Optimisation of Elephant detection

Conor Worthington

MSc in Machine learning and Autonomous systems
The University of Bath
September 2020

Optimisation of Elephant detection

Signed:

Optimisation of Elephant detection

Submitted by: Conor Worthington

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of Machine learning and Autonomous systems in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Signed:

Optimisation of Elephant detection

Abstract

The purpose of this project is to expand on existing techniques for identifying small objects as well as applying techniques which have not been applied to the specific domain of tracking elephants. This approach is currently focussed on handling the problem of resizing satellite imagery and the opportunities certain techniques can provide in terms of performance as well as the challenges in dealing with such small objects to detect.

Contents

List of Figures

List of Tables

Acknowledgements

My tutor Olga Isupova for assisting me consistently throughout this difficult period with weekly research seminars and biweekly meetings for this project

My friends and family for helping me out mentally making this thesis during lockdown

Chapter 1

Introduction

For an exceptionally long period humans in many cultures have viewed ivory as having either medicinal properties or as a material to generate art as such a lucrative trade has been created. Ivory is extracted from the tusks of elephants by killing them, which in and of itself is brutal. With only 400000 left in the wild and an ever increasing amount of conflict occurring with humans as habitats are diminished the species is currently listed as vulnerable (WWF, 2020). The threat to the elephant population is so vast that many countries such as the United States have imposed sanctions on the ivory trade to remove financial incentives for hunting elephants (Drayman-Weisser, 2018).

To allow a species such as this to disappear would not be just a humanitarian tragedy, such an intelligent species should not exist only in captivity, but also an ecological one which could vastly upset an already precariously balanced African savannah. As such it is clear that more should be done, current approaches to tracking elephants are limited to devices such as GPS collars - whilst results have been positive and successful methods for containment such as water manipulation have been used to good effect the sample size is small (Thomas Bindi et al, 2012). As such an approach must be devised which can handle far greater amounts of elephants to allow for greater inferences to be drawn so that actions planned will have the greatest effect. Therefore by having a better approach more elephants could avoid coming into conflict with nearby humans, saving elephant lives as well as allowing conservationists to work more effectively to prevent poaching for the ivory trade. As such any performance gained constitutes an improved chance at life for a vulnerable species. With the recent proliferation of satellites and the readings from them with an appropriate recognition technique there is an opportunity to vastly increase the performance of elephant detection. This does however pose challenges as due to the limits of current resolution the elephants whilst large are small on the satellite image, which by nature makes them hard to detect. As such techniques must be developed which can optimally find the elephants in satellite data to allow for greater quality of conservation.

Thankfully with recent advances in satellite technology the sensors available for us to search for elephants using commercial resources are greater than ever – but this also poses significant challenges. These images will be more detailed but still won’t provide all the detail necessary as a result we will need to find optimal ways to process the images. Furthermore, the images will as singular files be vastly too large to process in one go. Instead images must be systematically divided up to ensure the optimal approach is taken while processing. If an optimal approach can be found and utilised to assist in conservation it could greatly improve the management of elephants.

Chapter 2

Literature Survey

Currently this problem is an object detection problem with the satellites providing the dataset for which to identify elephants within. The labelled dataset we will be working with has been utilised before to good effect which means this problem is certainly solvable (Duporge et al, 2019). Furthermore there is a large amount of existing studies which have analysed satellites utilising supervised datasets to train models to identify small objects. Although the success has been limited as sensors can only pick up so much data. Attempts at utilising machine learning to process satellites have been ongoing for decades now (Segl et al, 2001) however in that time a plethora of more advanced machine learning techniques have been devised which can vastly increase performance.

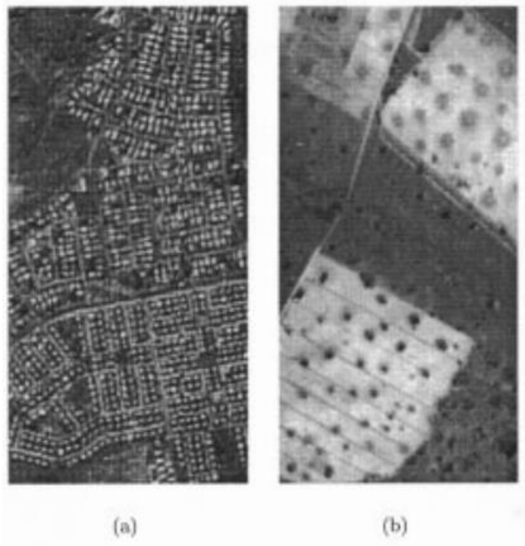


Figure 1- Satellite overview of a town (a) and a field (b) - showing how small even buildings are (Seql et al, 2001)

The graphics shown in figure 1 though depict a much older satellite of much lesser capability than the ones we shall collect data on. The satellites from 'Maxar technologies' have full RGB, high quality sensors with 30cm resolution and infrared sensors.

Figure 2 – Example of dataset being utilised for elephant detection

As can be seen in figure 2 the current dataset available presents an even larger challenge than the one referenced in figure 1. With an image as large as 8kx14k pixels for there to be any success the models must use the most up to date machine learning models.

One of these is the ‘You Only Look Once’ or ‘YOLO’ object detection model. This model is a machine learning model which boasts better accuracy than traditional fully connected convolutional networks but also boasts incredible speed (Redmon Joseph et al, 2016). The YOLO network much like other contemporary models utilises a region system to increase accuracy but unlike other models distinctly avoids re-comparison of bound regions for efficiency, naturally this sacrifices accuracy.

An alternate model which could boast different benefits for the purpose of elephant detection is ‘Faster R-CNN’ which is noted to be more accurate when tested on the COCO dataset (S. Ren et al, 2017). This can still be seen to be true as one of the most recent variations of ‘Faster R-CNN’ dubbed ‘Mask R-CNN’ is still one of the top performers on the COCO dataset (COCO, 2020). Unfortunately we do not need the semantic segmentation which is the main variation between the two. Faster R-CNN works by utilising a region proposal network before feeding through features to a pooling layer which provides classifications to regions in bounding boxes. This approach whilst considerably slower than YOLO is still fairly quick and is incredibly accurate. As such it makes sense to proceed in the investigation of how to optimise elephant detection presuming to be building from a network such as ‘Faster R-CNN’ as the network performs well in the COCO dataset of which 50 percent of the images contain a small object to be classified.

When considering potential roadblocks - one becomes very clear, given the resolution of the photos - whilst the images may be large and full of viable information the elephants are going to be exceptionally small. This means elephants may consist of only a few pixels - potentially a 15x20 bound box. This is vastly smaller than that which was measured utilising the COCO dataset - especially when considering how small of a region this is of the image as a whole. Furthermore when looking at other research which has considered ‘Faster R-CNN’ performance on small objects we start to notice some discrepancies such as the network not being sensitive enough leading to large amounts of false negatives. This leads to a ‘recall’ value of 46 percent when a tweaked ‘Faster R-CNN’ can easily perform at 90 percent (Cao et al, 2019). This leads to the notion that the small images do not provide enough information from which ‘Faster R-CNN’ can accurately determine an object - as such the expectation is that detecting elephants with just a stock ‘Faster R-CNN’ will not yield optimal results.

As such the size of the images currently provide a challenge - to just scale up the images by traditional resizing is unlikely to help on its own as it is likely to introduce more noise and deviate further from the expected shape of the object. As such when we consider scaling up the images to improve performance as an avenue, we must consider a more nuanced approach which seeks to interpolate the image in the best possible way. This is currently a very exciting field of research with many publications showing promising results - one approach which is currently commercially used in video gaming is Nvidia’s ‘super resolution’ which has been shown in research to vastly improve microscopy by upscaling and demonising objects which cannot be easily resolved with the current microscope resolution (Fang et al, 2019).

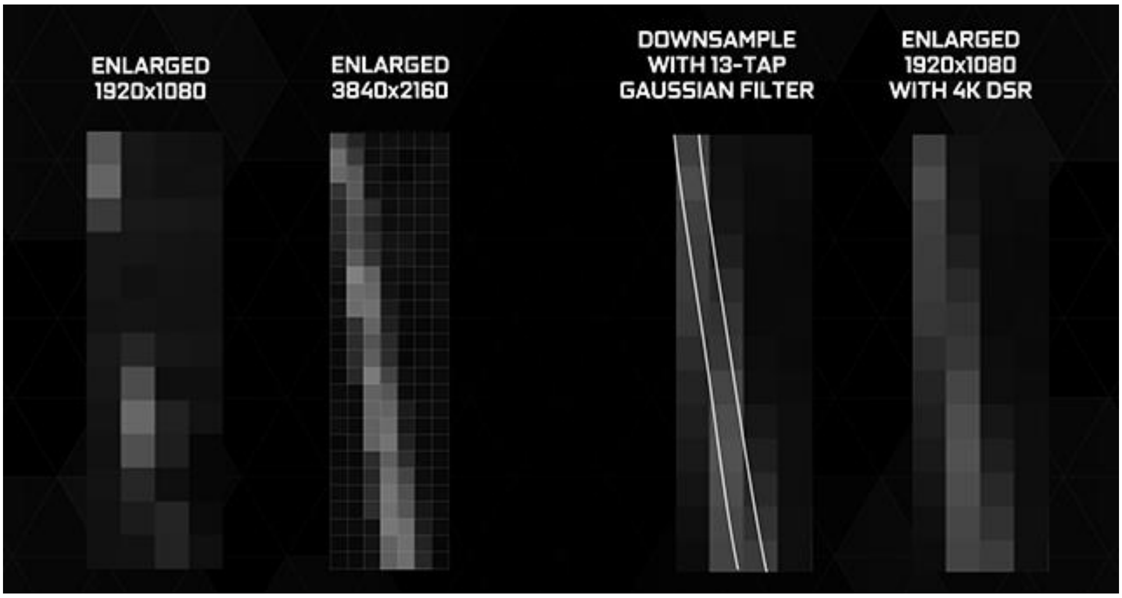


Figure 3- Example of Nvidias commercial application of this research for up scaling gaming (Burnes, 2014)

As such with the potential issues that come with dealing with such small images considered, it is imperative all current research which could optimise our elephant detection model by either optimising the ‘Faster R-CNN’ model for smaller objects or can handle resizing of objects in a more intelligent manner is considered. Figure 3 shows an example of upscaling and how it could more intelligently interpolate values so as to provide higher quality elephant images.

The first paper to address a relevant issue is a conference paper that deals with the domain of identifying faces in large images that are of variable size (Hu, 2017). This size according to the paper varies from 3 pixels tall to 300 pixels tall. The paper finds two key improvements which dramatically increase performance. The first is allowing the network to consider context - by allowing the network to attempt to classify by considering a larger amount of the space around the area which is being examined the accuracy is greatly increased - this potential improvement adds up to 18.9 percent performance. The second tweak is to adapt the resolution by training on a series of templates - for different sizes of faces. These templates are each respectively evaluated with each exemplifying increased performance both for the sets of small, medium and large faces as well as overall performance. As for the problem of identifying elephants with a satellite - the context awareness could be useful and could increase performance. The separate classifiers for size though will likely be of little help - as the sensors for the elephant photos are very similar and take photos at a roughly uniform distance - as such size variance should not be as large that distinct groups can be composed. It does however add more evidence to the idea that resolution of the objects we are attempting to classify is incredibly important.

Another application of 'Faster R-CNN' has been an attempt to track aircraft and boats using satellite images. The paper initially does acknowledge the merits of approaches which utilise variable resolution however it stipulates that it intends to take another approach as to avoid the computational cost of higher detail regions (Ren et al, 2018). This paper similarly focuses on improving contextual detection by passing on the information from the region proposal network (RPN) into the latter stages of the Fast R-CNN network. The paper also proposes to grow its relatively small dataset by adding with a set probability a chance to rotate the image by a random angle theta. By doing so it very greatly increases the size of the dataset when generating. This paper once more provides us with an idea for providing contextual information to our model to greatly improve performance furthermore it gives us a data augmentation strategy which could provide additional training data should that be seen as relevant to improve performance.

Another approach to handling the small object detection problem is outlined in a paper much like the prior paper. This domain is very similar to ours the domain once more is with aircraft detection (Pang, 2019). However in a stark change the fundamentals of 'Faster R-CNN' are moved away from - this is done in this paper to move to a more lightweight neural network. This paper instead determines using a 'light detection' to find if an object may contain an image before continuing with its main detection. The other main addition in this paper is the splitting of the image down to slices with an overlap to assist with processing. This in theory allows for context to be retained for boundary images whilst decreasing the global area which requires attention at any given time. This overall produces a network which is faster than the competing networks however it does not significantly perform any better than any other untuned networks including 'Faster R-CNN' with these tweaks. As such these tweaks have not amounted to much other than an increase in frames per second, which is not a goal in this task to be increased.

Another paper of importance when considering how to improve elephant detection is a generic paper unspecific to any given domain but does utilise examples of training with regards to satellite data sets. The paper like many others fully utilises the 'Faster R-CNN' model although goes into more detail of the specific challenges posed by small objects as smaller objects struggle to be downsized and have high level features extracted (Quan, 2018). This network varies in attempting to modify the network by normalising the features so as higher level features which will not be present on our small objects do not take too great a precedence. The model boasts minor increases with up to 10 percent increases over the default 'Faster R-CNN' although its difficult to estimate how much it could improve our network as the elephants are fairly uniform in size throughout.

Another paper which grows once more on the optimisation of 'Faster RCNN' training is a paper that mainly focuses on the optimisation of the loss function. This suggests a new loss function "IIOU". It is a comparison of the relative area in the target bounding box and the proposed one by the network. With other considerations being made such as attempting to retain features of small objects this paper ultimately finds no major improvement over other models which have deviated from the 'Faster R-CNN' design but does overall perform significantly better when considering recall accuracy with respect to 'Faster R-CNN' (Cao et al, 2019). Overall though this could be the conscious decision to modify the algorithm so as to retain features as seen in prior papers this does boast an improvement in performance. As such it's hard to ascertain how much gain is provided solely by this new loss function nor its importance over traditional loss functions. A further paper which explores ‘Faster R-CNN’ with the specific domain of satellite imagery is a paper which once more applies the model to the domain of detecting aircraft. This algorithm once more differs from Faster RCNN in a couple of ways - the first is similar to the previous papers identified by modifying feature connection so that lower level features are persistent at the terminal state with a high amount of importance in detection (Zhu et al, 2019).

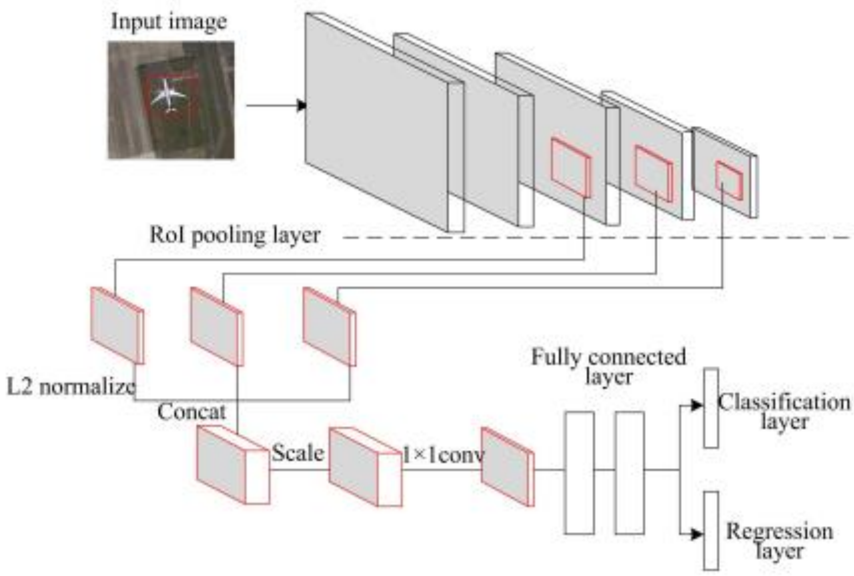


Figure 4- Example of Airplane detection network architecture (Zhu et al, 2019)

It does differ further though in that instead of providing a singular prediction this model provides an output of multiple bound boxes. This then makes use of an algorithm called NMS or non maximum suppression which iterates through all the possible boxes and checks the values against an intersection threshold. This terminally results in one, more accurate bounding box. Although this can cause issues if two objects are too close as the NMS may prune and lead to missed positives which have actually been identified. As such, knowing that elephants are social creatures and tend to travel in herds, utilising a feature such as NMS - which has minimal performance gains could cause positives to be missed as the elephants may be too close and the bounding boxes pruned.

Another paper which makes use of trying to increase context to increase performance of the ‘Faster R-CNN’ model can be seen making use of Faster R-CNN’s region proposal network to subsequently build up 8 neighbouring regions which are of equivalent size (Fang et al, 2018). A non maximum suppression algorithm is used to reduce the amount of proposals neighbouring and leaves the region proposed with only neighbouring regions with significant overlap. These regions which overlap likely contain additional contextual clues and are therefore useful when deciding which region to pass through for detection. The network was tested on a reduced COCO dataset which did see a performance uptake - although not very much. This could be a valid approach for trying to find other contextual clues around the elephants such as additional elephants however it is unclear from this research if this provides a significant gain over other contextual methodologies.

A paper which is also worth considering is one which attempts to solve the problem of small object detection when handling images from satellites, which is remarkably close to this project. This approach, unlike other papers, does move away from the 'Faster R-CNN' model and instead utilises a fully connected model of convolutional networks (S. Jiang et al, 2020). Like prior papers - this papers key optimisation strategy for its network over the traditional approach of 'Faster R-CNN' is to utilise a greater amount of contextual information by once more utilising NMS and misplacing boundary boxes so that a greater amount of contextual information is considered around the region of interest when attempting to detect an object. This result allows the more simple convolutional network to perform very well, so much so it roughly equals to mildly surpasses an unedited version of 'Faster R-CNN' when detecting objects, however the dataset is not always attempting to detect small objects sometimes including things such as baseball pitches in this example. As such it's hard to ascertain how much benefit can be produced by following this methodology and trying to optimise a detection network for detecting elephants using this context approach.

Taking a very different approach - now considering the aspect of improving the images by upscaling, one paper chooses to handle the problem of how to best upscale images by using arcs. This paper attempts to remedy primarily - jagged edges or the artifacting that can occur when shapes are replicated as they are increased in size. This algorithm tries to find arcs and aliasing these arcs to reduce jaggedness and maintain a more appropriate colour (Xu et al, 2019). Ultimately this algorithm is not particularly suitable for our domain - firstly it was not primarily designed with ultra small images in mind - this is noted in the paper where it states this algorithm is to assist resizing not attempt to improve image quality markedly. Furthermore this algorithm may fail to detect any patterns which it could assist with without creating artefacts whilst looking at the low pixel count elephants. Finally this algorithm was designed with speed whilst running on a CPU in mind - as such it notes that it is inferior in quality to machine learning techniques but does not need a GPU, a benefit that is of no use to us as our object detection will be trained on a GPU.

An approach which might prove more fruitful for the challenge of improving the images could be an approach outlined by Google in their paper (Romano et al, 2016) which proposes an outline for a machine learning model which can optimise filters during a resolution upscale to dramatically improve quality of textures. This is done by applying a machine learning model to a dataset of low resolution training data with appropriate high resolution expected results - the model is expected to learn an optimal mapping between a series of available filters and the expected result. As such this allows for a previously ineffective traditional upscaling to be turned into a far higher quality one - allowing for texture quality to be improved.

Figure 5- Example of RAISR upscaling on a turtle (Milanfar, 2016)

One of the downsides is that this algorithm can cause aliasing artefacts, these are always plausible and may be something that could be prevalent on the ultra small images of the elephants. However overall this provides an incredible viable approach for a more successful resizing approach which could solve the problems of features not being able to be identified due to the small size of the elephants. The algorithm clearly from figure 5 can manage with larger more distinct objects such as this close of up a turtles face.

One paper which seems to fully bridge the gap between just utilising a machine learning detection model and utilising optimised upscaling to improve performance could provide the solution for elephant detection. This paper is the first paper seen so far which makes use of a generative adversarial network - this network architecture is a design which pits two networks against one and other - the generator network must produce new images the discriminator network must attempt to detect fake images(Xing et al, 2019). As such they are pitted against one another to train simultaneously. This eventually leads to a generator network which can provide upscale estimates which will allow for the traditional ‘Faster R-CNN’ model to work. Overall this model seems to produce very strong results - with a noticeable increase in performance with regards to both mAP (mean Average Percentage) and missing rate indicating that this method can increase performance with regards to small target detection very effectively. This specific usage of rescaling on already very small images and subsequently measuring performance is a good indicator that such an approach will be very valuable when applied to detecting elephants.

The concept of applying a GAN is taken even further in another paper which applies the use of GAN upscaling to the domain of satellite images which is perfect for us - the paper even further benefits us by utilising the ‘Faster R-CNN’ for the object detection tasks (Jakaria et al, 2020). This model whilst using the same fundamental GAN model as the prior paper does make some distinct changes. First and foremost it incorporates an edge model into the generation - this model seeks to detect coarse edges and seeks to smooth these features out - this improves the possible visual information gain and visual pleasantness of the images after the upscaling the process.

Figure 6- Example of GAN image resizing (Jakaria et al, 2020)

Overall this approach produces the best results seen so far. With this approach models are assessed after being trained on either low resolution, high resolution or super resolution. Super resolution being the resolution produced by the network - high resolution just the training target data and low resolution is the downscaled data. This can be seen in figure 6 with the middle images looking vastly sharper than the original images. When moving from low resolution to the super resolution a performance delta of 40 percent can be observed with super resolution data achieving near perfect scores. This indicates that this approach could be the best approach so far in terms of optimising elephant detection. This is then developed even further in a paper which whilst not specific to the domain of satellite imaging is once more a paper which builds from the model of ‘Faster R-CNN’. This model once more re-introduces the discriminator model and the generator. Which like the previous implementation is trained using images versus downscaled images so that the model can learn how to both detect fake upscaling and produce fake upscaling (Li et al, 2017). In this model when compared with a traditional ‘Faster R-CNN’ without any upscaling the performance uptake is far more staggering. The initial results of ‘Faster R-CNN’ for recall and accuracy respectively are 50 percent and 24 percent. The model when trained on data using a trained upscaling GAN produced results of 89 percent and 84 percent respectively when considering small data. The differences were overall nonexistent when considering the dataset for larger items indicating that the upscaling is working correctly so that features can correctly be mapped. Furthermore this paper showed the GAN performance and thus the amount of improvement that can be gained is very dependent on the size of the dataset as such considerations should be made to optimise the dataset. This is a further increase into potential for the application to elephant detection as this implementation could see a dramatic increase in model performance.

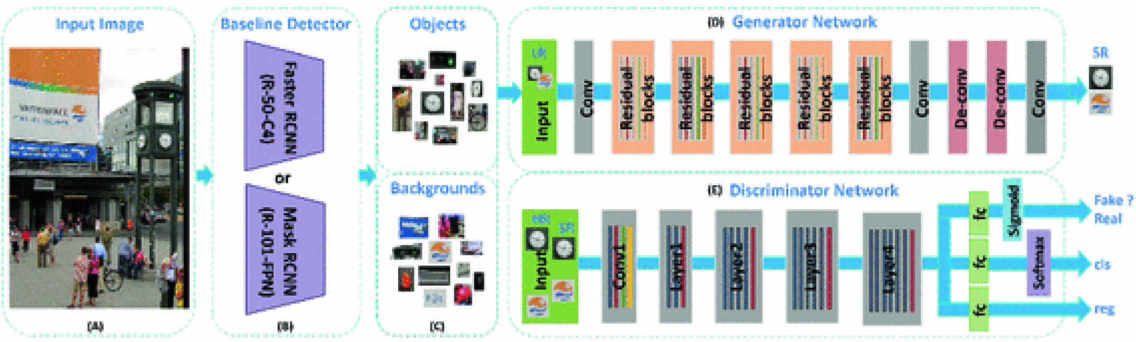


Figure 7– SOD-MTGAN Network diagram (Bai, 2018)

In yet another paper this concept of GAN optimisation of small object datasets is further explored. This paper utilises the same adversarial network structure as previously noted but places a hard limit to prevent artifacting of a 4x increase in the ‘super resolution’ by the GAN. Unlike prior models the losses from the detection model are also back propagated through the GAN in an effort to produce higher quality images in the regions which are going to have the largest impact on detection (Bai, 2018). The model architecture is still very similar to other GANS as can be seen in figure 7.Overall much like prior papers this does cause a noticeable increase in performance with the new ‘SOD-MTGAN’ being the top performer in every performance metric for small object detection. This adds further weight to the idea of pursuing an intelligent upscaling technique to optimise elephant detection.

The final paper which shall be considered - works with a very similar domain in that it utilises UAVs or unmanned aerial vehicles which have limited sized cameras but naturally are not quite as high as the satellites being used. As a result the datasets are fairly similar and the challenges of small object detection are still relevant. As much like the prior GAN papers the same basic GAN architecture is utilised with the explicit purpose of producing super resolution images which can then be trained on using Faster R-CNN to produce higher quality results (Y. Chen, 2019). The findings in this paper are very much in line with the prior papers - finding a 10 percent uplift with this specific implementation which is still significant and could prove crucial in detecting elephants.

When looking at all of the literature some common threads begin to develop. Firstly context is exceptionally important as such, when building up the 'Faster R-CNN' model a modification early on to allow for a greater degree of context would likely be a very sensible choice. This could allow for detection of other features in the nearby area which could clearly distinguish if this object was an elephant or not.

Furthermore when looking at the literature another issue is the amount of data present and how to optimise this. One approach which was used in a similar domain was the idea of randomised rotation of images and labels to augment the dataset to make it far larger whilst still providing ample learning experience for the model. Such an approach could be used on both the detection model and any additional upscaling models which would be used during this.

Another approach outlined in this is the attempts when processing the small images to modify 'Faster R-CNN' to make greater considerations of the smaller features. These features are often disregarded and as a result smaller objects tend to get missed by this model - to counter this modifying the model slightly to preserve or process differently these features could allow for greater performance.

The final major beneficial approach which was identified during this literature review was the concept of attempting to rescale the image in a manner which could assist with training. Whilst the model provided by Google was promising with the dynamic filter selection, ultimately it was unproven in this domain and as such should likely be left untouched for now. However the gains offered by some of the potential GAN structures were fantastic - as such a GAN to provide upscaling to target regions with a few minor tweaks could provide an optimal solution to elephant detection using the satellites. This would mean providing upscaled images with higher quality textures as the GAN thought was correct as training data for our model detection and model identification.

Although it is also clear than avenues for more traditional algorithms should be explores so as to provide variety and explore performance of up and coming machine learning approaches to upscaling versus more traditional approaches so as to see how they impact the performance of the model.

Optimisation of Elephant detection

Overall a clear focus on optimising the image upscaling process has the greatest potential to increase performance in the most significant manner. As such it shall take up the bulk of the work in optimising elephant detection. Should adequate time be available considerations should be made to alternative approaches which could also boost network performance.

Chapter 3

Implementation and Testing

To begin with implementing this project means first processing the satellite images – as these images too large for any model to handle they require being split up. For this implementation the images were decided to be split up 600x600 as this provides a large amount of contextual information, doesn't potentially split up herds too often and still retains enough detail whilst allowing a lot of space for upscaling.

Figure 8– example of a sliced region at 600x600 gathered around watering hole

The quality of the splitting can be seen in figure 8, in which a herd can be seen almost entirely from every angle around a watering hole. That is not to say that just simply splitting every 600 pixels is the approach that will be taken. Instead each split will have a 5% buffer or 30 pixels. This means that each snippet will overlap with the previous – ensuring that any cropped elephants can be displayed fully in one dataset preventing any potential information loss around edge cases.

For the specific choice of network the model of ‘Detectron2’ was chosen. This model was originally designed around instance segmentation however as it was built up on top of ‘Faster-RCNN’ this enables us to utilise it for bounding box segmentation which is useful for this project as that is the dataset we will be working with.

As such the optimal baseline was selected from model zoo to begin training with by selecting utilising AP performance on the coco dataset. The best performer in this regard was a model named ‘X101-FPN’ and as such that is the baseline which will be used throughout all experiments.

One of the major advantages of utilising Detectron2 was that the data augmentation in Detectron2 by default runs performing flip augmentations on all sides.

Figure 9– Example of elephant dataset being rotated

This is a significant advantage for this usage case as with spatial data the angle of the elephants doesn’t matter – as the photos are from above as such this augmentation approach just guarantees a better training set for the models to learn from. This is visualised on some elephant spliced data in figure 9.

As can be seen in figure 9 the elephants are currently very low resolution as they represent a small area of the total satellite image. One of the key considerations is how to subsequently upscale the data and to what size. By experimentation it was found that 2400x2400 was the limit of what can be trained in a reasonable time on a modern GPU – as the resolution increases beyond this it becomes too slow to realistically test as many models as would be needed.

Furthermore tests will be required to ensure that models which are trained in counter intuitive ways to our models do not outperform them. This means assessing affects of aspect ratio and downscaling to ensure that an easy optimisation strategy is not inadvertently looked over.

Define RDN – PSNR

The method being primarily used for the upscaling of our models is one of the current best super resolution algorithms (Zhang, 2018). This model makes use of extremely deep convolutional neural networks and a memory system via block designs to allow for the model to dynamically learn features and decide how to best upscale them.

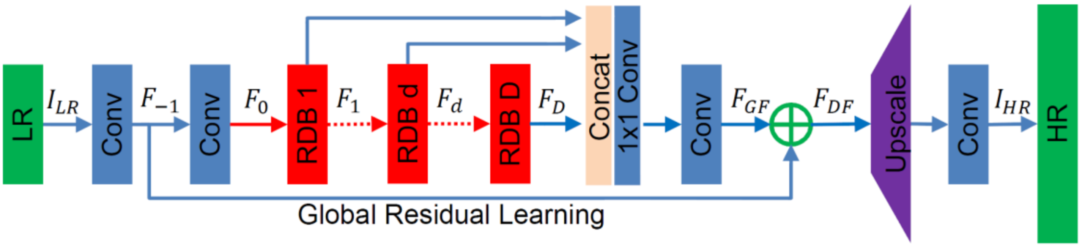


Figure 10– Example of some of the block architecture underpinning RDN (Zhang, 2018)

The full architecture of the RDN is displayed in figure 10, showing the simplicity but also with how well the network scores relative to other super sampling techniques this is likely one of the best approaches.

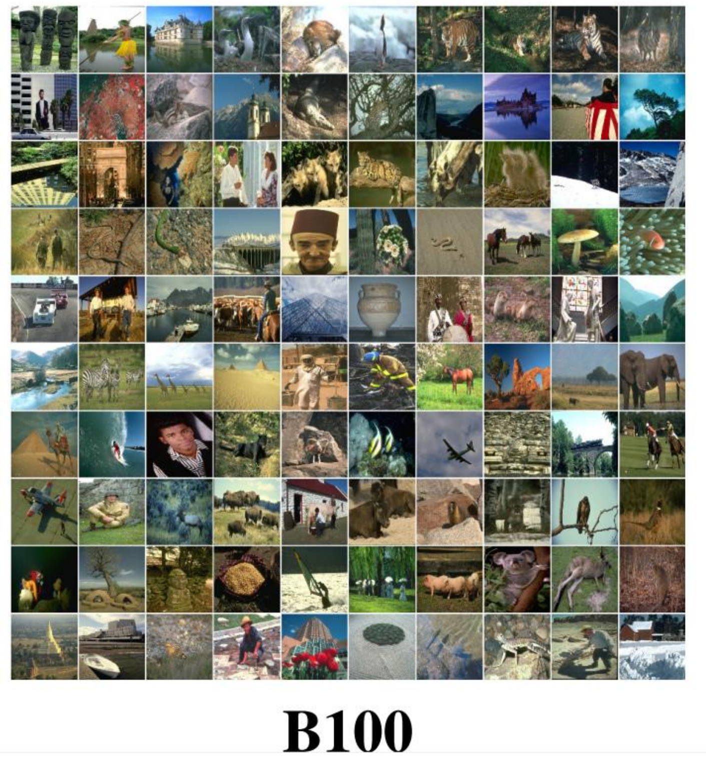


Figure 11– Dataset example from DIVerse 2K resolution (Eirkur, 2017)

Another advantage is part of the dataset which this model was trained for super resolution on includes some aerial shots and a variety of natural environments of various colours.

There are two variations of these RDN networks. Each has a different pooling configuration and separate weights – small or large. These two models functionally are the same algorithm and were trained on the same training set which is close to our set but should diverge in performance significantly as they are set up to produce different amounts of noise.

Another model which we are using to upscale our base data is RRDN which is a trained version of ESRGAN (Wang, 2018). This is a GAN which makes use of additional VGG weights to attempt to minimise the losses.

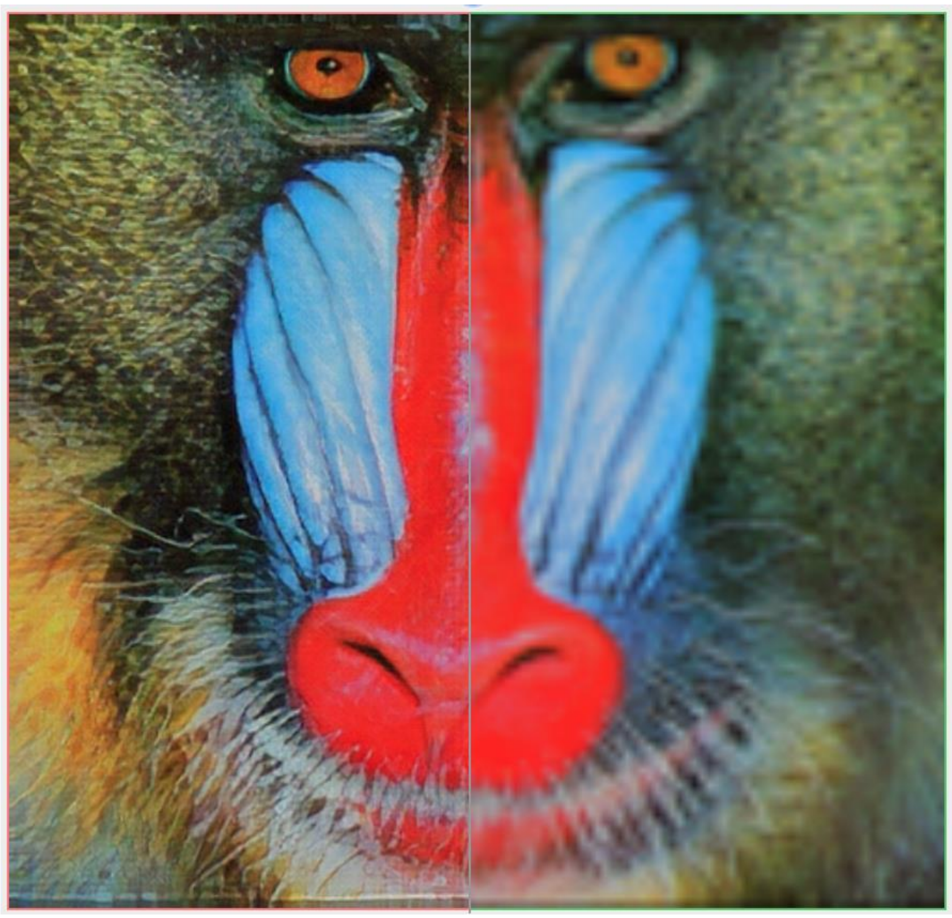


Figure 12 – RRDN Upscaling of a monkey – Upscaled left, source right

The model we are using is also pre-trained as can be seen of a test image performed by the network on a monkey demonstrating the enormous capabilities of RRDN. In theory this could be the best network however it is unclear how well this model can adapt without a dataset to train on specifically of satellite images of elephants.

Bicubic interpolation is one of the best conventional algorithms for up sampling images as a result of the amount of data it takes into account for a single interpolation. Bicubic interpolation takes into consideration a mixture of features from 16 different nearby pixels.

$$p(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j.$$

Figure 13– The equation for the interpolates surface after bicubic interpolation

The equation in figure 13 is used to resolve the coefficients which are used to produce the surrounding interpolating pixels.

Bilinear interpolation in contrast to prior approaches is a much more simplistic algorithm considering only the weighted average of all pixels in each column and row of the pixel to be estimated. This means that the weight decreases and the contribution becomes less significant the greater the distance from the pixel being estimated.

Another consideration which has been key when considering testing has been the consideration of how to measure performance of the networks. The first two measures of performance which are going to be used are AP and AP50.

AP or Average precision is the default measure of performance in detectron2 for bounding boxes and is the basis from which we will base most of our results.

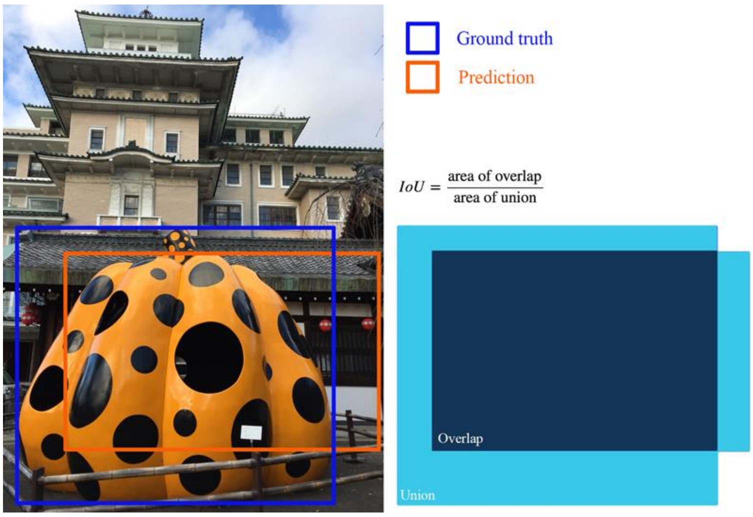


Figure 14– Showing IOU (Hui, 2018)

In figure 14 the IOU can be shown as a measure of how much overlap is present between the suggested label and the actual label. If this passes a threshold and it was an elephant labelled this is a true positive. In AP the score is a mean of a bunch of scores at various IOU threshold values where as in AP50 the IOU is always at 0.50.

The final measures being used are false detection percentage and accuracy percentage. Accuracy is the (total number of detection – false detections) /Actual number of elephants – this results in a score which gives a good estimate for how well these models perform according to the human eye – as there will no IOU but simple judgements to evaluate performance. Furthermore, the false percentages is just the false detections as a fraction of total detections. These two methods must be done by hand and as such will be limited to a dataset of about 10 test images to make testing all the models feasible.

All models will subsequently be trained on the appropriate dataset with all AP and AP50 scores being appropriately generated and the subsequent accuracy and false detection being investigated by humans after.

Chapter 4

Results

Once the images have been split and a dataset has been created processing can begin to be undertaken. The first step of this is running the upscaling algorithms and visualising the results of the upscaling algorithms. These can then be analysed for anomalies at a visual level by comparing the upscaled images to the ground truth image.

The first of these algorithms to consider is RDN-SMALL, one of the deep learning based methods of upscaling – this network naturally being the smaller of the two RDN based models being evaluated.

Figure 15– Left ground truth region at 600x600, right RDN-SMALL resized region at 2400x2400

Figure 15 shows a small section of a larger image – and is selected as it shows a variety of colours as well as vehicles and most importantly elephants. The results demonstrated in Figure 15, show that the model does successfully upscale the image as per expectations – but a hatching can be noticed across some of the water. This doesn't seem to affect the quality of the elephants pictured but this slight noise is persistent throughout all RDN-SMALL generated resolutions and could be detrimental to performance compared to an ideal upscale.

The most similar of the algorithms to this, RDN-LARGE, also performs very similarly to RDN-SMALL in that we once more can observe a noise being generated but this time of a much more distinctly small variety which is harder to notice.

Figure 16– Left ground truth region at 600x600, right RDN-LARGE resized region at 2400x2400

This isn't distinctly too different from the results generated by RDN-SMALL, the noise is still present in the image although slightly subdued – however the quality of the elephants appears to be equivalent. Boundary regions between foliage and water seem consistent with expectations and the shrubbery in the middle appears to also be equivalent.

The third of the deep learning networks being evaluated for upscaling is RRDN, which is the largest network and also the highest performing network according to multiple benchmarks of the three adversarial networks for upscaling being evaluated.

Figure 17– Left ground truth region at 600x600, right RRDN resized region at 2400x2400

Unlike the prior networks, this one has a peculiar habit in that the noise seems to be generated much more aggressively than in previous networks and in waves. This can be seen in Figure 17 showing a much more noisy image than previous models produced. This doesn't seem to be symbolic of network collapse as the outputs between pixels are valid interpretations which do seem moderately consistent however the network seems to produce blotches of colour while upscaling leading to details becoming very unclear.

Figure 18– Left RRDN, Middle RDN-LARGE, Right Bicubic Interpolation

The issue of details becoming significantly lower in quality can be seen in this zoomed in comparison image where the details of the RRDN generated elephants suffer a large drop in quality and recognisability versus the prior RDN generated elephants. As such this is likely to significantly affect the detectability of the elephants in the upscaled images produced by this model.

In contrast to the prior approaches to upscaling we have the Bicubic interpolation method. This method is a more traditional algorithm which has eclipsed a lot of other conventional upscaling techniques in lieu of its superior performance.

Figure 19– Left ground truth region at 600x600, right Bicubic interpolation resized region at 2400x2400

This algorithm seems to lack a lot of the shortcomings of the other algorithms, whilst some noise can still be seen the pattern is nowhere near as severe as the RRDN method. Furthermore as can be seen in figure 19, the elephants are clearly visible and the image once more has been created in line with the expectations of it.

As can be seen in figure 19, the issue is no longer noise but rather individual colour pixels being able to be able to be picked out from the elephants as the bicubic approach has placed them in shadows – creating occasional erroneous pixels. However these mistakes are hard to distinguish and require a lot of focus to notice.

Figure 20– Left RDN-LARGE 2400x2400, Middle Bicubic interpolation 2400x2400, Right Bilinear interpolation 2400x2400

As can be seen in figure 20, the image does not ever lose the elephants to the blending but certain details can become slightly more subdued when elephants are on a background which matches the colour of their skin.

The final algorithm which is being utilised for resizing is bilinear interpolation. This once more a conventional non machine learning algorithm which is an extension on linear interpolation.

Figure 21– Left ground truth region at 600x600, right Bicubic interpolation resized region at 2400x2400

This algorithm much like the results from bicubic interpolation produces very little noise, as can be seen in figure 21. However when examined much closer as seen in figure 20, the top half of the elephant is likely blended too much mixing the edges with the ground disguising it due to the similarity with the background and skin. This could potentially result in degraded performance.

Training results

When training models on these datasets there were a variety of scores that were tracked. One of the metrics used was an internal metric tracked on Tensorboard known as class accuracy. This measured the accuracy of the detection when a region proposal box has a class prediction inside.

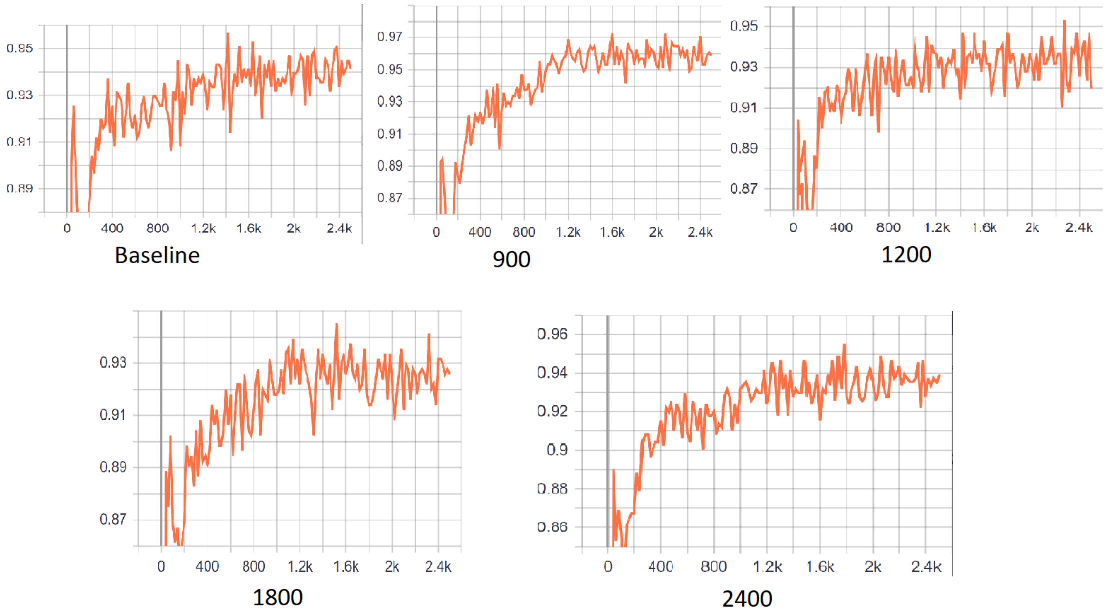


Figure 22– class accuracy scores, scored from 0 to 1 using RDN-LARGE as the model with resolution denoted beneath each graph

As can be seen in figure 22, the accuracy for class accuracies are consistent throughout all models trained using RDN-LARGE. All models finish within a few percentage points of each other and well within the random variations that can be seen occurring when the models have plateaued. This is not just a story for RDN-LARGE upscaled models, this trend is consistent with the results observed across every upscaling model.

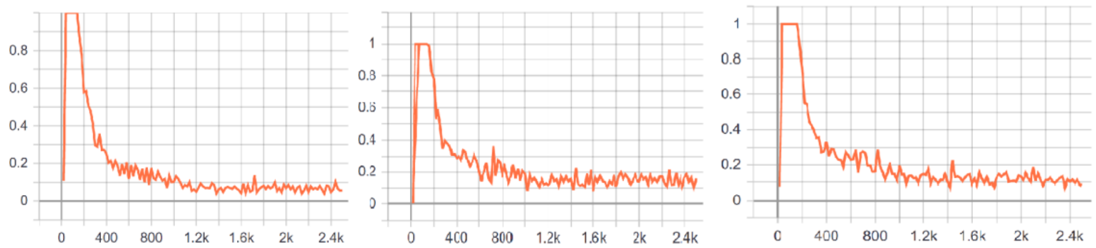


Figure 23– Class false negative scores, scored from 0 to 1 using RDN large, with baseline 600x600 on the left, 1200x1200 in the middle and 2400x2400 on the right

Similarly as can be observed in figure 23, the false negative scores which monitors the likelihood of detecting an elephant when none are present in the bounding box converges for all resolutions of all models in a similar fashion and within margin of error.

The main metric which is used however to evaluate the performance of bounding box classifiers is AP or average precision. For measuring we use the default detectron2 measure of mean average precision denoted as AP which takes an average AP from a set of IOU values ranging from 0.5 to 0.95 as well as AP50 which is the average precision when the IOU has a value of 0.5.

	Scoring	600x600
Baseline	AP	7.112
	AP50	26.951

Table 1- Showing baseline performance for detectron2 on images without any upscaling

The baseline performs reasonably well and provides a model which can very reasonably detect elephants – the scores it provides whilst low compared to detection on a COCO dataset represent scoring for exceptionally small objects which can also blend very easily into the surrounding environment in aerial shots.

Figure 24– A test image with contrast edited and baseline detections from trained model shown on left, 600x600

As can be seen the model provides a reasonable estimate and roughly does bind the elephants – if not slightly a miss of the original labels. But does still miss the less obvious 4 elephants in the image. This is however the smallest model and should if the literature is correct be superseded by models which have been upscaled using an accurate algorithm to do so.

Upscaling:	Scoring:	900x900	1200x1200	1800x1800	2400x2400	Linear correlation
RDN-SMALL	AP	5.915	3.996	6.674	12.778	0.864962
	AP50	23.111	14.767	27.149	50.905	0.871975
RDN-LARGE	AP	9.299	5.659	9.651	11.662	0.680345
	AP50	36.286	20.922	34.942	45.26	0.640882
RRDN	AP	10.388	4.19	5.186	5.658	-0.49953
	AP50	34.717	17.539	20.856	17.370	-0.66965
Bicubic	AP	10.781	8.64	6.691	11.334	0.079872
	AP50	43.739	31.778	26.553	42.298	-0.03444
Bilinear	AP	8.845	12.591	6.755	8.392	-0.45694
	AP50	34.359	41.798	30.852	33.633	-0.44769

Table 2– Showing performance of models trained on upscaled datasets

When comparing the results of the upscaled datasets as seen in table to the results of the baseline model in table we see a dramatic uplift in performance. Only 8 models perform worse when measuring by AP scores, of which only one is in the 2400x2400 pixels resolution. Furthermore only 7 of the models are worse by AP50 scores.

This in combination with the best performing model achieving an AP as well as an AP50 score which is almost twice as good we can begin to see that resizing has had an affect on the performance.

Furthermore when a linear correlation is ran between the resolution and the upscaled models AP scores we see a stark result models such as RDN-SMALL and RDN-Large show a very clear correlation between AP and resolution meaning that as resolution increases so to does the end performance of the elephant detection. Bicubic however shows a very mixed bag – because it initially performs so well at 900x900 (a result that was retested and replicated to ensure accuracy) as well as performing so poorly at 1800x1800 it throws off its correlation for a model which otherwise holds true to the same rule with the highest resolution model scoring the best in AP.

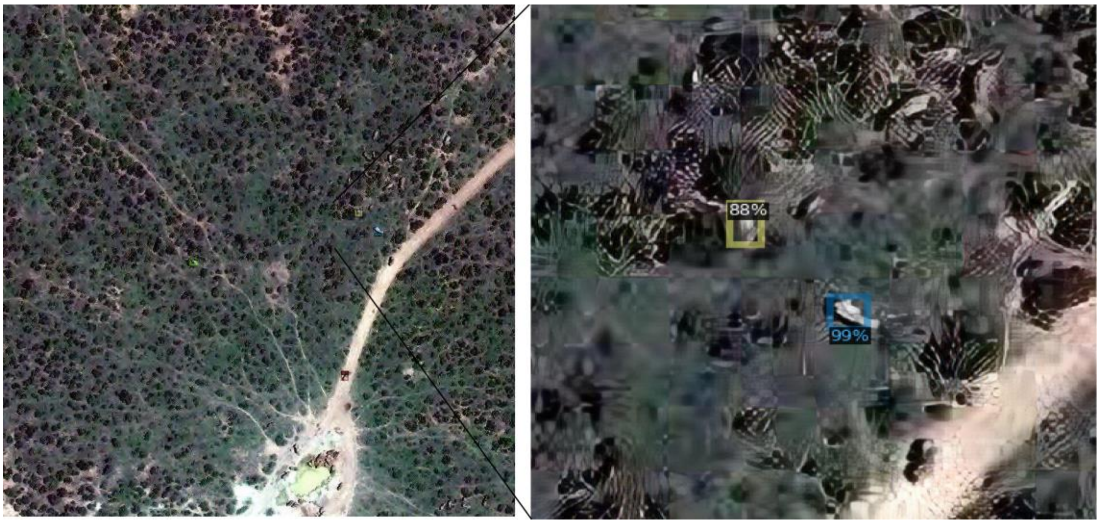


Figure 25– An example with an upscaled image from RRDN at 2400x2400 with a slice of the image zoomed in on the right

RRDN however holds true with the inverse of this. This might have been to be expected however, the model had been demonstrated earlier producing a much greater amount of image noise making the images look visibly distorted when visually judged. This is demonstrated clearly in figure 25, in which a region which had a lot of foliage and elephants present is massively distorted during upscaling to 2400x2400 to the point at which the image becomes more like a dream when examined up close than a satellite image. The noise brought through as resolution increases and the model causes more significant noise likely is what leads to less information being available leading to a decrease in performance by the model as resolution increases.

Figure 26-, ground truth with bounding boxes, middle Bilinear upscaled to 2400x2400 and right – zoomed in region showing false detections amidst real elephants

Bilinear in contrast to RRDN does not fail as obviously – the model performs better than baseline but does not scale very well. This seems to be as a result of the model blending details out too often and as a result random lumps or shadows can be misclassified. This happens especially often for the model incorrectly finding elephants in other elephants shadows as shown in the figure 26 as this area often contains a calf and as such this behaviour is positively reinforced in the data.

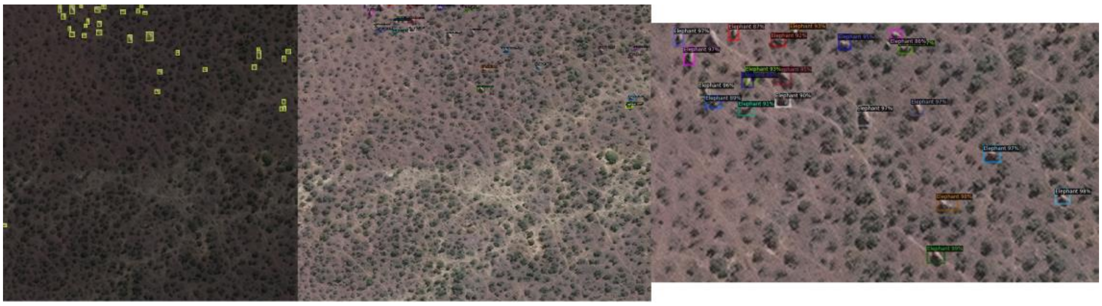


Figure 27-, showing ground truth on left with elephants bound yellow and ground discoloured for clarity, upscaled image in middle using RDN-SMALL at 2400 and zoomed in detections on the left

As can be seen in figure 27, in contrast the performance exhibited by the models which score highest with regards to AP – in this case RDN-SMALL at 2400x2400 allow for fantastic detections finding almost every elephant in an image with very few false positives. This represents a large step forward over the baseline.

In the prior models demonstrated – the retaining of aspect ratio was key. This seems like an obvious ideas as to change aspect ratio would mean elephants from different angles would have vastly different characteristics, however in order to ensure this idea underpinning the models tested so far was accurate two models were tested stretched to a 4:1 aspect ratio using bicubic interpolation, one of the best algorithms tested so far.

	Scoring	600x2400
Height stretched	AP	6.263
	AP50	23.679

Table 3– Showing performance of models trained on a height stretched dataset

The results of the first aspect ratio stretch were as expected. As can be seen in table , the model underperforms compared to baseline despite having been increased by bicubic interpolation dramatically in size. As a result, this model is slower to run, requires more memory and performs worse by every AP metric than the baseline.

	Scoring	2400x600
Width stretched	AP	5.676
	AP50	22.299

Table 4– Showing performance of models trained on a width stretched dataset

The prior results indicating altering aspect ratio are further supported by the results of table . This finds an even worse score for altering the aspect ratio by changing the width with a 4x stretch using bicubic interpolation. This is indicative that the practices used in our testing by maintaining aspect ratio were positive towards building the best possible model.

However, that does not mean every avenue has been exhausted. In the dataset certain models did perform better at lower resolutions, even bicubic interpolation had an incredibly good run at 900x900. As such lowering the resolution by downscaling the base images should be investigated as a possible avenue of improving performance.

	Scoring	300x300
Downscaled	AP	1.4423
	AP50	8.461

Table 5– Showing performance of models trained on a downscaled dataset

This model however does give further credence to the hypothesis that more information, given by a higher resolution does mean better performance as the scores are woefully low in table when compared to the baseline and upscaled scores.

A novel idea which I also investigated was an idea to improve the RRDN model by reducing the amount of interpolation it has to do. This was done by having a bicubic interpolation method turn the 600x600 images into 1200x1200 images which will have more information and we know from training provides a good baseline. The theory behind this was that by allowing RRDN to have more context to work with perhaps the noise could be reduced and the model which could have been the fastest could produce higher quality images.

Figure 28– Contrast edited ground truth showing elephants left, detections on combined approach middle, zoomed in detection and missed elephants right

Unfortunately this model did not remedy the issues present. The noise is still very strong and this reduces the visibility of elephants significantly – as can be seen in figure 28. The right of this figure has 3 elephants in that region but it's almost impossible to find any bar the one highlighted as the distortion makes them impossible for even a human to detect.

	Scoring	2400x2400
Combined approach	AP	9.272
	AP50	36.684

Table 6– Showing performance of combined approach method

The performance overall of the combined approach is acceptable. Table shows a tangible increase from the RRDN performance of AP5.658 and AP5017.37 at 2400x2400. This does suggest that this has improved the model greatly however it still performs worse than interlinear, RDN-SMALL and RDN-LARGE making this unexceptional at the very best.

The final model being investigated is a modified version of RRDN which utilises an additional network to identify regions of an image which have high amounts of noise generated during the upscaling process and subsequently attempt to correct for them. This model on paper seems like the perfect solution to the significant issues with upscaling utilising RRDN.

	Scoring	2400x2400
Noise cancellation	AP	5.414
	AP50	21.493

Table 7– Showing performance of noise cancellation method

This model theoretically could have solved the issues with noise in the images however it appears based on the results in table that the model has done nothing to remedy the issues and in fact made it worse. The RRDN model by default at 2400x2400 had AP 5.658 and AP50 17.37 respectively – whilst this model underperforms on AP and does not significantly improve on AP50 to even beat baseline performance.

Figure 29– Contrast edited ground truth showing elephant left, detections on noise cancelled middle, zoomed in detection and missed elephants right

Unfortunately, noise cancellations almost entirely removed elephants from visibility to combat the noise. Whilst the noise is still present it is clearly not as severe however as can be seen in figure 29 the elephant in the right photo can no longer be seen as it has been blended to aggressively as such the network can not detect it either. This results in far lower performance than otherwise would have occurred.

Whilst AP and AP50 were the measures chosen to perform the majority of the evaluation of our models with that is not to say they are wholly indicative of the performance of the models nor are they the only measure with which the models can be assessed. Two alternative methods which were found when going through the results of the output images were to assess the percentage of the false positives and the percentage correct of all potential elephants in the image. Whilst these seem similar to class accuracy and false negative those values were assessed during training whereas these values were done by hand on a series of 10 test images per model. These had to be done by hand as detectron2 does not offer support currently for reporting statistics such as false positives and accuracy on a test dataset. As such every image was check twice over for every elephant and false detection that by eye was not reasonably overlapping with an elephant bounding box.

Upscaling:	Scoring:	900x900	1200x1200	1800x1800	2400x2400	Linear correlation
RDN-SMALL	False positive	7.407	10.769	13.888	7.936	0.1065
	Accuracy	46.913	27.692	44.444	57.142	0.619
RDN-LARGE	False positive	23.255	28.571	11.627	5.882	-0.911
	Accuracy	39.534	17.142	34.883	66.666	0.727
RRDN	False positive	10.294	17.948	19.562	23.076	0.9067
	Accuracy	44.117	38.461	36.956	36.538	-0.822
Bicubic	False positive	24.390	6.666	6.153	6.250	-0.690
	Accuracy	36.585	46.666	32.307	47.916	0.302
Bilinear	False positive	14.634	11.111	13.888	11.76471	-0.362
	Accuracy	48.780	53.333	41.666	42.647	-0.765

Table 8– Showing percentage false positives and accuracies of various models as well as linear correlation of percentage versus resolution

Overall the results in table don't seem too far from the results seen in the prior AP results table with the dominant models once more being RDN-SMALL, RDN-LARGE and bicubic interpolation.

RDN-SMALL once more is a top performer with one of the lowest false positive rates at 7.407% and an accuracy score of 57.142%. Unlike the previous model the 900x900 model of RDN-SMALL has performed better in some ways such as the false positive rate, albeit to within a margin of error. Overall though this model has too many sporadic readings in false positives for it to correlated as expected – which would be in a negative trend which is surprising. However, this is likely accounted for in the very strong correlation of accuracy with a coefficient of 0.6 showing that the model whilst still making mistakes at the same very low rate as initially, does become better at finding even more elephants as the resolution increases.

This contrasts with RDN-LARGE which performs precisely as expected with the initial hypothesis that increasing resolution will increase performance with a good interpolation algorithm. This can be seen in table with a false positive rate of 5.882 and an accuracy of 66% vastly surpassing any competing algorithm.

Figure 30-, showing ground truth on left with elephants bound yellow and ground discoloured for clarity, upscaled image in middle using RDN-LARGE at 2400 and zoomed in detections on the left

Whilst this model is considered by these metrics to be the best, it still can be seen as in figure 30 that the model is not perfect – detecting only 3 out of the 5 elephants present in that image. However the model does considerably better than other models and it strongly supports the idea that increasing resolution when an 'upscaler' is able to make information more clear for the object detection algorithm provides massive benefits for performance.

Once more opposing the trends set by the RDN based models is the similar RRDN model. This model once more is one of the worst performers consistently regarding performance in table .With both very accuracy percentages and very high false positive percentages this model consistently underperforms expectations.

Figure 31- Contrast edited ground truth showing elephant left, detections on RRDN middle, zoomed in detection and false positive elephants right

This can be seen very clearly in figure 31, which demonstrates the RRDN model detecting the one elephant in the image but also incorrectly detecting a tree as two elephants. This is likely an issue of it seeing one large object resembling an elephant and falsely placing a second label for a calf, but these failures are very common on this upscaling technique. Furthermore the noise issue leads to a negative correlation with resolution and accuracy and a positive correlation with false positives. This means the network in every single metric from AP to false positives or accuracy gets definitively worse with resolution increases as the noise makes the images too hard to work with.

Bicubic interpolation also performed very well once more with its final false positive rate being incredibly low whilst the accuracy never quite got up to par with RDN-LARGE. However the model followed the predicted trend and seems to show that it can reliably as an upscaling algorithm increase performance with regards to false positive rate and accuracy.

Finally Bilinear interpolation also performed similarly to prior tests by defying the expectations of the model – likely by degrading feature quality too much. The model does however manage to get a good correlation of false positive rates – if not a little high when compared to other models but loses correlation with the accuracy being vastly inconsistent as the images increase in resolution.

	Scoring	600x600
Baseline	False positive	17.105
	Accuracy	44.315

Table 9– Showing percentage false positives and accuracies of the baseline model

The baseline model is rather surprising – whilst the false positive ratio is very much larger than anything produced by RDN-LARGE, RDN-SMALL, bicubic interpolation or even bilinear interpolation the accuracy does surpass certain models. As such it may mean depending on the purpose whilst a model could score a higher AP there could be validity in moving away from it if catching every potential elephant is the goal and occasionally getting false positives is fine.

The model still is clearly surpassed though by some upscaled models showing that even in this metric the upscaling still is providing significant performance advantages.

Chapter 5

Conclusions

Overall this experiment has proven itself to be a success. Satellite data which is pushing the limits of civilian image capturing capabilities has been processed in an efficient manner and systematically evaluated.

The model which was built as baseline served its purpose and it successfully provided a benchmark score which was not poor – but was certainly beatable. By running the models on the same Detectron2 model and making use of the FASTER-RCNN model it utilises we were able to have in excess of 30 models tested in the course of a few weeks. This is not to say that merely using detectron2 is the best method going forwards – as object detection improves it's likely this experiment could be repeated with vastly improved scores using future object detection models as and when they become available to use.

A potential improvement could be making use of a greater variety of resolutions and being able to make greater use of retests. Whilst we managed 4 resolutions of upscaling ranging from 1.5x to 4x as GPU technology improves it should become feasible to dramatically increase size past 2400x2400 without suffering vast increases in training time. Furthermore with more resolutions just been 600x600 to 2400x2400 a more nuanced analysis could be performed – the current analysis using just AP and AP50 with linear correlation whilst accurate could likely be more nuanced to better describe ways future models could be optimised.

On top of this the second method used to assess the models utilising manual measurements of false positives and accuracy could be done better. With more time a tool could likely be implemented into detectron2 to provide a more accurate answer alongside being able to do this analysis on a larger dataset such as all of testing instead of a sample of images.

The novel idea to try feed in conventional interpolation techniques to try and start off poorer performing GAN based models is a point which could be of interest. This did increase performance – although not more than the bicubic interpolation it could be used in the future for better performing GAN models as while this model did reduce noise considerably it wasn't enough.

Figure 32– Width stretched image

One thing these models did help with was providing analysis for negative cases in detail as additional evidence. Varying from aspect ratio stretching to see how the model can handle the aspect ratio changing appearance of elephants such as the width stretch in figure 32 to height stretching and downscaled models to show that we weren't accidentally missing optimal performance.

Another key idea in this paper which seems to be quite clear is that not all rescaling models are equal. Some of the algorithms can dramatically increase performance as a result of making contextual clues clearer to the model however some can hide contextual clues or change the image in such a manner so that elephants are simply no longer visible. As such blind rescaling using a potential 'best' model is not a good approach.

One way in which other people can improve upon this vein of research would be to experiment with contrast and colour filters as a method of altering the dataset – either for augmentation or to improve performance. Often some of the more naïve algorithms when upscaling images can accidentally blend features if the background around an elephant is too similar. As such transforming the dataset with regards to colour and contrast could be a valid strategy to overcome this and improve performance.

Other improvements noted in the literature review included ideas such as changing how the model calculated loss. Other new metrics such as improved intersection of union offer benefits in performance over traditional IOU based approaches. Whilst the AP score in detectron2 is a mean AP over a set of various IOU values and is good it stands to reason that better loss models will soon be able to surpass its performance.

A wasted piece of potential with this project was the lack of use of the infrared channels made available from the satellites. These channels could have proven useful and provided an extra piece of information to the models which could vastly improve accuracy. However the choice was made to not try increase channels being passed into detectron2 and the challenges that would bring in assessing functionality.

In conclusion, this paper provides an interesting insight into an optimisation technique for handling very small objects which provides a considerable performance increase in many metrics. Furthermore it shows that satellite data with only simple tricks is capable of being analysed in great detail allowing for insights into wildlife and the world which may not have otherwise been feasible.

Bibliography

WWF, African elephant, May 2020, Accessed on 1/05/2020, at https://www.worldwildlife.org/species/african-elephant

Drayman-Weisser, Terry. ”An Art Conservation Perspective: Saving the African Elephant and Ivory Cultural Heritage.” Curator 61.1 (2018): 161-86. Web.

T., Bindi, John D. Holland, and Edward O. Minot. ”Seasonal Home Ranges of Elephants (Loxodonta Africana) and Their Movements between Sabi Sand Reserve and Kruger National Park.” African Journal of Ecology 50.2 (2012): 131-39. Web.

I. Duporge, Olga Isupova, Steven Reece. “Using machine learning to protect elephants”. In PositionIT, March/April 2019. pp. 15-16.

K. Segl and H. Kaufmann, ”Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 9, pp. 2080-2083, Sept. 2001, doi: 10.1109/36.951105.

J. Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. ”You Only Look Once: Unified, Real-Time Object Detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (2016): 779-88. Web.

S. Ren, K. He, R. Girshick and J. Sun, ”Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137- 1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

COCO, Detection leaderboards, May 2020, Accessed 1st of May 2020, available at http://cocodataset.org/

Cao, Changqing, Bo Wang, Wenrui Zhang, Xiaodong Zeng, Xu Yan, Zhejun Feng, Yutao Liu, and Zengyan Wu. ”An Improved Faster R-CNN for Small Object Detection.” IEEE Access 7.99 (2019): 106838-06846. Web.

Fang, Linjing, Monroe, Fred, Kirk, Lyndsey, Schiavon, Cara, Yu, Seungyoon, Zhang, Tong, Wu, Melissa, Kastner, Kyle, Kubota, Yoshiyuki, Zhang, Zhao, Pekkurnaz, Gulcin, Mendenhall, John, Harris, Kristen, Howard, Jeremy, and Manor, Uri. ”Deep Learning-Based Point-Scanning Super-Resolution Imaging.” BioRxiv (2019): BioRxiv, Oct 24, 2019. Web.
Burnes Andrew, dynamic super resolution improves your games with 4k quality graphics on HD monitors, Nvidia, 2014, accessed on 1/5/2020 available at https://www.nvidia.com/en-us/geforce/news/dynamic-super-resolutioninstantly-improves-your-games-with-4k-quality-graphics/

Hu, Peiyun, and Deva Ramanan. ”Finding Tiny Faces.” Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017 (2017): 1522-530

Ren, Yun, Changren Zhu, and Shunping Xiao. ”Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN.” Applied Sciences 8.5 (2018): Applied Sciences, May 2018, Vol.8(5).

Pang, Jiangmiao. “R2 -CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images”. IEEE Transactions on Geoscience and Remote Sensing 57.8 (2019): 5512-525. Web.

Quan, Wei, “Small Object Detection with Multiscale Features”, International Journal of Digital Multimedia Broadcasting, Hindawi, 30/09/2018, available at https://doi.org/10.1155/2018/4546896

Zhu, Mingming, Xu, Yuelei, Ma, Shiping, Li, Shuai, Ma, Hongqiang, and Han, Yongsai. ”Effective Airplane Detection in Remote Sensing Images Based on Multilayer Feature Fusion and Improved Nonmaximal Suppression Algorithm.” Remote Sensing 11.9 (2019): Remote Sensing, Jan 2019, Vol.11(9).

Fang, Pengcheng, and Yijie Shi. ”Small Object Detection Using Context Information Fusion in Faster R-CNN.” 2018 IEEE 4th International Conference on Computer and Communications (ICCC) (2018): 1537-540.

S. Jiang, W. Yao, M.S Wong, G Hong, Z Kuc, T.Y and Tong X ”An Optimized Deep Neural Network Detecting Small and Narrow Rectangular Objects in Google Earth Images,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1068-1081, 2020, doi: 10.1109/JSTARS.2020.2975606.

Xu, Jian, Meng Li, Jiulun Fan, and Wen Xie. ”Discarding Jagged Artefacts in Image Upscaling with Total Variation Regularisation.” IET Image Processing 13.13 (2019): 2495-506.

Romano, Yaniv and Isidoro, John and Milanfar, Peyman. (2016). “RAISR: Rapid and Accurate Image Super Resolution.”, IEEE Transactions on Computational Imaging. PP. 10.1109/TCI.2016.2629284.

Milanfar Peyman, Google AI Blog, Enhance! RAISR Sharp Images with Machine learning, November 2016, Accessed on 1/05/2020, Accessed at https://ai.googleblog.com/2016/11/enhance-raisr-sharp-images-with-machine.html

Xing, Chen, Xi Liang, and Zhiyan Bao. ”A Small Object Detection Solution by Using Super-Resolution Recovery.” Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT) (2019): 313-16.

Jakaria Rabbi, Nilanjan Ray, Matthias Schubert, Subir Chowdhury, and Dennis Chao. ”Small-Object Detection in Remote Sensing Images with Endto-End Edge-Enhanced GAN and Object Detector Network.” Remote Sensing 12.1432 (2020): 1432.

Li, Jianan, Xiaodan Liang, Yunchao Wei, Tingfa Xu, and Jiashi Feng. ”Perceptual Generative Adversarial Networks for Small Object Detection.” In Proceedings of the IEEE Conference on Computer Vision and Pattern recognition 1222-1230, ArXiv.org (2017): ArXiv.org, Jun 20, 2017.

Bai, Yancheng, Yongqiang Zhang, Bernard Ghanem, and Mingli Ding. ”SOD-MTGAN: Small Object Detection via Multi-task Generative Adversarial Network.” Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11217 (2018): 210-26.

Y. Chen, J. Li, Y. Niu and J. He, ”Small Object Detection Networks Based on Classification-Oriented Super-Resolution GAN for UAV Aerial Imagery,” 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 4610-4615, doi: 10.1109/CCDC.2019.8832735.

Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, "Residual Dense Network for Image Super-Resolution," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 2472-2481, doi: 10.1109/CVPR.2018.00262.

Augustsson, Eirikur and Timofte, Radu, “NITRE 2017 Challenge on Single image super-resolution: Dataset and study”, In proceedings of The IEEE Conference on Computer vision and Pattern recognition (CVPR) workshops, July 2017

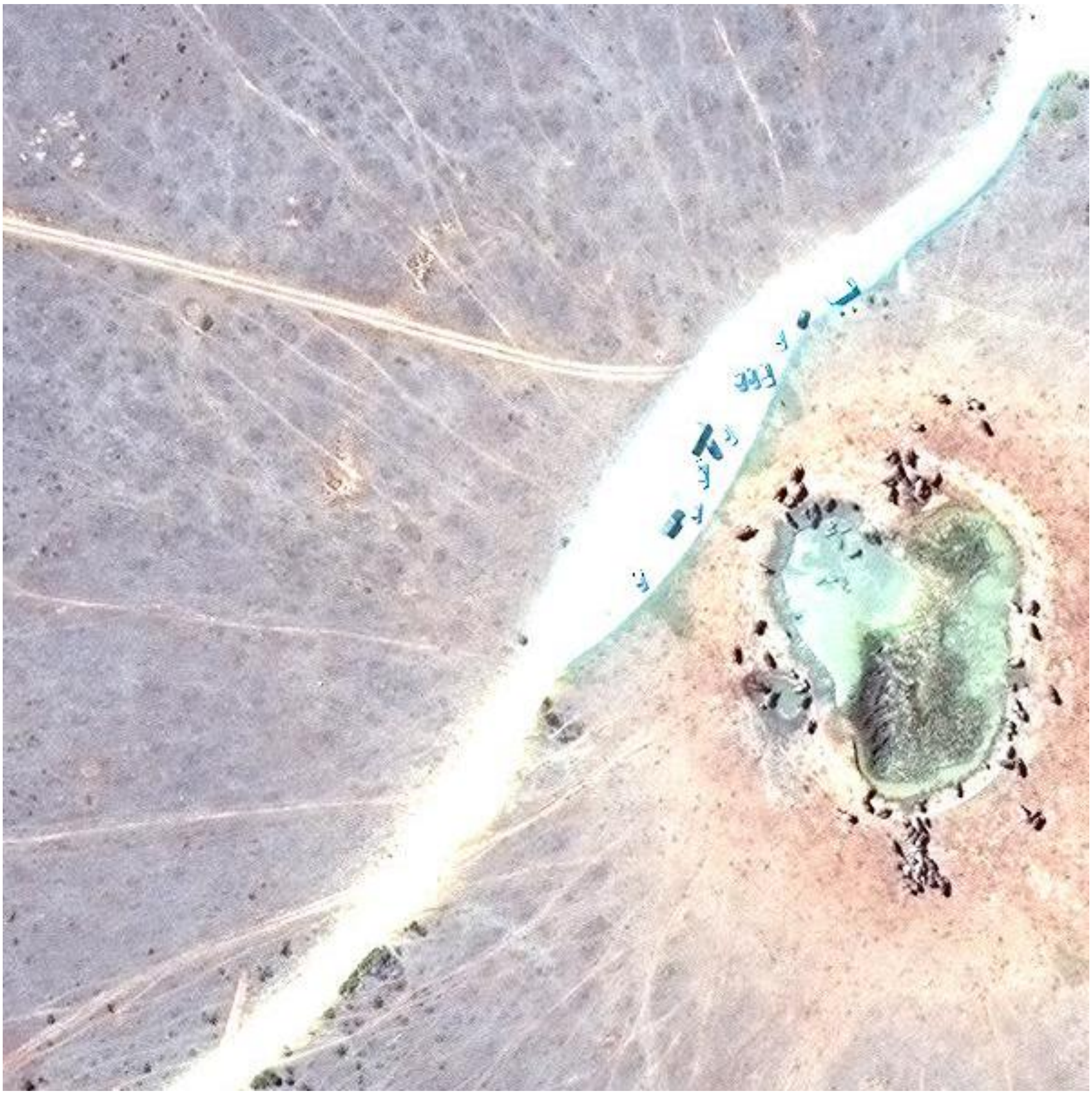
J. Hui, “mAP (mean average precision) for object detection”, medium, available at https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173 accessed on August 2020, published march 7th 2018

Appendix A

Raw results output

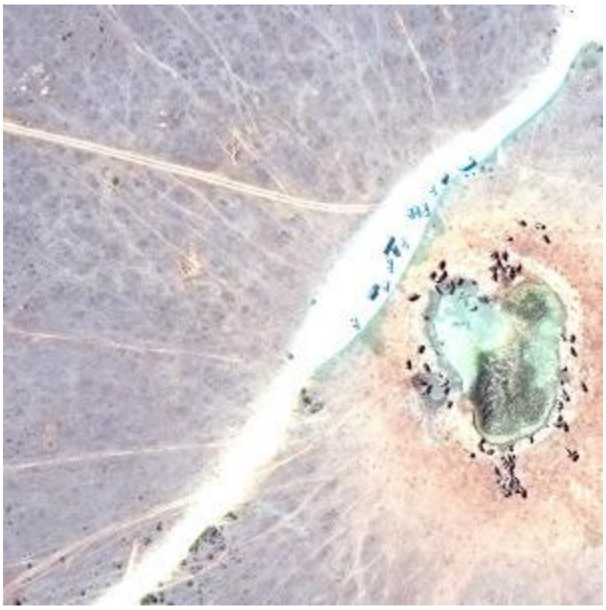
Optimisation of Elephant detection

BASELINE

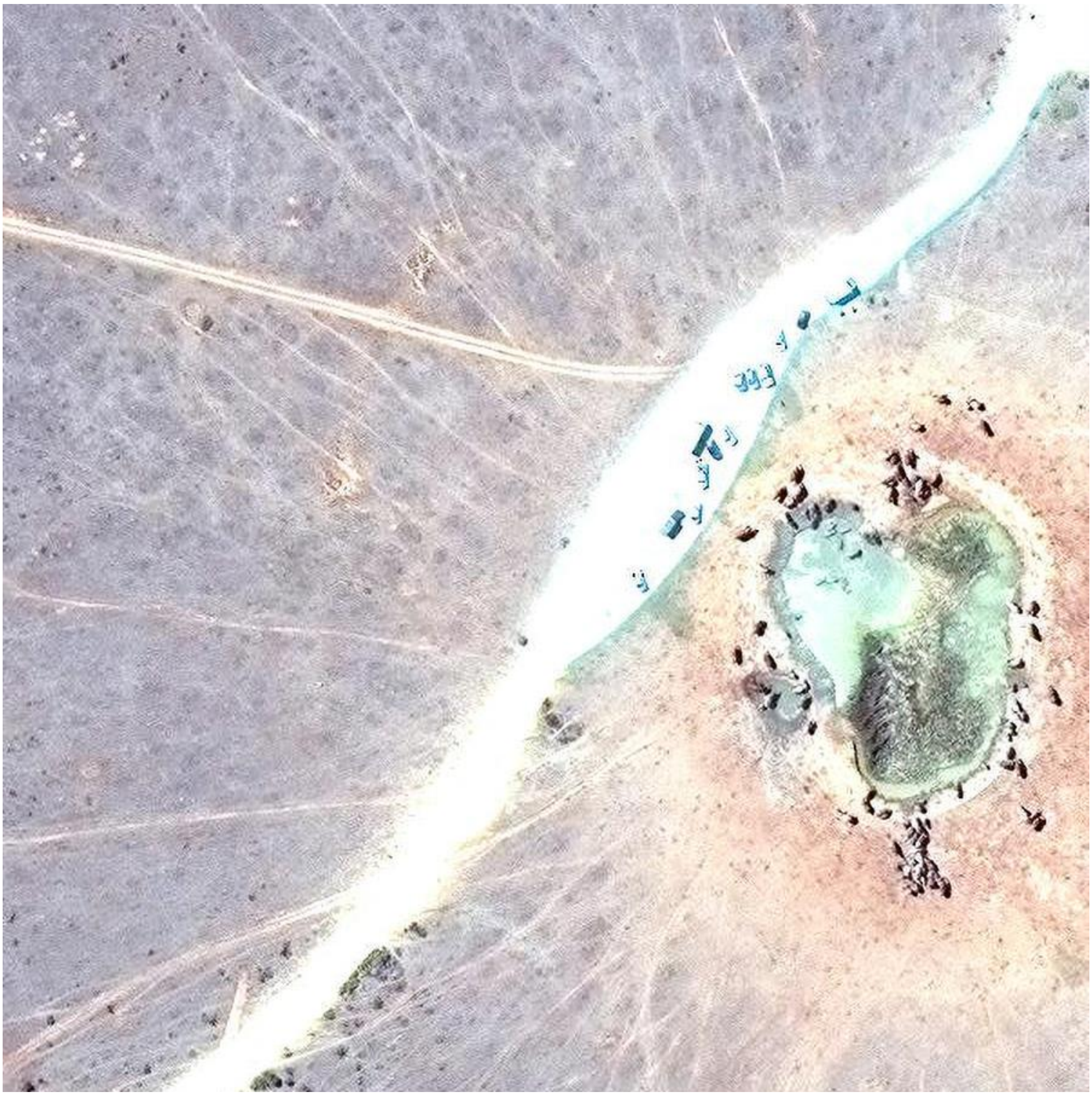


Optimisation of Elephant detection

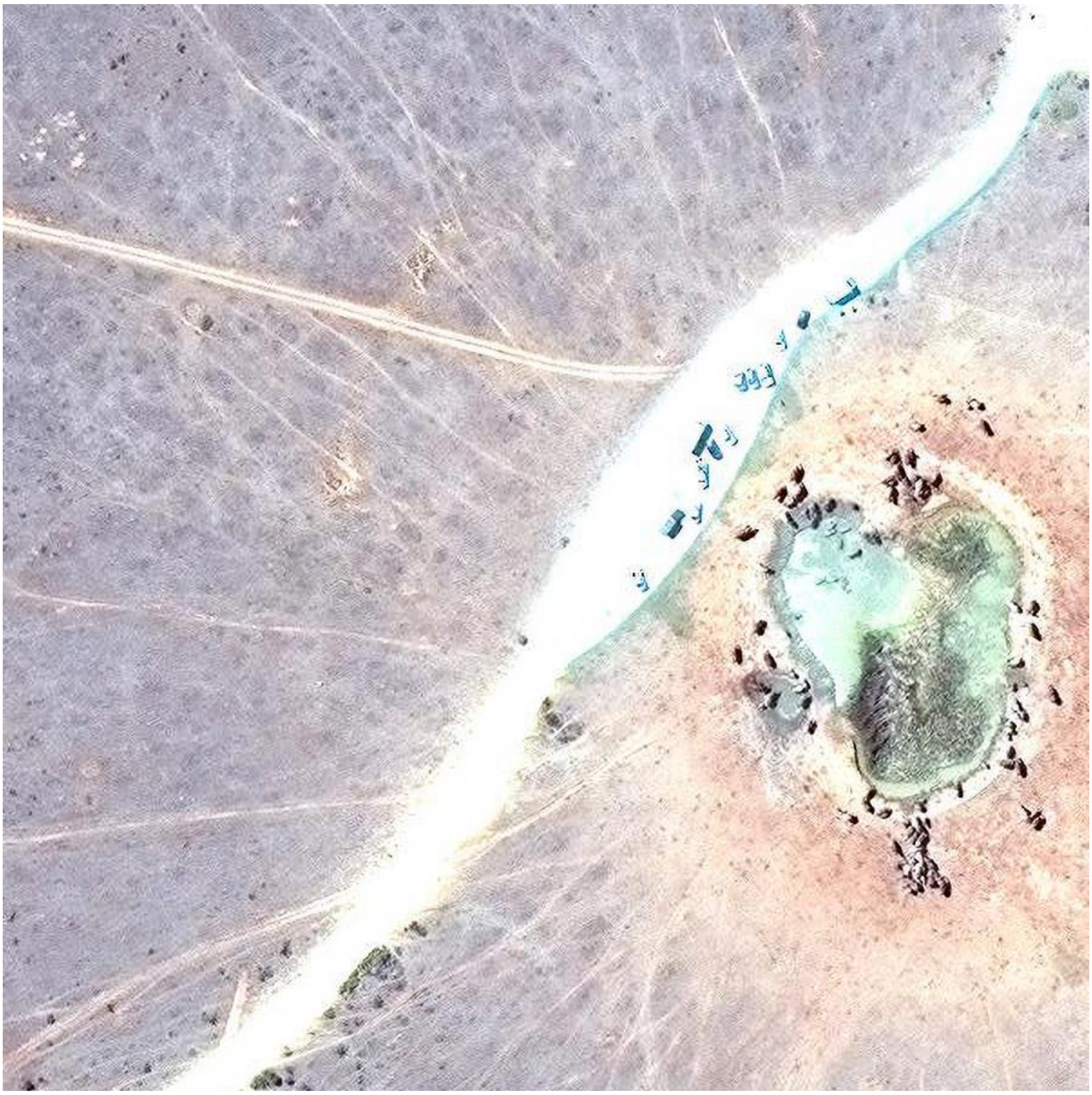
DOWNSCALED



RDN-LARGE-900

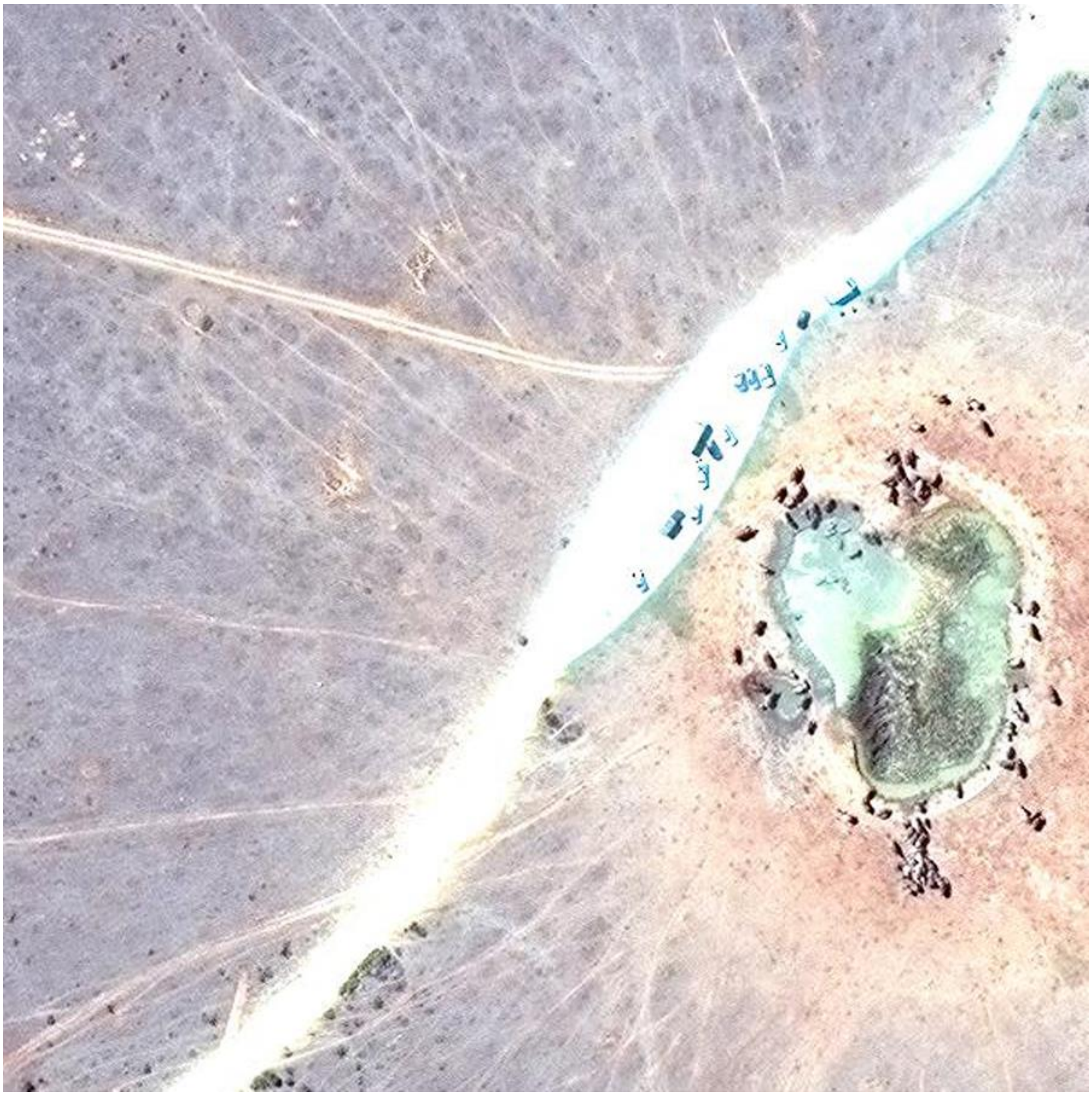


RDN-SMALL-900

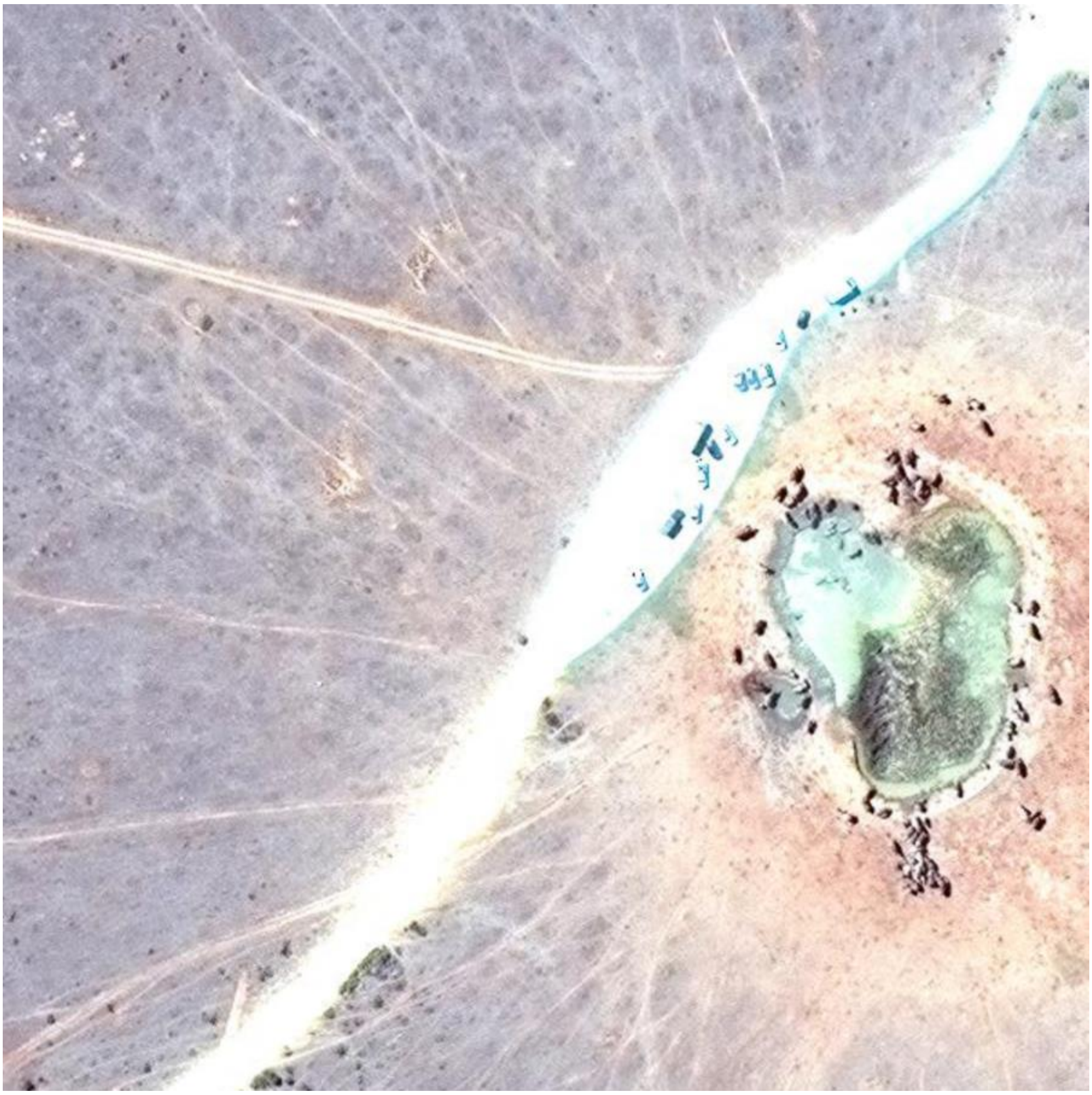


RRDN– 900

BICUBIC – 900



BILINEAR - 900



This form must be attached to the dissertation as an appendix.

Department of Computer Science

12-Point Ethics Checklist for UG and MSc Projects

Student Conor Worthington

Academic Year **1st Year Masters or**

Project Title

Supervisor Olga Isupova

Does your project involve people for the collection of data other than you and your supervisor(s)?
NO

If the answer to the previous question is YES, you need to answer the following questions, otherwise you can ignore them.

This document describes the 12 issues that need to be considered carefully before students or staff involve other people ('participants' or 'volunteers') for the collection of information as part of their project or research. Replace the text beneath each question with a statement of how you address the issue in your project.

1. *Have you prepared a briefing script for volunteers?*
YES / NO

Briefing means telling someone enough in advance so that they can understand what is involved and why – it is what makes informed consent informed.

2. *Will the participants be informed that they could withdraw at any time?* YES / NO

All participants have the right to withdraw at any time during the investigation, and to withdraw their data up to the point at which it is anonymised. They should be told this in the briefing script.

3. *Is there any intentional deception of the participants?* YES / NO

Withholding information or misleading participants is unacceptable if participants are likely to object or show unease when debriefed.

4. *Will participants be de-briefed?* YES / NO

The investigator must provide the participants with sufficient information in the debriefing to enable them to understand the nature of the investigation. This phase might wait until after the study is completed where this is necessary to protect the integrity of the study.

5. *Will participants voluntarily give informed consent?* YES / NO

Participants MUST consent before taking part in the study, informed by the briefing sheet. Participants should give their consent explicitly and in a form that is persistent –e.g. signing a form or sending an email. Signed consent forms should be kept by the supervisor after the study is complete.

6. *Will the participants be exposed to any risks greater than those encountered in their normal work life (e.g., through the use of non-standard equipment)?* YES / NO

Investigators have a responsibility to protect participants from physical and mental harm during the investigation. The risk of harm must be no greater than in ordinary life.

7. *Are you offering any incentive to the participants?*
YES / NO
The payment of participants must not be used to induce them to risk harm beyond that which they risk without payment in their normal lifestyle.

8. *Are you in a position of authority or influence over any of your participants?* YES / NO
A position of authority or influence over any participant must not be allowed to pressurise participants to take part in, or remain in, any experiment.

9. *Are any of your participants under the age of 16?*
YES / NO Parental consent is required for participants under the age of 16.

10. *Do any of your participants have an impairment that will limit Their understanding or communication?* YES / NO
Additional consent is required for participants with impairments.

11. *Will the participants be informed of your contact details?*
YES / NO
All participants must be able to contact the investigator after the investigation. They should be given the details of the Supervisor as part of the debriefing.

12. *Do you have a data management plan for all recorded data?* YES / NO All participant data (hard copy and soft copy) should be stored securely, and in anonymous form, on university servers (not the cloud). If the study is part of a larger study, there should be a data management plan.