Idaho State University

Deep Learning Takes Flight:

Detecting Jackrabbits with Thermal Drone Imagery

Christina Appleby

GEOL 5508: Geotechnology Seminar

Professor Delparte

December 13, 2024

Abstract

Detecting wildlife efficiently and accurately is crucial for wildlife abundance surveys that support ecological monitoring and conservation efforts. This study evaluates the application of a YOLOv5 object detection model to identify black-tailed jackrabbits (*Lepus californicus*) in thermal video acquired with an uncrewed aerial vehicle in the Morley Nelson Snake River Birds of Prey National Conservation Area in Southwest Idaho. We trained and evaluated two YOLOv5 models, optimized with SGD and Adam algorithms. After hyperparameter evolution, the Adam-optimized model outperformed the SGD model with a mAP50 of 0.991 and zero false positive detections. While the results demonstrated the effectiveness of using a YOLOv5 for detecting jackrabbit, distinguishing them from cottontails still presents a challenge due to similar thermal emissivity. This study highlights the potential for integrating a YOLOv5 model with spotlight surveys for estimating jackrabbit abundance.

1. Introduction

The black-tailed jackrabbit is an important prey to many species and is the prey of choice for golden eagles (Steenhof & Kochert, 1988). Many studies have examined the role of jackrabbits in golden eagle diets and habitat (Bedrosian et al., 2017; Brown et al., 2021; R. K. Murphy et al., 2023), particularly at the Morley Nelson Snake River Birds of Prey National Conservation Area (NCA) (Steenhof et al., 1997; Kochert & Steenhof, 2002; Heath et al., 2021). Golden eagle diet composition is determined through the examination of prey remains and pellets, with increased prey diversity indicating a decline in jackrabbit abundance (Bedrosian et al., 2017; Brown et al., 2021; Heath et al., 2021). However, researchers have not estimated jackrabbit abundance through surveys or counts at the NCA since 1997 (Knick & Dyer, 1997).

1.1 Estimating Jackrabbit Abundance

Traditional methods of non-invasive wildlife surveys include human observations, camera traps, and acoustic readers. Human observations are prone to bias depending on the experience and training of the individual. Due to their sensitivity, the camera traps capture tens of thousands of images each year, with only a small percentage of

images containing wildlife (A. Chen et al., 2022). Additionally, the placement of camera traps greatly affects the reliability of abundance estimations when using the random encounter model (S. M. Murphy et al., 2024).

Common jackrabbit survey methods include spotlight transects (Knick & Dyer, 1997; Simes et al., 2015), flushing transects (Langbein et al., 1999; Simes et al., 2015), foot transects (Cypher et al., 2018; Smith et al., 1981), drive counts (Driscoll, 2009; Simes et al., 2015) and pellet counts (Dunagan & Karels, 2018; Simes et al., 2015). Flushing transects and drive counts require significant human resources and are not practical for surveying large areas (Driscoll, 2009; Langbein et al., 1999). Spotlight transects are relatively easy and cost effective (Cypher et al., 2018; Driscoll, 2009) and capture jackrabbit counts at night when they are most active (Lechleitner, 1958). However, habitat conditions can produce data biases, particularly in areas with shrubs (Cypher et al., 2018). As a result, spotlight surveys can lead to inaccurate abundance estimations in NCA shrublands.

Another non-invasive method for wildlife surveys is using uncrewed aerial vehicles, commonly referred to as drones, to capture high-resolution imagery and video. Many drones can capture not only RGB (red, green, blue, i.e., visible light) imagery, but also multispectral or thermal imagery (Mpouziotas et al., 2023). Thermal and multispectral imagery is particularly useful for monitoring camouflaged wildlife that can be hard to detect in RGB imagery (A. Chen et al., 2022). Additionally, drones enable data capture in remote areas (Mpouziotas et al., 2023).

Obtaining high-resolution imagery via drones is gaining popularity because of its cost-effectiveness (Lawrence et al., 2023). However, drones also provide large amounts of data, which can be time-consuming to analyze (L. Chen et al., 2024). Object detection, a field of computer vision, automates this process, allowing for efficient processing and analyzing of large datasets, including wildlife detection, species classification, and tracking (L. Chen et al., 2024; Mpouziotas et al., 2023). This use of deep learning significantly reduces the time required to analyze drone imagery, as well as camera trap imagery and acoustic data. Furthermore, object detection algorithms have high accuracy rates and reduce manual tasks (L. Chen et al., 2024).

1.2 Computer Vision for Automated Detection

Convolutional neural networks (CNN) are a popular object detection algorithm that provides exceptional accuracy but at the cost of speed. Their relatively slow detections prohibit their use for real-time detection and monitoring (Lawrence et al., 2023). This slowness is because CNN uses two-stage object detection. One-stage detection algorithms, such as You Only Look Once (YOLO) are faster, have acceptable accuracy, and are capable of real-time detection (L. Chen et al., 2024).

Povlsen et al. (2023) used thermal drone imagery to train a YOLOv5 model for hare and roe deer detections, achieving an accuracy of 0.99 and 0.96, respectively. Schütz et al. (2021) trained a YOLOv4 model to detect red foxes in videos to analyze movement patterns and animal welfare, as well as examine behavioral activity. A. Chen et al. (2022), Lawrence et al. (2023), and Mpouziotas et al. (2023) also used this YOLO detection algorithm for bird detection.

1.3 YOLOv5 Object Detection Pipeline

The main steps in the YOLOv5 object detection pipeline are dataset creation, model training, and model validation. Optionally, model improvement and subsequent model reevaluation can occur before model deployment.

Dataset creation includes acquiring still frame images or extracting image frames from a video and annotating the images. Manual image annotation consists of placing bounding boxes around the object and labeling the object class. In addition to annotated images, background images provide a baseline to help the model learn to distinguish between the objects and the background environment, resulting in a more accurate and reliable model. Image augmentation is a technique used to improve computer vision model performance by creating a more robust dataset by applying various transformations to the original dataset images. Datasets are broken down into training, validation, and test (optional) datasets with the training set consisting of 70 to 90 percent of the images. Model training occurs on the test dataset, and YOLOv5 automatically validates the model using the validation set, which is unseen by the model

during training to ensure an unbiased evaluation. The user can run the model on the test dataset, which is also unseen by the model, for further evaluation.

The user can specify different parameters that affect model training, such as the optimizer used, model weights, and number of epochs. The optimizer is the algorithm used to change the model weights during training with the goal of finding the best weights for learning to increase the model accuracy. Model weights are the parameters that define the model's internal structure and decision-making logic. For YOLOv5, there are different pretrained weights to choose from depending on the size of the dataset. An epoch is a complete cycle of data processed through the model. Training datasets are typically broken down into smaller datasets, specified by the batch size; one epoch is complete when all batches are processed. Increasing the number of epochs can increase the model accuracy as it gives the model more time to learn. Conversely, it can also lead to model overfitting.

Hyperparameters are the settings that control the training process of a YOLOv5 model. Initial model training uses a set of default hyperparameters. Hyperparameter evolution finds optimal hyperparameter combinations to increase model accuracy. Mean average precision (mAP) is an object detection model performance metric calculated using the precision, recall, and intersection over union (IoU). Precision is the proportion of all positive detections that are truly correct. Recall is the proportion of positive cases that the model correctly identifies. IoU measures the overlap between the predicted bounding box and the annotated bounding box.

The goal of this study is to train a YOLOv5 model to detect jack rabbits in thermal imagery and deploy the model on a live video stream. However, if a live stream is not available, the model deployment can be on an existing video that does not contain any frames used in the training data.

2. Methods

2.1 Thermal Video Data Collection

The 485,000-acre NCA has the highest density of nesting birds of prey in North America. Our study area lies within the Orchard Combat Training Center inside the

northern part of the NCA. In 2022, we conducted a pilot study to determine the optimal drone altitude, speed, and size of transects for jackrabbit drone surveys. Spotlight transects performed on multiple nights between 2200 to 0600 revealed that jackrabbits were most active between 2200 to 0200. Not only are jackrabbits more active at night, but the cooler night temperatures maximize the contrast between the body temperature of the jackrabbits and the background environment.

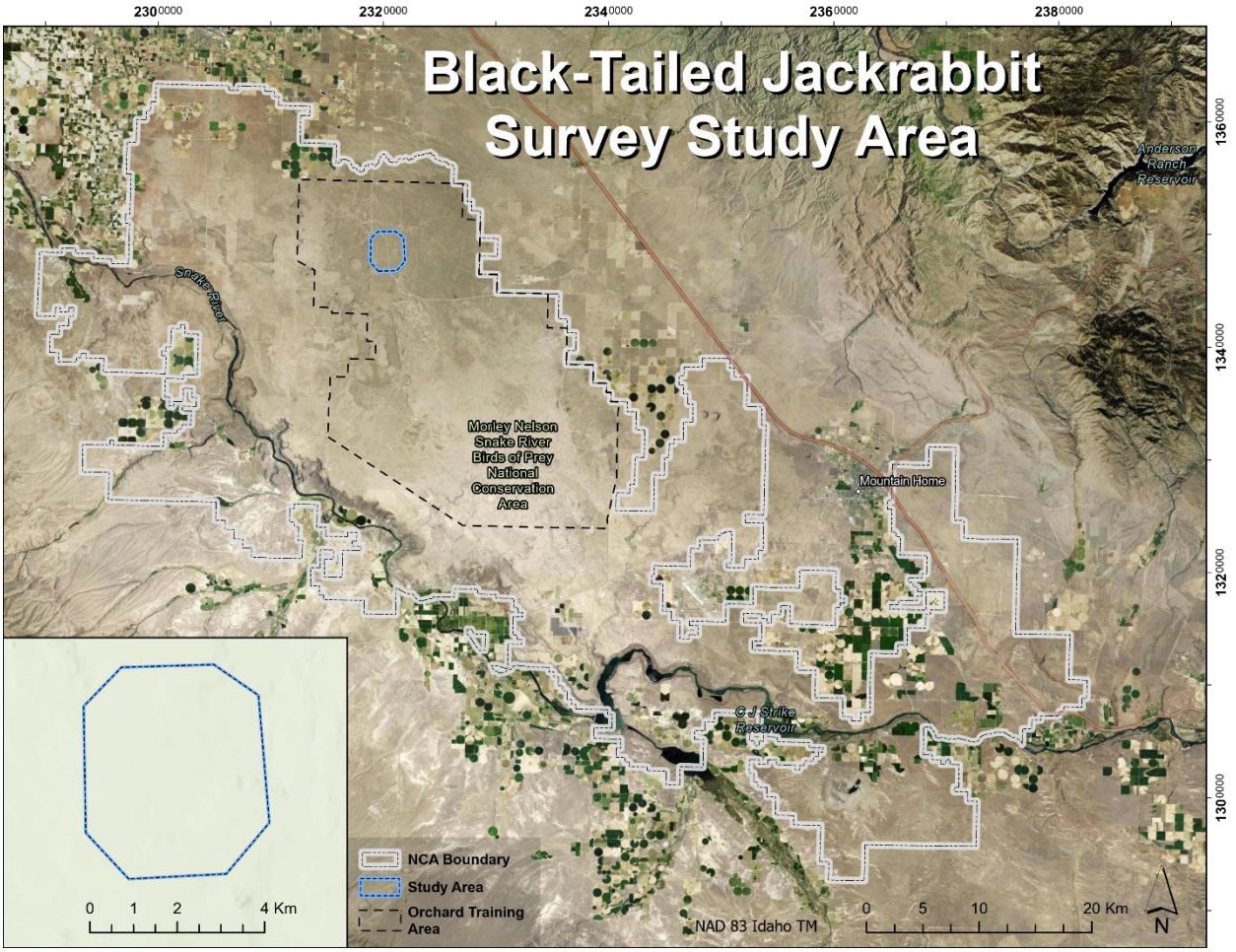

Figure 1. Map of the study area where thermal video was captured, located within the Orchard Military Training Center (Orchard Training Area) inside the Morley Nelson Snake River Birds of Prey National Conservation Area.

In 2024, an Inspired Flight IF 750 drone equipped with a FLIR DUO PRO R thermal camera was used to collect thermal video data. We flew three to six transects each night for four nights in late May; Table 1 shows the flight parameters. Because of the

limitations of flying a drone in windy conditions, approximately half the flights occurred outside the optimal 10pm to 2am jackrabbit detection window.

Table 1. The drone flight details for the four nights of thermal video collection, including number of flights, times, speed, altitude, and transect distance.

Night	# of Flights	Times	Speed (m/s)	Altitude (m)	Transect (m)
1	3	2330 - 0230	3.5	40	250
2	5	2015 - 2315	3.5 - 5	40	250
3	5	0345 - 0715	5 - 7	40	500
4	6	0000 - 0245	7	35 - 40	500

2.2 Image Annotation

A researcher watched numerous thermal videos at slow speed to train themselves to identify jackrabbits in the videos. The researcher then watched all the thermal videos and recorded jackrabbit observations, including the number of jackrabbits, start and video and end time, their certainty of the observation (certain, likely, or uncertain), and the coordinates (found using the video time and drone flight data). They distinguished jackrabbits from cottontails based on their size.

Using a Jupyter Notebook and the cv2 Python package, we used the video start and end times of jackrabbit detections recorded during manual identification and marked with a certainty of “certain” or “likely” to extract video frames for image annotation. With a video frame rate of 30 frames per second, the start and end times were not accurate enough to ensure that we extracted all frames containing jackrabbits. Therefore, we extracted 10 frames prior to the start time and 10 frames after the end time. In the event the start time and end time were the same, we extracted an additional 30 frames, for a total of 40 frames after the end time. In addition to the 2024 video frames, we also extracted frames from one of the 2022 videos that showed two jackrabbits in motion.

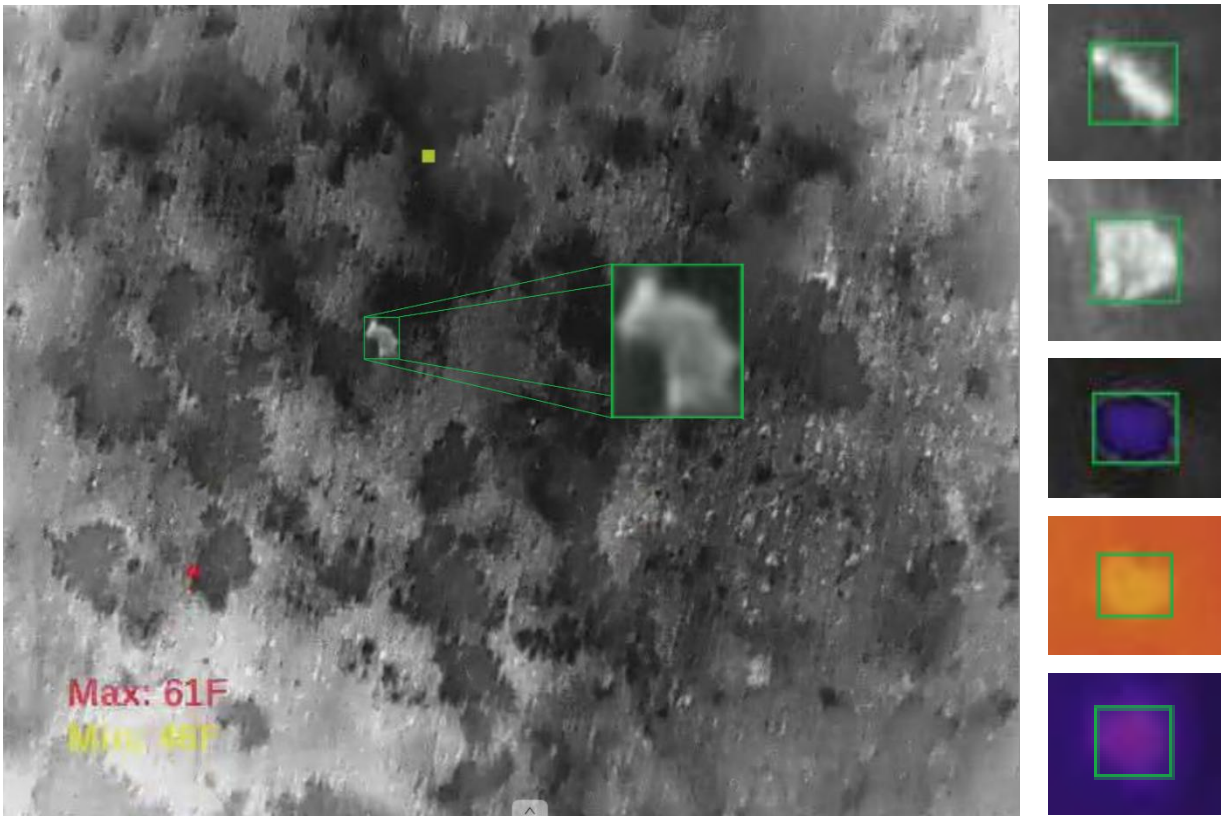

Figure 2. Examples of manual image annotation bounding boxes (green lines).

Roboflow is a popular platform offering tools for each stage of the computer vision pipeline, and we initially planned to use Roboflow for this study. Ultralytics, the software company that has enhanced the open-source YOLO algorithm and made it popular, partnered with Roboflow to simplify the YOLO workflow. Video capture occurred on Orchard Military Training Center grounds, requiring the dataset to remain private. However, a paid Roboflow subscription is required to keep your data private. To minimize project costs, we used the open-source Computer Vision Annotation Tool (CVAT) installed locally in a Docker container.

In our study, we only used one object class for image annotation, jackrabbit. In 2022, we flew at a higher altitude than in 2024, resulting in the jackrabbits appearing smaller in the 2022 imagery than in the 2024 imagery. The final dataset consisted of 590 frames: 105 with jackrabbit labels from 2022, 438 with jackrabbit labels from 2024, and 47 background images. The user can apply augmentations to images prior to model training; however, we relied on YOLOv5's built-in augmentations applied during model

training (Figure 3). Lastly, we divided the dataset into training, validation, and test images using an 80/10/10 split.

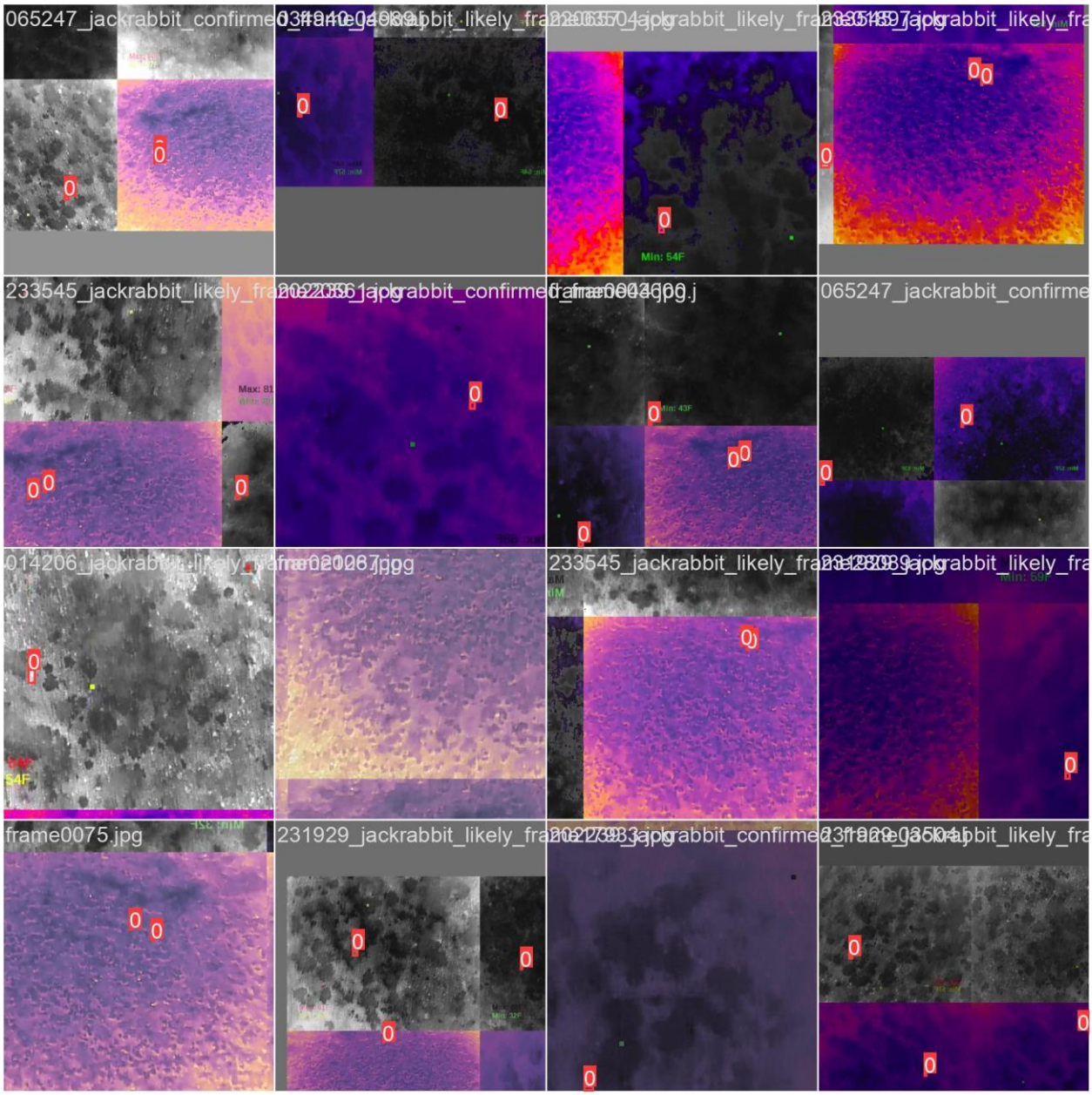

Figure 3. Examples of the YOLOv5 built-in image augmentations applied to the training dataset. The zeros in the red box indicated jackrabbit labels.

2.3 YOLOv5 Training on Custom Data

While designed for execution in the terminal, we chose to execute the YOLOv5 code in a Jupyter Notebook for repeatability. We trained two different models for comparison;

one using the SGD optimizer and one using the Adam optimizer. The initial model training parameters used for both models are as follows: image size of 640 pixels, batch size of 16, 300 epochs, yolov5s weights, and the default hyperparameters (scratch-low). We ran the models on the test dataset with a confidence threshold of 0.80 to determine its accuracy. To optimize the model accuracy and object detection, we evolved the hyperparameters for 507 and 355 evolutions for the SGD and Adam models, respectively. Each evolution ran for 10 epochs. Using the evolved hyperparameters, we trained each model for an additional 150 epochs. Again, we ran the models on the test dataset with a confidence threshold of 0.80.

Since we could not deploy the models on real-time video, we extracted video clips from the thermal video footage with a certainty of “uncertain” to run through the Adam model. To further test the model, we chose clips containing only jackrabbits, both jackrabbits and cottontails, and only cottontails. and ran the models on those clips. We used the video start and end times recorded during the manual identification with an additional two seconds of video at the beginning and end.

3. Results

3.1 Jackrabbit Detections

Although we could not help the timing of the video acquisition flights because of wind, the flight parameters varied by night in terms of speed and transect distance (Table 1). Flight paths and jackrabbit detections are in Figure 4. The three flights on the first night covered only a small portion of the study area and yielded the lowest number of detections, with only two jackrabbit detections during the initial flight. The highest number of detections, 30 jackrabbits, occurred on the second night even though we only covered half of the study area. Flights on the third and fourth nights covered the entire study area, but only yielded six and seven jackrabbit detections, respectively.

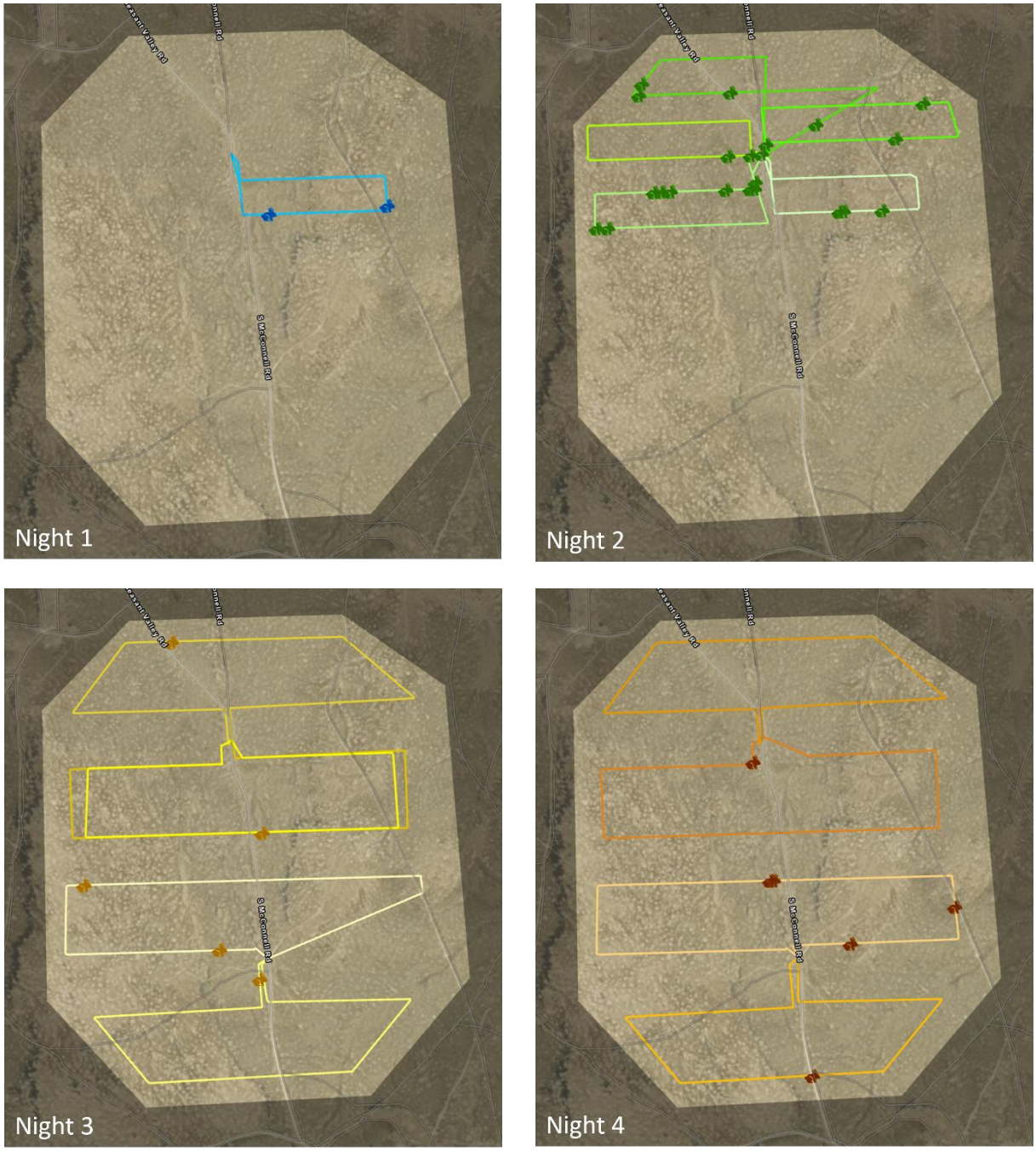

Figure 4. Flight paths for the four nights of video acquisition with each flight path presented in a different color. All three flight paths for the first night were identical. The rabbit symbols indicate jackrabbit detections.

3.2 YOLOv5 Model Performance

Both models performed better after additional training with evolved hyperparameters, resulting in a mAP50 of 0.949 for the SGD model and a mAP50 of 0.991 for the Adam model (Table 2). The SGD model struggled with jackrabbit detection in the 2022 frames

even after hyperparameter evolution, missing three of the detections, whereas the Adam model detected all jackrabbits in 2022 frames after hyperparameter evolution.

*Table 2. The number of epochs for model training, the optimizer used for the training, model accuracy (precision and recall), performance metrics (mAP50 - mean average precision with an IOU of 50), and number of false positive (FP) detections using a confidence threshold of 0.8 on 59 test images. *Additional 150 epochs of training used evolved hyperparameters.*

Epochs	Optimizer	Precision	Recall	mAP50	FP
300	SDG	1	0.797	0.898	1
300 + 150*		1	0.898	0.949	0
300	Adam	0.926	0.847	0.914	4
300 + 150*		1	0.983	0.991	0

Because of its higher accuracy, we ran the video clips through the Adam model. Since all the observations in the video clips were uncertain, it was difficult to determine the accuracy of the model detections. Many of the detections did appear to be jackrabbits; however, the smaller size of some detections could indicate that the model mistakenly identified cottontails as jackrabbits (Figure 5).

4. Discussion

The lack of visual context and reduced resolution in thermal imagery can make object detection more difficult in general. When using thermal imagery for wildlife detection, the lack of visual details, such as fur color, pattern, and physical characteristics, make it difficult to identify individual species. Because of this, similar thermal emissivity between jackrabbits and cottontails could have led the model to identify cottontails as jackrabbits. The model mistaking cottontails for jackrabbits could also be a result of using training data obtained from different altitudes.

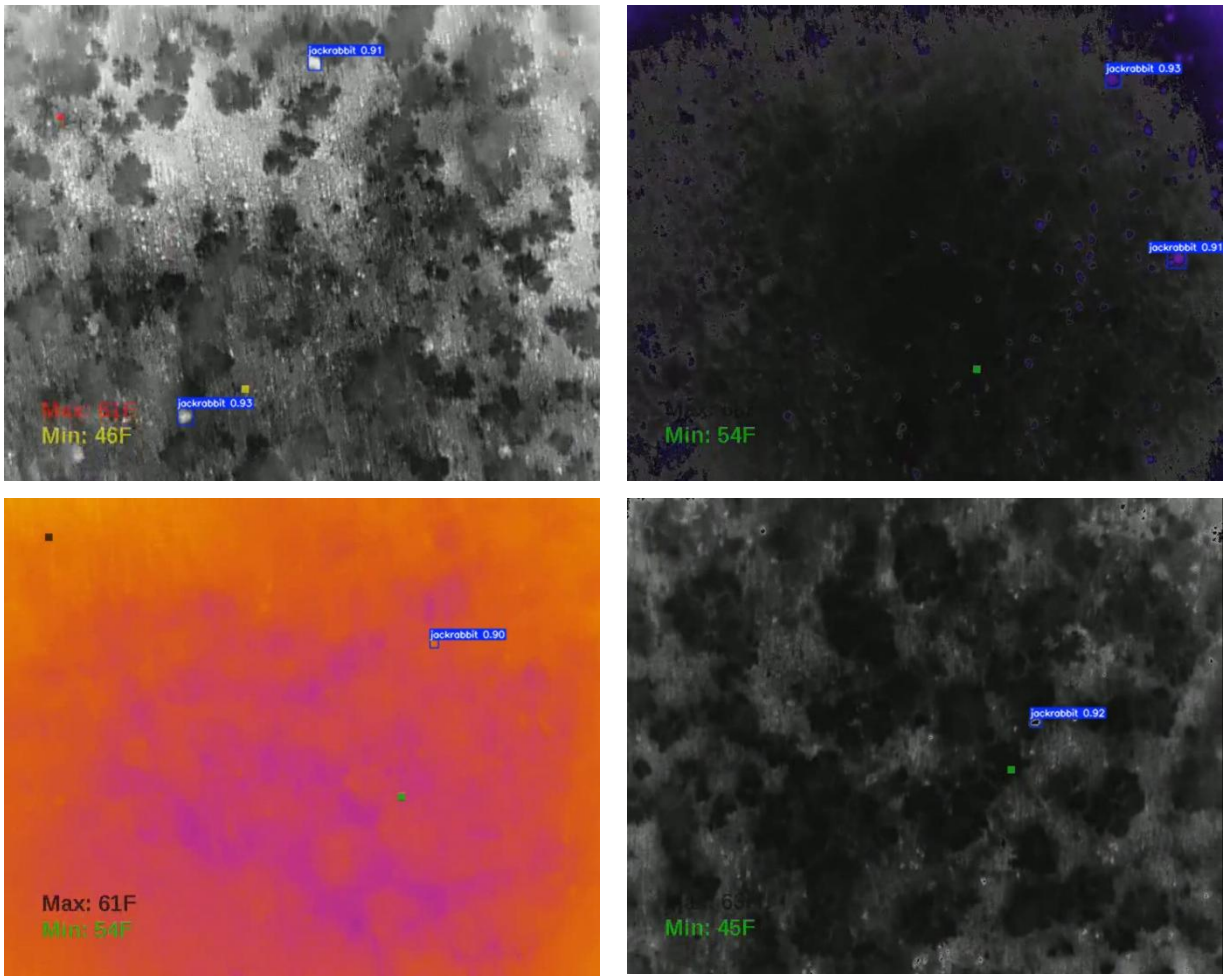

Figure 5. The top images are likely positive jackrabbit detections, whereas the bottom images are likely cottontails due to their size.

There are two solutions for distinguishing cottontails from jackrabbits. When implementing the model on real-time video footage, we could limit the model results by a minimum number of pixels based on the flight altitude. The other solution is to train a model for each flight altitude, ensuring that the jackrabbit size is consistent throughout the training data. Further research would provide insight into which solution is most viable. Additional research in the form of flight is also needed to determine the optimal video acquisition speed and transect distance, as indicated by the highest number of detections on the second night of flights. The slower flight speed and shorter transect distance of the second night's flight could explain the increased jackrabbit detections over the third and fourth nights.

While the goal of this study was to train a YOLOv5 model to detect jackrabbits in thermal imagery, the ultimate goal is to integrate the model with spotlight surveys to estimate jackrabbit abundance. Many object detection model users try to minimize the number of false negative and false positive detections; we were only concerned with minimizing false positives. Wildlife abundance models already account for missed detections, or false negatives. Traditional abundance models rely on correctly identified human observations or human-identified counts and do not account for “extra” detections, or false positives. Both models had zero false positive detections after hyperparameter evolution, making either model suitable for integration with spotlight surveys to enhance traditional abundance models.