Summer 8-1-2023

The Impacts of Transfer Learning for Ungulate Recognition at Sevilleta National Wildlife Refuge

Michael Gurule
University of New Mexico - Main Campus

Michael Gurule
Candidate

Geography & Environmental Studies
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Christopher Lippitt, Chairperson

Lipping Yang

George Matthew Fricke

THE IMPACT OF TRANSFER LEARNING FOR UNGULATE RECOGNITION AT SEVILLETA NATIONAL WILDLIFE REFUGE

by

MICHAEL GURULE

B.S., GEOGRAPHY,
UNIVERSITY OF NEW MEXICO, 2020

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Geography

The University of New Mexico
Albuquerque, New Mexico

August, 2023

The Impact of Transfer Learning for Ungulate Recognition At The Sevilleta National Wildlife Refuge

By

Michael Gurule
B.S., Geography, University of New Mexico, 2020
M.S., University of New Mexico, 2023

Abstract

As camera traps have grown in popularity, their utilization has expanded to numerous fields, including wildlife research, conservation, and ecological studies. The information gathered using this equipment gives researchers a precise and comprehensive understanding about the activities of animals in their natural environments. For this type of data to be useful, camera trap images must be labeled so that the species in the images can be classified and counted. This has typically been done by teams of researchers and volunteers, and it can be said that the process is at best time-consuming. With recent developments in deep learning, the process of automatically detecting and identifying wildlife using Convolutional Neural Networks (CNN) can significantly reduce the workload of research teams and free up resources so that researchers can focus on the aspects of conservation.

Table of Contents

List of Figures

List of Tables

The Impact of Transfer Learning for Ungulate Recognition at Sevilleta National Wildlife Refuge

Michael Gurule

Department of Geography and Environmental Studies

Spring 2023

1.0 Introduction

With an increase in camera trap availability, researchers from a range of fields have been able to effectively deploy camera traps to observe wildlife remotely. Data collected in this manner can provide researchers with accurate, detailed, and up-to-date information regarding the location and behavior of wild animals and can improve the ability to study and conserve ecosystems (Norouzzadeh, 2018). Camera trap imagery needs to be labeled so that the wildlife in the imagery can be categorized and counted, and this has traditionally been done by groups of researchers and volunteers, which can be described as time-consuming at best (Norouzzadeh, 2018). The use of these camera traps in New Mexico has resulted in millions of unlabeled camera trap images waiting to be analyzed (Sanderson & Harris, 2013). With recent developments in deep learning, the process of automatically detecting and identifying wildlife using Convolutional Neural Networks (CNN) can significantly reduce the workload of research teams and free up resources so that researchers can focus on the aspects of the project (Islam and Valles 2020).

Figure 1 (A) Image of an African Oryx taken from a camera trap. (B) Image of a camera trap being deployed.

For CNNs to be able to accurately classify objects in images, they must first be trained with a large amount of labeled training samples that will produce updated weight parameters for learning (Han et al 2018). While camera trap imagery is in ample supply, the labels that identify the location of bounding boxes are not. Transfer learning allows researchers to take advantage of neural networks that have already been trained on larger datasets and then fine-tune them on a much smaller, domain-specific dataset. The method of training allows additional flexibility and increases the functionality of CNNs when classifying imagery, which can help researchers learn more about places that only have a small amount of data. This strategy can also help researchers who utilize camera traps for operational wildlife monitoring to improve species identification in a variety of environments based on the accumulated knowledge represented in labels from other sites. Furthermore, researchers may do more complicated detections like counting individuals or determining distance by utilizing CNNs that use bounding boxes for training rather than images with the entire sample labeled.

This research will evaluate the performance of the You Only Look Once version 5 (YOLOv5) algorithms in the detection and classification of animals located in the Sevilleta wildlife refuge. Due to its speed and precision, YOLOv5 was chosen for this project over other algorithms because, in a number of varying detection tasks, YOLOv5 has been demonstrated to achieve comparable performance with faster speeds (Kuznetsova et al., 2020) (Sa'Doun et al., 2021). YOLO divides the image into regions, applies a single neural network to the entire image once, and instantly determines the range of objects and probabilities of classes for each item

(Kuznetsova et al., 2020). The first model type will be trained on a domain-specific data set and be tested on imagery solely from the Sevilleta Wildlife Refuge. This is referred to as the domain model (DM). The second model type will utilize transfer learning for creating customized weights. This involves using a larger, regionally based data set instead of one that focuses only on wildlife in New Mexico. This method will allow the model to learn from a data set consisting of similar animals and be able to take advantage of a large amount of annotated imagery. The best-fit model (TM) will be further fine-tuned by exploring different learning rates that can impact the time needed for convergence. Therefore, this study assesses the potential of using transfer learning from national sources to enhance the detection and classification of camera trap imagery from the Sevilleta National Wildlife Refuge based on a limited collection of locally generated annotations.

2.0 Background

2.1 Introduction

Habitat is becoming depleted as the human population expands (Bar-Massada et al, 2014). This encroachment into adjoining natural lands drastically inhibits natural species movement and disrupts other natural processes (Butler, 2006). Because of this increase, human and animal populations alike have found themselves in numerous potentially dangerous situations they would not have found themselves in otherwise. As urbanization moves further into animal habitats, new anthropogenic mortality causes become more prevalent, producing major changes in natural ecology (Collins & Kays, 2011). The existing and increasing global scale and influence of human population expansion is incompatible with the survival of biological diversity, and the sixth mass extinction cannot be reversed (Naggs, 2017).

One of the most common methods for monitoring wildlife in its natural environment is to analyze imagery collected from camera traps (Beery et al., 2019). Since traps can be and often are deployed for extended periods, data sets collected in this manner tend to be very large, containing thousands of images per class. These data sets are often labeled manually by either a research team or a group of volunteers, which is described as time-consuming at best (Evans et al., 2020). Following recent advances in deep learning, researchers have been able to start exploring the process of automatic species identification through CNNs (Willi et al., 2019). The integration of camera trap imagery and Convolutional Neural Networks have been shown to be a promising tool for a potential method for persistent monitoring of

species presence, evaluation of species populations, and animal behavior without the need for a physical gathering of the subject (Evens et al., 2020).

This section focuses on how developments in deep learning, specifically CNNs, can allow researchers to non-intrusively analyze animal populations through camera trap imagery. The first key theme focuses on areas that are being used for wildlife conservation and areas with an increase in human-animal interaction (wildlife urban interface). The second key theme of this review examines the advantages and disadvantages of using camera trap imagery for wildlife monitoring. The third key theme of this review discusses how deep learning algorithms can be used for conservation and discusses how CNNs can enable scaled sampling using camera traps. The fourth and final key theme looks at how integrating transfer learning and data augmentation methods into domain-specific camera trap projects can increase classification accuracy.

2.2 Areas That Can Benefit from Monitoring Wildlife

Human development has had a significant impact on natural species movement, population dynamics, and other natural processes in the wilderness (Butler, 2006). The rate of human settlement expansion does not appear to be slowing down anytime soon, so one can only expect the wildlife-urban interface boundary to expand (Sella Veluswami, 2021). To limit the amount of damage to neighboring ecosystems, it is necessary to monitor the repercussions of increased human-animal contact and assess the effects humanity has on animal populations, habitats, and behavior.

With such dramatic growth in human population and urbanization, wildlife habitats are becoming even more constrained (Butler, 2006). The boundary between human communities and wildlife is referred to as the wildland–urban interface. This area consists of either interfaced housing developments, which are typically placed along the edges of continuous swathes of uncultivated land, or intermixed housing,

which is housing that is surrounded by natural or seminatural areas (Bar-Massada et al., 2014). Residential areas near natural boundaries can affect neighboring ecosystems in many ways, including exotic species introduction, wildlife subsidization, disease transfer, landcover conversion, fragmentation, and habitat loss (Bar-Massada et al., 2014).

The National Wildlife Refuge System is a designation for certain protected areas of the United States managed by the United States Fish and Wildlife Service. Fish and wildlife refuges make up 95 million acres, and it is estimated the total value of ecosystem services provided by the Refuge System in the United States is approximately $26.9 billion/year (Ingraham & Foster, 2008). The objective of wildlife refuges is to sustain and promote biodiversity-focused conservation initiatives as well as conservation activities that protect ecological integrity (Fischman, 2003). Some Fish and Wildlife programs, for instance, have focused on restoring native biological variety by assessing wildlife refuges for foreign species, evaluating the impacts of exotic wildlife on native animals, using removal strategies, and calculating the benefits of successful removal (Veitch & Clout, n.d.).

For conservation to be effective, a flexible and responsive recovery strategy, such as adaptive management, should include provisions for monitoring, allowing

researchers to track species throughout the recovery process (Campbell et al., 2002)(Gillson et al., 2019).By monitoring animals through this imagery, investigations into natural systems have provided new insights into migration patterns and species counts that can benefit the fields of ecology, zoology, and many others. Furthermore, conservation initiatives supported by CNNs can assist in the understanding of the complexities of natural ecosystems and act as a catalyst for the transformation of ecology (Islam & Valles, 2020).

2.3 Camera Trap Imagery for Monitoring Wildlife

While biologists and ecologists use many different methods of monitoring animal behavior, such as radio-tracking, wireless sensor network tracking, and animal sound recognition, one of the most common tools to monitor wildlife around the world is the camera trap (Islam & Valles, 2020) (Monterroso et al., 2009) (Li & Wu, 2015) (Handcock et al., 2009). Camera traps are heat- or motion-activated cameras placed in the wild to monitor and investigate animal populations and behavior (Beery et al., 2019). Existing camera-trap systems for wildlife monitoring have developed as a result of technological improvements in hardware and embedded software and are now commercially available at a reasonable cost, rapidly deployable, and easy to maintain, allowing them to be utilized by a wide variety of organizations (Chen et al., 2014). Moreover, while the cost of employing camera traps is relatively low, extracting information from these photographs remains an expensive, time-consuming, and manual task. (Norouzzadeh et al., 2018). Before the widespread use of CNNS, researchers devised a number of innovative and successful methods for automating the interpretation of animals from camera traps using raw pixel data

from images (Schneider et al., 2018). Earlier approaches to species classification required a domain expert to identify significant features for the desired classification, design an algorithm to extract these features from images and compare individual differences using statistical analysis (Schneider et al., 2018).

Images from camera traps can be used to help researchers learn more about an animal's behavior and social structure, which are not visible when using other monitoring techniques like GPS collars. The ability to study and conserve ecosystems would be enhanced if researchers had precise, detailed, and up-to-date information regarding the location and behavior of wild animals (Norouzzadeh et al., 2018). For example, imagery collected from camera traps can be used to identify, and locate threatened species, identify important habitats, monitor sites of interest, and analyze wildlife activity patterns (Beery et al., 2019). To create a better understanding of wildlife environments, it is essential to be able to acquire and analyze fine-scale, non-intrusive imagery. Other methods of monitoring wildlife include Manned Just in Time (MJIT) flights and Global Positioning System (GPS) collars, both of which collect wildlife observations in unique ways. MJIT refers to a technique for tracking wildlife with aerial vehicles over a predetermined area while tracking and monitoring wildlife in real-time, but it has been demonstrated that in some circumstances, the temporal resolution of sUAS collected images is constrained by the relatively short flight times, limited swath-width, and logistics of operating under existing regulations (Lippitt and Zhang, 2018) Hodgson et al., 2016). On the other hand, GPS collars provide highly precise position data by satellite triangulation, and researchers can analyze gain insights into a range of ecological

questions, such as animal migration patterns, home range size, and social structure (Naidoo et al., 2016) (Foley & Sillero-Zubiri, 2020). That being said, GPS collars have been known to fail prematurely (Johnson et al., 2002). In contrast, camera traps cover a small area in space, but they do so at infinite temporal resolution and for long durations when compared to the just-in-time counts that UAS are capable of.

Although camera trap imagery can be useful to researchers in general, there are a number of challenges in processing and interpreting it. For example, the vast majority of camera trap imagery that is gathered is unlabeled and does not contain any objects of interest (Zhu et al., 2011). To make the imagery useful for analysis, researchers or volunteers must go into the imagery and manually identify objects, which is a time-consuming operation. With the growing number of camera trap studies, it is becoming increasingly difficult to recruit enough participants to complete all projects on time (Willi et al., 2019). In addition to having to manually identify objects, roughly ~70% of camera trap images do not contain any objects, and this can be caused by false triggers (Beery et al., 2021). The manual labeling of the Snapshot Serengeti collection, which consists of 1.2 million images, has taken 28,000 registered and 40,000 unregistered volunteer citizen-scientists 2–3 months to identify each 6-month batch of photos (Norouzzadeh et al.,2018). As digital cameras improve and become more affordable, researchers will begin to use camera traps in their own domains, putting more strain on volunteer resources that identify imagery.

2.4 Deep Learning as an Alternative to Manual Classification

Deep Learning, a subfield of machine learning, is based on Artificial Neural Networks, a computational architecture inspired by how the human brain learns and

recognizes objects (O'Mahony et al., 2020). These networks consist of many neurons that conduct a simple operation and interact with one another to produce a decision, just like the human brain (O'Mahony et al., 2020). Neural networks, hierarchical probabilistic models, and unsupervised and supervised feature learning algorithms are all part of deep learning and have shown the potential to outperform the previous state of the art machine learning techniques in several tasks (Voulodimos et al., 2018). Deep Learning has expanded the boundaries of what was previously thought to be achievable in the field of digital image processing.

A Convolutional Neural Network (CNN) is a specific deep learning method that requires the user to train algorithms to recognize objects (e.g., animals, humans, and cars) in photos and classify the objects (e.g., species) present (Vélez et al., 2022). CNN platforms come in several different forms, and new ones are released on a regular basis, but they all have similar structures. CNNs are comprised of three main types of neural layers called convolutional layers, pooling layers, and fully connected layers (Voulodimos et al., 2018). Each layer of a CNN eventually converts the input volume to an output volume of neuron activation; this creates a fully connected layer that results in a 1D feature vector that can be categorized and classified (Voulodimos et al., 2018)

The convolutional layer consists of weighted filters that match the input data and helps identify features of varying complexities. Each filter identifies specific patterns or traits such as edges, corners, bright spots, dark spots, and shapes (Hussain et al. 2018). By stacking multiple convolutional layers, CNNS are able to learn more complex features and eventually will be able to identify complex objects like animals,

vehicles, or faces (Hussain et al. 2018). The max pooling layer utilizes a window (ex 2x2/3x3 pixels) that travels across the image, storing the maximum value of each step, reducing the spatial dimensions of the feature maps to leave only the most significant data, which can then be merged into one (LeCun et al., 2015). Each image's dimension is reduced while the most crucial information is retained (Hussain et al. 2018). The outputs of convolutional layers and max pooling layers are processed by the fully connected layers (see figure 2). Every neuron in one layer is connected to every other neuron in the following layer by fully connected layers (Voulodimos et al., 2018). The goal is to learn nonlinear combinations of information from previous layers that are relevant to identifying objects in an image.

Simply put, an image from a camera trap dataset is fed into the neural network, where it moves through different filters that creates feature maps and pooling layers that reduce the spatial dimensions of the image. Once the foregoing processes are completed, the final output can be identified and classified, and this can automatically be done to thousands of photos, saving researchers time improve the understanding of the population dynamics (Schneider et al., 2018a).

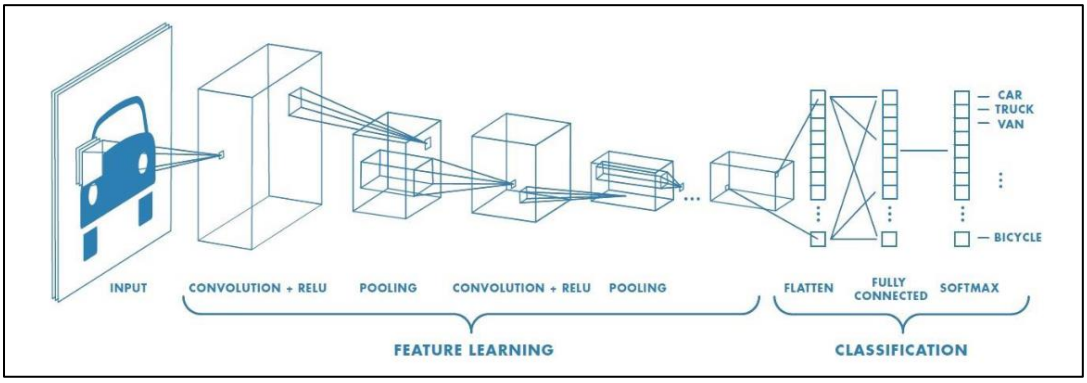

Figure 2 Basic structure of a Convolutional Neural Network (Hussain et al. 2018)

A CNN designed for species identification will output a set of activations, representing the observation of a particular species at a particular location and time, and are often used as a way to calculate population sizes in different regions (Evens et al, 2020). This is beneficial because it allows ecologists and other experts to observe animals in their native habitat without producing avoidance behaviors, habituation, or physiological changes in reaction to human-induced interactions (St. John, 2022). This has demonstrated great potential, but it has also revealed some limitations with CNNs that are particularly caused by camera trap images, such as varying illumination, weather, seasons, and a cluttered background (Evans et al., 2020) (Favorskaya & Pakhirka, 2019).

While CNNS may be trained to perform a variety of tasks, some are better than others at specific operations. For example, MegaDetector, created by Microsoft AI for Earth, provides a JSON file as an output, which indicates the locations of detected objects in the form of bounding boxes and is associated with confidence values for each detection (Velez et al., 2022). While this architecture can accurately detect objects, it is unable to classify species into distinct categories (Velez et al., 2022). While somewhat limited, this specific detector can still be beneficial in many fields. It was shown that 75% of the Snapshot Serengeti dataset, which consists of 3.2 million-images, were labeled empty by humans; therefore, automating the first stage alone can save 75% of human labor (Norouzzadeh et al., 2018). Additionally, MegaDetector detections can also be cropped to the predicted bounding box and

utilized in a different species classifier, reducing the species classifier's ability to overfit to the image background (Evans et al., 2020).

While the structure of CNNs is complex and training them requires users to have extensive training, the impact they can have on ecological studies is apparent (PIRES DE LIMA et al., 2020)(Schneider et al., 2018a). By incorporating CNNs into camera trap projects, the number of volunteers needed for labeling could be significantly reduced, allowing researchers to focus on other aspects of these types of projects. Manual labeling is a time-consuming task that improves in deep learning, and CNNs specifically, are well positioned to automate.

2.5 Enhancing Model Predictive Performance

The confidence of a Convolutional Neural Network's (CNN) detection and categorization is impacted by the excessive data hunger of deep learning models, which require massive amounts of high-quality data to function properly. (Pires de Lima et al., 2019). Neural networks can learn to identify spatial characteristics by adjusting their weights during the training phase due to the backpropagation of errors (Willi et al., 2019). Once training is complete, users then use additional imagery for validation. While this appears to be a simple process, complications can arise when working with various animal classes. When the model works well with the training data but fails to generalize validation data, it is referred to as overfitting (Islam & Valles, 2020). If there is an unbalanced amount of imagery used to train the model, it may learn to focus on specific patterns in the training data that are not relevant for other classes in the camera trap collection, which can also lead to overfitting (Willi et al., 2019).

2.5.1 The Power of Transfer Learning and Its Impacts on Computer Vision

Inspired by human beings' capabilities to transfer information across domains, transfer learning is the process of preconditioning a CNN on a preexisting repository of data that already consists of labeled imagery (Zhuang et al., 2021). For deep learning classification projects with insufficient data or computational limits, transfer learning has gradually become the preferred method of training CNNs (Liu et al., 2019). This process has shown that models that take advantage of transfer learning can increase output classification accuracy by 10.3 percent (Willi et al., 2019). Rather than starting the learning process on the model from scratch with random weight initialization, transfer learning allows users to start with learned features from weights generated from larger image collections and then adapt these features to suit the new set of imagery (Hussain et al., 2018). This process allows researchers who use domain specific data sets the benefits of higher accuracy classification without the need to collect the thousands of images needed to train accurate CNNS.

CNNs and transfer learning have reshaped the field of computer vision and have been effectively used for a variety of applications, such as the interpretation of camera trap data (Hussain et al., 2018). By leveraging pre-trained CNN models on large image datasets such as Snapshot Serengeti, transfer learning enables the efficient use of limited camera trap data for the automation of data analysis (Schneider et al., 2018b). This approach has enabled the development of CNN models capable of classifying and detecting species in camera trap images with high reliability and accuracy (Sharma et al., 2020). It has even been demonstrated that a

CNN can learn to recognize distinctive aspects of individual animals, such as their coat pattern or facial markings, by being trained on a sizable collection of annotated photos (Nepovinnykh et al., 2018). This makes it possible to create a database of distinct species and individuals for future population analysis. While transfer learning can be used to minimize overfitting and strengthen classification predictions, the primary benefit is not having to wait for more images to be labeled for training (Han et al., 2018).

2.5.2 Improving Deep Learning Performance Through Data Augmentation

To increase the accuracy of outputs in a localized region that doesn't have the benefit of using a domain focused repository to train networks, researchers can use preprocessing methods like data augmentation to artificially expand classes in a data set. Data augmentation is a regularization technique that uses label-preserving modifications to artificially enhance the data set by adding altered domain-specific pictures to an already existing collection of imagery. (Taylor & Nitschke, 2017). This process can expand existing datasets in a wide range of image variations that will help the model recognize objects in various shapes and forms (Islam & Valles, 2020). The basic augmentation process is accomplished by creating a copy of the original image that has been shifted, zoomed in/out, rotated, flipped, deformed, or tinted with a hue (Perez & Wang, 2017 Researchers in a variety of fields may employ generic data augmentation as a cost-effective substitute if transfer learning is not an option due to object rarity or distinctive settings (Taylor & Nitschke, 2017). By increasing the amount of domain specific training data artificially, the chance of

overfitting is significantly reduced. It was shown that data augmentation techniques that take advantage of transformations that alter the geometry of images were shown to improve neural network classification accuracy when compared to alterations of lighting and color (Taylor & Nitschke, 2017 (Perez & Wang, 2017). Data augmentation alone can aid future camera trap projects aimed at observing species that are difficult to photograph or in locations that are difficult to access, such as beneath water.

2.5.3 Optimizing Model Training with Learning Rates

The learning rate is a hyperparameter that specifies how much the model should change in response to the predicted error each time the model weights are updated. The learning rate parameter in gradient descent learning methods, such as error backpropagation, can have a substantial effect on generalization accuracy (Wilson & Martinez, 2001). In this case, using different learning rates translates to how fast YOLOv5 is able to learn what a mule deer looks like and how well it is at recognizing mule deer in new images. It has been shown that too small a learning rate will make a training algorithm converge slowly, while too large a learning rate will make the training algorithm diverge (Smith, 2017). Lowering learning rates below those which achieve the fastest convergence can improve generalization accuracy considerably, especially on complicated tasks (Wilson & Martinez, 2001). An unoptimized learning rate can require orders of magnitude more training time than one that is in an appropriate range. According to conventional thinking, the learning rate should be a single value that declines monotonically over training, and users should experiment

with a variety of learning rates so that optimal performance of the model can be achieved (Smith, 2017).

Deep learning systems function best when the training data has a balanced distribution of samples across all classes, but realistically camera traps will not collect an equal amount of imagery for each class (Schneider et al., 2020). This means that CNNS trained only on a large general repository will underperform when classifying images from different areas (Schneider et al., 2020). For researchers to achieve higher recognition accuracy, a combination of general repository transfer learning and data augmentation of domain specific classes is recommended (Perez and Wang 2017) (Hussain et al. 2018).

2.5.4 You Only Look Once

YOLOv5 is a resilient object detection deep learning network that has been utilized in a variety of applications, including pedestrian detection for self-driving cars, surveillance systems, and conservation (Vikram Reddy & Thale, 2021) (Sa'Doun et al., 2021). YOLOv5 is more user-friendly and requires far fewer processing resources than other Deep Learning neural networks while maintaining comparable results and operating faster than other networks (Choiński et al., 2021). Yet another explanation for why YOLOv5 is so effective is that it is a one-stage detection technique, which means it predicts bounding boxes and class probabilities for each object in a single forward pass of the network, whereas other networks use a two-stage method (Jiang et al., 2022). It has been demonstrated that, when comparing the speed and accuracy of the detection of wildlife from imagery, the YOLO architecture can perform just as well as other high-performance CNNs (Sa'Doun et

al., 2021). That is partly due to the fact that YOLO can either be trained from scratch or the user can take advantage of the Common Objects in Context (COCO) pre-trained model. The COCO dataset is a collection of challenging, high-quality images and labels that can be used for object detection or segmentation (Adamczyk, 2020). This repository includes 91 object categories and 330K images, and 200K labels. Stop signs, zebras, eyeglasses, elephants, and fire hydrants are among the categories included (Admin, 2018).

2.6 Conclusion

Despite the fact that CNNs have been readily accessible for several years, there are still no trained models that specialize in the detection of ungulates in New Mexico. Processing such vast amounts of data with CNNs necessitates substantial computer resources and knowledge, which may not be widely available. However, as the need for data analysis utilizing CNNs grows, more resources and experience will most likely become accessible, making it easier to process the data and extract valuable insights.

By training YOLOv5 with larger general image repositories, expanding domain specific data sets artificially, and optimizing learning rates, deep learning can provide new tools for researchers to label and categorize camera trap imagery in wildlife refuges without the need for extensive labeling campaigns. The process of collecting and analyzing large amounts of imagery from camera traps in an automated, accurate, and cost-effective manner can help accelerate the transformation of the fields of ecology, wildlife biology, zoology, conservation biology, and animal behavior into big data sciences (Norouzzadeh et al., 2018). This combination has the potential

to produce new information that can aid wildlife conservation efforts around the world and help create a balance between human-wildlife relations.

3.0 Research Design

This study aims to determine the effectiveness of using large collections of annotated camera trap imagery that can influence YOLOv5 predictions when determining the location and classification of ungulates in camera trap imagery from the Sevilleta wildlife refuge. In addition to evaluating the impact of transfer learning, YOLOv5 was fine-tuned using varying degrees of data augmentation and learning rates for practical employment in support of the USFWS**.**

3.1 Data

The primary image repository used for this project consists of 4 camera traps that are distributed around the Sevilleta wildlife refuge and consist of ~ 2000 images. These camera traps were set up in 2019 by the University of New Mexico (UNM) and the USFWS. Common species in this collection include but are not limited to mule deer, coyotes, elk, pronghorns, and African oryx. For this imagery to be used for training the CNN, it must contain separate bounding box information (category id, x, y, width, and height). For this portion of the project, images were labeled by volunteers from Zooniverse, which is a citizen science web portal that allows volunteers to participate in a wide range of research projects (see fig 3). Animal classification, x, y, width, and height of bounding box in are just a few examples of the label information that is provided (see fig 4). The Sevilleta labels on Zooniverse are 15 times redundant, majority rule, implying that a minimum of 15 volunteers must either build a bounding box around an object or say that there is no object in the image for it to be classified.

Figure 3 Zooniverse Labeling Interface

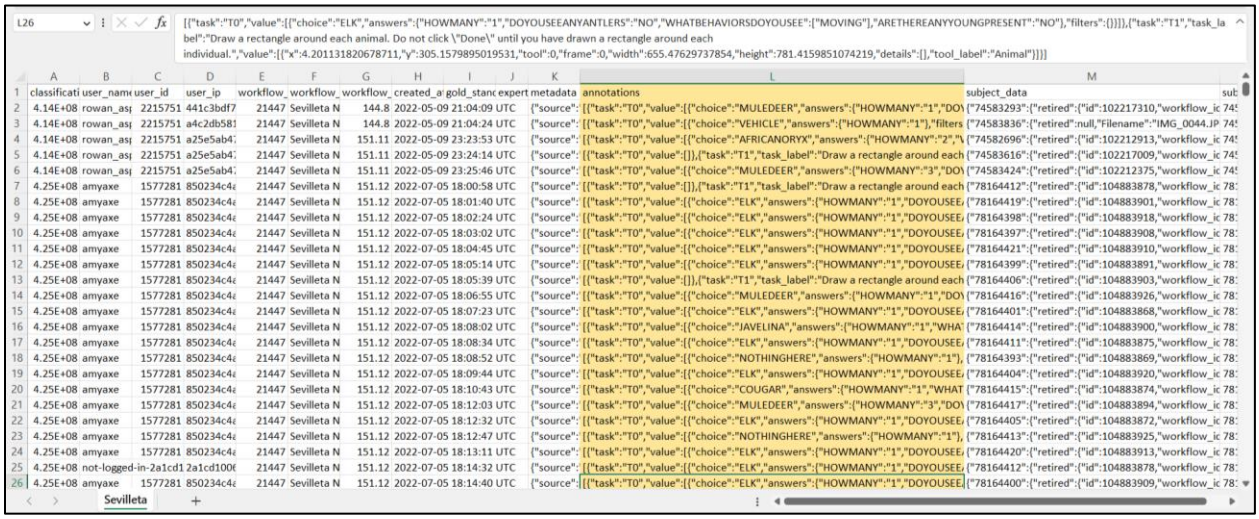

Figure 4 Zooniverse label example.

This Caltech Repository (Caltech Camera Traps), located in the Labeled Information Library of Alexandria: Biology and Conservation (LILA BC), contains 243,100 photos from 140 camera locations across the Southwest, with labels for 21 animal categories, primarily at the species level (the most common labels are opossum, raccoon, and coyote), and approximately 66,000 bounding box annotations, with approximately 70% of photos labeled "empty." The zipped data set is approximately

105 GB and can be downloaded from the Alexandria Labeled Information Library: Biology and Conservation ("Camera Traps Archives," n.d.).

Name of class	Category ID	Number in class
Pronghorn	0	439
African Oryx	1	694
Mule Deer	2	343
Other Than Ungulate	3	35

Table 1 Sevilleta Class Count

3.2 Data Pre-processing

The Caltech bounding box annotations are in JSON format and include the image id, class id, x coordinates, y coordinates, height, width, and information about the classes. YOLOv5 needs a single .txt file for each image where the annotation is normalized and only needs the bounding box x coordinates, y coordinates, height, width, and class id. The required single txt files were produced using Python and values and converted from COCO format ([x_min, y_min, width, height]) into YOLO format ([x_center, y_center, width, height]).

The Sevilleta labels provided by Zooniverse are 15x redundant, majority rule. This means that for an object to be labeled, it must meet the agreed-upon threshold of 15 identifications, with the majority agreeing on the category id. Of these, the 15 bounding boxes created median value was taken to create a new bounding box (Swanson et al, 2016). To do this, a script is used to count the number of labels for each potential class before designating the majority class as the class label for each training example (Sheng et al, 2019). Similar to the Caltech JSON, a Python script was used to filter out extraneous information and produce single Txt files and values

were converted to YOLO format. As previously stated, these single Txt files contain class ID, x coordinates, y coordinates, height, and width.

Location/Data Set	Number of Camera Traps	Image Count	Number of Labels in Data sets	Source	Label format
Caltech Repository	140	243,100	66,000	Library of Alexandria	COCO.csv
Sevilleta Wildlife Refuge Repository	4	2,292	2,292	United States Fish and Wildlife	COCO.json

Table 2 Camera Trap Data Description

3.3 Increasing Recognition Accuracy

CNNs trained solely on a large general repository will underperform when classifying images from different domains (Schneider et al., 2020). To mitigate this, data augmentation and fine-tuning of learning rates were employed. By utilizing data augmentation, the model can leverage a larger number of samples for training, thereby potentially enhancing overall performance. Fine-tuning the learning rate of the model facilitates faster convergence. Ultimately, employing both of these methods can stabilize the training and validation phases.

4.0 Project Execution

4.1 Data Management

For this project, the Caltech imagery was zipped and stored on an external hard drive. The imagery provided by the USFWS was compiled on one drive and backed up on an external hard drive. The Python scripts that were used for YOLOv5 are also accessible on a local machine, and the original code is also backed up on an external hard drive. Additionally, changes made to the data sets, including but not limited to data augmentation processes, were noted. The labels generated are open-sourced and located in the Wild Southwest repository and stored in the labeled Information Library of Alexandria: Biology and Conservation website.

4.2 Research Question

To progress towards the reliable use of Convolutional Neural Networks applied to camera trap imagery for monitoring wildlife in New Mexico, the proposed research is guided by the following question:

How does transfer learning based on regional wildlife samples affect the classification and detection accuracy of Convolutional Neural Networks for the detection and identification of wildlife from camera trap imagery at the Sevilleta National Wildlife Refuge?

Sub Questions

How do learning rates impact the robustness of transfer learning?

How does data augmentation affect transfer learning performance?

4.3 Study area

Imagery that was classified during the detection phase comes from the Sevilleta National Wildlife Refuge. The refuge itself is situated about 50 miles south of Albuquerque, NM, and consists of 230,000-acre refuge that includes four different biomes (Piñon–Juniper Woodlands, Colorado Plateau Shrub–Steppe lands, Chihuahuan Desert, Great Plains Grasslands) that intersect and support a wide array of biological diversity (“Sevilleta National Wildlife Refuge-Visitor Center Area,” 2022). Unlike other protected areas, the Sevilleta refuge was intended to preserve natural systems, allowing this sanctuary to remain as close to its original state as possible. This has led to a refuge where hundreds of species can thrive (“Sevilleta National Wildlife Refuge-Visitor Center Area,” n.d.).

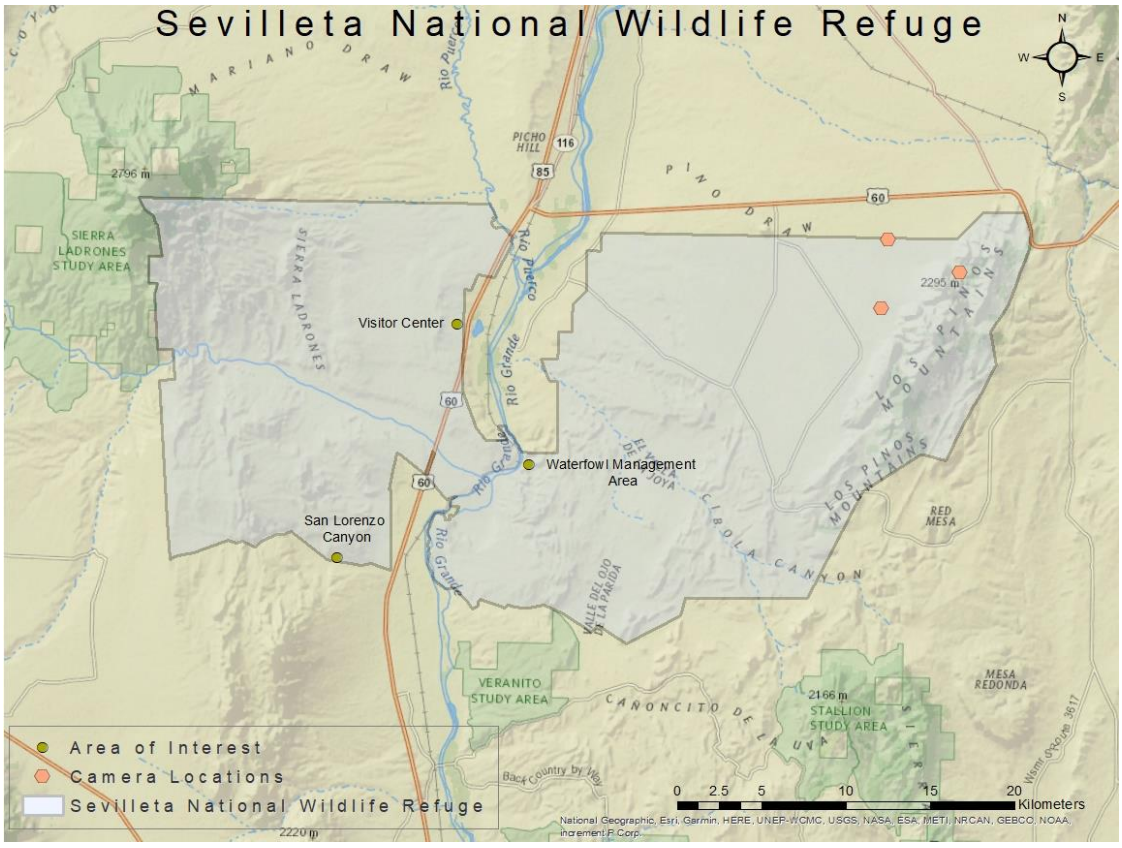

Figure 5 Sevilleta National Wildlife Refuge

4.4 Methods

This project systematically evaluates the effects of transfer learning on the calibration of YOLOv5 to detect and classify large ungulate species. Models that take advantage of both data augmentation and transfer learning were run 10x with a variety of learning rates so that training times are reduced, and overall performance is increased (Smith 2020). The optimal learning rate is identified by using different orders of magnitude, and effectiveness is determined by classification accuracies and mean average precision (mAP) (Wilson & Martinez, 2001). Transfer learning is used in two ways; the first method uses a pre-trained neural network that uses either COCO weights or Caltech weights. The second method also used the weights generated from either the Caltech data set or the COCO weights, but instead of using the entire network, the first ten layers of the backbone are frozen, and only the head of the network is trained. Models that use data augmentation are further fined tuned by using varying degrees of augmentation.

Model #	Model Description	Transfer Learning Pretrained	Transfer Learning Freeze layers	Data Augmentation	Learning rates
1	Sevilleta from scratch no data augmentation	-	-	-	0.01
2	Sevilleta COCO pretrained weights no data augmentation	X	-	-	0.01
3	Sevilleta COCO pretrained model data augmentation low	X	-	-	0.01
4	Sevilleta Caltech freeze 10 no data augmentation	-	X	-	0.01

Table 3 Model Description

Model #	Model Description	Transfer Learning Pretrained	Transfer Learning Freeze layers	Data Augmentation	Learning rates
5	Sevilleta Caltech pretrained model data augmentation low	X	-	X	0.01
6	Sevilleta Caltech pretrained model data augmentation high	X	-	X	0.01
7	Sevilleta from scratch data augmentation high	-	-	X	0.01
8	Sevilleta Caltech pretrained model data augmentation high Lr 0.001	X	-	X	0.001
9	Sevilleta Caltech pretrained model augmentation low Lr 0.0001	X	-	X	0.0001
10	Sevilleta Caltech pretrained model data augmentation low Lr 0.1	X	-	X	0.1
11	Sevilleta COCO freeze 10 no data augmentation High Lr 0.01	-	X	X	0.01
13	Sevilleta Caltech pretrained model data augmentation High Lr 0.001	X	-	X	0.001
14	Sevilleta Caltech pretrained model data augmentation High Lr 0.0001	X	-	X	0.0001
15	Sevilleta Caltech pretrained model data augmentation High Lr 0.1	X	-	X	0.1
16	Sevilleta Caltech pretrained model No data augmentation Lr 0.1	X	-	X	0.1

Table 3 Model Description Continued

4.5 Implementation of YOLOv5

To successfully run YOLOv5, the first thing needed to do is install its dependencies, which consist of python 3.7, pytorch 1.7, and base requirements (see figure 6)(*Ultralytics/Yolov5*, 2020/2022). They were installed on the Center for Advancement of Spatial Informatics Research & Education (ASPIRE) local BISON,

GPU-enabled computer (16 x 2.1 Ghz cores, 256GB RAM, 2 x 11GB 1080ti GPUs), using pip, a Python package installer. Initiation of models takes place in Anaconda3 command prompt.

```
# pip install -r requirements.txt

# Base ------------------------------------
matplotlib>=3.2.2
numpy>=1.18.5
opencv-python>=4.1.2
Pillow>=7.1.2
PyYAML>=5.3.1
requests>=2.23.0
scipy>=1.4.1
torch>=1.7.0
torchvision>=0.8.1
tqdm>=4.41.0

```

Figure 6 YOLOv5 base requirements

4.6 Hyper parameter configuration

Data augmentation, adjustments to learning rates, and other hyperparameters are defined via a *yaml* file. A *yaml* file is a human-readable data serialization language that is commonly used to create configuration files. The data folder in the master folder contains *hyp.Objects365.yaml*, *hyp.scratch-low.yaml*, *hyp.scratch-high.yaml, hyp.scratch-med.yaml, hyp.VOC.yaml.* All of these are preconfigured

hyperparameter settings that can be used for the model. In this file, data augmentation techniques like mosaic, shear, saturation, and translate are provided.

4.7 Training

All models are trained using 70 percent of the labels from the camera trap collections, with a randomly selected 20 percent held out for validation and 10 percent for detection/testing. The beginning segment of the model makes use of the Python training.py script. This is where the models are presented with the images and the corresponding labels so that weights can be updated and generated. Once completed, the model's performance is evaluated on mAP, bounding box regression loss, objectness loss, and classification loss (PhD, 2022) (Solawetz et al., 2020). The transfer learning models employ either the COCO weights or the Caltech best.py weights, and the newly trained model employs random weights. Data augmentation and learning rates that are to be used during training are applied in the hyperparameter yaml file. This file allows for the usage of high, mid, low, or custom augmentations and learning rates. All of which are executed through the training command line for training.

4.8 Validation

After training, the network is applied to 20 percent of the imagery with known labels and performance. Once validation is completed, mAP, bounding box regression loss, objectness loss, classification loss, and precision and recall are compared between the training set and the validation set. Overall performance is assessed using a confusion matrix of known species labels to modeled labels, in which each prediction label is compared to the appropriate ground truth label (Sa'Doun et al., 2021).

Commission errors identify false positives, and omission errors represent false negatives in the actual modeled results when compared to known examples.

4.9 Detection

Traditionally, the detect.py script is used for inference, but this does not generate useful statistics like mAP values, so instead, the val.py script is used for the detection phase. This is only possible because the test imagery has a corresponding annotation. It is necessary to do this so that all the model's inference capabilities can be compared. Once all models have been run, the confusion matrix for each run is compared.

5.0 Results

The best fit model was derived from a combination of using a pre-trained model that was based on the Caltech repository (transfer learning), high levels of data augmentation, and a learning rate of 0.01 (Table 4). The mAP values produced from the test set showed that the average for all classes was 0.567. Since the OTHERTHANUNGULATE class accounted for less than 5% of the whole training set, an additional average was calculated with it removed. This raised the mAP value to 0.735. P values were 0.756, indicating that the model's predictions were right 76% of the time. R values were 0.627, reflecting how well the true bounding box was predicted. Values for mAP@.5:95 were 0.384, indicating that the model could not effectively predict bounding box overlap (IOU).

Model number	mAP@.5	mAP@.5 W/O OTHERTHANUNGULATE	mAP@.5:.95	Precision(P)	Recall(R)
1	0.475	0.611	0.276	0.706	0.603
2	0.52	0.699	0.35	0.676	0.642
3	0.545	0.707	0.373	0.604	0.615
4	0.501	0.656	0.327	0.712	0.618
5	0.527	0.688	0.36	0.748	0.563
6	0.526	0.692	0.339	0.712	0.649
7	0.542	0.699	0.366	0.739	0.657
8	0.537	0.706	0.371	0.702	0.653
9	0.425	0.558	0.284	0.598	0.542
10	0.531	0.698	0.346	0.756	0.601
11	0.514	0.671	0.339	0.75	0.633

Table 4 Results of All Models

Model number	mAP@.5	mAP@.5 W/O OTHERTHANUNGULATE	mAP@.5:.95	Precision(P)	Recall(R)
12	0.567	0.735	0.384	0.756	0.627
13	0.553	0.727	0.383	0.744	0.617
14	0.47	0.62	0.338	0.691	0.592
15	0.544	0.71	0.377	0.738	0.652
16	0.493	0.643	0.335	0.704	0.632

Table 4 Results of All Models continued.

The model was able to detect objects from 3 out of the 4 classes, which can be seen in Figure 7. Detection among the 3 classes had varying degrees of accuracy, but all fell within an acceptable range. African oryx had the highest precision scores (0.98). While pronghorn and mule deer had significantly lower scores (0.89, 0.73). This might be the case given the particular unique characteristics of the African Oryx and the inclusion of more training samples—about 350 more in total (see Table 1). The fourth category, other than ungulate, had no detection, but this could be due to the extremely small sample size that was used. Additionally, this model had many issues with background detection, meaning that for many of the images, the model did not think there were any objects present when in fact, there were. Furthermore, excessive detection in each class afflicted the model as a whole. The values presented in the confusion matrix indicate that the model mistook portions of the backdrop vegetation for animal features.

For this project, transfer learning proved to be greatly advantageous. Models that only used this approach saw an increase in mAP of ~ 9%, but the major benefit was that mAP, mAP .5: .95, Precision, Recall, Bbox loss, Object loss, and Class loss all

remained stable and consistent during training. Without data augmentation and without freezing the first ten layers, the model that used the COCO dataset (2) slightly outperformed the Caltech repository model (16). Given that the COCO data set has ~ 200,000 annotations whereas the Caltech datasets have ~ 60,000, this suggests that researchers can get similar conclusions using only roughly 25% of the data.

Freezing the models' backbone and only training the head of the network showed increases in overall performance and prediction accuracy in both models. The model trained on the Caltech repository, model (4), had a slightly higher overall score in performance when compared to model (19), which froze the first ten layers and used COCO weights. That being said, class predictions were similar. The most significant advantage of this method was decreased training time, but since this data set was small, it did not make a significant difference. The most benefit would be seen when applied to a data set that has tens of thousands of images or when one is willing to trade minor drops in precision for time. Overall using transfer learning via freezing the backbone is beneficial, but this particular method still fell short when compared to the network when it was first pre-trained on Caltech weights.

Unsurprisingly the data augmentation techniques applied to these data sets significantly increased overall performance. This can be seen in the overall performance of model (01) and model (07) (table 4), which is about an 8% increase in the mAP@.5 W/O OTHERTHANUNGULATE. The effectiveness of the best-fit model (12) can also be linked to this method. Ultra-high augmentation was

attempted, but distorted images in such a way that the performance of the model was greatly decreased, and overall training became erratic and unreliable.

The best-fit model was negatively impacted by a change in learning rates of an order of magnitude. The model (15) using a larger learning rate (0.1) provided adequate performance but still fell short when compared to the best-fit model (12). Models that used a lower learning rate (0.001, 0.0001) showed significant decreases in prediction accuracy and overall performance during training. This is mainly due to the fact that YOLOv5 is already a proven model, as well as the fact that the data set is so small there were few changes in the amount of time the model needed to converge. However, if researchers adjusted the learning rate of a model that was analyzing a dataset with tens of thousands of samples, the model's time to converge would differ noticeably.

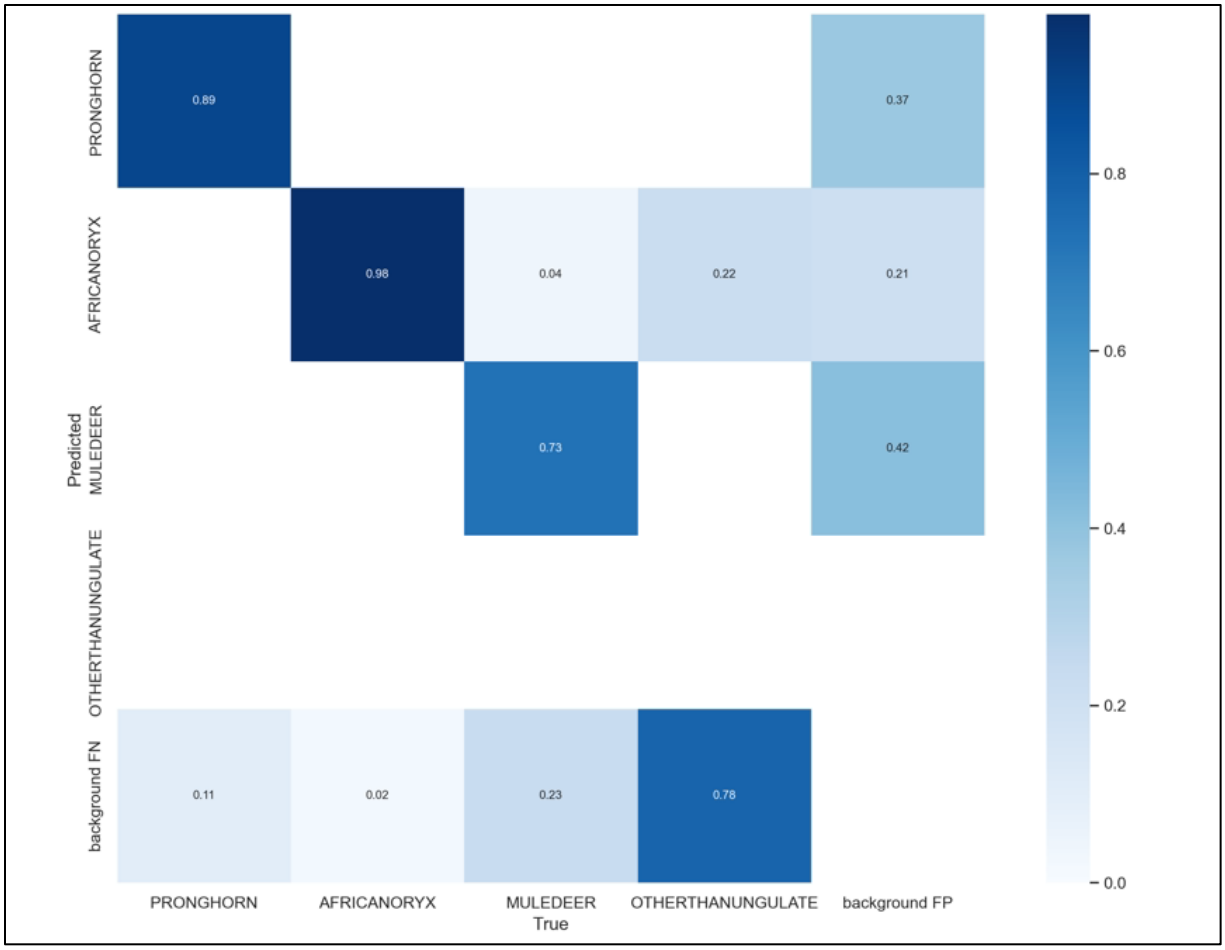

Figure 7 Best fit model confusion matrix from detection. This matrix can be used to evaluate the accuracy of object detection by showing the true positives (correctly detected objects), false positives (incorrectly detected objects), false negatives (missed objects), and true negatives (correctly rejected non-object areas).

6.0 Discussion

The purpose of this research was to use cutting-edge techniques to improve the state-of-the-art CNN, YOLOv5s, dependability when trained on infinitesimal data sets. It was demonstrated that YOLOv5 could produce acceptable detection and classification results using a combination of transfer learning and data augmentation. Since the Sevilleta collection is so small, we cannot completely rule out the possibility of overfitting, but we believe that this will have a minimal consequence because new images will still originate from the same camera traps.

Additional future work includes retraining the model on a new larger domain-specific repository called The Wild Southwest. The Wild Southwest is hosted on a citizen science platform called Zooniverse. Zooniverse is currently being labeled by ~6,500 volunteers. In total, it contains 87,000 images and as of May 2023 it is 17% completed. Doing this can increase the model's understanding of domain specific features, such as the shape of antlers that are only in New Mexico. Using a domain specific repository also reduces the bias and variance, which keeps the model from learning features that are not relevant to the task, and it also helps with the model's generalization (Shallu & Mehra, 2018).

Additionally, approaches like active learning can be utilized to improve the model's detection/classification capabilities and further raise performance. An active learning strategy automatically chooses additional images that, when annotated and incorporated, will improve the recognition performance after starting the learning process with an original model that was only trained with a small, labeled training dataset (Auer et al., 2021). Active learning can greatly outperform baseline models

because it uses the most useful examples and selects them so that only a few annotations are required (Auer et al., 2021). It was shown that binary region-specific active learning-based classification models outperformed well-known binary classifiers like Megadetector (Auer et al., 2021). A similar approach might show promise when integrated with the ungulate classifier.

We believe that YOLOv5 can be employed to detect ungulates in images that have no annotations and that have been taken at the Sevilleta Wildlife Refuge. This will provide Southwest researchers insight into ungulate populations and interactions. Population monitoring in this region is crucial for establishing the efficiency of conservation initiatives and can be utilized as an early warning system so that these systems can be protected.

7.0 Appendix YOLOv5 Performance

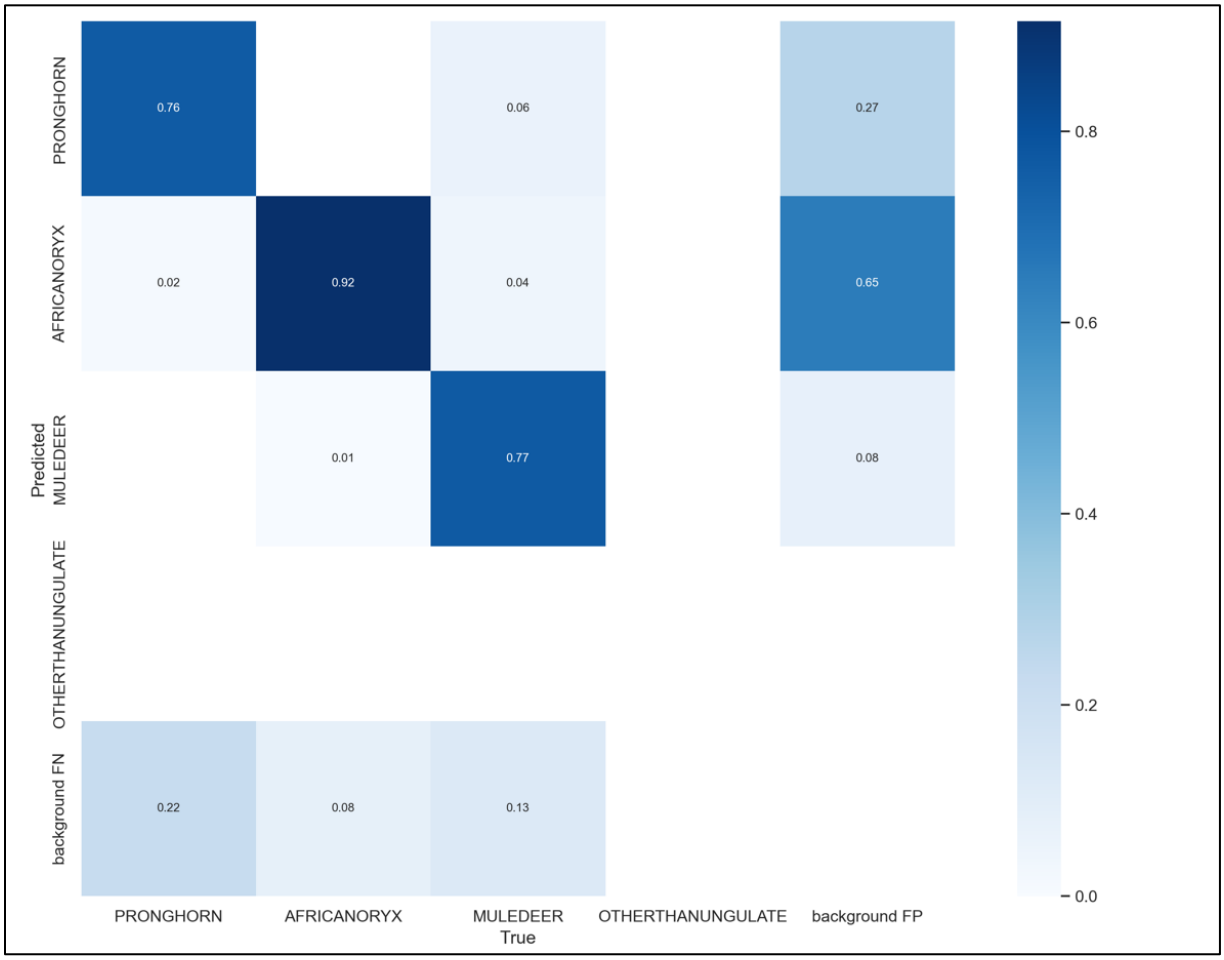

Figure 8 Confusion matrix that was generated from the training and validation phases.

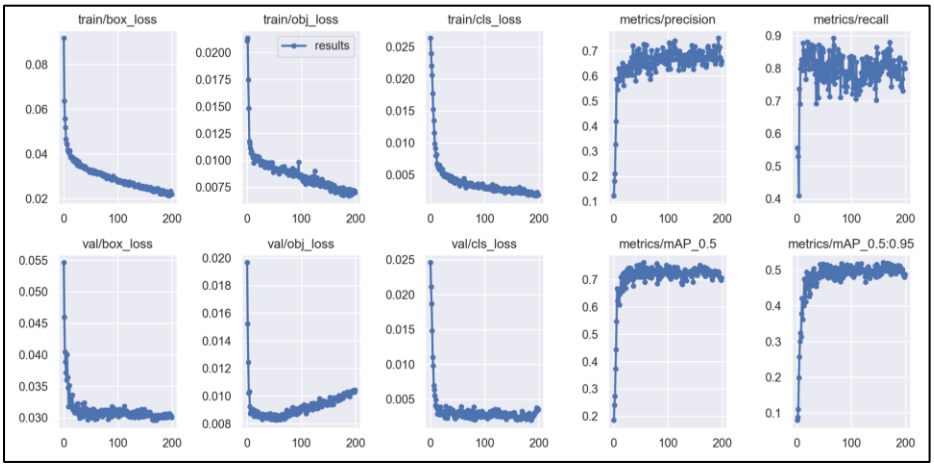

Figure 9 Detailed analysis of YOLOv5 performance from the training and validation phases.

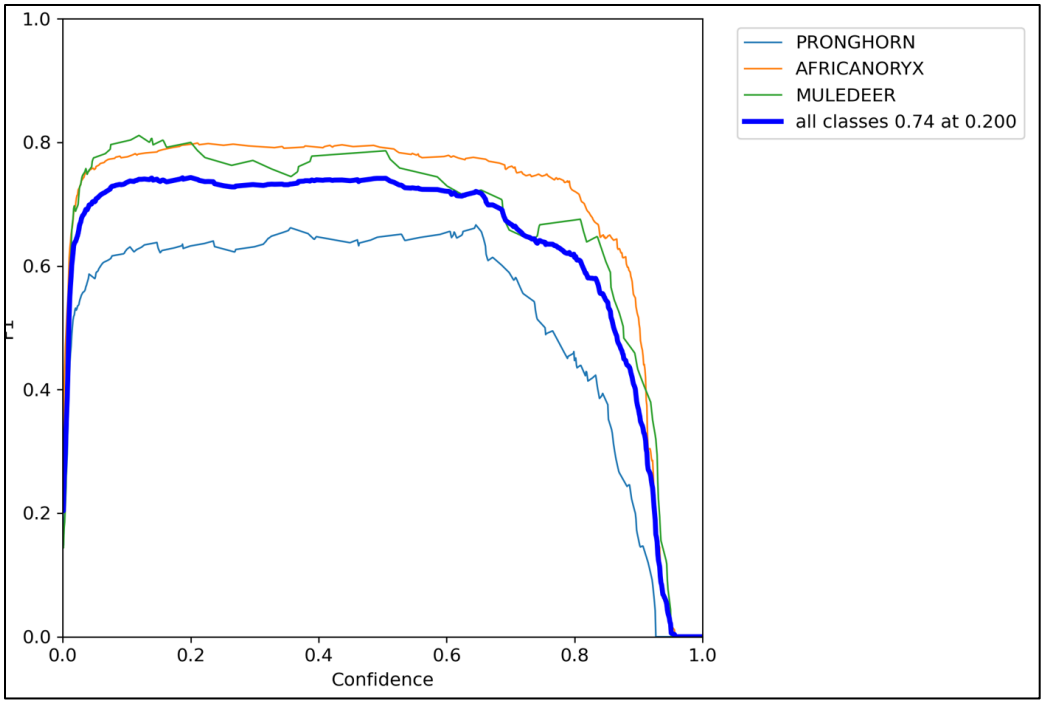

Figure 10 F1 curve provides insights into the overall detection performance of the mode at different confidence thresholds, enabling model selection, optimization, and comparison across different experiments or variations of the algorithm.

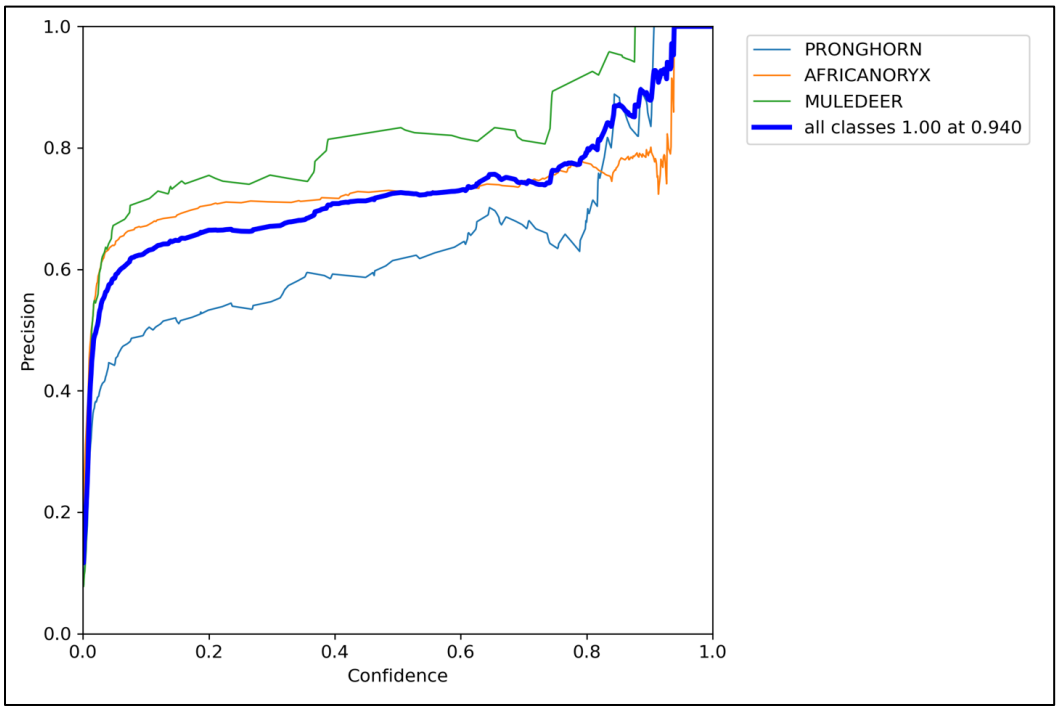

Figure 11 Precision curve shows the precision achieved by YOLOv5 at different confidence thresholds during the training and validation phases.

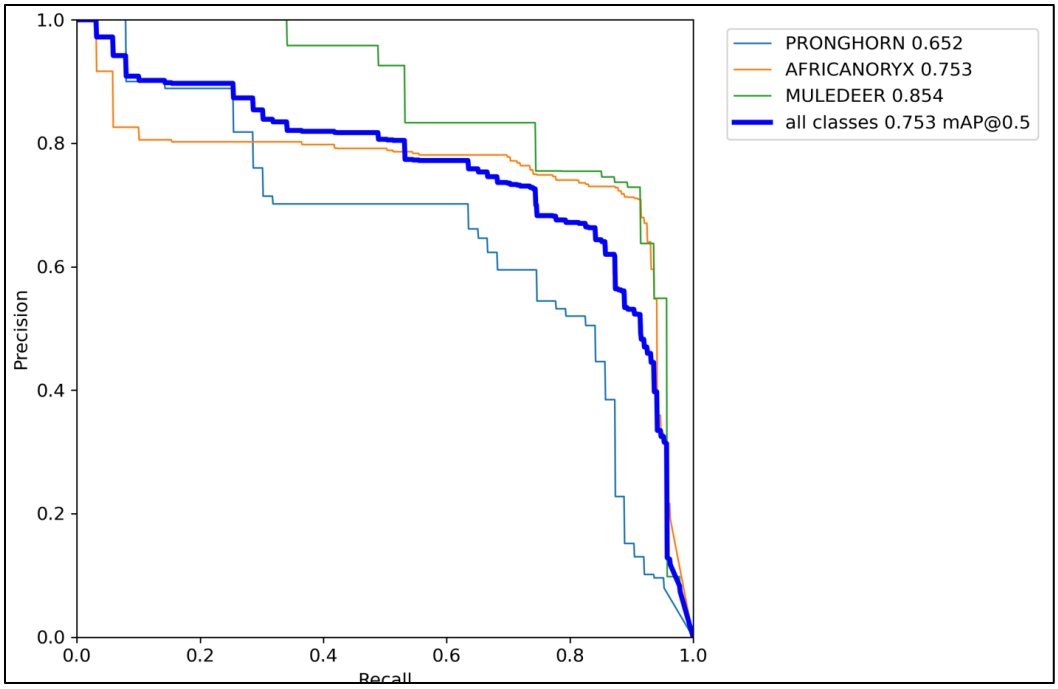

Figure 12 Precision and recall (PR) curve shows the relationship between precision and recall at different confidence thresholds.

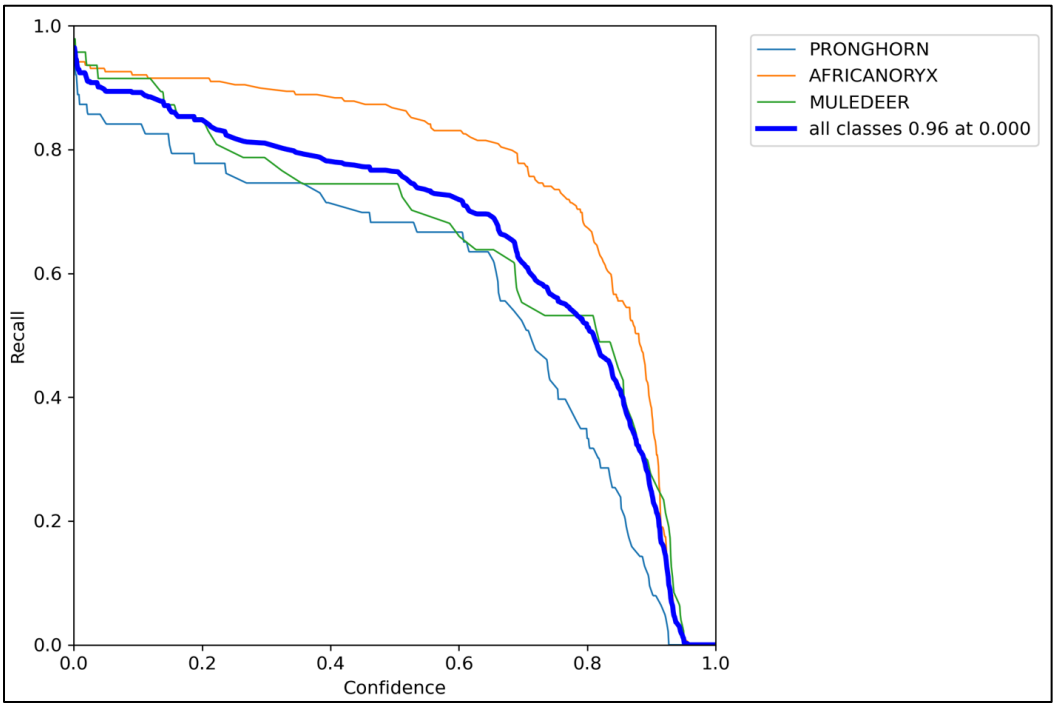

Figure 13 Recall curve displays the recall achieved by YOLOv5 at different confidence thresholds.

8.0 References

Adamczyk, J. (2020, July 30). Getting started with COCO dataset. Medium. https://towardsdatascience.com/getting-started-with-coco-dataset-82def99fa0b8

Admin, T. (2018, April 12). What Object Categories / Labels Are In COCO Dataset? Amikelive | Technology Blog. https:a//tech.amikelive.com/node-718/what-object-categories-labels-are-in-coco-dataset/

Auer, D., Bodesheim, P., Fiderer, C., Heurich, M., & Denzler, J. (2021). Minimizing the Annotation Effort for Detecting Wildlife in Camera Trap Images with Active Learning. Gesellschaft für Informatik, Bonn. https://doi.org/10.18420/informatik2021-042

Bar-Massada, A., Radeloff, V. C., & Stewart, S. I. (2014). Biotic and Abiotic Effects of Human Settlements in the Wildland–Urban Interface. BioScience, 64(5), 429–437. https://doi.org/10.1093/biosci/biu039

Beery, S., Agarwal, A., Cole, E., & Birodkar, V. (2021). The iWildCam 2021 Competition Dataset. ArXiv:2105.03494 [Cs]. http://arxiv.org/abs/2105.03494

Beery, S., Morris, D., & Yang, S. (2019). Efficient Pipeline for Camera Trap Image Review. ArXiv:1907.06772 [Cs]. http://arxiv.org/abs/1907.06772

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), Computer Vision – ECCV 2018 (Vol. 11220, pp. 472–489). Springer International Publishing. https://doi.org/10.1007/978-3-030-01270-0_28

Butler, D. R. (2006). Human-induced changes in animal populations and distributions, and the subsequent effects on fluvial systems. Geomorphology, 79(3), 448–459. https://doi.org/10.1016/j.geomorph.2006.06.026

Camera Traps Archives. (n.d.). LILA BC. Retrieved April 10, 2022, from https://lila.science/category/camera-traps/

Campbell, S. P., Clark, J. A., Crampton, L. H., Guerry, A. D., Hatch, L. T., Hosseini, P. R., Lawler, J. J., & O'Connor, R. J. (2002). An Assessment of Monitoring Efforts in Endangered Species Recovery Plans. Ecological Applications, 12(3), 674–681. https://doi.org/10.1890/1051-0761(2002)012[0674:AAOMEI]2.0.CO;2

Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014). Deep convolutional neural network based species recognition for wild animal monitoring. 2014 IEEE International Conference on Image Processing (ICIP), 858–862. https://doi.org/10.1109/ICIP.2014.7025172

Collins, C., & Kays, R. (2011). Causes of mortality in North American populations of large and medium-sized mammals. Animal Conservation, 14(5), 474–483. https://doi.org/10.1111/j.1469-1795.2011.00458.x

Choiński, M., Rogowski, M., Tynecki, P., Kuijper, D. P. J., Churski, M., & Bubnicki, J. W. (2021). A First Step Towards Automated Species Recognition from Camera Trap Images of Mammals Using AI in a European Temperate Forest. In K. Saeed & J. Dvorský (Eds.), Computer Information Systems and Industrial Management (pp. 299–310). Springer International Publishing. https://doi.org/10.1007/978-3-030-84340-3_24

Handcock, R. N., Swain, D. L., Bishop-Hurley, G. J., Patison, K. P., Wark, T., Valencia, P., Corke, P., & O'Neill, C. J. (2009). Monitoring Animal Behaviour and Environmental Interactions Using Wireless Sensor Networks, GPS Collars and Satellite Remote Sensing. Sensors, 9(5), Article 5. https://doi.org/10.3390/s90503586

Evans, B. C., Tucker, A., Wearn, O. R., & Carbone, C. (2020). Reasoning About Neural Network Activations: An Application in Spatial Animal Behaviour from Camera Trap Classifications. In I. Koprinska, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R. P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P. M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, E. Masciari, … J. A. Gulla (Eds.), ECML PKDD 2020 Workshops (pp. 26–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-65965-3_2

Favorskaya, M., & Pakhirka, A. (2019). Animal species recognition in the wildlife based on muzzle and shape features using joint CNN. Procedia Computer Science, 159, 933–942. https://doi.org/10.1016/j.procs.2019.09.260

Fischman, R. (2003). The National Wildlife Refuges: Coordinating A Conservation System Through Law. Island Press.

Gillson, L., Biggs, H., Smit, I. P. J., Virah-Sawmy, M., & Rogers, K. (2019). Finding Common Ground between Adaptive Management and Evidence-Based Approaches to Biodiversity Conservation. Trends in Ecology & Evolution, 34(1), 31–44. https://doi.org/10.1016/j.tree.2018.10.003

Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications, 95, 43–56. https://doi.org/10.1016/j.eswa.2017.11.028

Hussain, M., Bird, J., & Faria, D. (2018, June 16). A Study on CNN Transfer Learning for Image Classification.

Ingraham, M. W., & Foster, S. G. (2008). The value of ecosystem services provided by the U.S. National Wildlife Refuge System in the contiguous U.S. Ecological Economics, 67(4), 608–618. https://doi.org/10.1016/j.ecolecon.2008.01.012

Islam, S. B., & Valles, D. (2020). Identification of Wild Species in Texas from Camera-trap Images using Deep Neural Network for Conservation Monitoring. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 0537–0542. https://doi.org/10.1109/CCWC47524.2020.9031190

Johnson, C. J., Heard, D. C., & Parker, K. L. (2002). Expectations and realities of GPS animal location collars: Results of three years in the field. Wildlife Biology, 8(2), 153–159. https://doi.org/10.2981/wlb.2002.011

Jiang, L., Liu, H., Zhu, H., & Zhang, G. (2022). Improved YOLO v5 with balanced feature pyramid and attention module for traffic sign detection. MATEC Web of Conferences, 355, 03023. https://doi.org/10.1051/matecconf/202235503023

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), Article 7553. https://doi.org/10.1038/nature14539

Lippitt, C. D., & Zhang, S. (2018). The impact of small unmanned airborne platforms on passive optical remote sensing: A conceptual perspective. International Journal of Remote Sensing, 39(15–16), 4852–4868. https://doi.org/10.1080/01431161.2018.1490504

Liu, X., Jia, Z., Hou, X., Fu, M., Ma, L., & Sun, Q. (2019). Real-time Marine Animal Images Classification by Embedded System Based on Mobilenet and Transfer Learning. OCEANS 2019 - Marseille, 1–5. https://doi.org/10.1109/OCEANSE.2019.8867190

Li, Y., & Wu, Z. (2015). Animal sound recognition based on double feature of spectrogram in real environment. 2015 International Conference on Wireless Communications & Signal Processing (WCSP), 1–5. https://doi.org/10.1109/WCSP.2015.7341003

Majeed, F., Khan, F. Z., Iqbal, M. J., & Nazir, M. (2021). Real-Time Surveillance System based on Facial Recognition using YOLOv5. 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), 1–6. https://doi.org/10.1109/MAJICC53071.2021.9526254

Monterroso, P., Brito, J. C., Ferreras, P., & Alves, P. C. (2009). Spatial ecology of the European wildcat in a Mediterranean ecosystem: Dealing with small radio-tracking datasets in species conservation. Journal of Zoology, 279(1), 27–35. https://doi.org/10.1111/j.1469-7998.2009.00585.x

Naggs, F. (2017). Saving living diversity in the face of the unstoppable 6th mass extinction: A call for urgent international action. The Journal of Population and Sustainability, 1(2), 67–81.

Nepovinnykh, E., Eerola, T., Kälviäinen, H., & Radchenko, G. (2018). Identification of Saimaa Ringed Seal Individuals Using Transfer Learning. In J. Blanc-Talon, D. Helbert, W. Philips, D. Popescu, & P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems (pp. 211–222). Springer International Publishing. https://doi.org/10.1007/978-3-030-01449-0_18

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences of the United States of America, 115(25), E5716–E5725.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2020). Deep Learning vs. Traditional Computer Vision. In K. Arai & S. Kapoor (Eds.), Advances in Computer Vision (Vol. 943, pp. 128–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-17795-9_10

Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. ArXiv:1712.04621 [Cs]. http://arxiv.org/abs/1712.04621

Petso, T., Jamisola, R. S., Mpoeleng, D., & Mmereki, W. (2021). Individual Animal and Herd Identification Using Custom YOLO v3 and v4 with Images Taken from a

UAV Camera at Different Altitudes. 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), 33–39. https://doi.org/10.1109/ICSIP52628.2021.9688827

Pires de Lima, R., Suriamin, F., Marfurt, K. J., & Pranter, M. J. (2019). Convolutional neural networks as aid in core lithofacies classification. Interpretation, 7(3), SF27–SF40. https://doi.org/10.1190/INT-2018-0245.1

PIRES DE LIMA, R., WELCH, K. F., BARRICK, J. E., MARFURT, K. J., BURKHALTER, R., CASSEL, M., & SOREGHAN, G. S. (2020). CONVOLUTIONAL NEURAL NETWORKS AS AN AID TO BIOSTRATIGRAPHY AND MICROPALEONTOLOGY: A TEST ON LATE PALEOZOIC MICROFOSSILS. PALAIOS, 35(9), 391–402. https://doi.org/10.2110/palo.2019.102

Sa'Doun, M., Lippitt, C., Paulus, G., & Anders, K.-H. (2021). A Comparison of Convolutional Neural Network Architectures for Automated Detection and Identification of Waterfowl in Complex Environments. GI_Forum, 1, 152–166. https://doi.org/10.1553/giscience2021_02_s152

Sanderson, J., & Harris, G. (2013). Automatic data organization, storage, and analysis of camera trap pictures. Journal of Indonesian Natural History, 1(1), Article 1.

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. Ecology and Evolution, 10(7), 3503–3517. https://doi.org/10.1002/ece3.6147

Schneider, S., Taylor, G. W., & Kremer, S. (2018). Deep Learning Object Detection Methods for Ecological Camera Trap Data. 2018 15th Conference on Computer and Robot Vision (CRV), 321–328. https://doi.org/10.1109/CRV.2018.00052

Schneider, S., Taylor, G. W., & Kremer, S. (2018). Deep Learning Object Detection Methods for Ecological Camera Trap Data. 2018 15th Conference on Computer and Robot Vision (CRV), 321–328. https://doi.org/10.1109/CRV.2018.00052

Sella Veluswami, J. R. (2021). Human Wildlife Conflict Reduction Technology using YOLO Machine Learning Model. International Journal of Natural Sciences, 12, 36327–36333.

Sharma, R., Pasi, N., & Shanu, S. (2020). An Automated Animal Classification System: A Transfer Learning Approach. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3545100

Shallu, & Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express, 4(4), 247–254. https://doi.org/10.1016/j.icte.2018.10.007

Sheng, V. S., Zhang, J., Gu, B., & Wu, X. (2019). Majority Voting and Pairing with Multiple Noisy Labeling. IEEE Transactions on Knowledge and Data Engineering, 31(7), 1355–1368. https://doi.org/10.1109/TKDE.2017.2659740

Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. ArXiv:1506.01186 [Cs]. http://arxiv.org/abs/1506.01186

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay (arXiv:1803.09820). arXiv. http://arxiv.org/abs/1803.09820

Solawetz, J., JUN 10, J. N., & Read, 2020 7 Min. (2020, June 10). How to Train YOLOv5 On a Custom Dataset. Roboflow Blog. https://blog.roboflow.com/how-to-train-yolov5-on-a-custom-dataset/

Sevilleta National Wildlife Refuge-Visitor Center Area. (n.d.). EBird Hotspots. Retrieved April 9, 2022, from https://ebirdhotspots.com/birding-in-new-mexico/usnm-socorro-county/usnm-sevilleta-national-wildlife-refuge-visitor-center-area/

St. John, A. (2022). Exploring the Relationship Between Human Disturbance and Wildlife Behavior in Protected Areas. Oregon State University.

Taylor, L., & Nitschke, G. (2017). Improving Deep Learning using Generic Data Augmentation. ArXiv:1708.06020 [Cs, Stat]. http://arxiv.org/abs/1708.06020

Ultralytics/yolov5. (2022). [Python]. Ultralytics. https://github.com/ultralytics/yolov5 (Original work published 2020)

Veitch, E. C. R., & Clout, M. N. (n.d.). Turning the tide: The eradication of invasive species. 420.

Vélez, J., Castiblanco-Camacho, P. J., Tabak, M. A., Chalmers, C., Fergus, P., & Fieberg, J. (2022). Choosing an Appropriate Platform and Workflow for Processing Camera Trap Data using Artificial Intelligence. ArXiv:2202.02283 [Cs]. http://arxiv.org/abs/2202.02283

Vikram Reddy, E. R., & Thale, S. (2021). Pedestrian Detection Using YOLOv5 For Autonomous Driving Applications. 2021 IEEE Transportation Electrification Conference (ITEC-India), 1–5. https://doi.org/10.1109/ITEC-India53713.2021.9932534

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 2018, 1–13. https://doi.org/10.1155/2018/7068349

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., & Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution, 10(1), 80–91. https://doi.org/10.1111/2041-210X.13099

Wilson, D. R., & Martinez, T. R. (2001). The need for small learning rates on large problems. IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), 1, 115–119 vol.1. https://doi.org/10.1109/IJCNN.2001.939002

Zhu, Y., Chen, Y., Lu, Z., Pan, S. J., Xue, G.-R., Yu, Y., & Yang, Q. (n.d.). Heterogeneous Transfer Learning for Image Classification. 6.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, 109(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555

Zooniverse. (n.d.). Retrieved April 26, 2023, from https://www.zooniverse.org/projects/rowan-aspire/the-wild-southwest