National College of Ireland

Deer Surveillance System in Public Parks Using Deep Learning

MSc Research Project
MSc. in Data Analytics

Mardwin Alejandro Cardenas Rodriguez
Student ID: 20144237

School of Computing
National College of Ireland

Supervisor: Majid Latifi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Mardwin Alejandro Cardenas Rodriguez
Student ID:	20144237
Programme:	MSc. in Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Majid Latifi
Submission Due Date:	31/01/2022
Project Title:	Deer Surveillance System in Public Parks Using Deep Learning
Word Count:	6699
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deer Surveillance System in Public Parks Using Deep Learning

Mardwin Alejandro Cardenas Rodriguez
20144237

Abstract

The importance of disease transmission between wild animals and people is one of the most serious issue while people trying to get too close to wildlife. This research developed a method for detecting persons who are not following park regulations by utilizing deep learning approaches to detect people and deers in photos or videos. The suggested system used a YOLOv5 object detection algorithm to detect humans approaching deer too closely. The presented method was designed to be used in a drone video, and the system was well-trained to spot people and deer on camera or videos. The most significant limitation of this study was the lack of drone videos from the Phoenix park, as flying a drone is not permitted. We could not turn the video into a top-down image for distance estimation in the 2-D area because there were no camera specifications to do it, Despite the limitations mentioned above, the model was effectively trained to recognize people and deer, as well as to provide a document with the labels and the coordinates of the object detected in the image for further processing.

Keywords: YOLOv5, Deep Learning, Surveillance System, Object Recognition.

1 Introduction

The Phoenix Park is a significant ecological location in Dublin City, as well as a major amenity for residents and visitors. It is residence of almost a third of Ireland's migratory birds and half of the country's mammal species. Almost all of the park's old fields and forests, as well as important ponds, have been conserved. Furthermore, Dublin Zoo is located in Phoenix Park, which attracts over 899,000 visitors every year who come to learn about biodiversity. The community is encouraged to do not give any kind of food to the fallow deer at Phoenix Park, according to OPW standards. Maintain a safe gap between people and deer of at minimum of 50 meters.

Large volumes of data about animal behavioural patterns may now be gathered over a much larger spatial and temporal dimension. Drone surveillance approaches and many other new technologies can also be used in ecological monitoring and management due to their low cost and convenience of use.Because wildlife preservation has increased, investigations such as examining the effects of climate change on animals, habitat changes, the impact of human engagement on animals, and biodiversity across times, locations, and species have become more common. Drone surveillance systems are also cost-effective, as they can simply fly over a larger region and require less maintenance; as a result, they are frequently used for environmental monitoring. We can use drone cameras to

collect data on animal visual attributes, which can help us better understand species behavior and biometric qualities, as well as vital information related to wildlife habitat and surroundings. The Aerial Surveillance System could be used to accomplish this duty without compromising people's lives. Drones will also allow for a faster pace of operation because they can fly for many hours at a time. As a consequence, drones are an excellent instrument for law enforcement from the air. The Phoenix Park became named as a Royal Deer Park when it was built in 1662 by the Duke of Ormond, with fallow deer imported from the Great Britain. About 600 deer make up the park's current wild population, most of whom are descendants of the very first herd that came in the 1660s.

1.1 Motivation

Wild animal tracking in their natural environments should be successful and dependable in order to maintain conservation and management strategies. Due to their usefulness and efficiency in recording data on animals in an unobtrusive, persistent, and massive amount of data, automatic video surveillance, such as camera traps, have become a more prevalent equipment for wildlife observation. Manually scanning such a huge number of photographs and recordings taken by a camera is extremely costly, time-consuming, and tiresome. This is a significant barrier for biologists and environmentalists who want to explore animals in their natural habitat.

1.2 Problem Statement

According to recent studies, COVID-19 has a wide-ranging influence with numerous unexpected consequences. As a consequence of this phenomenon, the increased number of tourists and leisure visitors in Phoenix Park is bringing wildlife species in the park under unnecessary stress, People are trying to get close to wildlife species to snap selfies or treat animals, which is exacerbating the situation.

The main problems when the humans get in close contact with wild deer are:

- **Malnourishment:** Different items, such as potatoes or biscuits, are harmful to the park's deer becuase the parl should provides an optimal feeding habitat.
- **Food diet:** Deer may compete for food, leading toan increase of stress levels and injuries.
- **Wild life interaction:** Disease transmission may occur as a consequence of close contact between wild animals and humans.
- **Hazards and Unpredictable Movemenst:** Deer are wild creatures who may be violent and unpredictable.

1.3 Research Question

RQ: To what extend the development of a surveillance Deep Learning based system can help to recognize and estimate the distance between deer and humans?

1.4 Research Objectives

The Following research objectives are pursued in order to answer the above research question:

- Study and understanding of the current problem that public parks are facing in their premises while people trying to get in close contact to wildlife because their try feeding animal, this is causing the animal are changing their internal anatomy in order to adapt for the new diet.
- Proposed a new methodology based on the data acquisition from difference internet sources and using a deep learning technique to create a system that can alert the operator people getting in close contact with wildlife.
- Build a system bases on deep learning/machine learning models (using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)) for living beings detection.
- Help to Identify people not complying with public park regulation
- Final evaluation of the performance and the accuracy of the built deep learning model to detect deers and humans using some selected statistical measures

The main contributions while creating the surveillance system is implementing in regions that are not easily accessible or have vast lands where it takes a long time to patrol all the terrain or it is very costly or mostly impossible to have ranges covering all the areas of the park at all times. This system could also help researchers who are conducting current research on how human feeding to deer is changing their internal and external body structure to identify people when they are getting to close to animal.

The following is the content of this document: Following this introduction, Section 2 delves into the related work in the field of animal monitoring and deep learning approaches. The methods used to construct the system in this study are described in Section 3. The experimental procedure and design procedure are shown in Section 4. Section 5 explains how to use a deep learning technique to identify and to try to estimate distance. The evaluation of outcomes from the different stages of the process is shown in Section 6. Lastly, in Section 7, the author's findings are presented.

2 Related Work

2.1 The Role of Human Interaction in Wildlife

The food that a creature consumes is indeed an important aspect of its metabolism. Tragically, there has been very little investigations on the evolution of animal feeding in all the species. Scientists looked at food statistics from mammals across the evolutionary chain to answer three important concerns. Firstly, can foods evolutionary conserved or very inert and vary across organisms?. In adaptive ecosystems, the question of if environmental features are historically preserved has now become a subject of dispute. Several researches typically looked at qualities across shorter time frames; however this study looks at a key environmental attribute for a very long time scale. Secondly, can nutrition have an effect mostly on pace at which animals proliferate, and thus on the

structures of variety across vertebrate animals? The biodiversity in each creature phylum ranges from little around five to more than million.

Earlier research has shown that nutrition (particularly a plant-based, herbivorous) affects individual development and biodiversity trends in ruminant animals groupings, for example, Mammals and invertebrates

Nutrition is amongst the most important parts of an organism's physiology. Mammalian predators who hunt enormous or deadly food, bug ranchers who concentrate in just few vegetation, even aquatic creatures which sieve graze on small creatures all seem to have a top reasons among meals or lives during the most evolutionary stages, however, the development of protein source is still remain unclear (Hickman et al.; 2012).

Laura Griffin is a Research scientist at the University College Dublin Institute of Wildlife Biology and Management, where she is researching the complicated relationships with Fallow deers, especially regards of supplemental nutrition around Phoenix Park, Ireland. The focus of this thesis is to see how visitor feeding affects deer mobility biology, hierarchy, vital signs, and reproductive effectiveness. This study will give a clearer picture of something like the possibility of vegetative propagation as a result of human activity. Laura is considering putting control tactics in place or evaluating overall efficacy, with the objective of creating a prototype monitoring system that may be used in many other places with comparable conflicts [1].

Humans have been advised not to engage with any wildlife around Phoenix Area, as new study has revealed because they're under greater stress over average owing to the large number of visitors inside the field.But according to the Department of Public Works, visitors attempting to somehow get near to the wildlife to snap photos with them or feeding animals which administers the Area and maintains the wildlife have caused this surge in tourists even worse (of Public Works; 2020).

Human willingness to really get near to the animals, according to Park Administrator Paul McDonnell, is reasonable. He also mentions the following statement: "They are beautiful wild animals, and as we experience the loss of wildlife around us I are naturally drawn to engage with it when I see it. Deer have been the subject of many books and films and have a special place in our hearts. However, these depictions often make them appear domesticated which they are not. It has led many of us to believe that we are being helpful when we give them carrots or bread but we are causing significant long term harm to the animals and potentially ourselves"

This study (Belotti et al.; 2012) look at how ecotourism affects Eurasian wildanimals migrations and resource use in the Czech Republic's Nature Reserve. This same appearance of lynxes through the day near the location where they would have hidden a wild animal product, such as a herbivore, who provide them with meals over many occasions, has been impacted by human perturbation, according to information recorded from 5 Navigation lynxes as well as assessments of vegetation characteristics. In these kind of circumstances, proper national tourist administration should include a dedication to biodiversity protection.

2.2 Wildlife surveillance

The paper presented by Hammam et al. (2019) provides a method for identifying domesticated animals in rural settlements. Pet creature tracking in IoT is just a difficult

[1] `https://sites.google.com/ucd.ie/wildl-ecol-behav-at-ucd/media-coverage/laura-griffins-msc-research-at-phoenix-park`

topic, as we already understand. Air tagging, Global positioning, and Radio-frequency are just a several of the traditional methods for identifying and monitoring animals. It fails to provide the complete amount of animal surveillance that is necessary. The above technologies have several drawbacks and are prohibitively expensive. The purpose of this study is to discover and categorize the item of concern by monitoring a domestic animal on a live stream utilizing deep learning methods. However, if the amount of dogs grows, this approach would become inefficient in relation to cost, making it impossible to keep track on abandoned animals with in open.

Nikam et al. (2018) emphasizes in their work on smart agriculture approach, which takes into account producers' basic requirements, such as assisting landowners with agricultural automation. The researchers has offered an accessible and inexpensive idea for building a smart ranch creatures protection surveillance device in the suggested protocol. The positions of the creatures are recorded for study. Wildlife movement, moisture, heat, and illness identification are now all monitored by using multiple sensors in cows, chicken farms, and horses. This program includes Arduino to keep track of the dairy farm surroundings and animals. The input collected by sensors is recorded and stored. In addition, an internet repository application has been developed to make it more visible. The information may be utilized to create a sensor module that could also track several metrics linked to creature well-being.

Throughout recent times, the usage of distant or still film CCTV in environmental study and administration has exploded (Nguyen et al.; 2017). Monitoring can be used for a variety of reasons, like identifying types of insects or issue conduct to assessing the quantity and variety of species of ecological importance, and they mostly have one similar theme: identifying specific target organisms.

Artificial algorithms for computerized wildlife detection have been extremely prevalent in ecological and fishery surveillance in past few years. Several innovations have enhanced the capacity to essential quality photographs in difficult locations, leading to more efficient environmental protection (Spampinato et al.; 2015).

This paper study conducted by Zin et al. (2017), where He suggested a monitoring architecture for tracking cow movement in a farm home or in the outdoors, not only for tracking but also for maintaining a control of animals. Producers, wildlife medical experts, and academics had also widely known that analyzing variations in cow social behaviors is a critical component of an effective animals health and social care control program. Furthermore, in today's dairy market, land holdings are rapidly expanding, resulting in fewer treatment strict deadlines for organisms. As a consequence, digital surveillance systems will aid current software in becoming self-contained tracking systems. This employs Markov and Embedded Markov systems to identify distinctive kinds of activity among the objects, as well as React In different and the Markov Management Decision Making to generate expected result. In this regard, machine learning is a potentially useful innovation for making a tough challenge simple to fabricate and execute.

Zotin and Proskurin (2019) used a sequence of photos taken during difficult suggests an approach to recognize animals. In this paper (Backs et al.; 2017) negative electrical sensors have been introduced to avert railway collisions with wildlife on the route. Researchers tested bears, wild boars, foxes, and deer in Bay and Yoho Nature Reserves, and vast rates of bears, grizzly bears, wolves, or deer was discovered to be murdered. The first strategy uses two independent sets of sensors, the majority of which are in charge of analysing incoming locomotives and sending signals to the other, which is placed at a high success signal frequency. The last stage involves forecasting railway estimated time

of arrival from a distance of at least yards and activating an incorporated alert system at the predefined period. A sensitivity of around 80 percentages was reported using Random Forest algorithms using tenfold cross validation.

2.3 Image and video processing with deep learning techniques

This publication by Jing et al. (2020) proposed a scheme storage-based approach to address the problem of efficiency in spatial data picture maintenance and classification. The researchers record the information inside the H-Base shared repository and then create the U-Net on that information. To build the system for such information, researchers effectively use cloud applications using Neural Nets.

The topic of strategic element in cyclic periodic bands in two-dimensional pictures is addressed in this paper(Kennedy; 2020). Interpreting things which does not take on different but seem differently that can be seen from different viewpoints is an illustration of this kind of difficulty. The issue seems clear when you realize that robots comprehend their surroundings using pictures portrayed in Cartesian coordinate system, even if things are much more naturally defined in linear form. Addressing a situation like this is difficult and morning. Considering polar separable answers of view from the top networks, the researcher proposes a deep learning model technique for handling such issues in this paper.

This publication by (Lai et al.; 2021) provides a significant machine learning algorithm that addresses the limitations of genuine tool for image monitoring, particularly with elevated cross pictures. Skillet approaches seek to fuse the data of reduced photos to make maximum visual information. But since most profound attempting to learn algorithms may accomplish during which, many will not utilize all of the data implicit inside the pictures' complex form.

A hybrid item recognition and approach typically for pictures collected by equipment mounted on aircraft (like drones) was examined in study by Ahmed et al. (2021) to allow the network of aircraft idea, which took some real internal computation at the objects level under account. This suggested program was made using a type of deep learning technology called the U-Net. The researchers' testing findings shows that the suggested scheme has an average precision of 92 percent in various circumstances.

The topic of population activity recognition in security camera footage was studied by Rezaei and Yazdi (2021). A 2 different process extracts reduced characteristics and conducts classification in traditional techniques. Learning algorithms had also already demonstrated superior efficiency when compared to traditional methods. It is mostly due to the deep convolution networks' intrinsic multi-level extracting features as well as processing. This is due to the CNN's networks intrinsic multi-level character recognition and interpretation capabilities. The investigators present a novel handcrafted approach that focuses purely on fully linked auto-encoder architecture and uses visual images as training.

There have been a lot of studies on general surveillance applications, but in this paper (Han and Choi; 2019) it has been explained that there have been only few publications that properly reviewed based on the surveillance systems using a convolutional neural networks. However, the majority of organisations design a questionnaire that are limited to a single feature. The study discusses several deep learning algorithms for graphics processing that may be used to identify intrusions like aberrant actions such as violence, rioting, traffic regulations breaches, stampedes, and odd objects including firearms and discarded suitcases.

Nonetheless, according to a study by Nayak et al. (2021), deep learning algorithms have revolutionized visual picture categorization. In another work, researchers developed frame interpolation for abnormal social psychology detection using a deep learning technique for video surveillance (Dhiman and Vishwakarma; 2019). In other publication by Borja-Borja et al. (2018), He evaluated state-of-the-art deep learning approaches in video processing to come up with a list of collaborative and demographic behaviors. Tong et al. (2020) focused on tiny picture categorization in based on neural machine learning for video processing. The authors explained how convolutional neural algorithms such as R-CNN, Fast R-CNN, Deconv-R-CNN, Expanded Faster R-CNN, as well as other derived characteristics have progressed over history. Sánchez et al. (2020) describe a study using deep neural networks with camera encoders to investigate community causal inference or crowd warning systems.

Facial identification utilizing deep convolutional neural network has advanced significantly over the years, frequently outperforming classic machine learning approaches considerably. Notably, Zhan et al. (2016) utilized Convolutional neural networks to train and generate feature representations for facial detection dynamically. In this research study (Li et al.; 2016) proposed a final multi-task racist and discriminatory education system that blends a Conv-Net with a 3-dimensional average facial image for pattern recognition in the environment. Jiang and Learned-Miller (2017) previously was using the Faster R-CNN (Ren et al.; 2015) , a state-of-the-art image recognition and classifier, with interesting results. Furthermore, a lot of time and effort into improving their Faster R-CNN design.

Wan et al. (2016) used a combination of the Faster R-CNN facial recognition techniques, strong critical mining, and ResNet to improve recognition rate on facial identification standards such as FDDB. By upgrading the Faster RCNN architecture, they offer a novel approach in facial recognition in this publication.

Only by using a single picture, this research (Masoumian et al.; 2021) proposes a deep learning system that comprises of two deep networks for background subtraction and object recognition. The system in the paper was trained on real pictures from external landscapes, the given distance estimate system was successfully assessed. The obtained findings demonstrate that the suggested methodology is feasible, with a 96 percent accuracy and an RMSE of 0.203 of the right distance measure.Using a supervised learning approach, the recommended item identification was created based on the YOLO algorithms group algorithms, whereas the deep convolutional neural network was self-supervised.

In the summary of the relate work I can conclude that the use of deep learning techniques, for example convolutional neuronal networks to monitor an specie, has become more important in recent years because most of the current research have used this techniques in order to archive better results. Also in the last two years have been and incredible growing in the surveillance application in order to track people following the social distancing rules that most of the governments have implemented in the cities in order to try to decrease the number of COVID cases or even to detect people that is wearing face mask properly.

In the follow table Table 1, we can see an overview of the methods analyzed and some of their advantages and drawbacks for each model.

Table 1: Summary of the Methods Analyzed

Method Name	**Purpose**	**Advantages**	**Drawbacks**
R-CNN Tong et al. (2020)	• Produce a set of bounding boxes • Object Detection	Bypass the problem of selecting a huge number of regions by only extracting just 2000 regions from a single image	Can not be Implemented Real Time
Fast R-CNN Tong et al. (2020)	• Produce a set of bounding boxes • trained end-to-end	faster than R-CNN	Region proposals cause disruptions
Faster R-CNN Wan et al. (2016)	• Region proposals generated by Selective Search • Based on Region Proposal Network	Faster R-CNN is much faster than it's predecessors	slow and time-consuming process
YOLO Masoumian et al. (2021)	• Object Detection • Real-Time Object Detection	Faster(45 frames per second) than other object algorithms	Have problems with small objects in the image

3 Methodology

3.1 Introduction

In all the research, the selection of an existing mythology must be completed in order to perform a real analysis; this will enable us to determine which stages must be followed. Currently there 2 well-known approaches that can be used to complete a proper study, the Knowledge Discovery in Database (KDD) and The Cross-Industry Standard Process for Data Mining (CRISP-DM). The objective of the project is to create a model that can identify people and deer in videos taken in public parks to help to estimate the distance between them. This could be achieved using various Deep Learning model such as CNN, RNN, R-CNN, fast R-CNN, YOLO. So, In order to answer our research question, as shown in the Figure 8, a costume research framework is used in order to follow our

research project. The next parts outline the steps that we are going to used in this project:

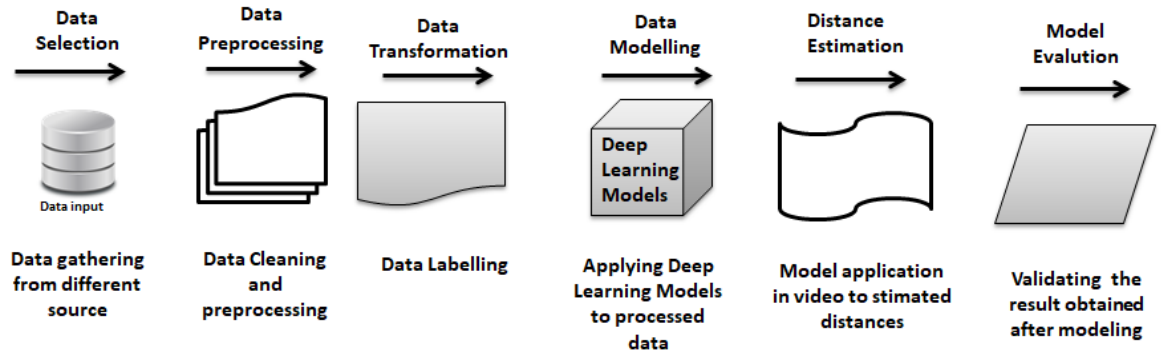

Figure 1: Proposed Research Flow

3.2 Data Gathering

This would be the initial phase of the methodology, in which the material needed during model computation and development is collected. Every image needed for segmentation is gathered from a variety of websites and applications. Wide variety of deer pictures should be gathered, such as deer in different poses and in different backgrounds in order to train a good model and the same criteria should be applied for the human data set as I need people under specific circumstances for example people walking , running, jogging in the park, Every of these pictures and video should be collected and saved in a local storage to be processed later. For human picture I have selected the MPII Human Pose Dataset as this dataset contains approximately 25000 photos with over 40000 persons with identified body positions. The collection includes 410 human actions, including an action tag attached to every photo. The video in which I are going to validate the creation of our model was collected from a website in order to cover our criteria specification [2].

In the case to get the deer pictures I have chosen a data base from kaggle, this data base includes a numerous photographs of deer and deer fawn in various locations, 744 picture can be found in this data set[3].

The data policies of both sites have been analyze and I have the rights to use this data set only to perform research studies and no make any profitable business with this. In the Figure 2, I can see a general preview of the 2 data set I have gather together for this research.

3.3 Data Processing

Many machine learning techniques rely on data analysis. However, when it refers to pictures segmentation in a collection, there is no a simple way to undertake the data exploration in an easy way. Based on the previous information in the fist step it is strongly advise to take extra time manually exploring the data, using a mix of solid measurements. Representations might help us acquire a better idea of the full scope of the work.

[2] `http://human-pose.mpi-inf.mpg.de/`
[3] `https://www.kaggle.com/faisalakhtar/images-of-deer-for-svm-classifier`

Figure 2: General Data Overview

The second step that should be performed is the image resize, because the collection comprises high-resolution photographs with varying resolutions, the learning model could be well trained incorrectly. To address this problem, I reduced each photograph to a constant dimension.

Data Augmentation methods like image rotation or changing the images contrast could be applied to increase the precision of the final results. Image modification with a CV2 package in Python might be used to conduct data augmentation. By using CV2 getrotationmatrix2D and warpAffine methods, I was able to flip pictures at any angle from the center. This could help us to enlarge the count of relevant images in our own data set.

3.4 Data Labeling

Pre-processing data is a necessary stage to train a model to have a better accuracy and precision. The photos included in the collection should be labelled for easy identification. In this case in order to identify the wild animal and the humans I must label all the photos where these are present. They are many tools that could be use in order to label our data set for example CVAT, Lableme, Labelimg and so on. I can introduce false negatives to the model if I do not identify the item in certain photographs that is why I need keep in mind that the labeling stage is a time-consuming procedure, however the more effort you dedicate to labeling photographs stage, the more precise my model would be.

3.5 Data Modelling

Computer vision and image processing problems involve in determining the presence, location, and type of one or more objects in a photograph. As a result, numerous latest innovations in deep learning approaches have emerged from attempting to improve the features of a neural network. The Convolutional Neural Network is among the most effective designs for image processing applications, and it is now the most extensively employed structure. Inside a Convolutional Neural Network, there are convolutional layers that train a number of parameters where describe the geographic relevance of features detected in the photo.

Deep learning algorithms previously achieved high state-of-the-art object detection results on prominent standard datasets and in video processing challenges. The YOLO (You Only Look Once) family of Convolutional Neural Networks, for example, is well-known for achieving near-perfect effectiveness with a single end-to-end model capable of object recognition. Image detection techniques in the state-of-the-art could well be divided into two general categories: one stage detectors and two stages detectors.

One-stage detectors: These kinds of detectors forecast the bounding areas over the pictures without the region proposal stage that are as follows:

- This object detection systems emphasize inference efficiency and are extremely quick, but they struggle to recognize oddly shaped items or a collection of tiny objects.
- The fundamental benefit of the single-stage detectors systems is that they are typically quicker and have a lower density over the multi-stage.
- YOLO(2016), SSD(2016), RetinaNet(2017), YOLOv4 (2020), YOLOv5(2020), are the most prevalent one-stage detectors.

Two-stage detectors: In this type of detectors the deep features are utilized to suggest estimated item areas before other attributes are employed for categorization and bounding box for the item proposal:

- Object region proposal use traditional Computer Vision algorithms or deep networks are accompanied by object categorization for the bounding-box.
- The approaches are the most accurate in terms of detection, but they are also the slowest. The efficiency is not as excellent as the one-stage detectors due to the multiple interpretation processes every picture.
- Faster RCNN , Fast RCNN, Mask R-CNN , G-RCNN(2021) are the most prevalent two-stage detectors.

3.6 Model Evaluation

TensorBoard is a visualization utensil that I can use in order to view the system metrics, Tensorboard could only display one measurement at a moment; it cannot display all the measurements in a single graph.

It must have been agreed that the trained model's performance will be validated against the testing data, which comprised the residual 20% of the data collected, for the

objectives of this work. The validity of the model's success on additional data could be the benchmark by which it can be assessed.

The standard Mean Average Precision (mAP) at a defined Intersection over Union criteria (eg. mAP at 0.55) should be used for the analysis. mAP is an information retrieval measure which is often used for measuring bias in sorting tasks and assessing object recognition tasks. Lets consider the IoU formula 1 as bellow:

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (1)$$

When the model missed an area that really should have been identified, or when you identify everything which does not appear, then mAP will penalized this behaviour.

4 Design Specification

4.1 YOLOv5 (You Only Look Once): Real-Time Object Detection

YOLOv5 is a real-time object identification technique that employs neural networks. The python library ImageAI, additionally, includes the ability to build a customized YOLOv5 network to recognize whatever types of objects. CNNs include examples of classifier-based systems in which the algorithm commonly filters or carry out tasks for recognition and uses the detection algorithms to a picture at many dimensions and regions. "Top rating" portions of the picture are deemed analyzed samples using this method. In simple words, the areas that most closely resemble our training pictures are definitely detected [4].

YOLOv5 conducts identification and bounding box regression just in one phase, made it extremely quicker than typical CNN. YOLOv5 object recognition, by contrast, is hundred times quicker than R-CNN but also hundred times faster than Fast R-CNN.

YOLOv5 hit the object recognition field.Video processing professionals were immediately interested in "You Only Look Once: Integrated, Real-Time Image or Video Segmentation," and the YOLOv5 algorithm determines all of the image's attributes and provides estimations for all of the elements.

Talking about innovative techniques, YOLOv5 have a overwhelming advantage. YOLOv5 has been implemented in Python instead than C, as it had been in previous versions. This makes easier the deployment and incorporation of Internet of things. However, PyTorch network is bigger than the Darknet society, making that PyTorch can get a bit more donations and have a lot of productively capacity.It could impossible to test YOLOv4 and YOLOv5s efficiency, because they have been developed in two separate programing languages and they have been run on two distinct systems. Nonetheless, in certain cases, YOLOv5 has proven to be more efficient than YOLOv4 across time, and has acquired some faith inside the field of computer vision in contrast wih the YOLOv4.

[4] `https://github.com/ultralytics/yolov5`

YOLOv5 design has merged the most recent advances in a way that is comparable towards the YOLOv4 structure, therefore there aren't many important distinctions in concept. The following points is a summary of the YOLOv5 framework :

- Backbone: CSP network
- Neck:SPP block
- Head: YOLOv3 using GIoU-loss

The YOLOv5 detector examines the entire picture at once, thus its estimates are influenced by the picture's surroundings. The supplied picture is divided across SxS grids via YOLOv5 methods.Every grid's features can be extracted. It calculates the confidence ratings again for expected categories inside the bounding areas and forecasts actual bounding areas, see Figure 3. Bounding areas and confidence ratings are detected for each data point.

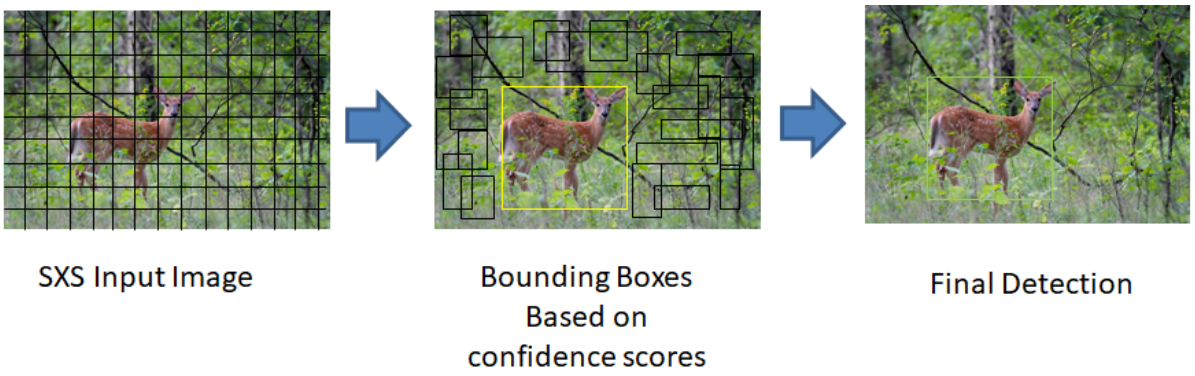

Figure 3: YOLOv5 Schematic Diagram

YOLOv5 anticipates multiple region recommendations for each bounding box. During the training process, one grid cell predictor is fully responsible for each categorization.The forecast gets responsible for predicting the contextual information entity with the maximum IoU. This technique refers to expertise in the estimation of coordinates. For the loss function, YOLOv5 systems employ the sum squared error (SSE) between underlying data and the boundary box estimations.The model's classification, localisation, and confidence losses are computed using this SSE.

In terms of object detection, YOLOv5 was built to extract characteristics from incoming photos and then transfer those characteristics into a prediction engine that draws rectangles over the entities and predicts their classifications.The fundamental elements of this architecture are the backbone, neck, and head. The backbone of YOLOv5 is the state-of-the-art network EfficientNet, which ensures that the algorithm is capable of learning the complicated properties of incoming photos.

To provide an effective and quick multi-scale hybrid algorithm, YOLOv5 used an enhanced PANet called bi-directional feature pyramid network (Bi-FPN) as its neck. Bi-FPN provides trainable weights, allowing the computer to work the relative relevance of various parameter characteristics, and performs top-down and bottom-up multi-scale feature selection technique iteratively. Finally, YOLOv5 has a complex amplification approach that adjusts the quality, width, and length of the backbone, feature structure, and crate forecasting models simultaneously, providing optimum effectiveness and precision

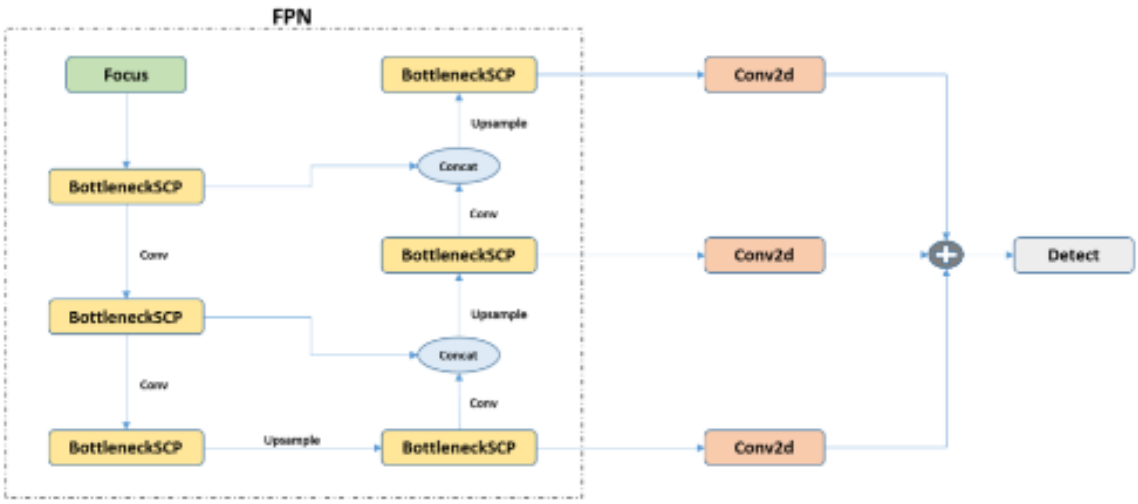

Figure 4: YOLOv5 Network Architecture Overview

with restricted computer resources. The structure of YOLOv5 which researchers utilized to identify entities is shown in the Figure 4.

5 Implementation

This section describes all the details on how the YOLOv5 algorithm are used to recognize deer and wild animal classes in aerial and phone videos as well as in a general pictures.

5.1 Development Environment

R and Python are usually by far the most popular programming languages for implementing Machine Learning and Deep Learning models.This academic study is entirely carried out in Python, utilizing the Jupyter Notebook environments, which can be integrated in Google Colaboratory as this operates as an online Integrated Development Environment (IDE). Python has been selected as the primarily programming language since it contains a number of packages that could assist in the development of a Deep Learning model. Keras, Tensorflow, and Pytorch are some of the libraries that have made it much easier to create Deep Learning models. NumPy, scikit, sklearn, and other libraries that assist in executing fundamental functions are utilized in this study.

5.2 Data Preprocessing

Before any other process stage, I have look at the picture to identify some fake picture and those image that are not good for our study, in the deer data set I could identified some picture of deer pictures that belong to toys or drawing, as I pretend to implement this system in real life, I have only chosen photographs of deer in their native surroundings. In other hand, the human data set contains over 40,000 persons with annotated bodily joints in over 25000 photos into 410 human activities. But most of these pictures were not relevant for the study so I must precede to selected those picture just on the categories like running, jogging ,walking or standing, so I end it up with a data set of 500 hundred images, I can see and example of some wrong images in the Figure 5.

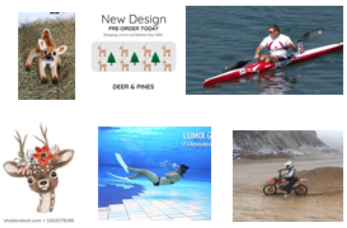

Figure 5: Data Cleaning Examples

5.2.1 Image Labeling

Computers are now having trouble detecting things in photos. While there are effective alternatives for some basic classes of items, like face recognition or trucks, its more fundamental challenge of identifying the several classes of things in the universe (e.g. instruments, containers, and smartphones) stays unresolved. Studies are being done studying ways for recognizing and localizing hundreds of distinct item categories in complicated situations using deep learning techniques.

Data collected to build the deep learning model of every item is an important part of all these methods. LabelMe's purpose is to develop an online annotation tool that will allow users to contribute to a vast collection of labeled photos. In order to annotated all our data set I have process each photo manual to have a better object recognition, We can see how this process have been done in the Figure 6. As a output, LabelMe have produced it a JSON file for each photos were the object is allocated.

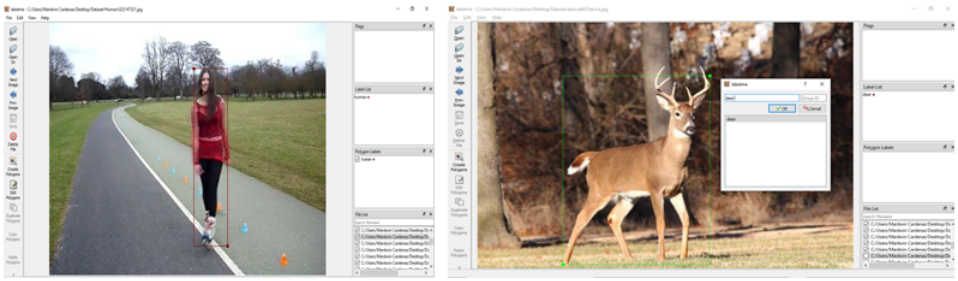

Figure 6: Data Labelling Tool

5.2.2 Data Preprocessing

Roboflow is a software platform for Computer Vision that allows for improved data collecting and preparation. Roboflow provides customers with accessibility to public databases as well as the ability to submit our unique customised data sets. From gathering data to implementation, Roboflow assists in each and every phase of a machine learning challenge.

After the labelling step all the annotation and images files generated by the previous step were loaded into Roboflow, This tool allows to prepossess all the data to applied data augmentation or data resizes, we can see all the images configuration applied in the Table 2.

The datasets used in this research are MPII Human Pose Dataset and Deer images from Kaggle which includes 505 images of deers and 500 images of humans and in total

Table 2: Data Prep-reprocessing Specifications in RoboFlow

PREPROCESSING	Auto-Orient Resize	Applied Stretch to 416x416
AUGMENTATIONS	Outputs per training example 90° Rotate Grayscale	3 Clockwise,Counter-Clockwise Apply to 50 % of images

1005 images. Following Figure 7 shows the distribution of the 2 combined datasets that I used in this research after I have applied the data processing in RoboFlow.

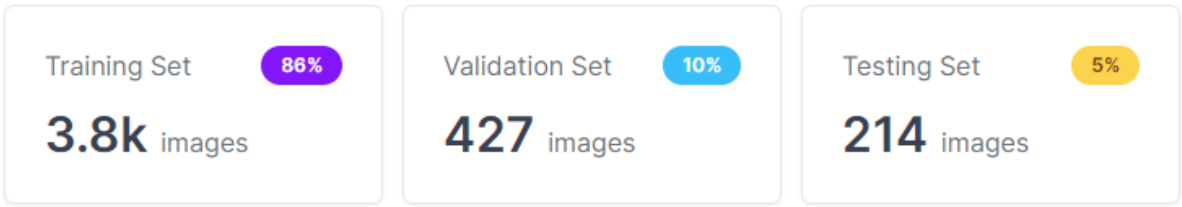

Figure 7: Data Split Information

5.2.3 Modeling

I start by downloading the YOLOv5 repository and establishing up the requirements necessary to execute YOLOv5 from the offical rebository on GitHub. [5].

For the collection of the dataset labelled, which is kept storage in WorkFLow, to be more specify in the work space that I have created before in Roboflow, I subsequently can access from Google Colaboratory trough the API that Roboflow provide us, and then I get the data in our Google Colab, After that, I can start training the YOLOv5 model.

Already I have imported the Data set, YOLOv5 training weights, and the configuration file into Jupyter. I observe the result after running the Item Detection, where the object is recognized using rectangular rectangles called as bounding boxes and the label - "deer" for animals or "people" for humans. The square areas that contain the item recognized in a photo are known as bounding boxes. In the section 6 I explore more into the findings.

5.2.4 Technical Configuration

The baseline software and hardware specifications applied in this study are listed in the Table 3.

[5] `https://github.com/ultralytics/yolov5`

Table 3: Summary of Technical Configurations Used

RAM	8 GB
SPEED	CPU@ 1.60GHz 1.80 GHz
PROGRAMMING TOOLS	GOOGLE COLAB AND JUPYTER
PROGRAMING LENGUAGES	PYTHON
Libraries	CV2,Glob,roboflow, yaml, torch, NumPy
Adittional Toolls	LabelMe and RoboFlow

6 Evaluation

After performing the YOLOv5 object detection approach over our entire data, the results are presented in this section. Even I have finished training, I should have a look at the verification metrics to see how good the learning technique worked.

The Figure 8 illustrates the YOLOv5 models Mean Average Precision (mAP), whereas Figure 9 shows overall loss numbers. The mAP number grew rapidly from 60 to 90 iterations, then gradually increased till it reached 100 iterations. My developed model which is based on the YOLOv5 architecture, has reached the 100th epoch of training. The training inaccuracy of the algorithm lowers as the number of epoch was growing. In the Figure 8 we can see how my model reach a mAP@.5:.95% of accuracy.

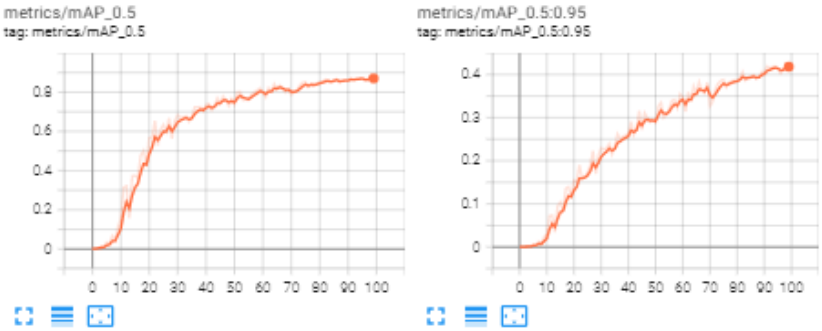

Figure 8: mAP of YOLOv5 Model

Figure 9: Total Loss of YOLOv5 Model

Even that I know the confusion matrix is generally unspecific in object recognition because the percentage of errors would be with the background class instead of another

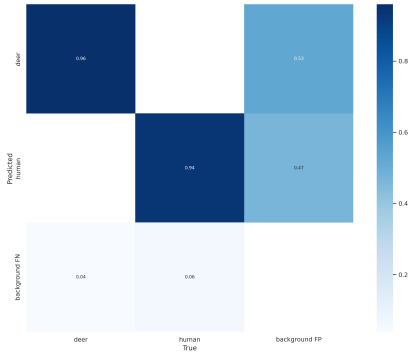

Figure 10: Confusion Matrix

category. In this study execution of the confusion matrix is accurate. However, everything will be determined by the amount of confidence, which is why the results may change. I can see our confusion matrix in the Figure 10.

7 Conclusion and Future Works

Inside the video processing world, there is a huge debate concerning the nomenclature and enhancements of YOLOv5. Regardless of this, the YOLOv5's effectiveness in terms of effectiveness and efficiency is comparable to other generations.This report described a study that was conducted to identify them and categorize them into different categories, such as deers and people.

Image detection techniques such as Fast R-CNN, Faster R-CNN or any other YOLO version were studied and contrasted in order to determine the most suitable and efficient approach. The choice to use YOLOv5 has been made because of the excellent properties and characteristics it offers over alternative techniques. Inside this upgraded YOLOv5, a framework called RestNet with four fusion layers is also used.

The experiments that was performed in this research was done on The MPII Human Pose Data set which belongs to human class. Following Figure 11 show the results of the model was applied in some frame taken from videos that were taken in the data gathering stage. This program correctly identified persons who were in proximity to the animal.

In conclusion, what I had created is a Deep Learning model that works on your testing data. As a result, I must run the "YoloV5/detect.py" script to evaluate our model on test data. Command to test the model on your data is shown in the Figure 11. Even the model has show good result while detect people and deers in the park, I have faced some limitations in order to fully answer our research question.

Model has show a good performance while it has been applied on the videos, I can specify this command –save.txt in the scrip in order to get a file with the coordinate of the exactly location of the bounding boxes for each object detect in each image. From an upper perspective, the distance in between human and deer could be estimated, and the measurements are adjusted by the correction factor obtained from camera alignment, But as I could not get any information about the camera where the videos were taken, the distance could not be calculated, as a future work I proposed to work with the government authorities to get a special permission from the park to setup a camera with our own calibrations to be able to estimate the distances between people and deers.

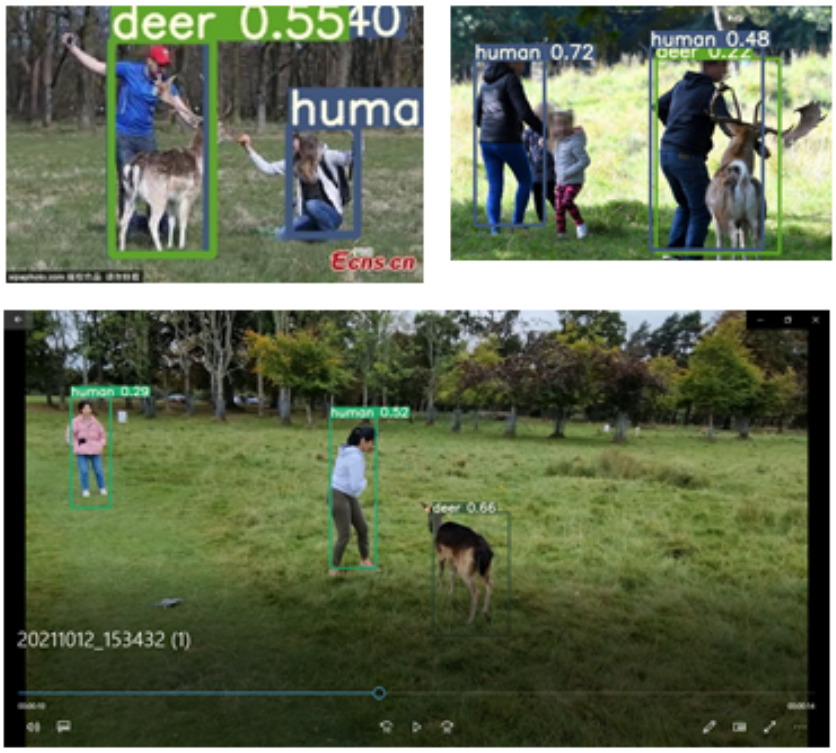

Figure 11: Video Testing with the Final Model

References

Ahmed, I., Ahmad, M. and Jeon, G. (2021). A real-time efficient object segmentation system based on u-net using aerial drone images, *Journal of Real-Time Image Processing 2021 18:5* **18**: 1745–1758.
URL: *https://link.springer.com/article/10.1007/s11554-021-01166-z*

Backs, J. A., Nychka, J. A. and Clair, C. C. S. (2017). Warning systems triggered by trains could reduce collisions with wildlife, *Ecological Engineering* **106**: 563–569.

Belotti, E., Bufka, L., Heurich, M., Kreisinger, J. and Šustr, P. (2012). Influence of tourism and traffic on the eurasian lynx hunting activity and daily movements.

Borja-Borja, L. F., Saval-Calvo, M. and Azorin-Lopez, J. (2018). A short review of deep learning methods for understanding group and crowd activities, *Proceedings of the International Joint Conference on Neural Networks* **2018-July**.

Dhiman, C. and Vishwakarma, D. K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition, *Engineering Applications of Artificial Intelligence* **77**: 21–45.

Hammam, A. A., Soliman, M. M. and Hassanein, A. E. (2019). Deeppet: A pet animal tracking system in internet of things using deep neural networks, *Proceedings - 2018 13th International Conference on Computer Engineering and Systems, ICCES 2018* pp. 38–43.
URL: *https://www.researchgate.net/publication/331106428*$_{DeepPet_{APet_{Animal_{Tracking_{System_{in_{I}}}}}}}$

Han, Y. K. and Choi, Y. B. (2019). Human action recognition based on lstm model using smartphone sensor, *International Conference on Ubiquitous and Future Networks, ICUFN* **2019-July**: 748–750.

Hickman, C., Keen, S., Larson, A. and Eisenhour, D. (2012). *Animal Diversity*, Vol. 9, McGraw-Hill.

Jiang, H. and Learned-Miller, E. (2017). Face detection with the faster r-cnn, IEEE, pp. 650–657.
URL: *http://ieeexplore.ieee.org/document/7961803/*

Jing, W., Zhang, M. and Tian, D. (2020). Improved u-net model for remote sensing image classification method based on distributed storage, *Journal of Real-Time Image Processing 2020 18:5* **18**: 1607–1619.
URL: *https://link.springer.com/article/10.1007/s11554-020-01028-0*

Kennedy, H. L. (2020). On the realization and analysis of circular harmonic transforms for feature detection, *Journal of Real-Time Image Processing 2020 18:5* **18**: 1621–1633.
URL: *https://link.springer.com/article/10.1007/s11554-020-01040-4*

Lai, Z., Chen, L., Jeon, G., Liu, Z., Zhong, R. and Yang, X. (2021). Real-time and effective pan-sharpening for remote sensing using multi-scale fusion network, *Journal of Real-Time Image Processing 2021 18:5* **18**: 1635–1651.
URL: *https://link.springer.com/article/10.1007/s11554-021-01080-4*

Li, Y., Sun, B., Wu, T. and Wang, Y. (2016). Face detection with end-to-end integration of a convnet and a 3d model, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9907 LNCS**: 420–436.
URL: *https://link.springer.com/chapter/10.1007/978-3-319-46487-9*$_2$6

Masoumian, A., Marei, D. G. F., Abdulwahab, S., Cristiano, J., Puig, D. and Rashwan, H. A. (2021). Absolute distance prediction based on deep learning object detection and monocular depth estimation models.
URL: *https://arxiv.org/abs/2111.01715v1*

Nayak, R., Pati, U. C. and Das, S. K. (2021). A comprehensive review on deep learning-based methods for video anomaly detection, *Image and Vision Computing* **106**: 104078.

Nguyen, H., Maclagan, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E. G. and Phung, D. (2017). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring, Vol. 2018-January, Institute of Electrical and Electronics Engineers Inc., pp. 40–49.

Nikam, M. C., Wagh, S., Mahajan, P., Waghmare, D. and Padewar, K. (2018). Animal welfare monitoring using iot : A review, *International Journal for Modern Trends in Science and Technology* **04**.
URL: *http://www.ijmtst.com/vol4issue10.html*

of Public Works, O. (2020). gov.ie - opw asks 'this christmas please don't feed the phoenix park deer'.
URL: *https://www.gov.ie/en/press-release/9382e-opw-asks-this-christmas-please-dont-feed-the-phoenix-park-deer/*

Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in Neural Information Processing Systems* **2015-January**: 91–99.

Rezaei, F. and Yazdi, M. (2021). Real-time crowd behavior recognition in surveillance videos based on deep learning methods, *Journal of Real-Time Image Processing 2021 18:5* **18**: 1669–1679.
URL: *https://link.springer.com/article/10.1007/s11554-021-01116-9*

Spampinato, C., Farinella, G. M., Boom, B., Mezaris, V., Betke, M. and Fisher, R. B. (2015). Editorial open access special issue on animal and insect behaviour understanding in image sequences, *EURASIP Journal on Image and Video Processing* **2015**: 1.
URL: *http://jivp.eurasipjournals.com/content/2015/1/1*

Sánchez, F. L., Hupont, I., Tabik, S. and Herrera, F. (2020). Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects, *Information Fusion* **64**: 318–335.

Tong, K., Wu, Y. and Zhou, F. (2020). Recent advances in small object detection based on deep learning: A review, *Image and Vision Computing* **97**: 103910.

Wan, S., Chen, Z., Zhang, T., Zhang, B. and kat Wong, K. (2016). Bootstrapping face detection with hard negative examples.
URL: *https://arxiv.org/abs/1608.02236v1*

Zhan, S., Tao, Q. Q. and Li, X. H. (2016). Face detection using representation learning, *Neurocomputing* **187**: 19–26.

Zin, T. T., Kobayashi, I., Tin, P. and Hama, H. (2017). A general video surveillance framework for animal behavior analysis, Institute of Electrical and Electronics Engineers Inc., pp. 130–133.

Zotin, A. G. and Proskurin, A. V. (2019). Animal detection using a series of images under complex shooting conditions, Vol. 42, International Society for Photogrammetry and Remote Sensing, pp. 249–257.