networks to Detect People for Search and Rescue Operations

Semaa Amin, Mitchell Lawson, Kaitlyn Lowen, John Maunder, Ian Yip

Abstract—For the purposes of this research project, we will be examining the use of thermal imaging, RGB (visual) imaging and composite imaging (RGB and thermal overlaid), in the context of Search and Rescue (SAR) missions utilizing Remotely Piloted Aircraft Systems (RPAS), commonly known as drones. In a search and rescue operation, there is a high probability that the lost or endangered person will need immediate medical attention. As there is an impending urgency to provide care, it is greatly important for the SAR team to safely make the first contact in the timeliest manner possible to ensure the mitigation of potential long-term health impacts. The two most common imaging methods for SAR operations, RGB and thermal, each have limitations and advantages within SAR applications. We assess whether using an RGB, thermal or dual-mode (composite images utilizing both RGB and thermal images with an equal weight factor of 0.5) dataset would serve as the best imaging technique for training a YOLOv4 image classification algorithm. To assess the feasibility and suitability of the image types and the application of the YOLOv4 algorithm in the context of SAR operations, we trained three YOLOv4 models. Each model was trained and tested on an annotated dataset containing either 1803 RGB training images, 1803 training thermal images, or 1803 created composite images. Each dataset has undergone three augmentations that include flips, shear, and rotate. Through this analysis, we conclude the dual-mode (hybrid) image dataset, which utilizes both RGB and thermal imaging, is most effective when training a model to classify a person from an aerial perspective. This is displayed by the superior false positive and F-1 score, while maintaining a false negative record in line with the alternative imaging models. The dual-mode imaging combining both thermal and RGB images is sufficient and feasible in SAR operations, although there may be hardware dependencies that prevent this application on embedded systems and instead may need to use a live video stream to a dedicated computer to perform the image processing. Additionally, we examined four different path-finding algorithms that can be used to direct the RPAS. This analysis concluded that the Quadrant Scanline Fill algorithm was consistently the fastest path-finding algorithm. By utilizing a composite dataset for the image recognition model and applying the Quadrant Scanline Fill drone algorithm, we can ensure a resource-efficient and effective image detection SAR strategy.

Index Terms—Neural Networks, People Detection, Search and Rescue, YOLOv4, drone, Unity simulation, Scanline Fill, Quadrant Scanline Fill, Pathfinding.

1 Introduction

The world of today faces an unrivaled availability of data, technology, innovation, and global cohesion. Innovative new developments and emerging technologies, like drones and machine learning algorithms, prove to be a promising industry and application of said technology in a variety of areas and industries [1].

Remotely Piloted Aircraft Systems (RPAS), commonly known in commercial applications as drones, are essentially any aerial vehicle that do not depend on an on-board human operator for flight and can be either entirely autonomous or remotely operated [2]. Specifically, drones that feature autonomous or semi-autonomous features, enabled through machine learning, are an expanding field of research because of the added benefit they provide to society through the enabling of valuable data retrieval and processing.

In 2006, the Federal Aviation Administration (FAA) officially issued the first commercial drone permit. This marked the beginning of a massive uptake and production of RPAS for civilian and commercial usage. As of May 2021, the FAA reports that the United States has 872,694 registered drones, of which 362,101 registrations account for commercial usage and 507,086 for recreational use [3]. Given the changes in RPAS regulations, the FAA speculates drones could generate more than "$82 billion for the U.S. economy and create more than 100,000 new jobs over the next 10 years" [4]. In 2014, the global expenditure on commercial drones amounted to $700 million, signifying a very large market still waiting to be tapped [2]. This untapped potential serves as a premise to explore how drones can be used to serve society.

Given this market potential and the accelerating uptake and advancement of RPAS algorithms and technology, it is no surprise that industries have begun to utilize drones for in a wide breadth of applications. The adaptable nature of RPAS has led to the application of these machines and algorithms in industries ranging from logistics, research, agriculture, photography and video filming, humanitarian, and medicine [1]. Aerial footage and image recognition technologies can lend themselves to multiple industries, as it allows the operators to access heights, locations, and images that would not be achievable by human traversal.

Compared to human-led search and rescue (SAR), drone SAR offers an alternative to the typical "by-foot" traversal of terrain. RPAS used for SAR are able to survey the area of interest by flight with a camera and utilize image processing techniques to classify humans in a given area. With the ability to widely scan an affected area, the use of RPAS allow for faster retrieval, minimized time-lapsed to first contact, and safer traversal for the SAR team.

We will discuss the usage, benefits, and limitations of three different imaging types when working with RPAS in the realm of SAR operations, specifically in the context of avalanche retrieval. By assessing the three different imaging types, RGB, thermal, and composite, we will offer insight into the ideal imaging dataset for an SAR operation. Additionally, we will examine different path finding algorithms that can be used to efficiently direct the drone. Given that RPAS have limited resources, such as battery life, both analysis combine to offer an overall strategy for a successful and optimal SAR operation.

2 Related Works

Drones usage in the context of SAR has proven to be effective in decreasing the time spent surveying an area and reaching the victim. In Turkey, Y. Karaca et al. [1] analyzed the potential benefit of using drones when conducting a search and rescue mission in the Zigana Mountain region. This study constructed ten blind search and rescue missions, where each location placed a mannequin laying on top of the snow. The experiment calculated the average retrieval time using a Drone-Snowmobile technique (DST) and a Classical Line Search Technique (CLT). The results of these two techniques found that the drone-led search (DST) had a median time to first human contact of 8.9 minutes, and the human-led search (CLT) with a median time of 57.3 minutes [1], showing a 146.22% decrease in time to first contact.

As SAR detection requires image processing and deep learning, one of the largest barriers to overcome in the model training is finding sufficient data sets to account for numerous situations where humans appear very small in the image. To approach this issue, B. Mishra et al. [5] developed a novel data set using six predetermined actions that signalled a need for help. The proposed data set resulted in a rich variety of images ranging in colour, height, people, and backgrounds, allowing it to be an effective proposed data set for model training. This research highlighted the importance of using a dataset that will closely resemble the types of images used in the models' application.

Given that we are exploring the possibility of applying thermal imaging in the context of SAR operations, one of the key components of our research will be comparing the results of the RGB trained model to the thermal trained model. Oliveira and Wehrmeister, conduct a similar comparison outside the scope of SAR and discuss the limiting factors of thermal imaging in their application. Two low-cost cameras, one RGB and one thermal are used to detect pedestrians. Traditional images contain extra data that needs to be processed to decrease the search area, whereas thermal images contain less data needing processing, as people stand out from the ground in thermal imaging [6]. However, this article highlights a potential downside of using thermal imaging, as an object's temperature and an object's environment temperature are key factors in the image capturing process. In thermal imaging, gray-scale imaging differentiates objects based on heat. Therefore, this imagining is not applicable when the climate is as hot, or hotter than the human body. Given that our research scope occurs in primarily cold climates, this limitation will not be a factor.

With thermal imaging, a downside to this imaging technique is the typically higher signal-to-noise ratio(SNR). A higher SNR can result in lower image quality, thus making the processing more difficult [7]. The article [8] discusses the impact of image resolution in SAR operations by looking at the accuracy of detection as well as processing speed. While the article does not directly discuss thermal imaging, our research will be assessing whether thermal images provide a higher level of detection, compared to RGB images, as a result of how subjects are highlighted and contrast their background environment, despite the lower image quality.

Blumenstein et al. [9] discuss how the traditional object detection systems are based on finding potential objects and their bounding boxes, completing feature extraction, and classification using a good classifier. The classifier process of this system has been replaced by more advanced classification algorithms [9]. In our research, we are only interested in classifying a person. Although lower image quality is a barrier to thermal imaging in image recognition, we are only interested in training our model to classify one object class, "person." This single-class classification model leads us to suspect that the accuracy of our bounding boxes will not decrement due to the lower image quality. As humans stand out from the background due to the temperature difference between people and the cold environment, we hypothesize that this gray-scale contrast between objects will counteract the disadvantages to thermal imaging.

Some of the more complex classification algorithms are convolutional-based neural network systems such as R-CNN, Fast R-CNN, Faster-CNN, YOLO and SDD. The majority of these neural networks are very fast but come with their limitations. For example, the article [9] states that using Region of Interests (ROI) and Region of Pooling Networks (RPN) techniques in R-CNN and Faster R-CNN respectively did increase the accuracy and speed of image recognition, however, it required a significant amount of computation power to complete. This trade-off made the overall process slower. Even though the Faster R-CNN was costly in terms of computing power, it was able to be accurate at a frame rate of 7 frames per second.

The researchers address how to overcome the time issue when using a Faster R-CNN by implementing the You Only Look Once (YOLO) algorithm. The YOLO algorithm tries to overcome this computational inefficiency by dividing the image into a grid of cells, where the bounding box and class are based on these individual cells. Lastly, the article discusses how the "YOLO-CNN system" or any CNN system can be tested on image data sets to test the accuracy of MS COCO, VOC2012, and PASCAL VOC systems. In the context of SAR missions, time spent processing is time lost for retrieval. For this reason, we will be implementing the YOLOv4 algorithm.

Trong et al. [10] explored methods in detecting people in sea rescue operations using an algorithm that searches for victims in concentric circles using deep learning algorithms. They built a sea simulation environment using AirSim and that provided a diverse data set for training deep learning algorithms. This paper is informative for building a simulation to test an algorithm.

A barrier to utilizing RGB images in image detection algorithms is when the object of interest is obstructed. Pengfei et al. [11] discusses the difficulties image detection algorithms encounter when detecting busses and other motor vehicles when the object classes were visually obstructed in the data set. Additionally, the algorithms had issues with class imbalances. For this reason, we are implementing the use of thermal imaging to address the object obstruction issue, and will also be detecting a single class, thus solving the problem with multiple classes and imbalances.

Božić-Štulić et al. directly examined how to use an aerial image taken from RPAS for search and rescue application and created a model that can identify a person. Using real-world images, the researchers pre-trained the convolutional neural networks (CNN) to identify a person. Their scope extended beyond the scope of this paper by obtaining over 68,750 images of wilderness acquired from an aerial perspective. An important aspect the paper identified is the ability for both thermal and RGB images can be taken from a drone and used in combination to achieve a better detection rate through the application of a bimodal model [12]. Thermal imaging may not be effective if the environment is hotter than the human body, as it displays heat differentials between objects. Some techniques or algorithms that address this barrier use a sliding window. The sliding window is a simple and effective solution, however, it may not be the most efficient technique when used with CNN models as classifiers. A proposed solution identified is the use of region-proposal-base convolutional neural networks (R-CNNs). These models consist of the components that select the regions that are likely to have the person or object being searched for, use a CNN to extract features, and apply a classifier. We will be exploring the possibility of training our YOLOv4 model with a hybrid image. The images used for training will be a composite image created by overlaying RGB and thermal images with equal weight.

Diving further into thermal imaging research, Rodin et al. [13] detail the task of object classification from thermal images when using CNNs for drone search and rescue missions. An important aspect of using a CNN is that an efficient model can convert 16-bit images into 8-bit images by normalizing the image's colour value from 0 to 255, corresponding from lowest to highest intensity [13]. In their dataset containing 22,000 images, issues like distinguishing the difference between a buoy in the water from a person drowning highlighted the importance of using real-world data, due to its high variability and the possibility of containing similar images and shapes. As we are only classifying one class within our model, the overlap of common shapes between different classes will be less of a barrier.

Bondi et al. [14] continued researching the applications of thermal imaging in the African Savannah. This research tested algorithms in an environment simulated using AirSims and Unreal Engine for wildlife conservation in an African Savannah. They also created a new thermal infrared model simulation that was later integrated as part of AirSim. They used API code expansions to pan the camera to follow the object of interest while capturing images of the object which are then converted to thermal infrared images [14]. Our research with thermal imaging will focus on using real-world thermal images, rather than utilizing a simulated environment.

3 The YOLO Algorithm

3.1 Detection

The algorithm starts by taking an input image and dividing it into sections. These sections represent a grid of $S \times S$ dimensions [15], [16]. Each grid cell has a set of responsibilities and properties. For example, each grid cell is responsible for detecting an object if the object falls into the center of the grid cell. Through a set of calculations each cell predicts B bounding boxes and a confidence score for the boxes. The confidence score reflects how confidant the model is that the corresponding box contains an object and how accurate it thinks the box it predicts is. Confidence is defined as:

$$\Pr(Object) * IoU_{pred}^{truth} \tag{1}$$

If no object exists, the confidence will evaluate to 0 for that cell, else the confidence score will be equal to the intersection over union (IoU).

The bounding boxes used in the YOLO algorithm encompass 5 predictions in total: x,y,w,h and *confidence*. (x, y) represent the center of the box with relation to the bounds of the grid cell. Conversely, the width and height are relative to the entire image [16].

The grid cell is also responsible for predicting the C conditional class probabilities:

$$\Pr(Class_i|Object) \tag{2}$$

The probability of the class calculated with relation to the object within the cell. Combining these equations yields one set of class probabilities per cell in the grid, irrespective of the number of bounding boxes, B [16]:

$$\begin{aligned}\Pr(Class_i|Object) * \Pr(Object) * IoU_{pred}^{truth} = \\ \Pr(Class_i) * IoU_{pred}^{truth}\end{aligned} \tag{3}$$

3.2 Convolutional Neural Networks

In the context of a CNN, convolution is a linear operation that involves the multiplication of a set of weights with the input. The input, in this case, is an image, an input array of data with a filter, a two-dimensional array consisting of weights. The filter is nearly always smaller than the image itself so that it can be applied to different portions of the image and in overlapping regions. The dot product of the input portion of the image and filter yields a single number that is then used to create a feature map of probabilities [15].

3.3 Activation Function

The activation chosen for YOLOv4 is called Mish and is part of a series of improvements in YOLOv4 called a bag of specials (BoS). BoS slightly increase the cost in terms of processing but simultaneously increases the accuracy of the prediction. Mish also avoids the issue of ReLu, a commonly used activation function, where a neuron becomes overfitted and the neuron dies stopping its learning [5].

3.4 Design and Training

A sum of squares error calculation, in addition to two parameters λ_{coord} and λ_{noob}, is used to increase the loss from bounding box predictions and decrease the loss from confidence predictions in the boxes where objects are absent. Redmon et al. [16] used weights of λ_{coord}=5 and λ_{noob}=0.5 to alleviate this and formed the equation below to optimize:

$$\begin{aligned}
&\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\ [(x_i-\widehat{x_i})^2+(y_i-\widehat{y_i})^2]\\
+\lambda_{coord}&\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(\sqrt{w_i}-\sqrt{\widehat{w_i}})^2+(\sqrt{h_i}-\sqrt{\widehat{h_i}})^2]\\
&+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-(C_i))^2 \qquad (4)\\
&+\lambda_{noob}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\ (C_i-\widehat{C_i})^2\\
&+\sum_{i=0}^{S^2}1_{ij}^{obj}\sum_{c\epsilon classes}(p_i(c)-\widehat{p_i}(c))^2
\end{aligned}$$

4 Search and Rescue Algorithms for Unknown Endpoints

4.1 Fill Algorithms

The flood fill algorithm is typically used for optimal path planning when the endpoint is unknown within the defined area [17]. As such, this type of algorithm is best suited for SAR operations. In most other SAR RPAS papers, researchers use a plethora of different algorithm types to find the shortest path to a known target. In SAR situations, two common cases are; the person's location is unknown, or the SAR team has the person's last known location. As our goal is to replicate this scenario in our simulation, we have selected two commonly favoured fill algorithms, a random fill algorithm, and a smaller, quadrant-based iterative algorithm to compare.

4.2 Types of Fill Algorithms

Our research explores four different algorithms using a Unity simulation modelled to reflect a Whistler-based location for a drone SAR operation. The pathway of the simulation starts at a fictional Search and Rescue centre in the town of Whistler. The drone will travel to the last known location of the randomly placed simulated victim and go to the edge of the 5 km radius. The four algorithms examined in this simulation are as follows:

1) Scanline Fill
   - This algorithm traverses the area in the x-axis until reaching the edge. It then translates by 1 unit in the y-axis and traverses to the opposite side and iterates until the victim is found or the area has been fully searched.
2) Quadrant Scanline Fill
   - This algorithm iterates in a similar manner as the Scanline Fill, but divides the area into four quadrants and traverses each quadrant in the Scanline Fill algorithm before moving forward to another quadrant.
3) Geocentric Fill
   - This algorithm traverses to the outer radius of the area and searches in a geocentric motion. After the completion of each circular path, the radius decreases by a set variable unit to a smaller area and iterates until victim is found or the area has been fully searched.
4) Random Fill
   - This algorithm creates random points within the area for the drone to search through until victim is found or the area has been fully searched .

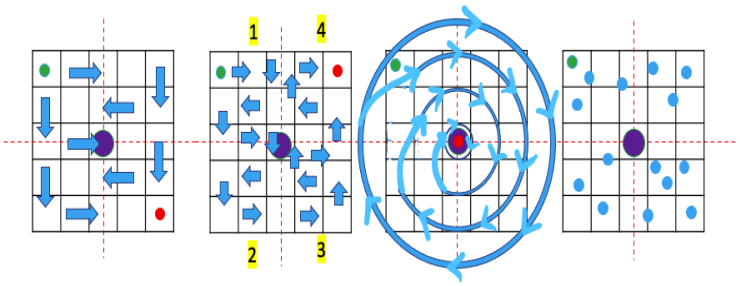

Fig. 1. Proposed Search algorithm patterns to be used in a search and rescue simulation using an environment created through Unity. From left to right, the patterns are Scanline Fill, Quadrant Scanline Fill, Geocentric Fill and Random Fill. The green circles indicate proposed start position of drone, the red circles indicate the projected end position of the algorithm should a victim not be discovered and the purple circles indicate the last known location of the victim. Note that the Random Fill algorithm does not have a defined start or end position.

5 Our Focus

For the purpose of this research project, we will be examining the use of thermal imaging, RGB (visual) imaging, and composite imaging (RGB and thermal overlaid), in the context of Search and Rescue (SAR) missions. Specifically, our focus will assess which imaging type makes for a more accurate and effective image detection model and whether the accuracy of this model is suitable for SAR operations.

The scope of our assessment will focus on evaluating the effectiveness of the three specified imaging techniques in the application drones-led SAR. Drones can assist in surveying mountainous terrains and cold climates, where SAR operations are highly dangerous for all parties involved. The use of drones and image recognition technologies to map the area before undertaking the rescue can significantly decrease the overall risk [1].

Given the possibility for extreme weather conditions or the need for medical attention in the mountains, the sooner a SAR team can mobilize and make the first contact with the victim, the more likely the team can mitigate the adverse health effects of the situation. In [1], researchers state that

standard CPR techniques must be performed within the first 60 minutes to achieve satisfactory results. Determining an efficient and effective drone path to survey the area and gather imaging data, ultimately leads to a faster retrieval. To find which of the four proposed path algorithms will consistently provide the quickest path to the victim, we will apply the different algorithms to identical simulations and determine the best flight path. The speed and accuracy at which the drone can survey the area and accurately detect the victim using image recognition can literally be the differentiating factor between a life or death situation.

Given these two crucial components of the SAR scope, our research will investigate the effectiveness of different path finding algorithms, and three different imaging techniques that can be used to train a YOLOv4 model.

We have chosen YOLOv4 as our image recognition algorithm because it is extremely fast and effective at accurately classifying when working with live recognition. This algorithm uses a single CNN to process images and has a increased detection speed [18]. For the scope of our research, a faster live image detection model adds the most value.

5.1 Imaging Types

In YOLOv4 applications, RGB images (colour images that have a red, green, and blue channel) are more commonly used in the training and testing of the model [19]. Although there are benefits to using RGB images, such as a better image quality, there are also constraints that limit their value in the context of SAR missions. As colour images are dependent on light and direct visual paths, situations that are likely to occur in the mountains such as high tree density, low light (i.e. night searches) and poor weather conditions (i.e. heavy snowfall or haze) can cause obstructions in the camera's direct visual path to the SAR victim [19]. In this scenario, the images being processed by the model would not classify the victim because the images do not show the person directly. This means a person may be in the frame, but due to the visual obstructions, the victim's presence would go undetected, leading to a false negative result. For these reasons, RGB imaging is greatly limited. To address this limitation, we want to research the possibility of utilizing thermal imaging.

Thermal imaging is dependent on heat instead of light [20]. For this reason, low light, image obstruction, and poor weather conditions do not impact the imaging in the ways it does for RGB imaging [19]. By detecting heat rather than light, thermal imaging can produce a black and white images (gray-scale using a single channel) in which warmer objects appear light and brighter and cooler objects dark and duller. This results in the thermal images having a flattened background, as the colour complexity that is present in RGB images is alternatively displayed on the gray-scale, based on the heat of the object, rather than displayed in colour, which is based on light. Given that our research is assessing this technology's use in the realm of cold, mountainous climates, the surrounding landscape appears more muted compared to a person, as the human body temperature averages to be thirty-seven degrees Celsius.

As image recognition is image dependent, often RGB images are preferred given their higher quality. However, knowing the limitations of RGB imaging in our scope, we will be comparing the results of a thermal, composite, and RGB dataset. We hypothesize that thermal imaging's benefits will outweigh the limitations caused by the output's quality issues, and additionally, contribute positively to the composite image dataset. Ultimately, by combining RGB and thermal datasets into a hybrid dataset, there is a possibility that merging RGB and thermal image techniques will reap the benefits of both imaging types and counteract their associated barriers; low quality output, light dependency, visual path obstructions, and background colour complexity.

6 Methods

6.1 The Model - YOLOv4

The algorithm we chose to work with to create an image classification model is the YOLOv4 algorithm. To expedite the creation the training and development of the machine learning model, we used the general YOLOv4 model provided by Roboflow. Roboflow provides free and formatted machine learning algorithms through the Google Colaboratory (Colab) platform. The YOLOv4 algorithm determines the batch number by declaring calculating 2,000 multiplied by the number of classes. As we are only working with a single class, our models completed 2,000 iterations. To ensure the dependability of our Google Colab, we subscribed to the professional version for $12, which guaranteed us access to a GPU and better processor. We were provided access to a Tesla V100 GPU through our Google Colab Pro account, which we used to train all three of our YOLOv4 models.

6.2 The Data

The greatest limitation in creating a good model for image classification is finding a good dataset and hardware for training the model. This proved to be especially difficult when one of the requirements is that the dataset must contain corresponding thermal and RGB images to make a fair comparison. As the contexts for this research paper involves SAR operations with drones, obtaining aerial images or imaging taken from a similar height at which a drone would fly was a primary priority in finding the most appropriate dataset. Given that image recognition models are commonly trained on the COCO image dataset [19], which is mainly RGB images taken at eye level, we wanted to custom train a model that was explicitly trained on aerial imaging. The angle of the drone surveying a SAR area of interest would most similarly relate to an aerial perspective, therefore, we selected a dataset composed of alternating RGB and thermal images taken from a high vantage point.

The dataset used for this research was found through the OTCBVS Benchmark Dataset Collection website [20]. We selected "Dataset 03: OSU Color-Thermal Database" to perform our research. The dataset is composed of both colour and thermal images, captured by cameras adjacently mounted on tripods. Each camera was positioned approximately three stories above ground, mimicking the height a drone would fly at, on a busy pathway of intersections on the Ohio State University campus with many people walking by. The thermal sensor used to build the dataset was a

Fig. 2. RGB Image from the Dataset 03: OSU Color-Thermal Database [20]

Fig. 3. Thermal Image from the Dataset 03: OSU Color-Thermal Database [20]

Raytheon PalmIR 250D, with a 25 mm lens and the colour sensor was a Sony TRV87 Handycam. The imaging format for the thermal images was an 8-bit gray-scale bitmap, and the colour images a 24-bit colour bitmap, where each image is 320 x 240 pixels.

6.3 Structure

The dataset is sorted by folders containing thermal images and RGB images based on two different locations. During the process of the image collection, the researchers alternated between taking a thermal image and an RGB image. This construction of the dataset is why we specifically chose this data to use in our research. As we are comparing the results of RGB, thermal, and composite images, this dataset allows us to have a more accurate representation of comparisons between the two image types because the datasets are nearly identical. We selected 751 RGB images and their corresponding 751 thermal images, as seen in Figure 2 and Figure 3, respectively. These images were manually annotated using labelImg, a python based annotation program, and Roboflow's built-in annotation feature. Both programs create a YOLOv4 formatted annotation based on the bounding boxes drawn for each image. Our research only classified one class, "person," because, in the context of SAR, this is the only class we are interested in identifying in an image.

Our third training set, a composite image dataset, was created using the same 751 images in the thermal and RGB training sets, as seen in Figure 4. To create these composite images, one thermal image and its corresponding RGB images were overlaid, where the RGB and thermal image were weighted equally by a factor of 0.5. The program used to build this training set is a python open-source, add-in library called the Python Imaging Library (PIL). PIL supports the manipulation and saving of different image file formats. The fused training set of 751 composite images was then annotated using the same aforementioned process.

After all three of our datasets were annotated, we uploaded the 751 images from each dataset into separate projects on the Roboflow platform. Roboflow allows the user to expand their datasets for image recognition by adding augmentations to the given dataset. The augmentations that were added to each of our original 751 images in the RGB, thermal, and composite datasets were as follows:

Fig. 4. Composite Image created using Python Imaging Library (PIL) [20]

1) Flip: Horizontal, Vertical
2) Shear: ±15° Horizontal, ±15° Vertical
3) Rotation: Between -15° and +15°

The Roboflow platform creates three outputs per training image based on the augmentations listed. The three resulting datasets totalled 1953 images; 1803 training images, 75 validation images, and 75 testing images. These were the ratios used to train our YOLOv4 models.

6.4 Unity Drone SAR Simulation

To find which proposed algorithms consistently provides the quickest way to find a potential victim, we created a simulation in Unity modelled after the town of Whistler, British Columbia, as can be seen in Figure 5. We used the OpenStreetMap add-on in Blender to pull and create a mesh out of the Geographic Information System (GIS) heightmap data. We then recreated the infrastructure of the town of Whistler using the OpenStreetMap data. Additionally, we created a fictional drone centre using Unity's primitive building meshes and materials in the outskirts of the town. Furthermore, we adopted a drone model from CGTrader to represent our SAR drone, as seen in Figure 6. To represent the lost victim, we used an avatar derived from the Unity Asset store, as seen in Figure 7.

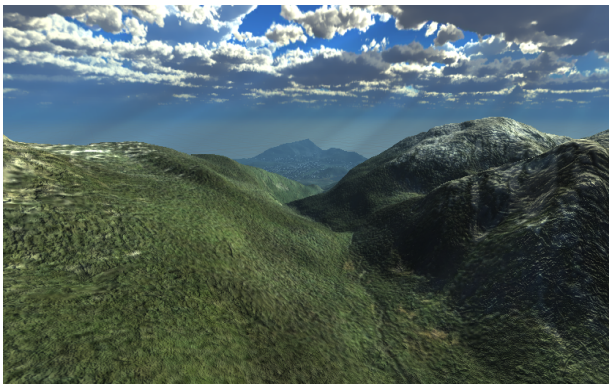

Fig. 5. Unity Whistler Terrain

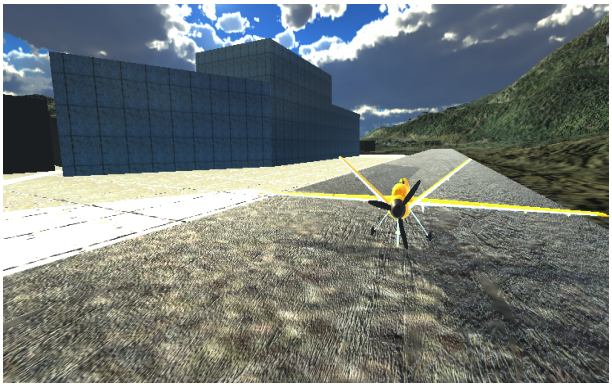

Fig. 6. Proposed drone centre and drone

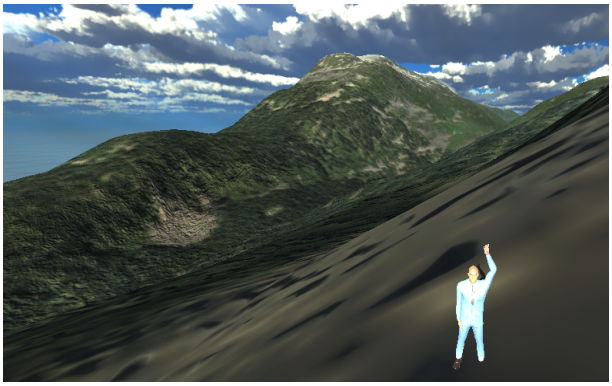

Fig. 7. Simulation of a missing victim

7 Results

7.1 Data

After completing our training and testing of the three models using the YOLOv4 algorithm, we are able to compare and assess the results of each imaging type. The results from our CNN focus on ten individual scores. Many of the testing results produced by the YOLOv4 algorithm are equations that contain other output variables from the model.

The four fundamental scores that highlight effectiveness of a given image recognition model are:

1) True Positive (TP)
2) False Positive (FP)
3) True Negative (TN)
4) False Negative (FN)

These four ratio scores form a confusion matrix. A confusion matrix is a 2x2 matrix that encompasses every outcome of our testing data. The outputs of our three different models and their associated scores can be viewed in Table 1. We have not included the True Negative values as this output does not contribute to our model assessment.

TABLE 1
Model Output Values.

Results	RGB	Thermal	Composite
True Positives	351	384	365
False Positives	11	6	5
False Negatives	4	4	4

Thus, for each human in our testing images, each human will be classified as correctly detected(TP), incorrectly detected(FP), correctly undetected (TN), incorrectly undetected(FN). Utilizing these four outputs, we can formulate the following two ratios:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

Precision will measure how accurately the model correctly predicts a person in the image (TP), compared to the sum of how many correct predictions (TP) and incorrect, positive classifications (FP) the model classified.

$$RGBPrecision = \frac{351}{351+11}$$
$$RGBPrecision = 0.9696 \tag{7}$$

$$ThermalPrecision = \frac{384}{384+6}$$
$$ThermalPrecision = 0.9846 \tag{8}$$

$$CompositePrecision = \frac{365}{365+5}$$
$$CompositePrecision = 0.9865 \tag{9}$$

Recall measures how accurate our model detects all people in our given data set based on the number of people annotated. Precision focuses on how many people were annotated in the dataset and compares this to the total number of people the model correctly and incorrectly detected. Comparatively, recall focuses on how many people our model correctly detects in the dataset, compared to the correct total number of people that were annotated in the dataset.

$$RGBRecall = \frac{351}{351+4}$$
$$RGBRecall = 0.9887 \tag{10}$$

$$ThermalRecall = \frac{384}{384+4}$$
$$ThermalRecall = 0.9897 \tag{11}$$

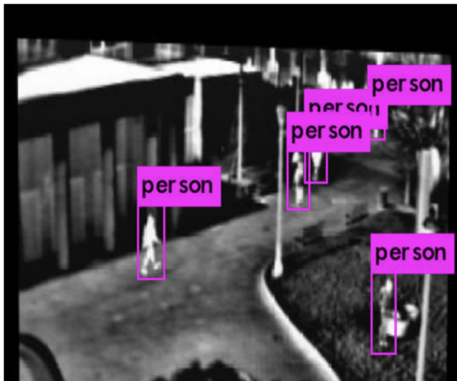

Fig. 8. YOLOv4 Model Image Detection: Thermal Image Example

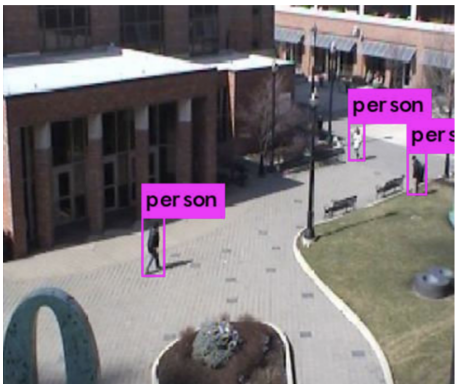

Fig. 9. YOLOv4 Model Image Detection: RGB Image Example

$$CompositeRecall = \frac{365}{365+4}$$
$$CompositeRecall = 0.9892 \quad (12)$$

The F1 score is calculated by combining the precision and recall scores. This formula considers the combination of the inaccuracies in the model, the False Positives and False Negatives. If the sum of these inaccuracies is large and the numbers are large, it can greatly impact the F1 score. This formula does not account for the true negatives, as we are only concerned about the correct classifications, the incorrect negative classifications, and the incorrect positive classifications. In our case, correct negative detection is not valuable information to consider.

$$F1Score = \frac{(2)(\text{Precision})(\text{Recall})}{\text{Precision + Recall}} \quad (13)$$

$$RGBF1 = \frac{(2)(0.9696)(0.9887)}{0.9696+0.9887}$$
$$RGBF1 = 0.9791 \quad (14)$$

$$ThermalF1 = \frac{(2)(0.9846)(0.9897)}{0.9846+0.9897}$$
$$ThermalF1 = 0.9871 \quad (15)$$

$$CompositeF1 = \frac{(2)(0.9865)(0.9892)}{0.9865+0.9892}$$
$$CompositeF1 = 0.9878 \quad (16)$$

The results to the precision, recall, and F1 score calculations for each image types can be seen in Table2.

TABLE 2
YOLOv4 Model: 2000 Iterations: RGB trained, thermal trained, and composite trained results.

Results	RGB	Thermal	Composite
F-1 Score	0.9791	0.9871	0.9878
Recall	0.9887	0.9897	0.9892
Precision	0.9696	0.9846	0.9865

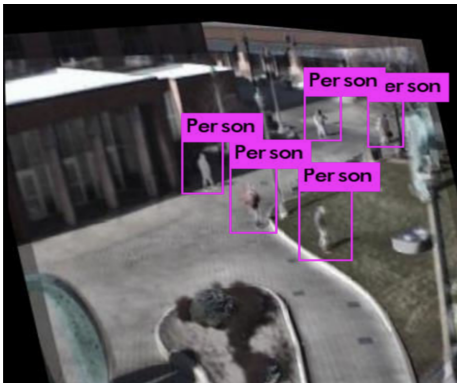

Fig. 10. Yolov4 Model Image Detection: Composite Image Example

8 Analysis

8.1 YOLOv4 Models

After completing the calculations for recall, precision, F-1 and gathering the values for the TP, FP, and FN, we can to conclude the composite dataset YOLOv4 model outperformed the thermal and RGB models. By referring to Table 1, we can see that the composite dataset compared to the thermal and RGB datasets, minimized the total number of false positives that our dataset produced.

Minimizing the number of false positives in the context of a SAR operation results in the more effective allocation of limited resources. As seen in Table 1, we can see the RGB model, thermal model, and composite model resulted in 11, 6, and 5 false positives, respectively. Overall, the composite dataset minimized the false positives count compared to RGB by 66% and the thermal model by 17%. As a minimized first contact time is one of the crucial elements to mitigating life-threatening injury [1], a smaller false positive count will result in fewer incorrect excursions for the rescue team, allowing the SAR operation to focus on the accurate classifications.

Additionally, the composite dataset offered the highest precision and F-1 scores, which can be seen in Table 2. The composite dataset offered approximately 1.7% and 0.2% improvement to the RGB and thermal datasets in the precision score, and approximately 1.0% and 0.1% improvement in the F1 scores. Given the individual datasets used to train our models consisted of 1953 images, we expect the differences

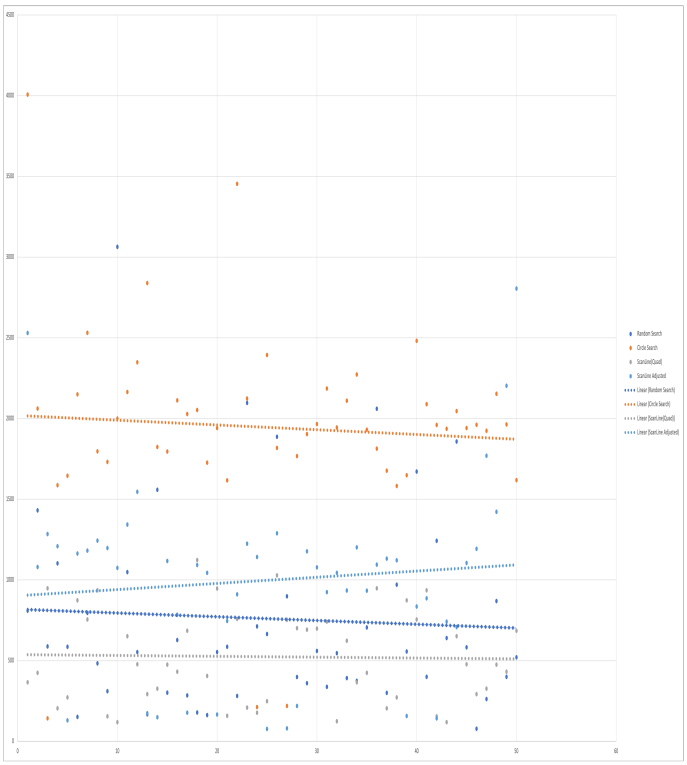

Fig. 11. Time it took each type of algorithm to find the missing person in a 50 trial simulation.

between the recall and precision score to further spread in favour of the composite dataset for larger datasets.

Looking at Table 1, we can see that the false negatives, a value that is hugely important to minimize in the context of SAR, do not vary among the three models. All three models wrongly classified 4 false negatives, we can conclude that there is not a significant impact of these different datasets on the YOLOv4 algorithm's calculation of false negative results. Examining Table 1 further, we can see that the composite dataset only failed to identify 1% of the population. To further minimize this percentage, the SAR operations would likely have multiple images from varying angles, and thus would likely decrease below 1%.

8.2 Search Algorithms Time

Each of the algorithms under examination; Scanline Fill, Quadrant Scanline Fill, Geocentric Fill, and Random Fill, were run 50 times in the unity simulation at 30 times the real-time speed. The time taken to find the victim was collected and scaled to real-time then plotted and compared to one another, as displayed in Figure 11. The summarized results can be viewed in Table3. This table shows the superiority of the Quadrant Scanline Algorithm, outperforming the second best algorithm by 36.76%.

TABLE 3
Path Finding Results in Seconds.

Algorithms	Scanline	QuadScanLine	Geocentric	Random
Average Time	999.9	523.7	1944.1	759.6

9 Discussion

Overall, the composite dataset performed better than its two counterparts. The simplification of backgrounds within thermal images combined with the higher image quality in RGB images, as a result of merging these two image types, resulted in a dataset that minimized the total number of false positives, while also most accurately classifying the objects. Furthermore, the composite dataset would also offer detectable images at night, as the thermal component of the composite image addresses the low light limitations of RGB imaging. In the application of this hybrid model, we would suggest adjusting the weightings of the thermal imaging and RGB imaging depending on the light and weather conditions. If there is low light or visual obstructions, adjust the weights of the composite imaging to reflect a heavier value on the thermal image (compared to our 0.5 weight factor). This hybrid dataset imaging recognition model will account for both the barriers associated with thermal and RGB imaging and will facilitate a more diverse range of SAR operations.

In a real SAR situation, where high accuracy is required in the critical operation, there is a problem with real-time detection on embedded systems where accuracy must be compromised for running YOLOv4-tiny, a more lightweight version of YOLOv4 used to run on small devices [21], [22]. A proposed solution to this issue would be to have a live relay of video to a central computer than can perform this processing, in a similar manner to how this paper performed the training and classification. Under no circumstances should a concession be made for performance over accuracy as the application in which this would be critical is for the well-being of lost or missing persons.

In conjunction with determining the dataset, selecting an optimal flight path for the drone can impact the efficiency and effectiveness of resource allocation. Interestingly, the Geocentric Fill algorithm was the slowest to find the missing person and of the four algorithms studied, the Quadrant Scanline Fill Algorithm was consistently the fastest to find the missing person.

10 Conclusion

This paper examined the use of thermal, RGB and combination images in the context of SAR operations where drones loaded with a pre-trained machine learning model could be utilized to help identify missing persons. It was found through the analysis of the three models that a composite image dataset, merging both thermal and RGB images into single images, yields the best precision score, F-1 score, and lowest number of false positives. The results shown in Table 1 distinctly show that the most critical measurement, false negatives, is the same across the different types of images that can be used to identify people meaning that the trained model will not let someone go unseen in many instances. However, false positives are lowest in the composite images indicating that objects or artifacts in the image are not falsely identified as people, resulting in fewer instances where the SAR team may explore without finding an actual person. The composite images also have a slightly better F-1 score than the other imaging techniques as viewed in Table 2. It has been surmised that the better performance of the

composite images in training the machine learning model for person detection in SAR applications because it has the added benefits of both RGB and thermal images. Although both images are the same resolutions, RGB images show objects with sharper edges, whereas thermal image objects make people more pronounced against their backgrounds and outperform RGB in low-light conditions. Figures 8, 9 and 10 excellently show how the composite image in Figure 10 exhibits details that would otherwise be missing by either of the other two imaging techniques.

As for determining the drone flight path algorithm, it is found that using small repeating search patterns is more effective at decreasing the time taken to find a missing person compared to large encompassing patterns. Ultimately, the quadrant scanline fill algorithm is the best choice for finding a victim in an unknown location. This algorithm divides the area into four quadrants and traverses each quadrant in the Scanline Fill algorithm before moving forward to another quadrant. The summarized results in Table3, shows the the quadrant scanline algorithm outperforming the second best algorithm by 36.76%. One of the reasons for this massive spread in performance could be attributed to the fact that this algorithm is more likely to find the target when searching in a smaller area compared to a larger area. Also, the quadrants are being reflected across the line of reflection which helps with covering a more diverse area in a shorter time.

11 Future Work

Moving forward in this research, we would build a custom aerial dataset utilizing a drone equipped with a thermal and RGB camera. Building a more comprehensive dataset with varying backgrounds, angles and heights in each type of imaging would contribute to a dataset that would more similarly compare to SAR operations. By training on a more customized dataset, we can create a more accurate weight file that would be configured to process real-time video. As current processing systems on drones cannot process the number of image frames produced from live video, we would focus on having an on-ground support station to remotely process images received from the drone.

The simulation created in Unity has the potential to be utilized to create an all encompassing simulation from finding the target to using the YOLOv4 algorithm for real-time image detection, given proper computer hardware. We can expand the drone flight path algorithm optimization by testing more patterns and by using multiple drones simultaneously. We can also further divide the patterns into smaller repeating patterns to further decrease the time taken to find the target. Another idea to implement is to have the drone first search historically known points of interest where people are known to have gone missing and program the drone to search those points first, before engaging in the search algorithm. After traversing through these points of interest, have the drone avoid iterating through the same path while still maintaining the flight search pattern. Furthermore, we can use the Unity layer mask system to create a hypothetical thermal vision for our drone camera to get an aerial dataset similar to the one used in this study. We can then have the Unity simulation collect RBG and thermal images from the drone flight simulation and then train the YOLOv4 to look for and identify potential victims found in the simulation based on different avatars used. Additionally, we can gather further data on typical typical RPAs used in SAR efforts and implement more realistic weight, speed, equipment and weather conditions to make the SAR results as realistic as possible. We could also take this further by examining the cost vs rescue time analysis for typical for utilizing different types of drones and multiple drones.

References

[1] Y. Karaca, M. Cicek, O. Tatli, A. Sahin, S. Pasli, M. F. Beser, and S. Turedi, "The potential use of unmanned aircraft systems (drones) in mountain search and rescue operations," *American Journal of Emergency Medicine*, vol. 36, pp. 583–588, 4 2018.

[2] B. Rao, A. G. Gopi, and R. Maione, "The societal impact of commercial drones," *Technology in Society*, vol. 45, pp. 83–90, 5 2016.

[3] U. S. D. of Transportation, "Uas by the numbers," May 2021.

[4] F. A. Administration, "Dot and faa finalize rules for small unmanned aircraft systems," *Press Release*, 6 2016.

[5] B. Mishra, D. Garg, P. Narang, and V. Mishra, "Drone-surveillance for search and rescue in natural disaster," *Computer Communications*, vol. 156, pp. 1–10, 4 2020.

[6] D. C. de Oliveira and M. A. Wehrmeister, "Using deep learning and low-cost rgb and thermal cameras to detect pedestrians in aerial images captured by multirotor uav," *Sensors (Switzerland)*, vol. 18, 7 2018.

[7] A. Sledz, J. Unger, and C. Heipke, "Thermal ir imaging: Image quality and orthophoto generation," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives 42 (2018), Nr. 1*, vol. 42, no. 1, pp. 413–420, 2018.

[8] M. B. Bejiga, A. Zeggada, A. Nouffidj, and F. Melgani, "A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery," *Remote Sensing*, vol. 9, 2017.

[9] M. Saqib, S. Daud Khan, N. Sharma, and M. Blumenstein, "A study on detecting drones using deep convolutional neural networks," pp. 1–5, Aug 2017.

[10] T. D. Trong, Q. T. Hai, N. T. Duc, and H. T. Thanh, "A novelty approach to emulate field data captured by unmanned aerial vehicles for training deep learning algorithms used for search-and-rescue activities at sea," pp. 288–293, Institute of Electrical and Electronics Engineers Inc., 1 2021.

[11] P. Zhu, L. Wen, D. Du, X. Bian, H. Ling, Q. Hu, Q. Nie, H. Cheng, C. Liu, X. Liu, *et al.*, "Visdrone-det2018: The vision meets drone object detection in image challenge results," pp. 0–0, 2018.

[12] D. Božić-Štulić, Željko Marušić, and S. Gotovac, "Deep learning approach in aerial imagery for supporting land search and rescue missions," *International Journal of Computer Vision*, vol. 127, pp. 1256–1278, 9 2019.

[13] C. D. Rodin, L. N. D. Lima, F. A. D. A. Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold, "Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems," vol. 2018-July, Institute of Electrical and Electronics Engineers Inc., 10 2018.

[14] E. Bondi, D. Dey, A. Kapoor, J. Piavis, S. Shah, F. Fang, B. Dilkina, R. Hannaford, A. Iyer, L. Joppa, and M. Tambe, "Airsim-w: A simulation environment for wildlife conservation with uavs," Association for Computing Machinery, Inc, 6 2018.

[15] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, vol. 2018-January, pp. 1–6, 3 2018.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, 2019.

[17] A. F. Golda, S. Aridha, and D. Elakkiya, "Algorithmic agent for effective mobile robot navigation in an unknown environment," pp. 1–4, 2009.

[18] D. Wu, S. Lv, M. Jiang, and H. Song, "Using channel pruning-based yolo v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments," *Computers and Electronics in Agriculture*, vol. 178, p. 105742, 2020.
[19] M. Ivašić-Kos, M. Krišto, and M. Pobar, "Human detection in thermal imaging using yolo," pp. 20–24, 2019.
[20] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 162–182, 2007.
[21] H. Liu, K. Fan, Q. Ouyang, and N. Li, "Real-Time Small Drones Detection Based on Pruned YOLOv4," *Sensors*, vol. 21, p. 3374, May 2021.
[22] L. Zhang, S. Wang, H. Sun, and Y. Wang, "Research on Dual Mode Target Detection Algorithm for Embedded Platform," May 2021.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Aibin for his direction, support and time.