

COUNTING WILDEBEEST FROM SPACE USING DEEP LEARNING

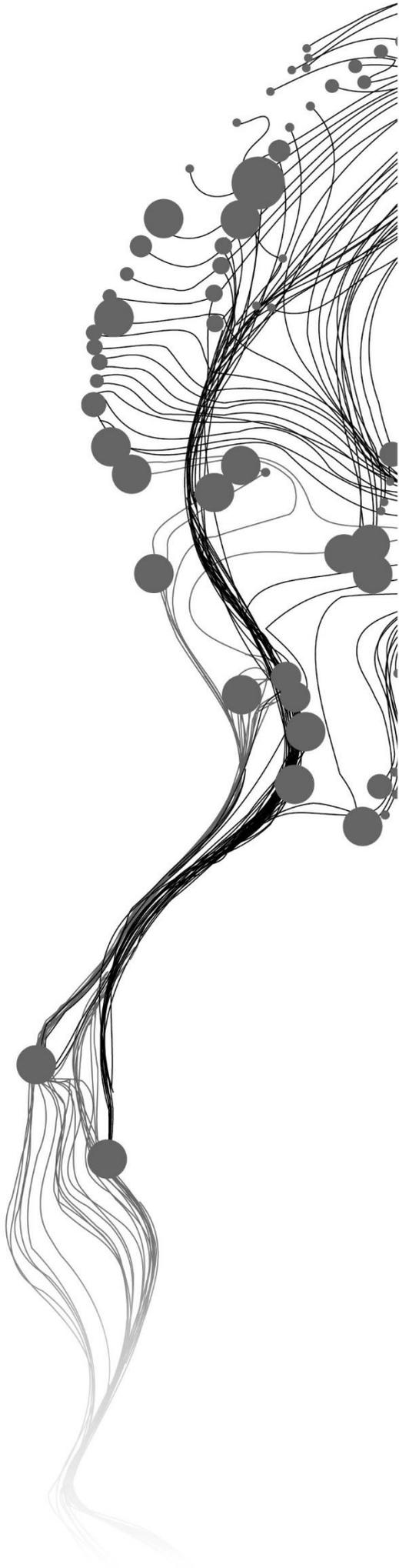
ZIJING WU

June, 2021

SUPERVISORS:

Assoc. Prof. Dr. Tiejun Wang

Prof. Dr. Andrew K. Skidmore



COUNTING WILDEBEEST FROM SPACE USING DEEP LEARNING

ZIJING WU

Enschede, The Netherlands, June, 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Resources Management

SUPERVISORS:

Assoc. Prof. Dr. Tiejun Wang (University of Twente)

Prof. Dr. Andrew K. Skidmore (University of Twente)

ADVISOR:

Dr. Ce Zhang (Lancaster University, UK)

THESIS ASSESSMENT BOARD:

Assoc. Prof. Dr. ir. Thomas A. Groen (Chair, University of Twente)

Dr. Xiaowei Gu (External examiner, Aberystwyth University, UK)

Assoc. Prof. Dr. Tiejun Wang (First supervisor, University of Twente)

Prof. Dr. Andrew K. Skidmore (Second supervisor, University of Twente)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Accurate, reliable, and up-to-date information on wildlife populations is crucial for species conservation planning in the face of unprecedented biodiversity loss worldwide. Deep learning techniques combined with high-resolution satellite images have proven successful in detecting medium- and large-sized animals. However, to date, no study has shown that this method can be used to detect and count animals with indistinct features (2-4 pixels in length) and low contrast to the background in satellite images across landscapes. In this study, I tested the ability of the U-Net deep convolutional neural network for detecting and counting the migrating wildebeest in the Mara Triangle in East Africa from GeoEye-1 satellite images. I also assessed the role of the near-infrared band in the accuracy of wildebeest detection. Moreover, I tested the model on a different area with varied landscapes and a temporally different satellite image to evaluate its transferability over space and time. The results showed that the U-Net model can be used to accurately detect and count a large number of wildebeest (more than 100,000 individuals) from the GeoEye-1 satellite image, with a high generalization accuracy (F1-score) of 0.87. Adding the near-infrared band to the RGB band combinations in the satellite image did not significantly improve the accuracy of wildebeest detection. In addition, the model was able to rapidly detect the animal clusters on the spatially and temporally different satellite images, suggesting that the U-Net wildebeest detection model has the potential to be applied to the entire Serengeti-Mara ecosystem. In conclusion, this study demonstrates an effective and efficient U-Net deep learning model for accurate and rapid wildebeest detection and counting from GeoEye-1 satellite imagery.

ACKNOWLEDGEMENTS

It was in June of 2021, not too hot a summer in China. The humming of the old fan mingled with the constant quietness of the afternoon under the bright sunlight with dry air, bringing me the cool breeze, and melting me into the lazy summer. But I still had to work on revising my thesis document. Wake up! A message from my old friend reminded me. Out of nowhere, I started to check our chat history. And then a coincidence just happened: I found that I was helping her with the English exam paper, and I answered this to one of the questions: “Detecting populations of animals and species is a challenge, where ...” This message was sent in May 2019, two years ago even before I decided to come to study at ITC, and I certainly had no idea of what research I am going to do. And now I’ve been working on animal detection and population counting for the past ten months! I must say, it is fate that brings me to this topic, and I was destined for it two years ago!

It was a magical start, and what happened later on is even more dreamlike. The mentor meeting during covid times brought my supervisor, Dr. Tiejun Wang, to me (and I am still so grateful for this opportunity). He is an acute observer: Tiejun found me with strong research interest and potential, but no clear direction in mind. He reached me with this extremely interesting topic that fits me in both the ecological side and technical side at the right time and guided me to the right road. During the past year, Tiejun offered great help and support in not only my academic research but also life planning. His supervision is clear, right-to-the-point, timely, and careful, which makes my MSc thesis research much much easier. His expertise is invaluable to me in solving the essential issues in the research and formulating the document as well as making presentations. In particular, with Tiejun’s help, I was able to identify and focus on the principle problems all along. This made my path direct and clear and saved me tons of time. His insightful feedback in our weekly meetings pushed me to think more and brought our work to the right track on a higher level. I would also like to thank his careful and detailed advice in my academic writing: this thesis document involved a great deal of his hard work. The experience I learnt from him will always shed the light upon me in my future career and benefit my life decisions.

I would also like to thank my second supervisor, Prof. Andrew Skidmore. I’ve been admiring him as a well-respected scientist and ecologist since I started my application to ITC. Again, fate brought him to me as my second supervisor! Thanks for his helpful advice in the whole process of my MSc thesis study. And I’m really grateful for his help in the proposal and final thesis writing. I really appreciate it that he worked up late till midnight to revise my document, especially with the time difference between the Netherlands and Australia. His professional dedication to work will always encourage me to overcome the obstacles in my life.

I would also like to express my gratitude to Dr. Ce Zhang and Dr. Xiaowei Gu. They provided me with invaluable technical support and research suggestions in our weekly meetings during my entire MSc study. They are very fast responsive 24/7 and always have a solution to my doubts. I will not forget our late meeting at 11:00 pm! The thesis research won’t be this smooth without their help. I am so lucky to have them as my advisors.

Moreover, I would like to thank my parents, who always support my decisions and give all they have to me to achieve what I want. Thanks to my mom. I never feel unfortunate or lost, and I know I am always protected and safe, and all of this is attributed to your selfless effort to support the whole family. I am sorry that I still haven’t done enough to care more about your feelings. Thanks to my friends and all the colleagues at ITC. I will always remember the night before my departure when we gathered to have the hot pot and hung out on the street in Enschede. We will all achieve what we hope and deserve in the future!

And of course, I would like to thank Mr. Kushanav Bhuyan, my study partner who studied deep learning (the total new field for both of us) together with me, my friend who is always there to listen to me and cheer me up when I feel frustrated and sad, my buddy who overcame all the obstacles in life with me and co-experienced all the good and fun moments, my boyfriend who gives to me his wholehearted love and support. My life in the past two years only became colorful because of your existence. I can't thank you more for being with me, growing up and making progress with me together.

Last but not least, I would like to thank myself. Thank you for being intelligent (:-p), strong, and mentally independent. I know you still have lots of weaknesses, but you're trying to live at the moment, cheer up the people around you, and pursue what you want (although you haven't figured out what exactly you want). It is difficult to live a life just by following your heart. You will get support, help, but also doubts and objections. But don't worry. Just be yourself. Everything is connected: the gap year that they thought to be "a waste of time", led you to this MSc research! Isn't it awesome? I believe you will continue trying to "suck out all the marrow of life", and live every moment deliberately. Carpe diem!

TABLE OF CONTENTS

1.	Introduction	1
1.1.	Background.....	1
1.2.	Research problem.....	3
1.3.	Research objectives	4
1.4.	Research questions	4
1.5.	Research hypotheses	4
1.6.	Research workflow	4
2.	Materials and methods.....	7
2.1.	Species and study area	7
2.2.	Satellite images	8
2.3.	Data preparation.....	9
2.4.	Deep learning model.....	13
2.5.	Locating and counting.....	17
2.6.	Transfer learning.....	18
2.7.	Accuracy assessment.....	20
2.8.	Statistical analysis.....	21
3.	Results	23
3.1.	Detecting the wildebeest	23
3.2.	Counting the wildebeest.....	36
3.3.	Contribution of the near-infrared band.....	37
3.4.	Spatial transferability.....	38
3.5.	Temporal transferability.....	41
4.	Discussion.....	43
4.1.	Wildebeest detection using U-Net deep learning model.....	43
4.2.	Spatial and temporal transferability of the wildebeest detection model	44
4.3.	The role of the near-infrared band in the detection of wildebeest.....	45
4.4.	Applicability of the wildebeest detection model to other animal species.....	45
5.	Conclusions	47

LIST OF FIGURES

Figure 1. Overall flowchart of the research.....	5
Figure 2. Location of the Mara Triangle in the Maasai Mara National Reserve, Kenya.....	8
Figure 3. Overview of the training and testing dataset in the study area.....	9
Figure 4. Wildebeest labelling on GeoEye-1 satellite image acquired in 2009.....	11
Figure 5. Satellite image with wildebeest and the rasterized labelled mask image.....	12
Figure 6. Flowchart of data processing.....	13
Figure 7. U-Net model architecture.....	14
Figure 8. The architecture of the U-Net-based ensemble model.....	17
Figure 9. The spectral characteristics of GeoEye-1 satellite image in 2009 and 2013.....	20
Figure 10. Learning curve during model training.....	24
Figure 11. The wildebeest detection output of the U-Net model and wildebeest locating product.....	24
Figure 12. Wildebeest density map generated by the preliminary model.....	26
Figure 13. Sensitivity analysis of searching distance in accuracy assessment.....	27
Figure 14. Precision-Recall curve of the U-Net model trained with different weights in Tversky loss.....	28
Figure 15. Predicting wildebeest on the GeoEye-1 satellite image.....	30
Figure 16. Predicting wildebeest on the GeoEye-1 satellite image.....	31
Figure 17. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on bare land).....	32
Figure 18. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on burnt area).....	33
Figure 19. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on grassland).....	34
Figure 20. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest crossing the Mara River).....	35
Figure 21. F1-score of the U-Net model under different thresholds.....	36
Figure 22. Precision-Recall curve of the U-Net model trained with different weights in Tversky loss using RGB+NIR bands.....	37
Figure 23. Transferring the wildebeest detection model to the highland area on GeoEye-1 satellite image in 2009.....	40
Figure 24. Detecting wildebeest on GeoEye-1 satellite image in 2013 (the plain area) using U-Net model trained with data in 2009.....	41
Figure 25. Detecting animals on GeoEye-1 satellite image in 2013 (the highland area) using U-Net model trained with data in 2009.....	42

LIST OF TABLES

Table 1. GeoEye-1 satellite images.....	8
Table 2. The size of the training and testing dataset	10
Table 3. Percentage of wildebeest and non-wildebeest pixels in the training dataset.....	12
Table 4. Pixel-level and object-level accuracy of wildebeest detection model trained using the full dataset and tested using the first testing dataset with 30 randomly selected patches	25
Table 5. Object-level accuracy of U-Net-based ensemble model on the first testing dataset with 30 randomly selected patches	25
Table 6. Updated testing data sampling strata	26
Table 7. Accuracy of wildebeest detection trained with different weights in Tversky loss.....	28
Table 8. The accuracy of the U-Net-based ensemble model on the updated testing dataset with the Tversky loss parameter β of 0.9	29
Table 9. Statistics of the wildebeest population count using the U-Net model from the GeoEye-1 satellite image.....	36
Table 10. The accuracy of wildebeest detection U-Net model trained with RGB+NIR GeoEye-1 satellite image.....	37
Table 11. The accuracy differences between U-Net models trained with RGB+NIR and RGB band combinations of GeoEye-1 satellite image using 5×2 cross-validation.....	37

1. INTRODUCTION

1.1. Background

Global biodiversity is declining at an unprecedented rate in human history (IPBES, 2019). According to the latest World Wildlife Fund (WWF) Living Planet Report, the global populations of vertebrates (i.e., birds, fish, mammals, amphibians, and reptiles) have declined by an average of 68% between 1970 and 2016 (WWF, 2020). To address the biodiversity crisis, the Aichi Biodiversity Targets were adopted by world leaders in 2010, however, none of the 20 Aichi Targets have been fully achieved by 2020 (Secretariat of the Convention on Biological Diversity, 2020). Although there could have been many reasons for this failure, the Aichi Biodiversity Targets are found to be comprehensive but not measurable, which jeopardizes the implementation and evaluation of these targets (“New biodiversity targets cannot afford to fail,” 2020). To quantify biodiversity consistently worldwide, scientists proposed a set of Essential Biodiversity Variables (EBVs) to identify what to monitor, including species population, species traits, and ecosystem structure and function (Pereira et al., 2013). Regarding how to monitor EBVs, Skidmore et al. (2015) identified ten variables that can be observed through satellite sensors to track biodiversity from space. Species population is one of the most fundamental EBVs (Jetz et al., 2019).

Africa’s biodiversity is of global importance. Yet, Africa is not immune to the rapid global biodiversity loss. The well-known annual migration of over 1 million wildebeest in the 40000-km² Serengeti-Mara ecosystem is one of the largest mammal migrations in the world (Serneels & Lambin, 2002). The great wildebeest migration is an important ecological phenomenon that shapes the terrestrial and aquatic ecosystems and also benefits the local communities (Holdo et al., 2013; Onyeanusi, 1989; Subalusky et al., 2017; Wishitemi et al., 2015). However, many wildebeest populations in East Africa are declining dramatically over the last several decades due to high human population densities, increasing urbanizations, and expanding agriculture and fences (Ogutu et al., 2011). This human-induced decline in the wildebeest population can alter the ecosystems and, in turn, exacerbate biodiversity loss and harm local tourism. Accurate, reliable, and up-to-date wildebeest population data are crucial to monitor the fluctuations of the number of wildebeest and determine the cause of species abundance change as well as ecologically functional changes (Ogutu et al., 2009; Subalusky et al., 2018).

Counting wildlife can be very challenging because animals are mobile and some species migrate over large areas. The most traditional way of conducting wildlife population census is through the ground survey with vehicle teams and a large number of people involved, however, this can be very expensive, labour-intensive, and time-consuming (Reid et al., 2003). More importantly, it is hardly possible to produce spatially continuous data in a ground survey, leading to inevitable sampling error (Betts et al., 2007). The ground surveys also cause disturbance to the animals. These crucial disadvantages hinder convenient and accurate animal abundance counting.

The rapid development of remote sensing technology enables us to reduce our dependence on traditional ground survey methods to a certain extent. Aerial surveys including manned aircraft surveys and Unmanned Aerial Vehicle (UAV) surveys are commonly used for wildlife population surveys. Benefiting from the high-resolution imagery (up to 2 to 5 centimetres) collected by the aircraft, the features of animals can be clearly recognized, and thus automatic species discrimination and population counts become feasible. There is a

wide range of applications, such as elephant, giraffe, and zebra counting using aerial photographs (Eikelboom et al., 2019), cow detection on a farm using UAV thermal images (Longmore et al., 2017), blacktip reef sharks and pink whiprays counting in a coral lagoon with UAV video (Kiszka et al., 2016), leopard seals counting and frigatebirds detection using UAV images (Goebel et al., 2015; Hodgson et al., 2016), and so on. However, aerial surveys still cannot avoid the sampling error caused by the movement of animals while taking aerial photographs, and the noise disturbance problem remains (H. Mduma et al., 2014). The coverage of aerial survey area is also limited as the flights can be expensive, laborious, and time-consuming.

Recent advances in satellite sensors offer great potential for detecting animals over large areas with lower labour costs. The availability of very high resolution (VHR) satellite imagery with submeter pixel size makes it possible to identify large animals from space (Wang, Shao, & Yue, 2019). Using WorldView, GeoEye-1, or Quickbird satellite imagery (with 31 to 65-centimetre resolution), researchers have detected polar bears in the Arctic, whales in the sea, wildebeest and elephants in the African savannahs, and so on (Duporge et al., 2020; Guirado et al., 2019; Larue et al., 2015; Stapleton et al., 2014; Yang et al., 2014). The acquisition of satellite remote sensing data costs much less than traditional aerial surveys in terms of labour and allows for large-scale wildlife censuses. Nevertheless, the limitation in spatial resolution compared with aerial photographs poses challenges for accurate and fast detection of animals from satellite imagery.

Many approaches for detecting animals have been explored in previous studies to count populations from satellite imagery. Manual counting by an expert is reliable, but very time-consuming, particularly in large regions (LaRue et al., 2011). Traditional supervised classification methods, such as the maximum likelihood classifier, require highly distinct spectral signatures of animals, which leads to poor accuracy when spectral noises are present in a heterogeneous environment (Barber-Meyer et al., 2007; Larue et al., 2015). The same problem occurs with the threshold approach when image noise causes a high misclassification and the decision of threshold is subjective and situation-dependent (Hollings et al., 2018). The use of a reference image provides prior knowledge that can facilitate correct animal detection on the target image through the image differencing technique (Larue et al., 2015). However, it is difficult to obtain multi-temporal satellite images with similar spectral characteristics under changing data acquisition conditions.

With the rapid development of computer vision technology, many machine learning algorithms have been developed that have proven to surpass traditional classification methods in remote sensing image processing, such as Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Networks (ANNs). Some algorithms have also been applied to animal detection tasks on satellite imagery. To classify large mammals from the GeoEye-1 satellite image, Yang et al. (2014) conducted a pixel-based ANNs approach on wildebeest detection and obtained more than 80% accuracy; Xue et al. (2017) utilized an Adaptive-Network-based Fuzzy Inference System (ANFIS) and achieved 79% accuracy. However, these studies are confined to small pilot areas. When it comes to a larger spatial scale, spectral noises from surrounding objects can cause challenges and affect the performance. The accuracy of large-scale animal detection using traditional machine learning methods remains a bottleneck.

Deep learning techniques have shown a fast-growing trend in remote sensing image analysis because of their remarkable performance over traditional machine learning approaches. Convolutional Neural Networks (CNNs) are a group of deep learning architectures that are designed for supervised image recognition tasks. Instead of classifying the pixels simply by spectral information, CNNs take the context information into account and detect the spatial features. CNNs consist of three basic types of layers, namely convolution layer, pooling layer, and fully connected layer. The convolution layer detects low-level features, such as edges at different directions across the image, and the following pooling layer extracts more complex local features.

The fully connected layer then connects the features to output neurons, which are eventually classified into the pre-defined classes. The combinations and concatenation of these three fundamental layers form different architectures, such as image segmentation classifier U-Net (Ronneberger et al., 2015a), SegNet (Badrinarayanan et al., 2017), and object detector Fast R-CNN (Girshick, 2015), You Only Look Once (YOLO) (Redmon et al., 2016). The pixel-based image segmentation architectures predict the class probability for every pixel in the whole image, while the object-based convolutional neural networks detect and classify individual objects in the image and generate the bounding box with a class label for each object.

Previous research presented many successful applications of deep learning on animal detection from remote sensing images, but mostly from aerial imagery (Han et al., 2019; Kellenberger et al., 2018a; Sahu, 2019; Torney et al., 2019). There are only four studies that detect animals completely from satellite imagery (both training and testing) using deep learning: whale detection with 82% F1-score and elephant detection with 78% F2-score, both using Faster R-CNN; seal detection with 30% accuracy using modified U-Net, and albatrosses counting with an F1-score of about 70% using U-Net (Bowler et al., 2020; Duporge et al., 2020; Gonçalves, Spitzbart, & Lynch, 2020; Guirado et al., 2019). These papers showed a promising direction towards automated animal detection using both pixel-based image segmentation and object detection deep learning methods from satellite imagery.

However, the feasibility of different types of deep learning methods highly depends on the body size of the animal. In these two studies using object-based Faster-RCNN above, the mature whales have a body length of more than 20 meters, and the African elephants are about 3 to 4 meters long (Christiansen, 2004; Leslie et al., 2020). Both have more than 10 pixels along the body length axis on the satellite imagery, and the features of the animal on an individual basis are distinct on the satellite image. It has been pointed out that current object detectors have limited performance when detecting very small objects with less than 10 pixels in one dimension, and the size limit of the most recent small object detector is more than 6 pixels (Pham et al., 2020). Therefore, animal detection using object-based deep learning methods will suffer from the small body size of animals relative to the pixel size of the satellite imagery (less than 6 pixels in one dimension).

In contrast, the albatross counting study showed that small objects, such as the 1.2m-long albatrosses can be successfully detected on WorldView-3 imagery by U-Net, the pixel-based image segmentation deep learning approach (Bowler et al., 2020). The albatross is around 4 to 5 pixels in length. Besides, the albatross detection was conducted in four islands with different types of environments, and the heterogeneity of the environment did not significantly affect the accuracy. However, the colour of the albatross is white, which can be easily distinguished from the greenish and black surroundings, thus the complexity of the background is not very problematic here. But this is not the case for animals with low contrast to the environment, such as brown, blackish wildebeests on open savannahs. Hence, U-Net has the potential to detect and count animals with a small size on satellite imagery, but no previous research has investigated its capability of detecting animals with indistinct features and low contrast to the heterogeneous background from satellite imagery.

1.2. Research problem

Deep learning methods have been successfully used to detect and count large and feature distinct animals at relatively small spatial scales from aerial photographs, UAV and satellite images. For example, deep learning object detectors are used to spot and count whales, elephants from satellite images (Duporge et al., 2020; Guirado et al., 2019), cattle, kiang, from UAV images (Peng et al., 2020; Xu et al., 2020), as well as large mammals in African open savannahs from aerial images (Eikelboom et al., 2019; Kellenberger et al.,

2018b; Torney et al., 2019). But most of the research only focuses on the large animal, typically more than 10 pixels in length on the images. The only study that detects albatrosses with smaller sizes using pixel-based U-Net deep learning is constrained by the colour contrast of the animals in the background (Bowler et al., 2020). Automated wildebeest detection from satellite imagery is crucial for rapid and cost-effective monitoring of wildlife populations and biodiversity conservation in the Serengeti-Mara ecosystem. However, the wildebeest only consists of 2 to 4 pixels in length on the very high-resolution satellite images such as GeoEye-1 and thus does not have distinct features. The similarity of the spectral characteristics between wildebeest and background objects results in low colour contrast, which may easily confuse the animal detection model. It is not yet clear whether U-Net deep learning can be used to detect and count wildebeest with high accuracy over a large area from very high-resolution satellite images.

1.3. Research objectives

The overall objective of this research is to test the ability of the U-Net deep convolutional neural networks for detecting and counting the number of migrating wildebeest in the Mara Triangle from GeoEye-1 satellite imagery. The specific objectives of this study are as follows:

- 1) To build and test a U-Net deep convolutional neural network for detecting wildebeest from GeoEye-1 satellite imagery.
- 2) To determine the role of the near-infrared band on the detection accuracy of wildebeest.
- 3) To assess the spatial and temporal transferability of the U-Net model for detecting wildebeest across landscapes.

1.4. Research questions

- 1) What is the accuracy in detecting and counting wildebeest from GeoEye-1 imagery using the U-Net model?
- 2) Can adding a near-infrared band in the RGB combination significantly improve the detection accuracy of wildebeest?
- 3) Does the U-Net model we are building in this study have the capacity for successful transfer in time and space?

1.5. Research hypotheses

Hypothesis 1:

H₀: There is no statistically significant difference in the wildebeest detection accuracy between the use and without the use of the near-infrared band of GeoEye-1 image.

H₁: Adding the near-infrared band to the RGB combination can significantly improve the wildebeest detection accuracy.

1.6. Research workflow

To answer these three research questions, this study is implemented in 5 steps. Firstly, the GeoEye-1 satellite images are processed to prepare the training and testing datasets required by the U-Net deep learning model. Then the datasets of the plain area are fed into U-Net deep learning model, and parameter tuning is conducted to achieve the optimal model with the highest accuracy, which is then used to classify the whole satellite image to count the wildebeest population. The first two steps are aiming to answer research question

1. To assess the influence of adding near-infrared band in the RGB combination on wildebeest detection accuracy (research question 2), the model training results of RGB data and RGB+NIR data are compared using a statistical test. Next, the model is transferred to the highland area with different landscapes to explore its performance in terms of spatial transferability. Lastly, the trained U-Net model is applied to the GeoEye-1 satellite image acquired in a different year and the temporal transferability is evaluated. Figure 1 presents the general workflow of this research.

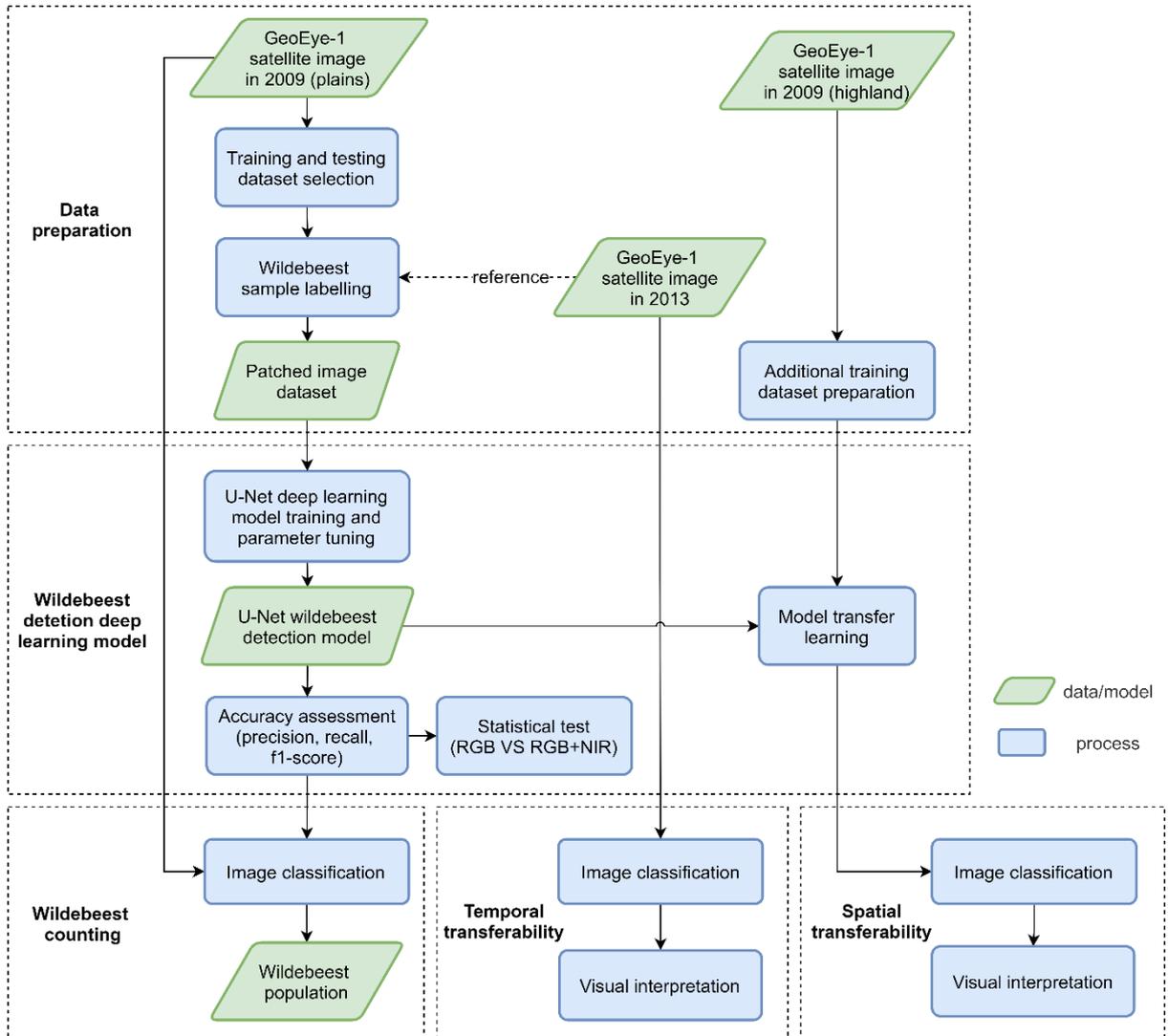


Figure 1. Overall flowchart of the research

2. MATERIALS AND METHODS

2.1. Species and study area

2.1.1. Species

Wildebeest (*Connochaetes taurinus*) are the dominant grazers across grassland savannahs in East Africa. There are over one million wildebeest in the Serengeti-Mara ecosystem, and they migrate over 1500 kilometres on average every year, tracking the nutrient-rich food resources (Ben-Shahar & Coe, 1992; Hopcraft et al., 2015). During June and August, the wildebeest migrate from the Serengeti plains in Tanzania into Maasai Mara National Reserve, and spread to the east and go across the Mara River in September, and then during November, they move south to Northern Serengeti (Thirgood et al., 2004) (Figure 2).

The adult wildebeest has a body mass of around 220 kg, and it is approximately 1.5 m to 2.5 m in length from head to tail, which is 3 to 4 pixels long and 1 to 2 pixels wide on the GeoEye-1 satellite image (Owen-Smith & Mills, 2008; Talbot & Talbot, 1963; Yang et al., 2014). The wildebeest has a black mane from the neck to the spin, a black beard, and a long black tail, and its pelage is mainly dark bluish or silver grey.

2.1.2. Study area

The study area locates in the Mara Triangle to the southwestern of Maasai Mara National Reserve, Kenya. The Mara Triangle is adjacent to the Serengeti national park of Tanzania to the southwest and Siria Escarpment to the northwest (Figure 2). The Mara River flows through Maasai Mara National Reserve and intersects with the Tanzania border and Escarpment, and forms the third border of the so-called Triangle, which covers an area of 520 km² (Allen et al., 2019). There are mainly two rainy seasons across the year, including April to May and November to December. The main land cover type in Mara Triangle is open grassland, with some plains covered by dwarf shrubs and bushes. Along the Mara River, the riverine forest appears to be the main landscape. Moreover, a regulated fire management program is conducted to maintain the grassland, thus a burnt area is witnessed in the reserve periodically (Eva & Lambin, 2000).

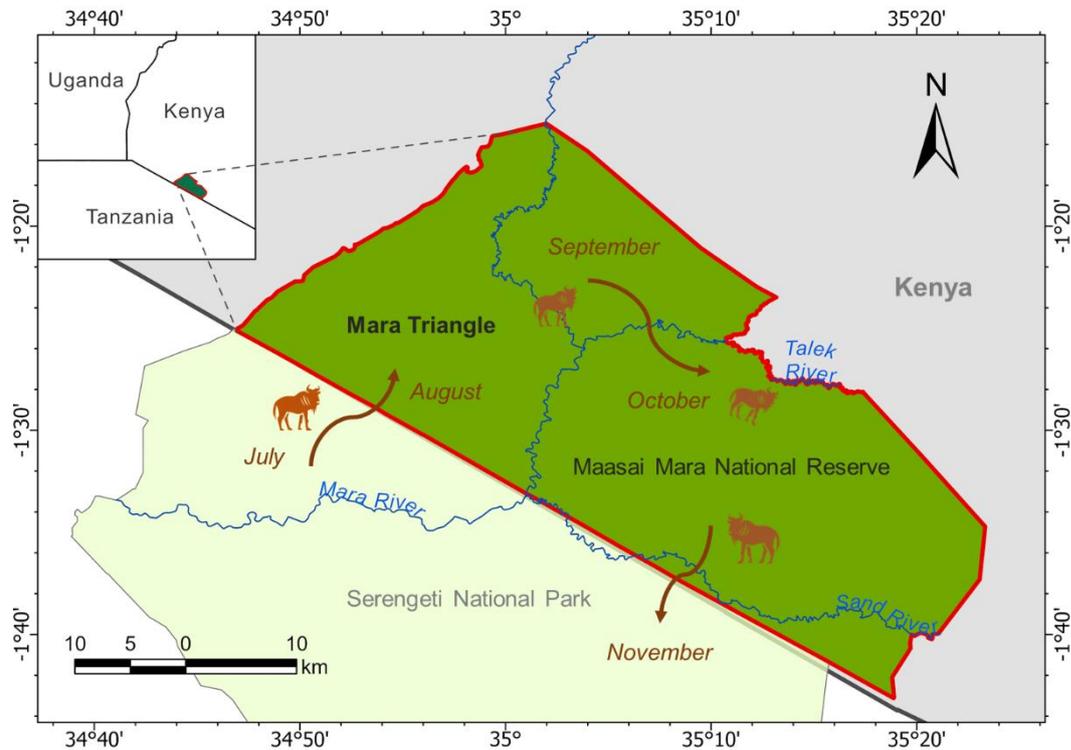


Figure 2. Location of the Mara Triangle in the Maasai Mara National Reserve, Kenya. The wildebeest migration route shown in this figure is based on the study by Thirgood et al. (2004).

2.2. Satellite images

Two GeoEye-1 satellite images supplied by Maxar Technologies company (<https://www.maxar.com/>) cover the same area that locates in the Mara Triangle in the Maasai Mara National Reserve, but they were captured on different dates. The first satellite image was captured on August 11th in 2009, and the second satellite image was captured on August 10th in 2013. Both images are cloud-free and the multispectral bands have been pan-sharpened according to the associated panchromatic band.

The total range of the satellite image is approximately 22 km × 13 km in the latitude and longitude directions. It covers around 280 km² of the study area, where there are around 140 km² of the lower plain land and 140 km² of the highland. More details about the satellite images are shown in Table 1.

Table 1. GeoEye-1 satellite images

Satellite Sensor	Acquisition Date	Spectral range	Spatial resolution
GeoEye-1	8/11/2009	Panchromatic: 450-800 nm Blue: 450-510 nm Green: 510-580 nm Red: 655-690 nm Near Infra-Red: 780-920 nm	Pansharpened 43 cm
GeoEye-1	8/10/2013	Panchromatic: 450-800 nm Blue: 450-510 nm; Green: 510-580 nm Red: 655-690 nm Near Infra-Red: 780-920 nm	Pansharpened 43 cm

2.3. Data preparation

2.3.1. Training and testing dataset

The wildebeest detection model uses the GeoEye-1 satellite image of 2009 where there are large numbers of wildebeest. On the satellite image, the lower plain land is where the wildebeest migrate across during the migration season. Thus, the focus of the wildebeest detection model is on the plain area. A 150 m by 150 m grid system was built for systematic sampling. Each grid represents one patch that consists of about 350 by 350 pixels.

For the training dataset, 191 image patches covering different types of landscapes and different wildebeest abundance were selected. The training dataset occupies nearly 3% of the plain area, and there are approximately 20,000 wildebeest individuals labelled in this dataset.

To explore the spatial transferability of the wildebeest detection model on the highland area, additional 20 patches on the highland were also selected as training data. The additional training dataset only covers the agricultural field, human settlements, and forest, and there are no wildebeests in it. The additional dataset was not used when training the model on the low-land plain area.

For the testing dataset, 30 patches were first randomly selected in the plain area, where there are 3,171 wildebeest individuals. To ensure its representativeness, the testing dataset was then updated using stratified proportionate sampling method according to the wildebeest density map produced based on the preliminary results. The details of the new testing data sampling method will be explained in Section 3.1.2. In the updated testing dataset, 100 patches that have different densities of wildebeest were selected (2,249 wildebeest individuals in total), occupying about 1.5% of the plain area.

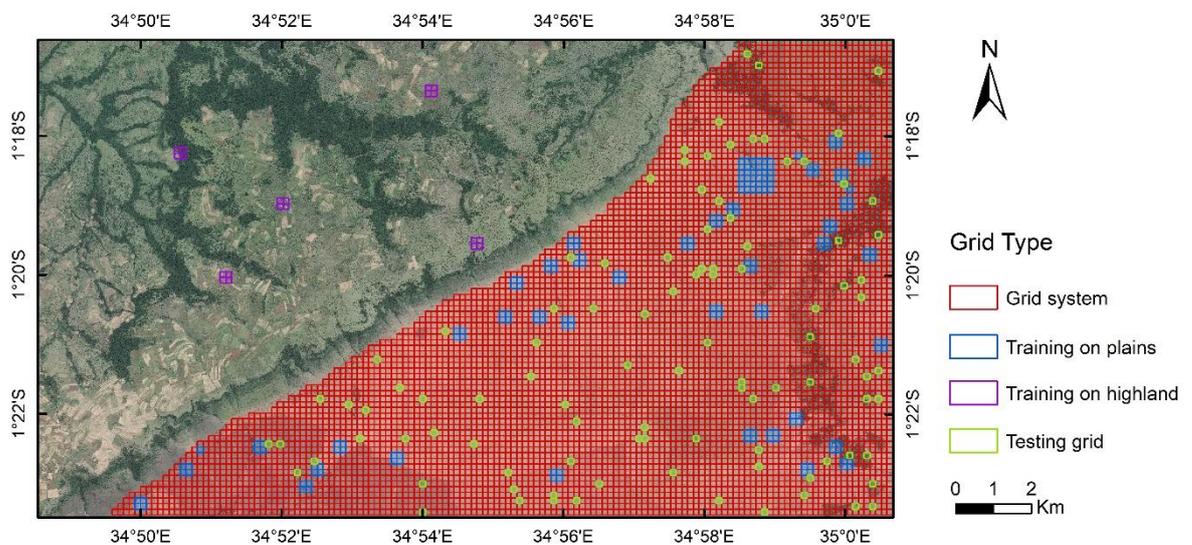


Figure 3. Overview of the training and testing dataset in the study area. The low-land plain area (to the lower right) and the highland area (to the upper left) on the whole satellite image are treated separately. The main wildebeest detection model is trained on the plain area using the training grids in red. The performance of the detection model is

tested on the testing grids in green. After this, to transfer the model to the highland, the additional 20 training grids in purple on the highland are introduced to train the model on the highland with different landscapes.

The satellite images both have four bands: red (R), green (G), blue (B), and near-infrared (NIR). Only the RGB bands were used for wildebeest detection and counting. After that, to explore the role of the NIR band in wildebeest detection, the NIR band was also added and the four-band satellite images with the same training and testing image grids were processed to detect and count the wildebeest.

To fit the input size requirement of the model, 336×336 pixels in each grid were cropped out of the 350×350 pixels to be fed into the model. The detailed information of the training and testing dataset is shown in Table 2.

Table 2. The size of the training and testing dataset

Dataset	Patch size	Number of patches	Area/km²	Number of wildebeest
Train	336×336	191	3.99	19,816
Test (updated)	336×336	100	2.09	2,249

2.3.2. Labelling wildebeest on satellite images

On the training and testing dataset, the individual wildebeest were labelled as points using the ESRI ArcGIS 10.7.1 environment. The wildebeest labelling was based on the majority votes of the visual interpretation from three observers on the GeoEye-1 satellite image in the year 2009 using satellite image in the year 2013 as the reference. Each wildebeest individual on the GeoEye-1 image is about 3 to 4 pixels in length and 1 to 2 pixels in width. On the true colour composite image, a wildebeest is a group of grey-brownish pixels with usually a dark black pixel in the centre representing its neck and spine with a black mane. The wildebeest form different patterns, such as lines, sparse groups, and dense clusters.

The second image acquired in 2013 was used to infer the correctness of labels on the first image acquired in 2009. By referring to the second image at the same location, the confusion caused by the surrounding objects like termite mounds and bushes can be cleared since the background objects are rather static between these two years. Examples of wildebeest point labels are shown in Figure 4.

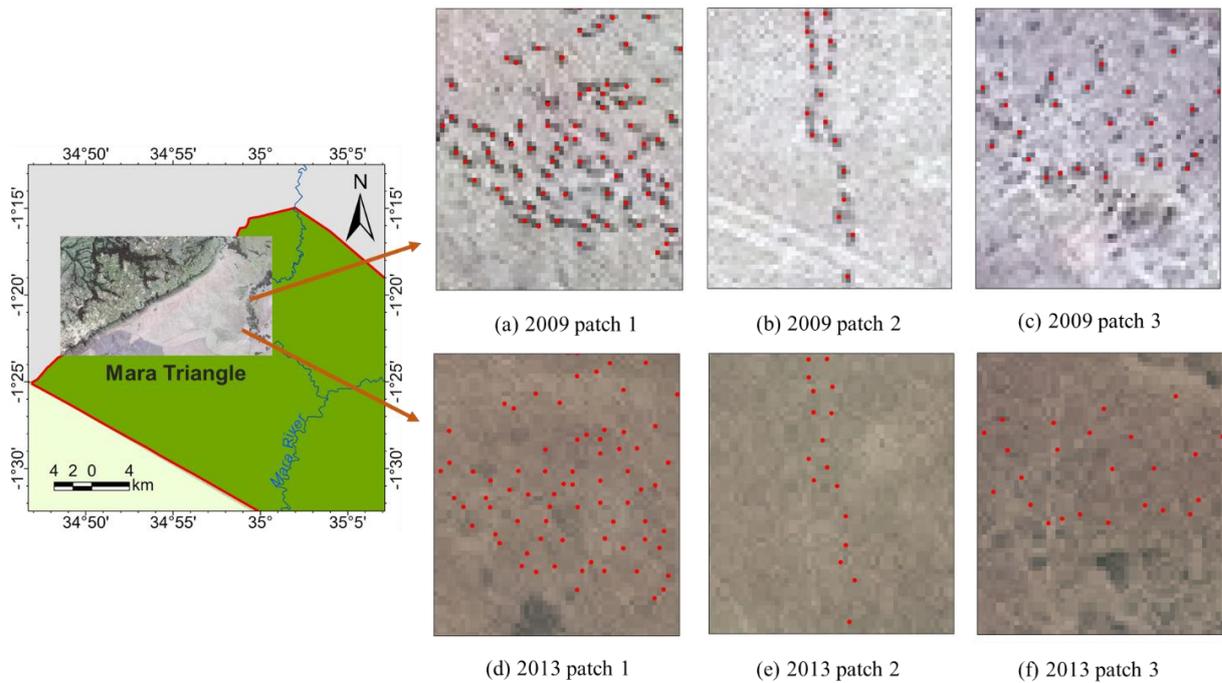


Figure 4. Wildebeest labelling on GeoEye-1 satellite image acquired in 2009. These six images on the right are three different patches cropped from GeoEye-1 images in 2009 and 2013. Each red point represents a wildebeest sample, and the samples are selected on GeoEye-1 2009 image. Image (a), (b), and (c) show three different distribution patterns of wildebeest, including dense group, line, and sparse group. Image (a) and (b) show the relative homogeneous environment, and (c) shows the heterogeneous environment, where shrubs confuse with the wildebeest. But when we compare image (c) with image (f) on the reference image in 2013, the shrubs are easily identified, and they will not be chosen as wildebeest samples in the 2009 image.

2.3.3. Wildebeest rasterization

In total, 19,816 wildebeest individuals have been labelled within the training image patches, and 2,249 wildebeest individuals have been labelled within the testing image patches, which were then rasterized into mask images consisting of wildebeest pixels and non-wildebeest pixels. Each wildebeest was represented by a segment containing 3×3 pixels that covered the full size of a wildebeest feature (Figure 5). Within the training dataset, the pixel percentage of wildebeest is less than 1 percent, and the non-wildebeest pixel percentage is more than 99 percent as seen in Table 3.



Figure 5. Satellite image with wildebeest and the rasterized labelled mask image

Table 3. Percentage of wildebeest and non-wildebeest pixels in the training dataset

	Number of wildebeest	Percentage of pixels/%
Wildebeest pixels	19,816	0.77
Non-wildebeest pixels	0	99.22

Figure 6 presents the workflow of data preparation.

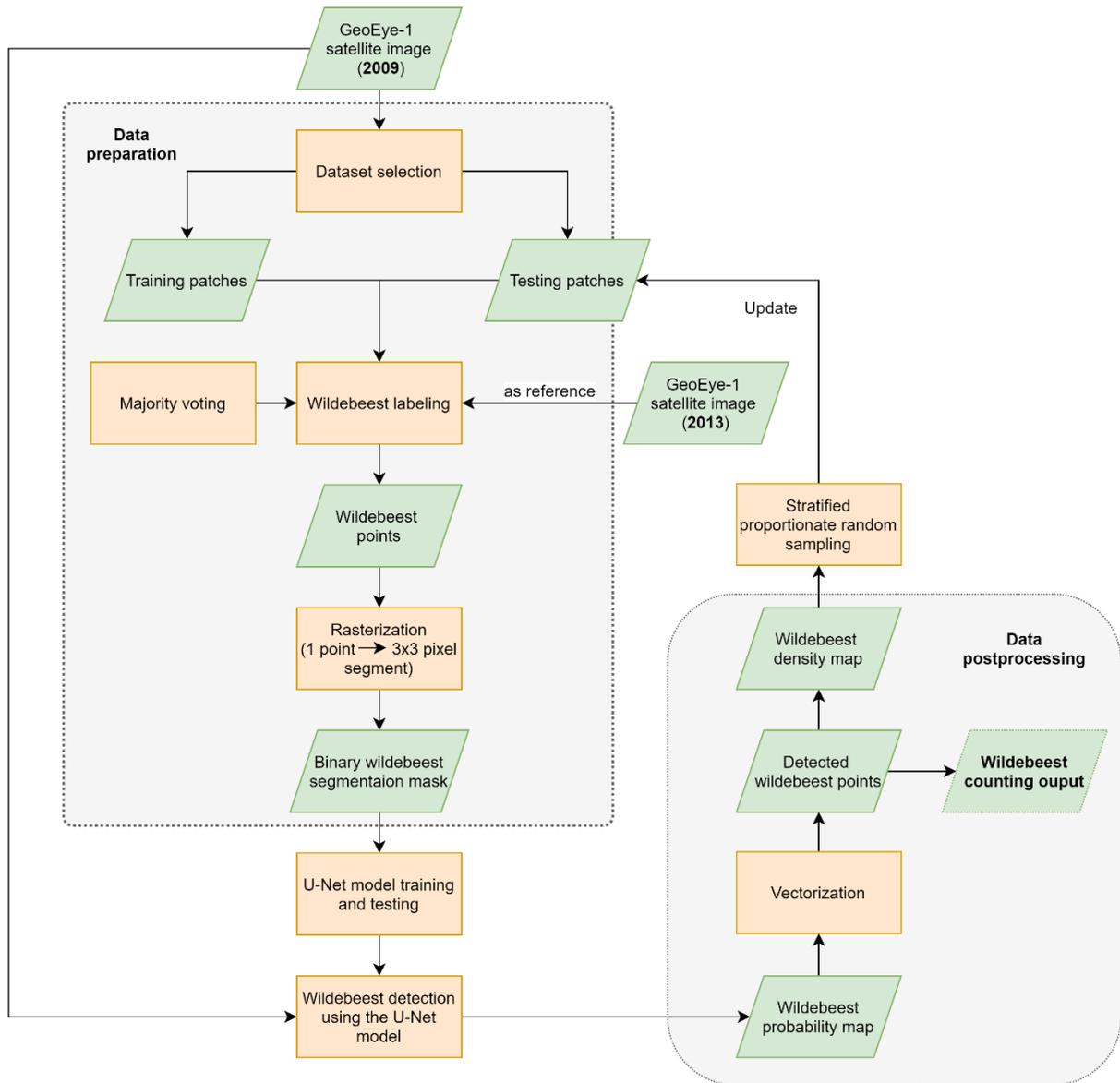


Figure 6. Flowchart of data processing

2.4. Deep learning model

2.4.1. U-Net model

The U-Net architecture was initially designed for biomedical image segmentation and then has been commonly used in other applications as well as remote sensing image segmentation (Ronneberger, Fischer, & Brox, 2015b). U-Net uses a U-shaped symmetrical encoder-decoder structure that consists of a contracting path on the left and an expansive path on the right (Figure 7). The contracting path encodes high-level contextual features step by step using successive layers, which generates lower and lower-resolution feature maps. Then the expansive path decodes the information of these feature maps and up-samples the image step by step to obtain the original resolution.

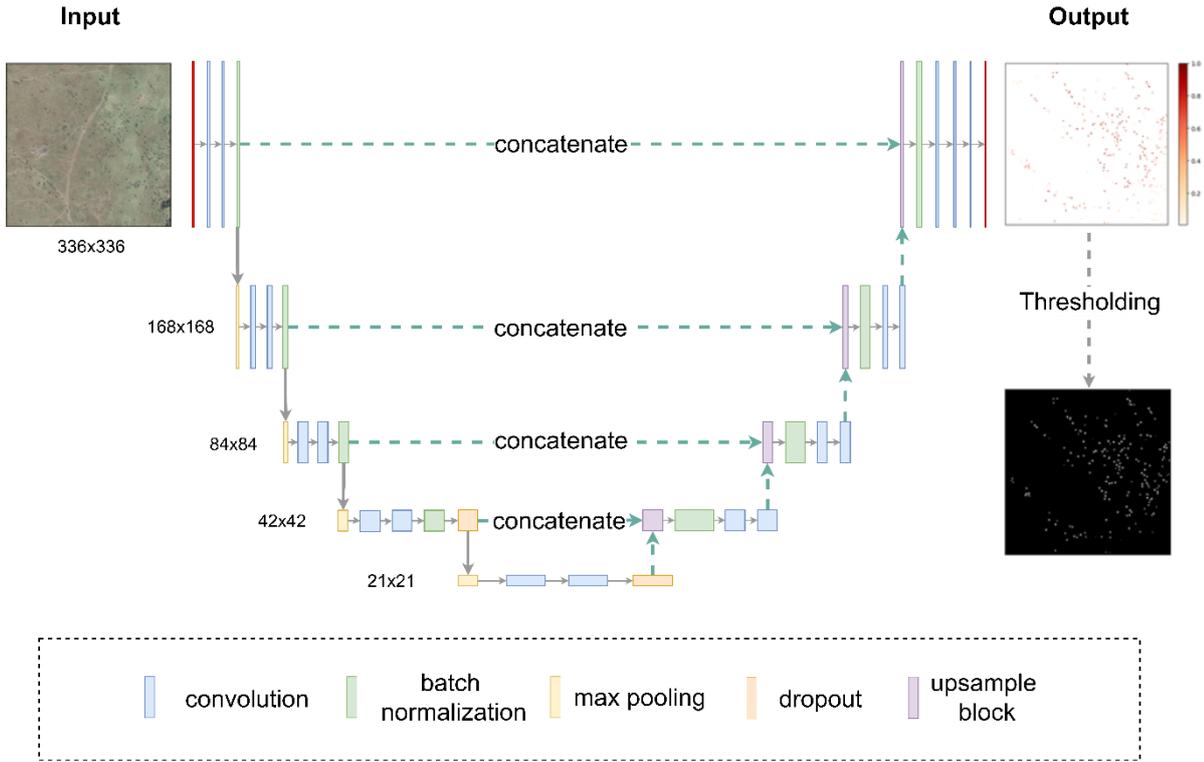


Figure 7. U-Net model architecture

On the left, each level consists of two convolution layers, one batch normalization layer, and one max pooling layer. The 3×3 convolution layer captures local features using the “same” padding followed by a rectified linear unit (ReLU) activation function, such as edge features. The weights in the convolution layers are initialized by the He_normal kernel initializer (He et al., 2015). Then the batch normalization layer normalizes the distributions of each training mini-batch, which can accelerate the training process by reducing the amplified change of data distributions when the parameters of every layer in the deep model are updated (Ioffe & Szegedy, 2015). The following 2×2 max-pooling layer down-samples the feature map to get higher-level features. The width and height of the feature maps are halved after pooling, and then the lower-resolution feature map is fed into the next block level. The same procedure is repeated 4 times until reaching the bottom level of the U-Net. A dropout layer is added to the last two blocks to prevent overfitting by randomly dropping the units and simplifying the network (Srivastava et al., 2014). With an input size of $336 \times 336 \times 4$, the output size is now $21 \times 21 \times 512$.

In the expanding path on the right, every level consists of an up-sample block, which has a 2×2 up-convolution layer that expands the size of the image by 2 times, a 2×2 convolution layer, and a batch normalization layer. The up-sampled output is then concatenated with the corresponding feature map with the same resolution in the contracting path on the left. This skip connection merges both low-resolution information that provides evidence for classification, and the high-resolution feature map, which supports precise localization. The following layers include one batch normalization layer and two 3×3 convolution layers. The process is repeated successively until it reaches the original image size. Finally, a 1×1 convolution layer with a sigmoid activation function is used to map the feature vector of each pixel to a value ranging from 0 to 1 that represents the probability of wildebeest existence on this pixel.

The discrepancy between the output probability map and the ground truth labelled image is represented mathematically by a loss function. To minimize the discrepancy and approach the ground truth, U-Net implements backpropagation using a gradient descent algorithm to adjust the weights in all the layers iteratively to reduce the loss.

2.4.2. Model setup

Before being fed into the model, the dataset was first cropped into patches with a size of 336×336 pixels to fit the requirement of input size by the U-Net model. Then all the data patches were augmented using flipping horizontally, flipping vertically, and rotating 90 degrees techniques to gain more variations into the dataset. Since wildebeests appear in different directions on the landscape, these data augmenters can help prevent overfitting and improve the generalization capability of the model on unseen data with different patterns.

A weighted loss function, Tversky loss, was used to measure the discrepancy between the predictions and ground truth labels (Salehi, Erdogmus, & Gholipour, 2017). Tversky loss function is calculated using this formula:

$$T(\alpha, \beta) = 1 - \frac{\sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i g_i + \alpha \sum_{i=1}^N (1 - p_i) g_i + \beta \sum_{i=1}^N p_i (1 - g_i) + \epsilon} \quad (1)$$

where p_i denotes the predicted probability of a pixel i to be a wildebeest pixel and g_i denotes the ground-truth value, which is 1 for a wildebeest pixel and 0 for a non-wildebeest pixel. Thus $\sum_{i=1}^N p_i g_i$ summarizes the wildebeest true positives (TP), $\sum_{i=1}^N (1 - p_i) g_i$ summarizes the false negatives (FN), and $\sum_{i=1}^N p_i (1 - g_i)$ summarizes the false positives (FP). α and β are the weights of penalties for FNs and FPs, respectively, and the sum of α and β is 1. For a highly imbalanced dataset, such as the wildebeest dataset (the percentage of wildebeest pixels is less than 1% in the training dataset as shown in Table 3), the model tends to predict all the pixels into non-wildebeest pixels to get high overall accuracy. Tversky loss function was used to address the issue of data imbalance. By using higher β , the emphasis is added on the wildebeest pixels to minimize the number of misclassified wildebeest pixels. The parameter β was finely tuned over a range of values (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99) to achieve the optimal trade-off between false positives and false negatives.

The model was trained with the Adam optimizer with an initial learning rate of 0.0001 (Kingma & Ba, 2015). The learning rate was reduced by a factor of 0.33 every time when the loss on the validation set stops improving after 20 epochs. The rate of the dropout was set to be 0 as preliminary experiments showed that a higher dropout rate did not significantly increase the model performance. The batch size was 12, and the model was trained for 100 epochs. The model that generated the lowest loss on the validation dataset among all the epochs was selected as the final model.

The direct output of the U-Net model was a probability map of wildebeest existence that varied from 0 to 1. To obtain the binary classification results, the predictions of each sub-model were first normalized and then classified into wildebeest and non-wildebeest classes using a threshold. All pixels with probability higher or equal to the threshold were wildebeest pixels. The default threshold was set as 0.5.

2.4.3. U-Net-based ensemble model

Ensemble model learning approach was employed to improve the generalization ability and robustness of the U-Net model. The advantage of the ensemble model of neural networks was first demonstrated in Hansen and Salamon's research that showed the improvement of classification accuracy compared with the individual model (Hansen & Salamon, 1990). Building an ensemble model comprises three steps: selecting data for each model, training the individual model, and combining multiple models to obtain a composite model (Polikar, 2012). The first step requires diversity in data sampling, which generates diverse "errors" in the ensemble system (Brown et al., 2005). Every single model has its inevitable variance, but they generally agree on the correct classifications. By combining the results of the individual model with an appropriate strategy, such as averaging the predictions or majority voting, the variance can be reduced, and thus a lower generalization error is achieved.

During the training process of this U-Net model, the training dataset is split into multiple batches in each epoch with a size of 12 sample images (in this research). The model is then trained on these batches one by one, and the weight parameters are updated accordingly. However, the wildebeest training dataset contains many variations such as heterogeneous landscapes and different characteristics of wildebeest. This splitting can result in imbalance and variation between the batches, and the randomness thus brings uncertainty to the model. The resulted model can be error-prone to certain landscapes. Besides, the weight initialization during model training can also introduce randomness and uncertainty. Ensemble learning approach can address this problem by the joint decision making of multiple similarly configured models that emphasize different variations of the dataset and thus reach a better final decision by averaging out the errors.

In this study, the K-fold cross-validation splitting method was conducted to assign the dataset for the individual model. The training dataset was split into k folds ($k = 10$ in this research), among which $k-1$ folds were used for training the U-Net model, and the remaining one was used for validation. During the training process of each model, the model that generated the lowest loss on the validation dataset was selected as the best model. In total, there was k number of separate models trained and validated with varied datasets. Then the predictions of k sub-models were summarized and averaged to get the predictions of the final ensemble model. The architecture of this ensemble model is presented in Figure 8. Tversky loss weight parameter tuning was also evaluated using the ensemble model approach.

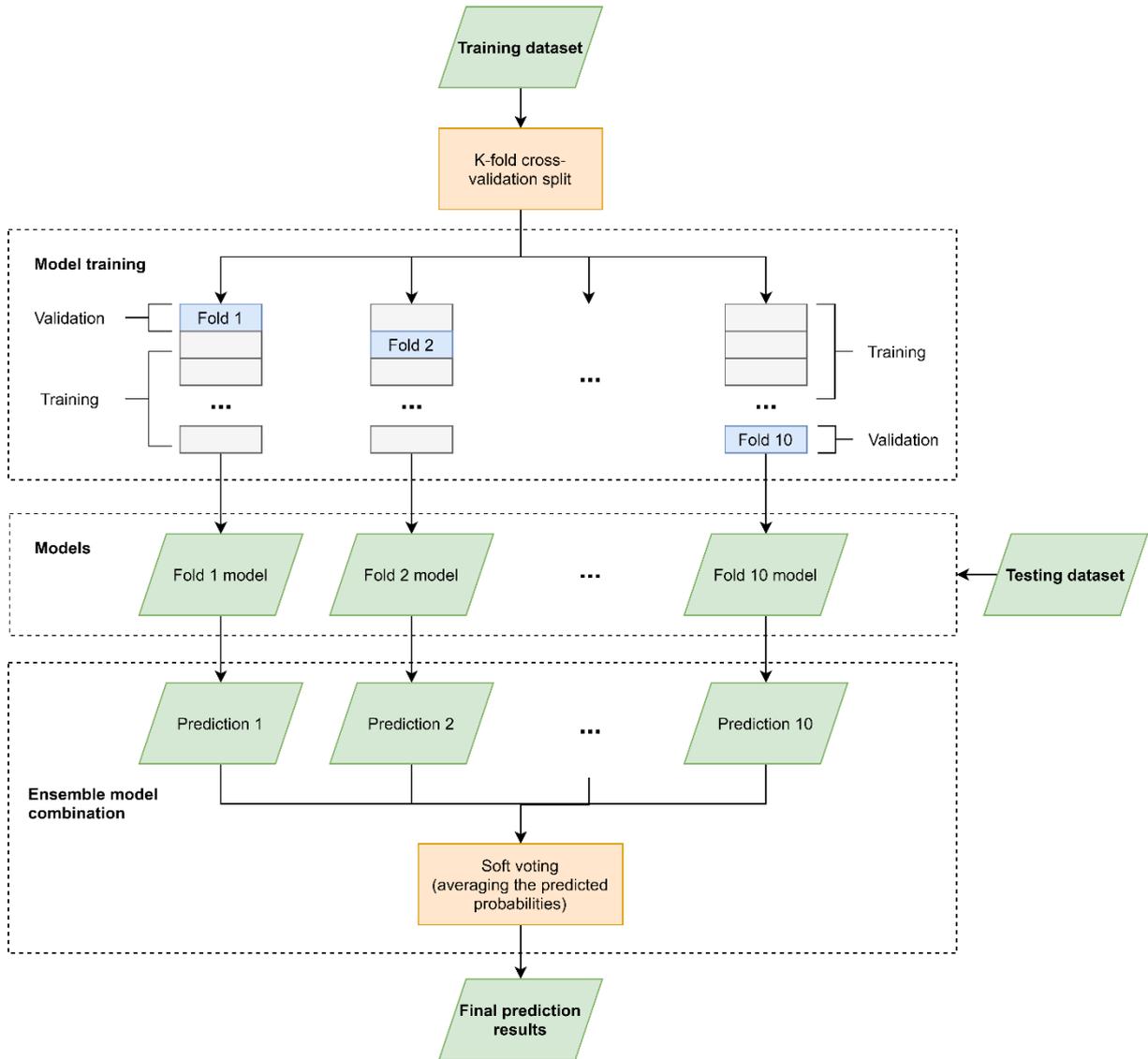


Figure 8. The architecture of the U-Net-based ensemble model

2.5. Locating and counting

2.5.1. Wildebeest locating

After thresholding, the segmentation map (wildebeest/non-wildebeest) was generated. Since the goal is to detect and count the wildebeest at the level of a single animal, the pixel-level results are still not enough. Therefore, the wildebeest segments were converted into vector points to locate the wildebeest individuals. This was done by extracting the centroid of each wildebeest segment to represent the whole wildebeest.

There are cases when the wildebeests are so close that one segment contains several of them together. For the aggregated wildebeest cluster that contains more than 9 pixels, K-means clustering was applied to partition the segment into k separate clusters, and the cluster centres were treated as wildebeests. The number of clusters, k , was calculated according to the ceiling division of the total number of pixels by the

size of one wildebeest (9 pixels). For instance, if there are 11 pixels in one segment, then the number of clusters is 2, and two wildebeest individuals will be extracted from this segment.

2.5.2. Wildebeest counting

The model trained on the training dataset with the optimal loss function weight parameter was run on the whole satellite image of 2009 to obtain the predictions. The predictions were normalized and then vectorized to count the total number of wildebeest points. To deal with the randomness and uncertainty of the deep learning model caused by the weight initialization and k-fold splitting, the U-Net-based ensemble model training process was repeated 10 times using the same set of parameters. The counting numbers generated by these ten separate composite ensemble models were averaged to get the estimation of the wildebeest population.

2.6. Transfer learning

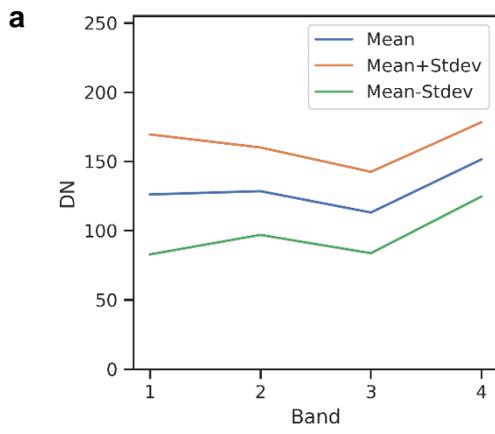
2.6.1. Spatial transfer learning

To test the spatial transferability of the wildebeest detection model, the U-Net model was applied to the highland area. As explained in section 2.3.1 and Figure 3, the training dataset with both plain and highland data was used for spatial transfer learning. No new testing dataset was selected because it is difficult to find wildebeest on the highland. There are animals in the field and forest, however, based on the expert knowledge of local people, they are more probable to be cattle instead of wildebeest. Thus, the performance evaluation of the transferred model was merely through visual interpretation of the final classification results of the whole area.

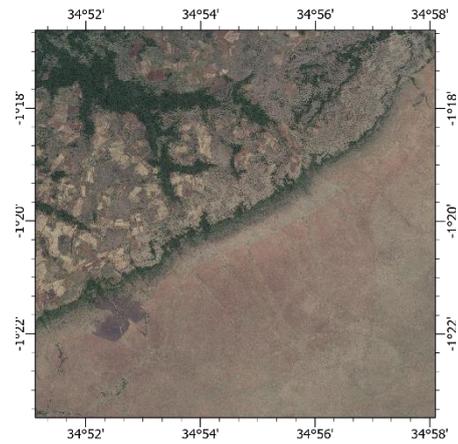
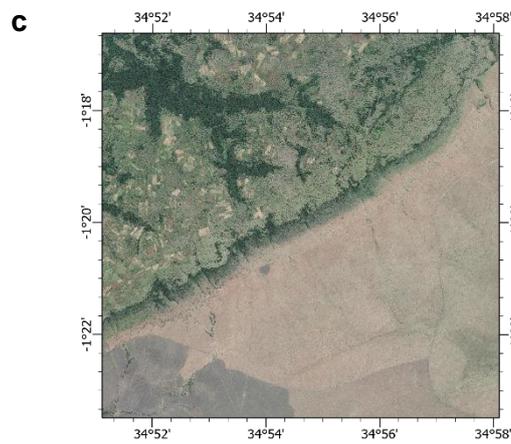
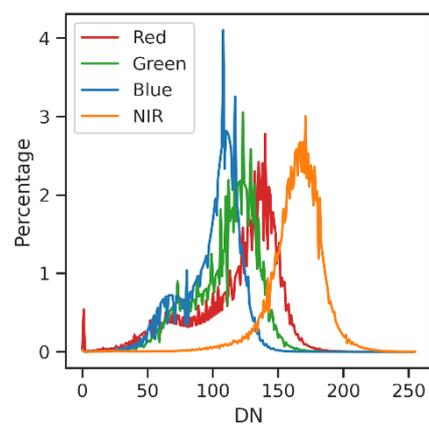
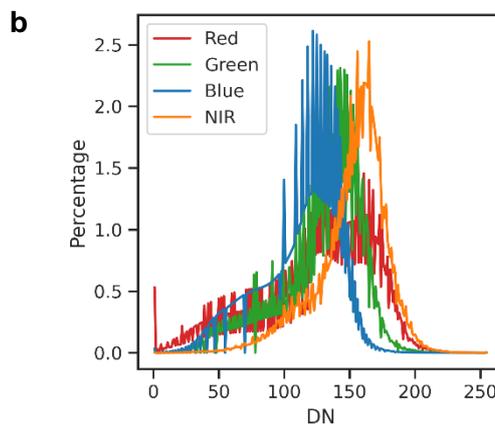
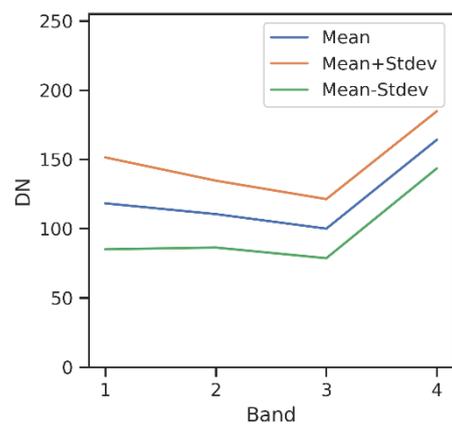
2.6.2. Temporal transfer learning

To test the temporal transferability, the GeoEye-1 satellite image in 2013 was utilized since it covers the same area with similar landscapes as the image in 2009, while the spectral characteristics are different due to the influence of the atmospheric conditions, sun elevation, and azimuth angle on different dates when the satellite images were acquired. Besides, the features of wildebeest pixels are also different. Figure 9 visualizes the spectral characteristics of both satellite images and how the wildebeest look on different images.

Satellite image in 2009



Satellite image in 2013



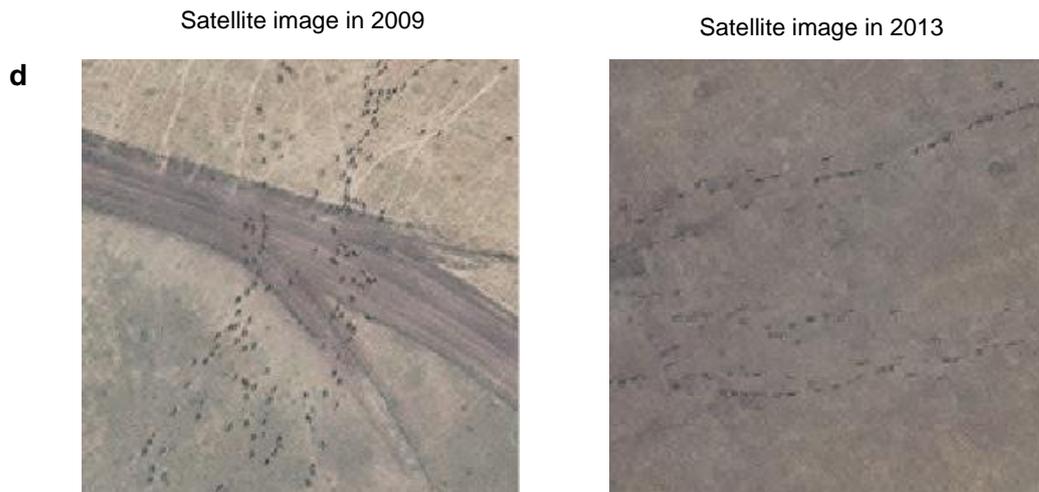


Figure 9. The spectral characteristics of GeoEye-1 satellite image in 2009 and 2013. **a**, the pixel DN value range of red, green, blue, and near-infrared band on GeoEye-1 satellite image in 2009 and 2013. The range is defined by the mean and standard deviation: mean \pm standard deviation. **b**, the histogram of DN values of the red, green, blue, and near-infrared band on GeoEye-1 satellite image in 2009 and 2013. **c**, an overview of the GeoEye-1 satellite image in 2009 and 2013 with true colour composites (red, green, and blue). **d**, a patch area with wildebeest on the satellite image in 2009 and 2013. The wildebeest on image 2013 look vaguer than it on image 2009.

Since the wildebeest population on the 2013 image is very limited, it is not possible to select enough training data to re-train the model, nor to select testing data to evaluate the performance. Therefore, the model trained on the 2009 image was directly applied to the 2013 image, and the transferability was assessed by visual interpretation.

2.7. Accuracy assessment

The accuracy of the wildebeest detection model was assessed at the individual level instead of pixel level. After vectorization, the ground truth mask images and the predicted images were both converted to wildebeest points. To determine whether the wildebeest point is predicted correctly, a searching distance was introduced to match the predicted point with the corresponding ground truth point. This is because the “ground truth” wildebeest segment (3×3) is not purely composed of wildebeest pixels. It also includes some background pixels, and the boundary between wildebeest and background is ambiguous. Therefore, the precise location of wildebeest should allow a slight shift within two pixels. The judgment of correctness followed this set of rules:

- Search for the ground truth point near a predicted point within the searching distance.
 - If a ground truth point is found within the searching distance, then this predicted point is a true positive (TP).
 - If there is no ground truth point near the predicted point, then this predicted point is a false positive (FP).
- All the remaining ground truth points that haven’t been matched after the searching process are false negatives (FN).

Sensitivity analysis was conducted to evaluate the influence of the searching distance setting on the detection accuracy. The distance range was 0 to 2 pixels (0 to 0.9 centimetres). The best searching distance chosen as the final parameter was the one that produces the highest accuracy.

To quantitatively assess the overall performance of the model, the following accuracy metrics were utilized: precision (2), recall (3), and F1-score (4). Precision measures how accurate are the predictions. It is calculated as the ratio between the number of detected true positives and all detected positives. Recall measures how good the model is at finding the true positives. It is the ratio between the number of detected true positives and all existing ground truth positives. F1-score is a harmonic mean of precision and recall, which reflects the overall accuracy.

$$Precision = \frac{\text{number of detected true positives}}{\text{number of detected positives}} = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{\text{number of detected true positives}}{\text{number of existing positives}} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

Precision-recall curve and Area Under the Curve (AUC) were adopted to compare the performance of the models trained with different U-Net model parameters. When applying different thresholds to the probability map, the binary classification results are different, and multiple pairs of precision and recall can be calculated. For threshold 0 and 1, the precision and recall were manually set to be (0, 1) and (1, 0), respectively. These precision-recall pairs were then plotted on the graph, and the area under the curve was calculated using the composite trapezoidal rule. The value of AUC is between 0 and 1. The higher the AUC is, the better the model performance is.

2.8. Statistical analysis

To explore the role of near-infrared (NIR) band of GeoEye-1 satellite image in wildebeest detection, another U-Net model was trained with 4-band (R, G, B, NIR) data, and the accuracy was compared with the model trained with 3-band (R, G, B) data. Due to the inevitable existence of randomness during deep learning model training, simply comparing the accuracy on the testing dataset is not sufficient. Statistical analysis is important to test the significance of the difference between the performance of two convolutional neural network models over the same testing dataset.

The commonly used approach is the K-fold cross-validated paired t-test. The training dataset is split into k folds, and the model will be trained on the k-1 folds and tested on the remaining one fold. In total k trials are conducted, and two series of testing accuracy from two models are obtained and compared using paired t-test. The K-fold cross-validated paired t-test ensures the independence of the testing dataset because they are not overlapping among all the k trails. However, the independence assumption of the training dataset is seriously violated. Between the cross-validation iterations, there is a high proportion of overlapping (89% overlapping in 10-fold cross-validation), which can cause high type I error (rejection of a true null hypothesis when there is no significant difference) (Dietterich, 1998).

To overcome this limitation, Dietterich proposed the 5×2 cross-validated paired t-test which then gradually replaced the k-fold method (Demšar, 2006; Dietterich, 1998). Instead of k folds, the dataset is split into 2 folds. In this way, the training sets are completely non-overlapping, as well as the testing sets. The cross-validations are then replicated 5 times to achieve a more stable estimate. Experiments showed that this method has a lower type I error compared to K-fold cross-validation.

The calculation of t statistic is as follows:

$$\tilde{t} = \frac{p_1^{(1)}}{\sqrt{\frac{1}{5} \sum_{i=1}^5 s_i^2}} \quad (5)$$

In replication i , the dataset is split equally into datasets S_1 and S_2 . Method A is trained on S_1 and tested on S_2 to get accuracy $p_A^{(1)}$, and trained on S_2 and tested on S_1 to get accuracy $p_A^{(2)}$. For method B, the resulting accuracy estimates are $p_B^{(1)}$ and $p_B^{(2)}$. By calculating the differences, we get: $p_i^{(1)} = p_A^{(1)} - p_B^{(1)}$, and $p_i^{(2)} = p_A^{(2)} - p_B^{(2)}$. The average of the differences is $\bar{p}_i = (p_i^{(1)} + p_i^{(2)})/2$, and the estimated variance is $s_i^2 = (p_i^{(1)} - \bar{p}_i)^2 + (p_i^{(2)} - \bar{p}_i)^2$. The t statistic is calculated from these estimates.

The application of paired t-test entails these three assumptions: (1) the samples are independent; (2) the differences follow a normal distribution unless the sample size is large enough; (3) the variances of two sets of samples are equal. For assumption 1, the estimates within 2-fold cross-validation are completely independent, but the 5 iterations do not fulfil the requirement of independence. However, research indicated that this issue is inevitable, and it gets less problematic after adjusting the formula (Dietterich, 1998). For assumption 2, the Shapiro-Wilk test was applied to test the normality of data distribution. For assumption 3, Levene's test was used to test the homogeneity of variances between the two groups of accuracy. Lastly, the two-tail t-test was implemented at a significance level of 0.05. The effect of the near-infrared band on wildebeest detection accuracy was evaluated statistically by the 5×2 cross-validated paired t-test. All the statistical tests were conducted using the SciPy library in python.

3. RESULTS

3.1. Detecting the wildebeest

3.1.1. U-Net model output and evaluation

3.1.1.1. Model trained on the full dataset

The U-Net model was trained on the Google Colaboratory platform (<https://colab.research.google.com/notebooks/intro.ipynb>). Using GPU P100, the training process of 100 epochs with the full dataset took approximately 0.8 hours. The training loss and validation loss decreased significantly from 1 to about 0.2 to 0.4 after the first 20 epochs and reach the lowest validation loss at epoch 10 to 40 usually. Then the training loss continued to decrease during training, while the validation loss kept relatively steady and slightly increasing.

Figure 10 shows one example of the learning curve (Tversky loss weight $\beta = 0.9$) while using the full training dataset, including the change of precision, recall, and f1-score on the training and validation dataset. In this training process, validation loss reaches the lowest at epoch 13. Therefore, the model at epoch 13 is selected as the best model.

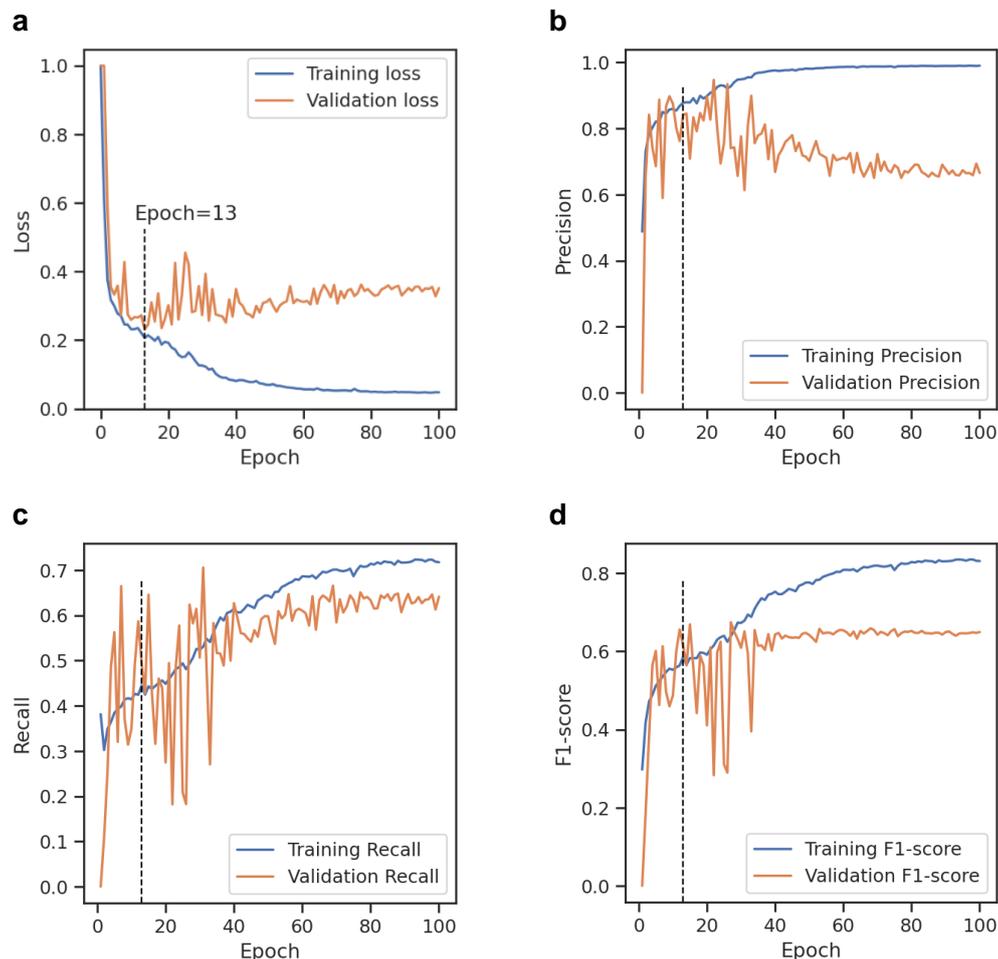


Figure 10. Learning curve during model training. **a**, Training and validation loss curves during model learning. **b**, the Precision curve on training and validation dataset during model learning. **c**, Recall curve on training and validation dataset during model learning. **d**, F1-score curve on training and validation dataset during model learning.

The direct output of the U-Net model is a probability map for each patch as shown in Figure 11.b. The value of each pixel is the probability of the existence of wildebeest in the corresponding pixel in the input satellite image in Figure 11.a. The range of probability is (0, 1). After thresholding (the optimal threshold is 0.5), the binary wildebeest segmentation map was obtained (Figure 11.c) and then was converted to a wildebeest point map (Figure 11.d). The wildebeest points were then compared with the ground truth labelled points, which produced true positives (green circles in Figure 11.d), false positives (red cross in Figure 11.d), and false negatives (yellow cross in Figure 11.d).

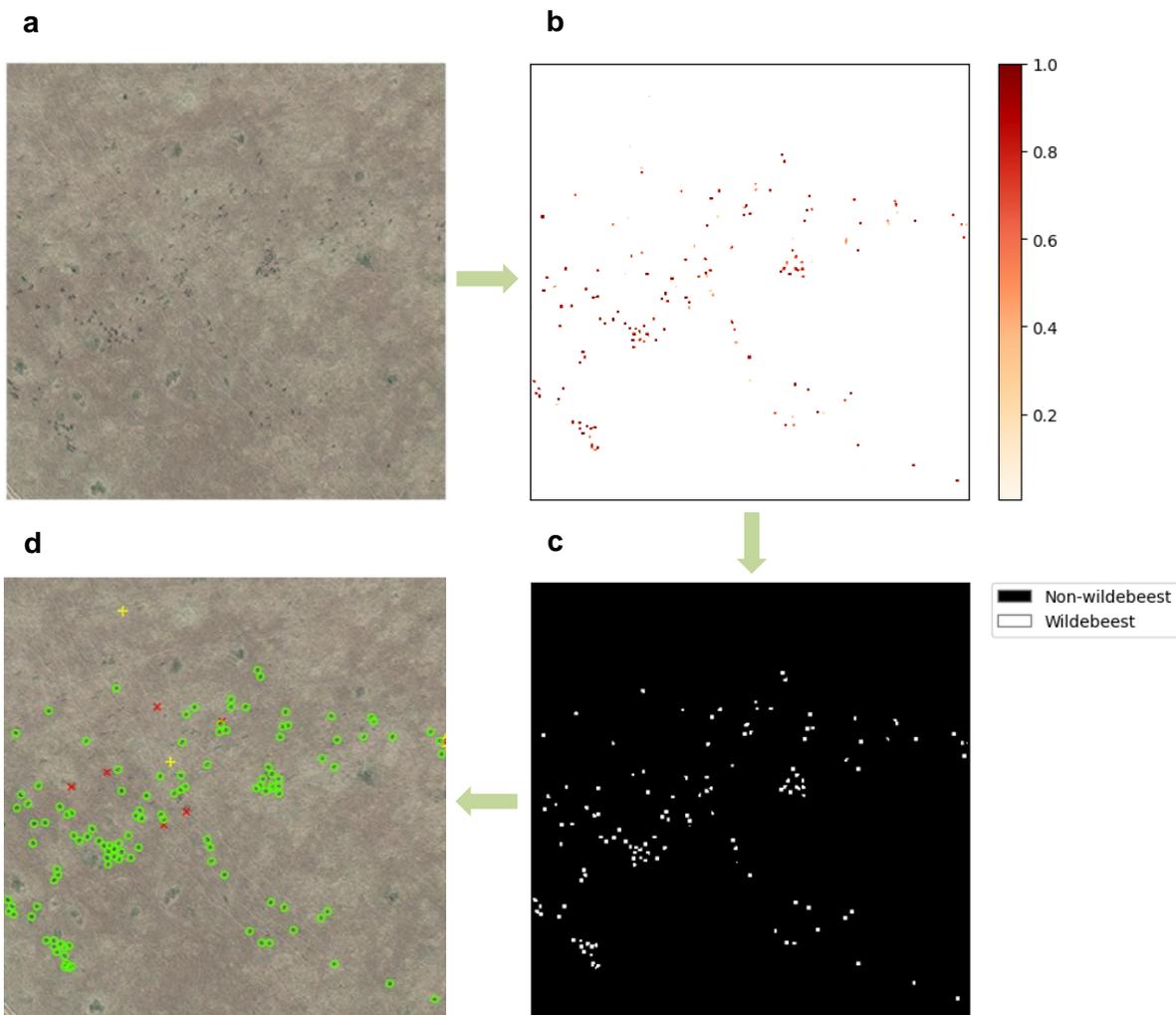


Figure 11. The wildebeest detection output of the U-Net model and wildebeest locating product. **a**, Satellite image input of the U-Net model. **b**, Probability map product of U-Net model. **c**, Wildebeest binary segmentation map after thresholding. **d**, The wildebeest detection results on the satellite image, including true positives (in green), false positives (in red), and false negatives (in yellow).

The pixel-level and object-level accuracy calculated based on the point output is presented in Table 4.

Table 4. Pixel-level and object-level accuracy of wildebeest detection model trained using the full dataset and tested using the first testing dataset with 30 randomly selected patches

	Precision	Recall	F1-score
Pixel-level	0.84	0.49	0.62
Object-level	0.83	0.69	0.75

3.1.1.2. Model trained using the ensemble approach

The U-Net-based ensemble model using the K-fold cross-validation splitting method was implemented (k=10). Using GPU P100, the training process of 100 epochs with 10 iterations for all the 10 combinations took approximately 10 hours.

The accuracy of all the 10 fold combinations is shown in Table 5. The predictions of all the 10 models were then averaged and the accuracy of the ensemble model was: precision 0.87, recall 0.84, f1-score 0.85, respectively. The ensemble model approach significantly improved the accuracy compared with using any single fold combination or using the full dataset.

Table 5. Object-level accuracy of U-Net-based ensemble model on the first testing dataset with 30 randomly selected patches

	Precision	Recall	F1-score
Fold 1	0.85	0.64	0.73
Fold 2	0.84	0.64	0.73
Fold 3	0.78	0.74	0.76
Fold 4	0.83	0.60	0.70
Fold 5	0.81	0.71	0.76
Fold 6	0.82	0.68	0.75
Fold 7	0.86	0.63	0.73
Fold 8	0.83	0.68	0.75
Fold 9	0.84	0.59	0.69
Fold 10	0.85	0.64	0.73
10-fold ensemble	0.87	0.84	0.85

3.1.2. Update the testing dataset

The preliminary experiments on the first testing dataset (30 randomly selected patches on the study area with 3,171 wildebeests) showed that the model achieved good accuracy: precision 0.87, recall 0.84, and f1 0.85 (Table 5). However, the wildebeest are not evenly distributed spatially, thus the accuracy generated based on random sampling is not reliable enough. To ensure that the testing dataset is well representing the whole study area, the testing dataset was updated using a stratified proportionate random sampling approach.

The strata were built according to the wildebeest density across the study area. The wildebeest density data were obtained using the preliminary model. As mentioned before, the accuracy of the preliminary model on

the first testing dataset is not representative and reliable enough. But still, it reflects the detection performance on more than 3,000 wildebeests and their background. Thus the wildebeest detection results can be used to illustrate the general distribution pattern of wildebeest density.

Firstly, the trained preliminary model was applied to the study area (lower land) to detect the wildebeest. Then the number of wildebeest within every grid was calculated, which resulted in the wildebeest density map (Figure 12). All the grids were classified into 4 classes using the standard deviation method according to the wildebeest count, and they were grouped into 4 density categories: low, medium, high density, and very high density. Table 6 shows the wildebeest count range and the proportion of grids in each category. Lastly, the testing samples were randomly selected among all the grids within each category separately according to the proportion. In total, 100 testing grids were selected (Figure 3) and the processes of wildebeest labelling and rasterization were implemented to get the new testing dataset. This new testing dataset was then used to evaluate the generalization performance of the deep learning model.

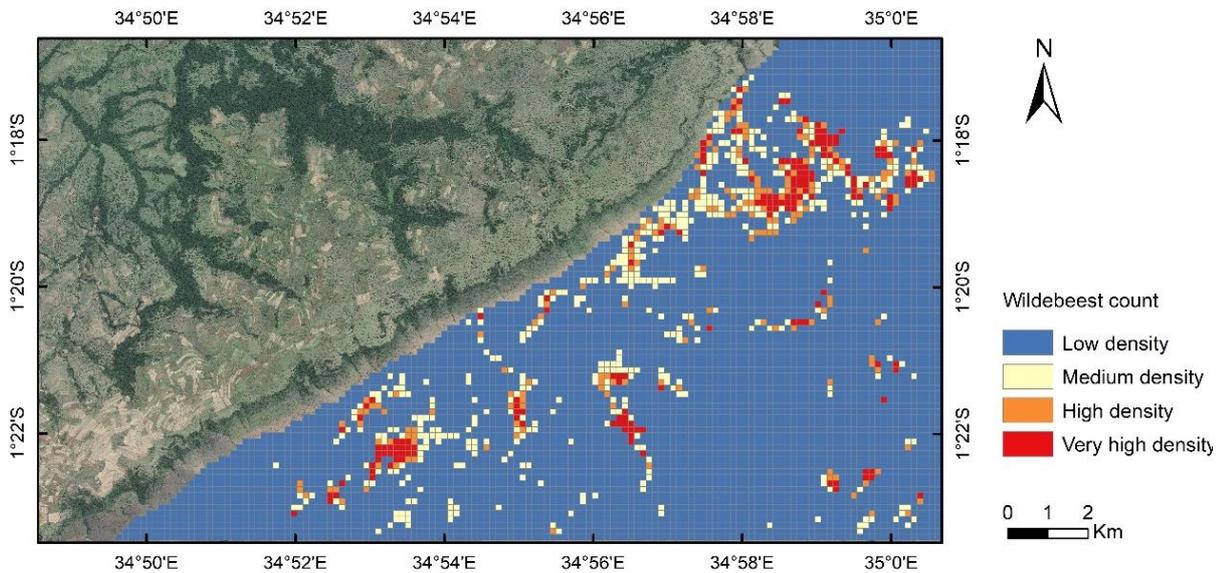


Figure 12. Wildebeest density map generated by the preliminary model

Table 6. Updated testing data sampling strata

Wildebeest count	Density group	Proportion	No. of sample grids
0 - 42	Low density	86%	86
43 - 89	Medium density	8%	8
90 - 135	High density	3%	3
136 - 526	Very high density	3%	3

3.1.3. Sensitivity analysis of searching distance

As discussed in section 2.8, a searching distance was introduced to match the predicted point with the corresponding ground truth point to evaluate the accuracy. Sensitivity analysis was conducted to evaluate

the influence of the searching distance setting on the detection accuracy and find the best searching distance. The distance range is 0 to 2 pixels (0 to 0.9 centimetres).

From Figure 13, it's observed that the detection accuracy increases when the searching distance increases. The accuracy increases from 0 to around 0.6 when the searching distance increases from 0 to 0.01 m. When the searching distance is between 0.01 to 0.42 m (which is about 1 pixel), the accuracy keeps stable. The accuracy rises again to around 0.8 at searching distance = 0.43 m and keeps stable after this until searching distance reaches 0.61 and the accuracy achieves above 0.85. When the searching distance is larger than 0.61m, all the accuracy metrics remained stable (i.e., reached an asymptote).

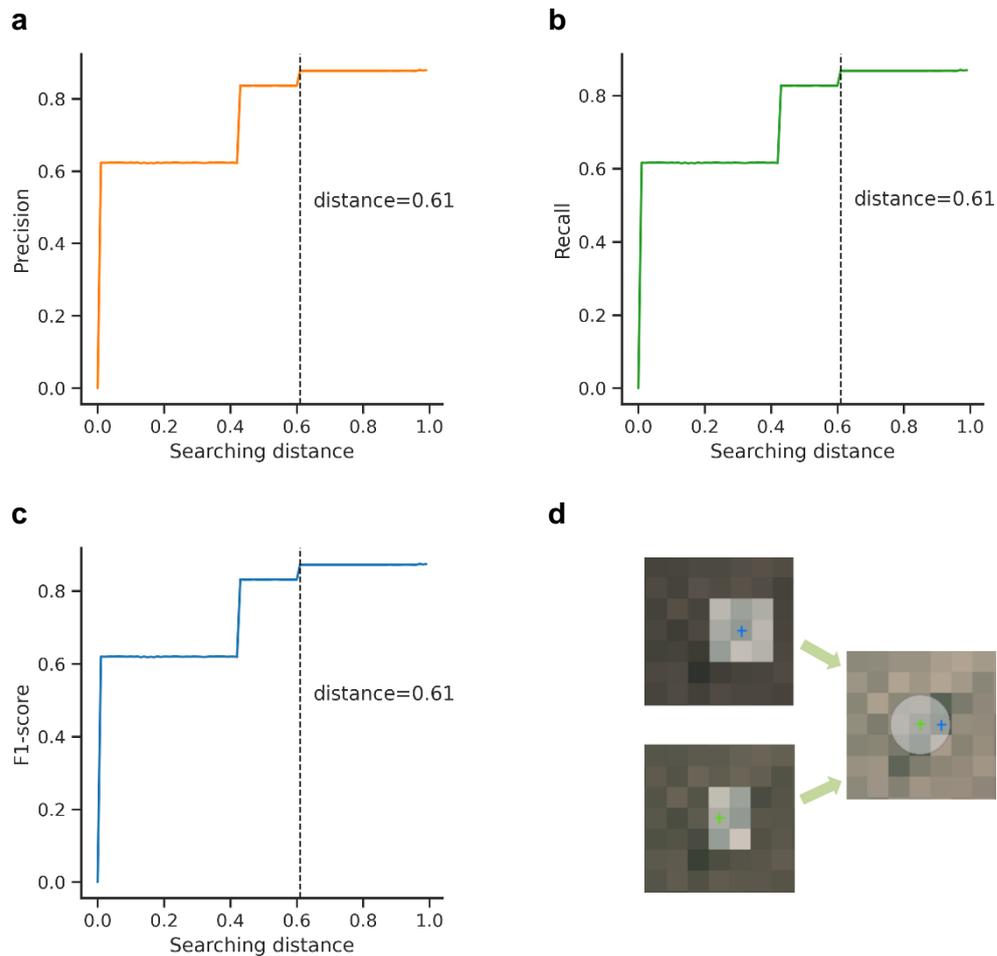


Figure 13. Sensitivity analysis of searching distance in accuracy assessment. **a**, Relation of Precision and searching distance. **b**, Relation of Recall and searching distance. **c**, Relation of F1-score and searching distance. **d**, An example of the application of searching distance in model evaluation. The cross point in blue is the ground truth centroid of the wildebeest segment, and the green cross point is the centroid of the predicted wildebeest segment. The transparent grey area in d.3 is the buffer area of the predicted centroid defined by searching distance. The point in green is a true positive when the distance to the blue ground truth point is smaller than the searching distance.

3.1.4. Parameter tuning

The parameter β , weight of false positives in Tversky loss, was finely tuned to achieve the balance of precision and recall. As seen in Table 7, with the increase of weight for false positives, precision rises from

0.19 to 0.97, while recall declines from 1.00 to 0.53. F1-score, the metric that compromises both precision and recall, increases from 0.32 to 0.87 and then drops to 0.68 when the weight increases to 0.99. The AUC (Area Under the precision-recall Curve) shows a similar trend as the F1-score (Table 7, Figure 14). The optimal weight is 0.9 when both F1-score and AUC reach the highest value.

Table 7. Accuracy of wildebeest detection trained with different weights in Tversky loss

Weight β	Precision	Recall	F1-score	AUC
0.01	0.19	1.00	0.32	0.64
0.1	0.32	0.99	0.48	0.74
0.2	0.40	0.99	0.57	0.81
0.3	0.42	0.98	0.59	0.84
0.4	0.48	0.98	0.64	0.87
0.5	0.54	0.98	0.69	0.89
0.6	0.59	0.97	0.74	0.90
0.7	0.72	0.95	0.82	0.92
0.8	0.75	0.94	0.83	0.93 (0.926)
0.9	0.88	0.87	0.87	0.93 (0.928)
0.99	0.97	0.53	0.68	0.89

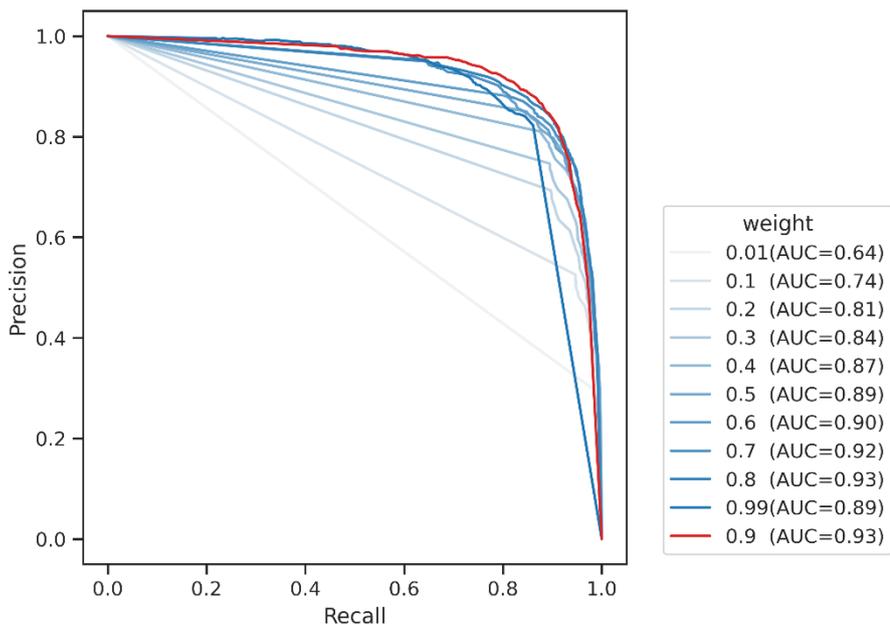


Figure 14. Precision-Recall curve of the U-Net model trained with different weights in Tversky loss

3.1.5. Performance of the final model

Using this weight that generates the highest AUC ($\beta=0.9$), the results of the U-Net-based ensemble model on the updated dataset are presented in Table 8. The individual models show a high variance: model Fold 8 is more advantageous in getting a high recall but it obtains a low precision; model Fold 3 and Fold 7 achieve a high precision but a low recall. The final composite model reaches a good balance among all the 10 separate models and the final F1-score is higher than every single model.

Table 8. The accuracy of the U-Net-based ensemble model on the updated testing dataset with the Tversky loss parameter β of 0.9

	Precision	Recall	F1-score
Fold 1	0.75	0.77	0.76
Fold 2	0.83	0.79	0.81
Fold 3	0.87	0.76	0.81
Fold 4	0.78	0.83	0.81
Fold 5	0.84	0.83	0.83
Fold 6	0.76	0.85	0.80
Fold 7	0.87	0.77	0.82
Fold 8	0.72	0.91	0.80
Fold 9	0.79	0.83	0.81
Fold 10	0.80	0.82	0.81
10-fold ensemble	0.88	0.87	0.87

For the final composite model, the precision of wildebeest detection is 0.88, the recall is 0.87, and the F1-score is 0.87. The model showed a high generalization accuracy on the testing dataset. Some examples of the predictions of wildebeest on the testing dataset are displayed in Figure 15 and Figure 16.

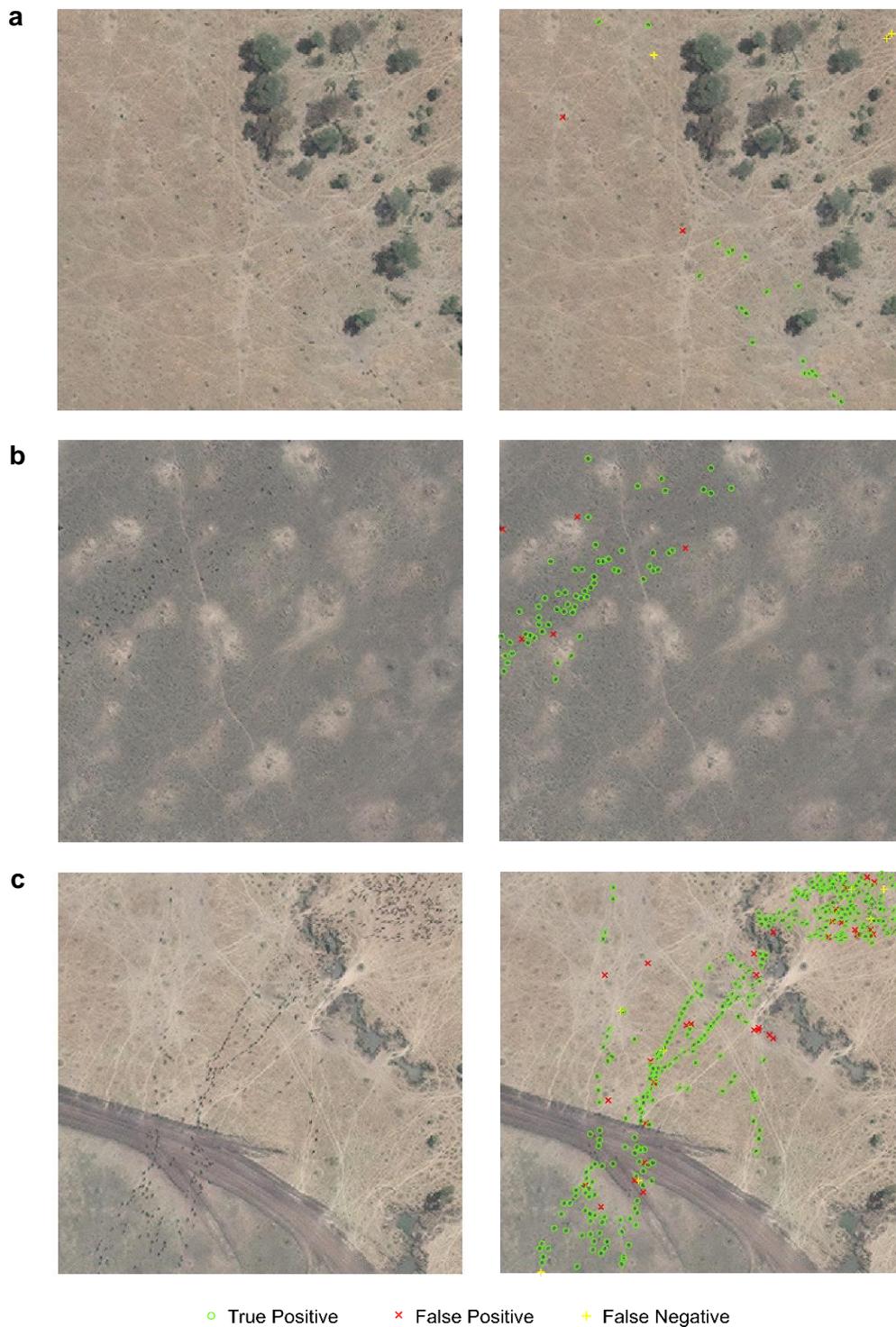


Figure 15. Predicting wildebeest on the GeoEye-1 satellite image. These example images of 336×336 pixels from the independent testing dataset show the capability of the U-Net convolutional neural network model to detect wildebeest from GeoEye-1 satellite imagery with 43-cm resolution. The left column shows the raw satellite images and the right column displays the wildebeest detection results on the image. **a**, The detection accuracy on patch a is: Precision 0.90,

Recall 0.86, F1-score 0.88. **b**, The detection accuracy on patch b is: Precision 0.92, Recall 1.00, F1-score 0.96. **c**, The detection accuracy on patch c is: Precision 0.90, Recall 0.97, F1-score 0.94.

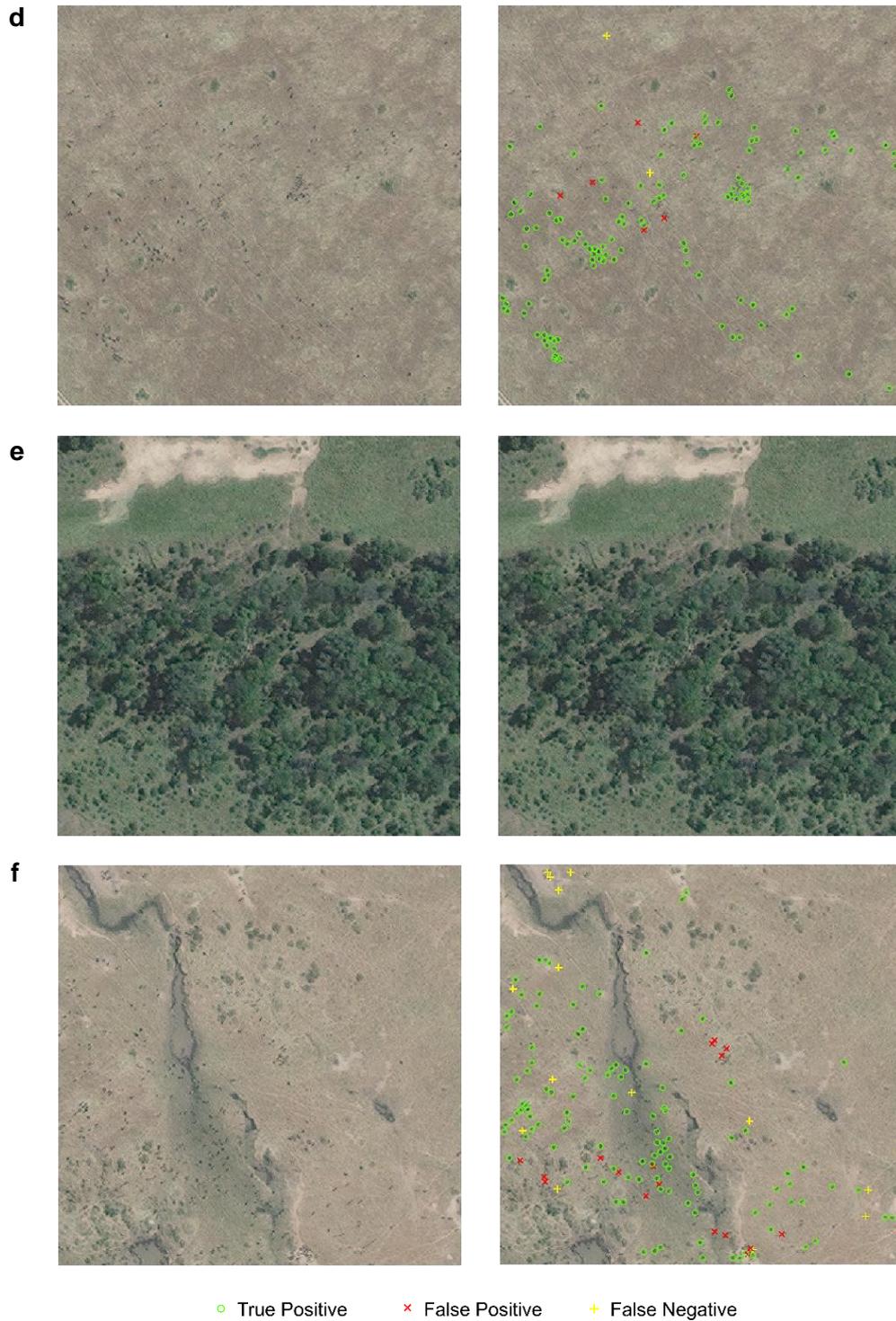


Figure 16. Predicting wildebeest on the GeoEye-1 satellite image. The left column shows the raw satellite images and the right column displays the wildebeest detection results on the image. **d**, the detection accuracy on patch a is: Precision 0.94, Recall 0.97, F1-score 0.96. **e**, the detection accuracy on patch a is: Precision 1.00, Recall 1.00, F1-score 1.00. **f**, the detection accuracy on patch a is: Precision 0.87, Recall 0.89, F1-score 0.88.

The model was then applied to the whole satellite image to detect the wildebeest over the entire study area. The results are shown in Figure 17, 18, 19, and 20.

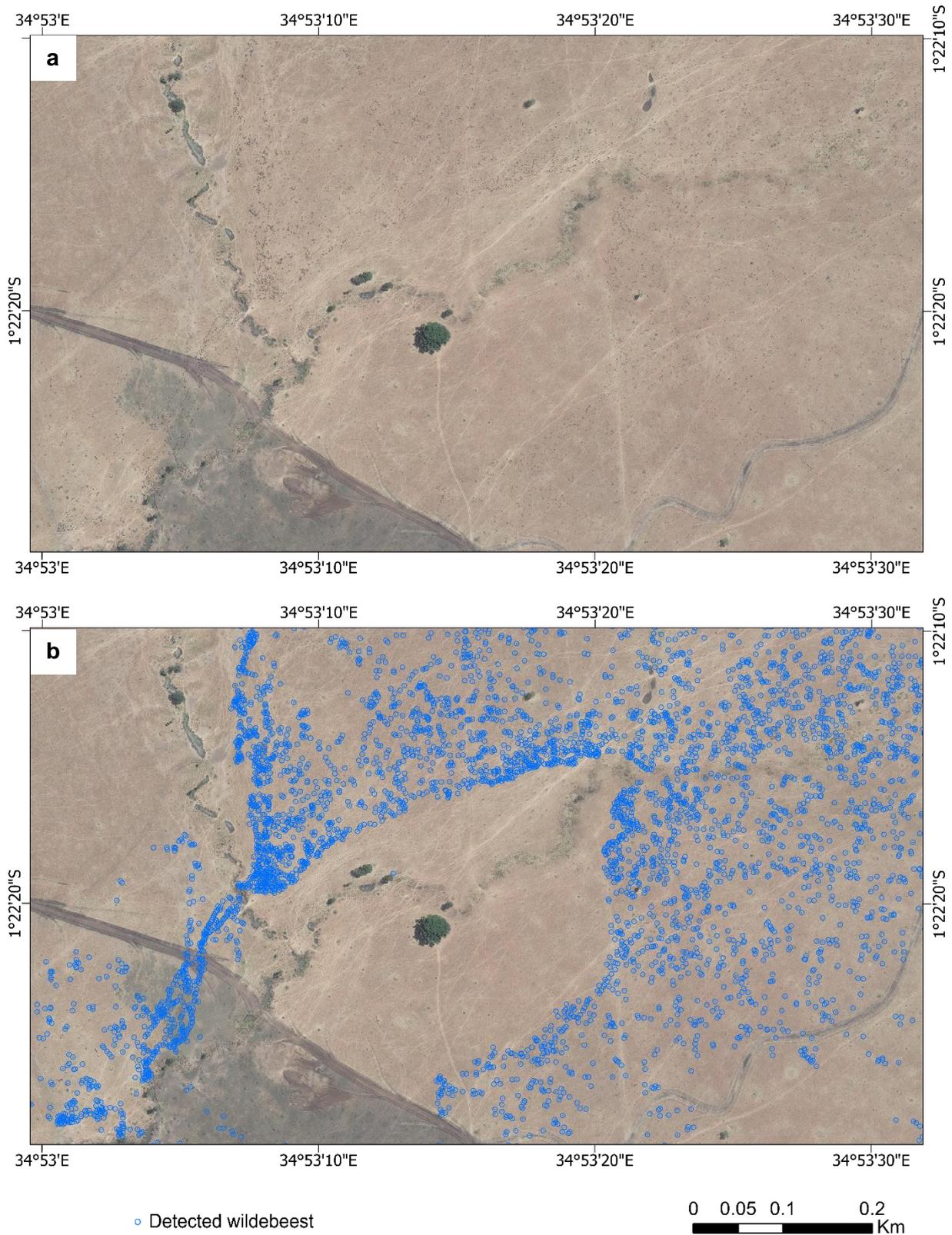


Figure 17. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on bare land). **a**, the GeoEye-1 satellite image. **b**, wildebeest detected on the GeoEye-1 satellite image.

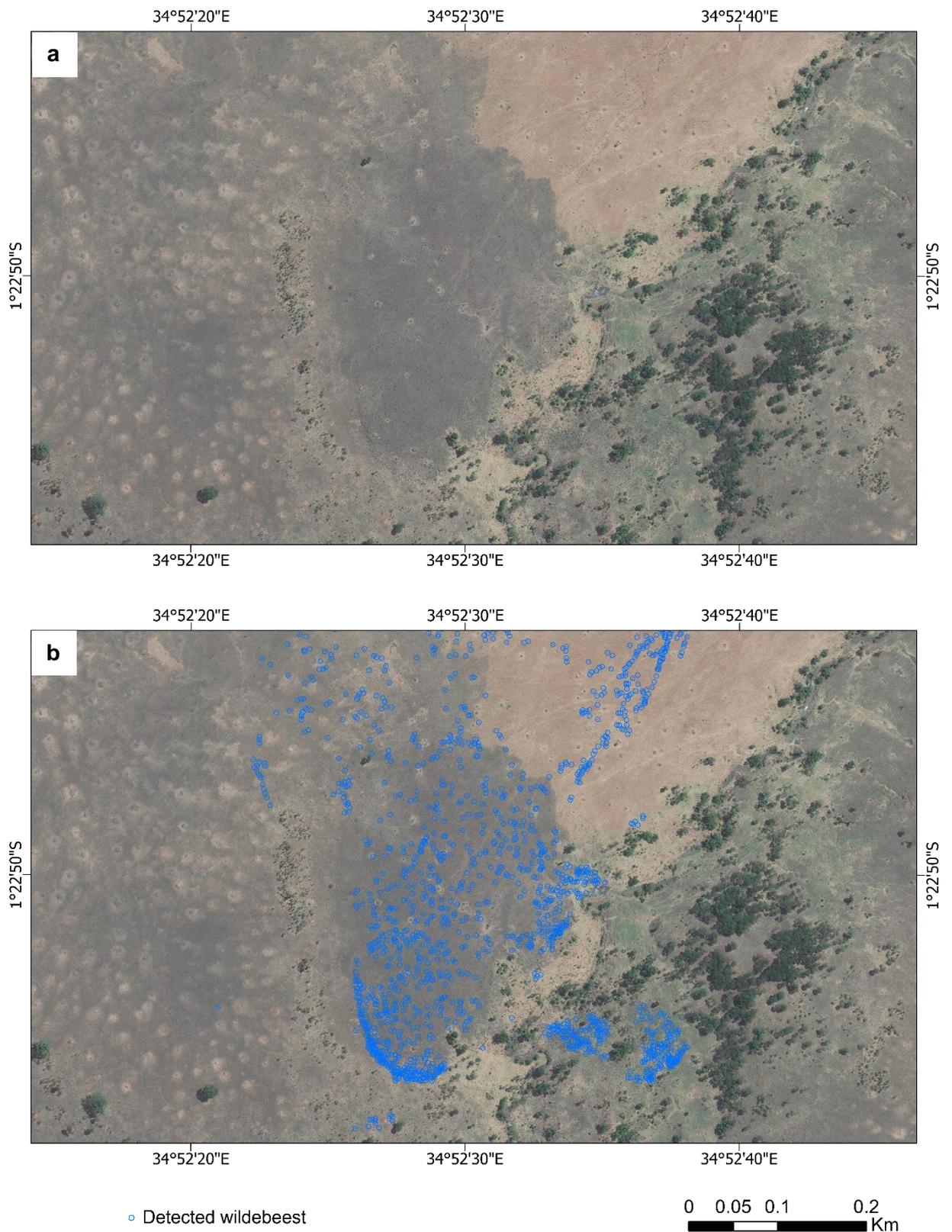


Figure 18. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on burnt area). **a**, the GeoEye-1 satellite image. **b**, wildebeest detected on the GeoEye-1 satellite image.

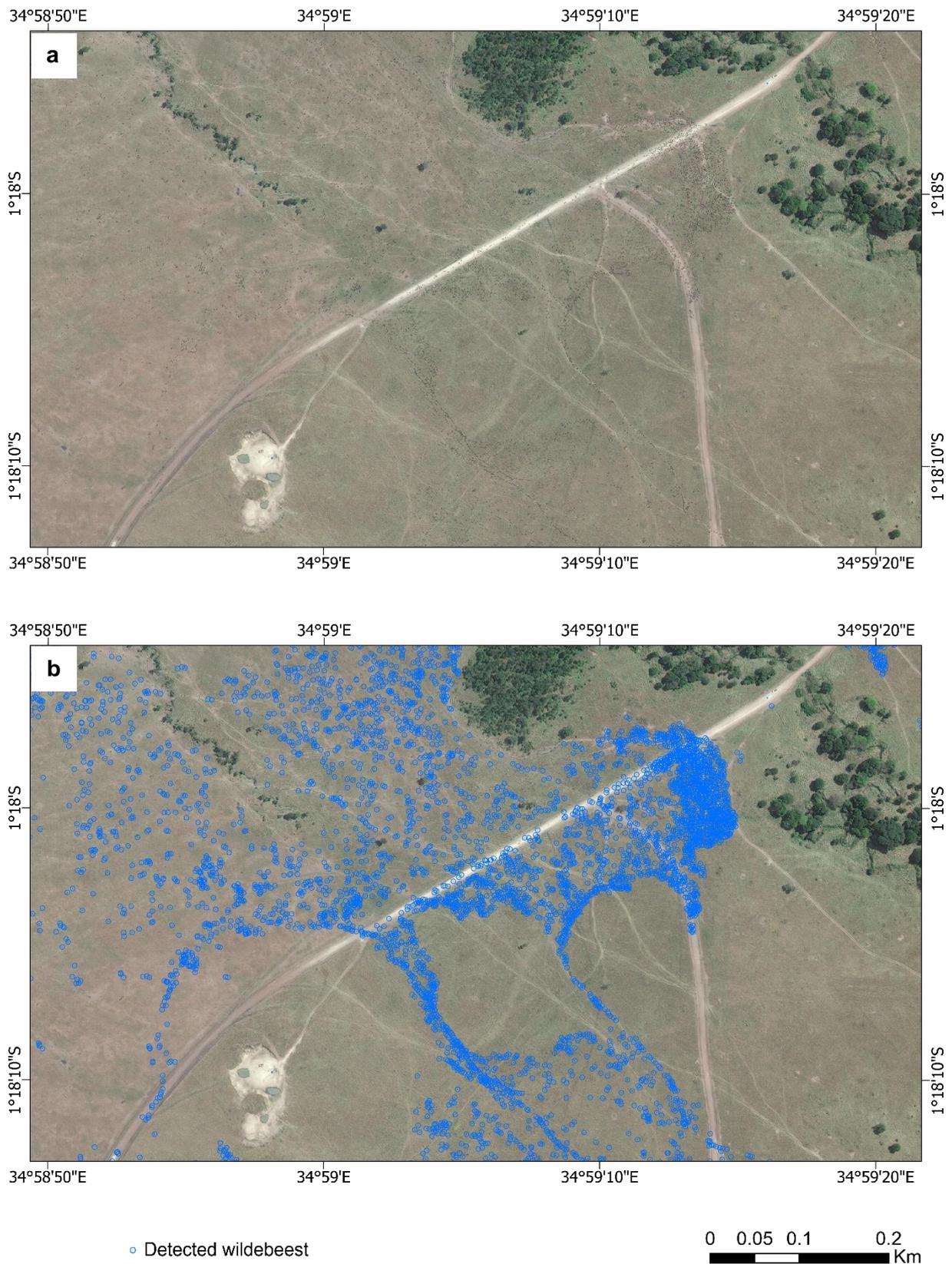


Figure 19. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest on grassland). **a**, the GeoEye-1 satellite image. **b**, wildebeest detected on the GeoEye-1 satellite image.

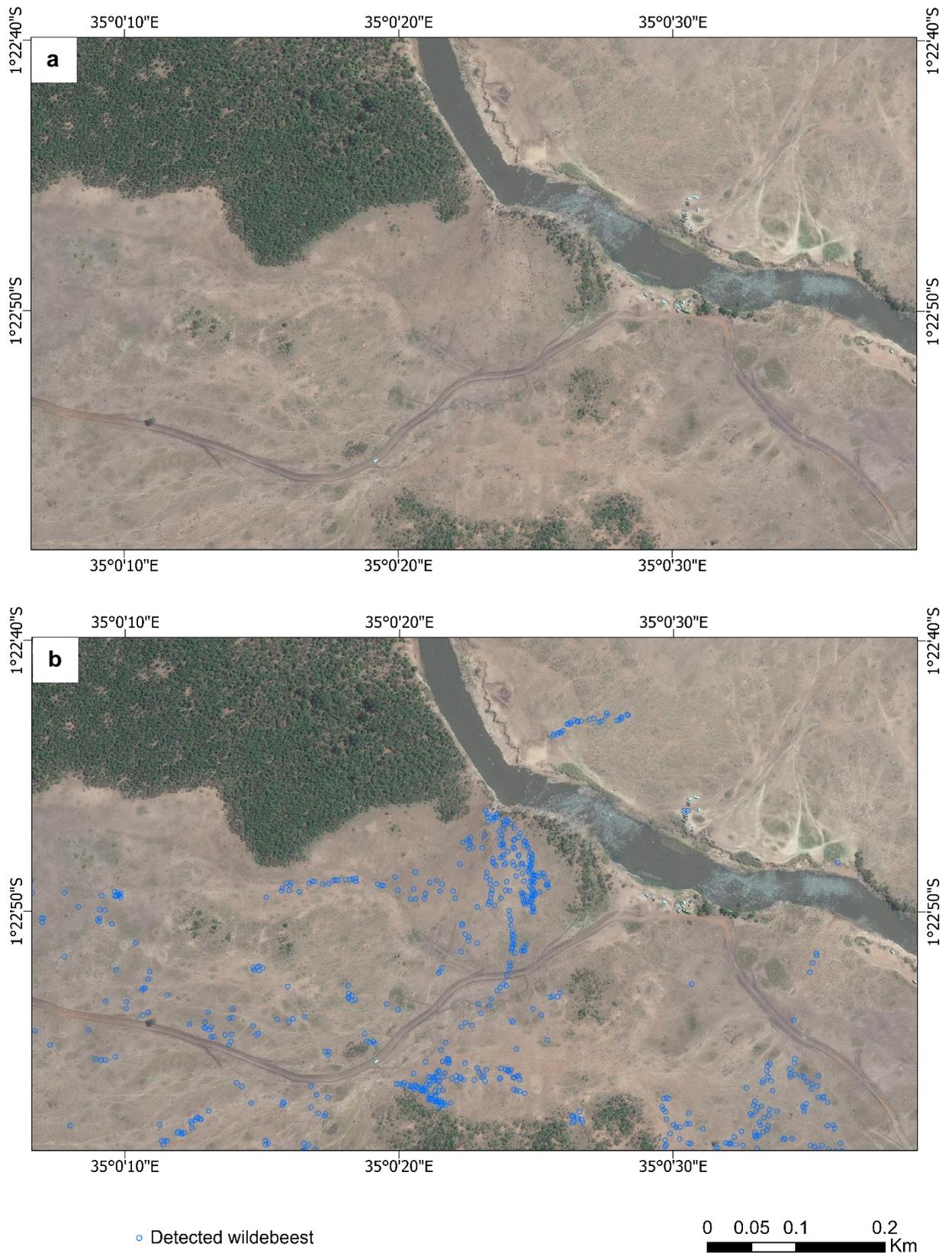


Figure 20. Detecting wildebeest from GeoEye-1 satellite image using U-Net model (example of wildebeest crossing the Mara River). **a**, the GeoEye-1 satellite image. **b**, wildebeest detected on the GeoEye-1 satellite image.

Using the weight parameter that generated the highest AUC (weight $\beta = 0.9$), the U-Net-based ensemble model training process was implemented 10 times. The accuracy varies among these 10 iterations. Figure 21 presents the F1-score of the ensemble model on the testing dataset using different threshold values. The results of all these 10 models are presented. The F1-score reaches the peak when the threshold is around 0.5. This suggests that the default threshold setting is appropriate. Besides, at threshold = 0.5, the confidence interval of the F1-score (the grey area) is very narrow, indicating the stability of the model on wildebeest detection.

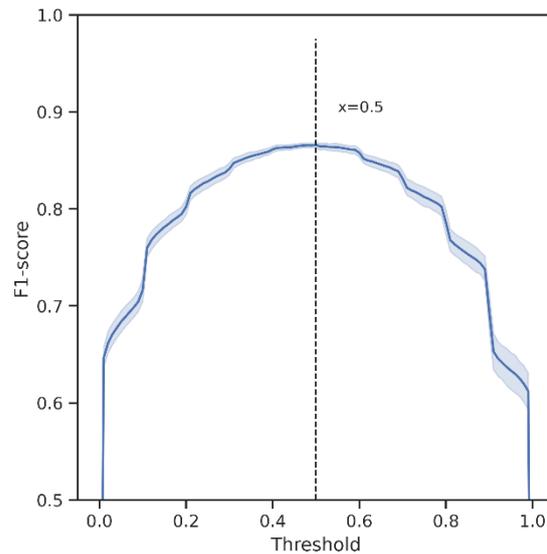


Figure 21. F1-score of the U-Net model under different thresholds. The model was trained 10 times separately, and the line plot aggregates multiple F1-score values of 10 models at each threshold. The solid line represents the average values of all the 10 iterations. The grey area is the 95% confidence interval of the F1-score.

3.2. Counting the wildebeest

The 10 models trained using 10 iterations were applied to the whole GeoEye-1 satellite image. After vectorization, the number of wildebeest in the entire study area were detected and then counted. The statistics of the wildebeest population count using the 10 different models are presented in Table 9. Since the sample size ($n=10$) is small, a Student's t statistic is used to estimate the confidence interval. Therefore, the 95% confidence interval of the population count on the GeoEye-1 satellite image (low land covering around 140 km²) calculated from the mean, standard error, and t statistic is (126122, 132792).

Table 9. Statistics of the wildebeest population count using the U-Net model from the GeoEye-1 satellite image. Df: degree of freedom; Std.: standard deviation; SE: standard error; CI: confidence interval.

df	Mean	Std.	SE	t statistic (95% CI)
9	129457	4422	1474	2.262

3.3. Contribution of the near-infrared band

The previous analysis was built on the GeoEye-1 satellite image with only the first three bands: red, green, and blue. To explore the role of the near-infrared (NIR) band, the 4-band satellite image was processed. The U-Net model was trained again using the same strategy, including the same selection of training dataset and testing dataset, U-Net-based ensemble model with K-fold cross-validation, as well as the parameter tuning process. The parameter tuning results are shown in Figure 22. The optimal weight of the new model that generates the highest AUC is also 0.9, the same as the model only using RGB. The wildebeest detection accuracy is presented in Table 10. The F1-score reaches 0.87, which is the same as the accuracy using only the RGB combination.

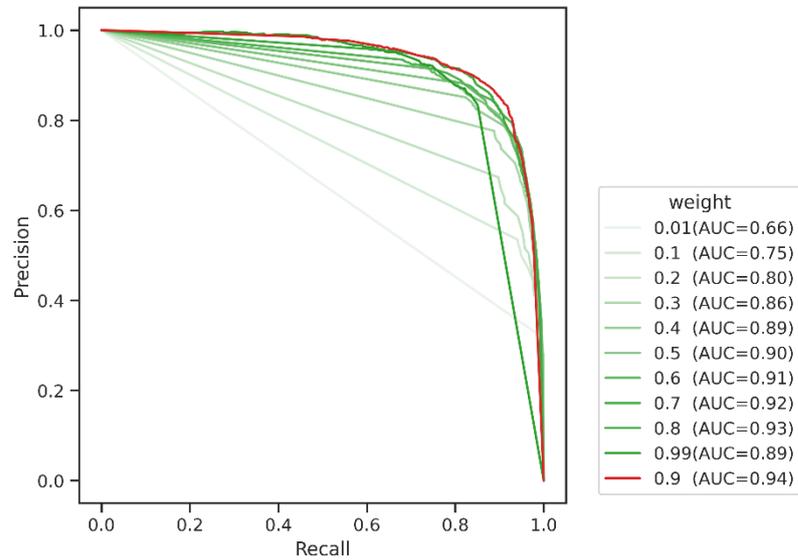


Figure 22. Precision-Recall curve of the U-Net model trained with different weights in Tversky loss using RGB+NIR bands. The AUC (Area Under precision-recall Curve) reached the highest when the weight is 0.9.

Table 10. The accuracy of wildebeest detection U-Net model trained with RGB+NIR GeoEye-1 satellite image

Weight β	Precision	Recall	F1-score	AUC
0.9	0.89	0.86	0.87	0.94

The 5×2 cross-validated paired t-test is implemented to test whether there is a significant difference between the wildebeest detection accuracy trained with RGB only and RGB+NIR of GeoEye-1 satellite image. The score differences are presented in Table 11.

Table 11. The accuracy differences between U-Net models trained with RGB+NIR and RGB band combinations of GeoEye-1 satellite image using 5×2 cross-validation

Iteration	Fold	Accuracy difference
1	1	0.06
	2	-0.01
2	1	0.07
	2	0.03
3	1	-0.03

	2	-0.03
4	1	-0.05
	2	-0.01
5	1	-0.01
	2	-0.01

With respect to the normality of the data distribution of the accuracy differences, the Shapiro-Wilk test generates the p-value of 0.17, which is larger than the α level of 0.05. Therefore, the null hypothesis cannot be rejected. Thus, the data follows a normal distribution.

The results of the variance equality test of the accuracy samples show that the p-value is 0.75, which is larger than the α level of 0.05. This indicates that the variances of both accuracy samples are homogeneous.

Regarding the results of paired t-test, the calculated t-statistic is less than the t-critical value at the significance level of 0.05 ($2.17 < 2.571$), therefore, the null hypothesis (there is no significant difference between the accuracy of these two models) cannot be rejected. There is no statistically significant difference in the wildebeest detection accuracy between the use and without the use of the near-infrared band of GeoEye-1 image.

3.4. Spatial transferability

In the highland area with different landscapes, the model shows the capability of avoiding false positives. As Figure 23.e and f present, the model is able to correctly ignore the wildebeest-like objects on the ground, such as the shadow of soil on the fallow field, and very small bushes that can be easily confused with wildebeest.

The model also detects some of the animals. For example, in Figure 23.a, b, and c, the distinct objects on the agricultural field and open forest are groups of animals, which are, according to the local expert, highly probable to be cattle or a mixture of cattle and wildebeest. The model can detect the black cattle that are very similar to wildebeest, while the cattle with white or light brown pixels are left out such as the example in Figure 23.a, b, and d.

a



b



c



○ Detected animal



◦ Detected animal

Figure 23. Transferring the wildebeest detection model to the highland area on GeoEye-1 satellite image in 2009. The left column shows the raw satellite images and the right column displays the animal detection results on the image. **a**, example patch 1 of open grassland with animals. **b**, example patch 2 of open grassland with animals. **c**, example patch 1 of the agricultural field with animals. **d**, example patch 2 of the agricultural field with animals. **e**, example patch of human settlements and agricultural fallow field without animals. **f**, example patch with bushes.

3.5. Temporal transferability

Despite the difference of spectral characteristics on GeoEye-1 satellite image in 2013, the model trained only on the satellite image in 2009 is still capable to discriminate wildebeest from complicated background objects, especially small bushes (Figure 24.a and b).

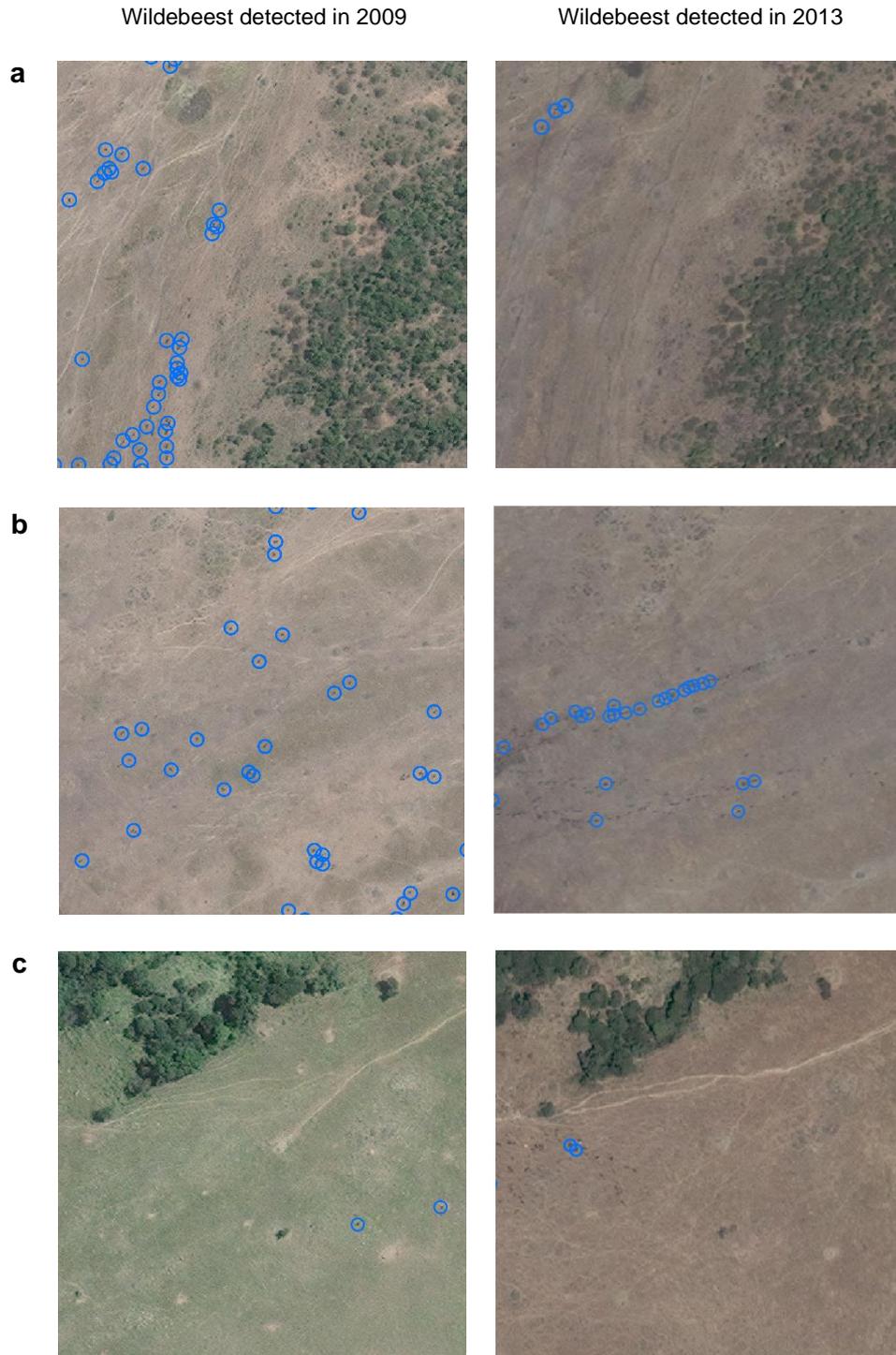


Figure 24. Detecting wildebeest on GeoEye-1 satellite image in 2013 (the plain area) using U-Net model trained with data in 2009. **a**, example patch with bushes. **b**, an example patch with linearly grouped wildebeest. **c**, an example patch with forest.

Animal detected in 2009

Animal detected in 2013



Figure 25. Detecting animals on GeoEye-1 satellite image in 2013 (the highland area) using U-Net model trained with data in 2009. **d**, **e**, **f**, example patches on the highland area.

4. DISCUSSION

4.1. Wildebeest detection using U-Net deep learning model

This research presents a successful model based on the U-Net deep learning architecture for detecting and counting wildebeest from GeoEye-1 satellite images. The most important breakthrough is the capability of detecting small animal targets with indistinct features across heterogeneous environment on sub-meter level resolution images at a large scale, a task that is empirically considered to be challenging (Hollings et al., 2018). This U-Net model can detect wildebeest with less than 9 pixels on a 43cm-resolution image, while these 9 pixels also contains a mixture of animal and background pixels. On the contrary, other commonly used object detection deep learning models are restricted either by the resolution of images or the size of animals, such as animal detection on aerial photographs or elephant and whale detection on satellite images (Duporge et al., 2020; Guirado et al., 2019; Peng et al., 2020). The generalization accuracy F1-score of the U-Net model on the independent testing dataset distributed on a large scale (area of 140 km²) reaches 0.87, while similar research on wildebeest detection in pilot areas with a much smaller scale (area of 0.04 km²) only achieves an accuracy of 0.79 (Xue et al., 2017). In addition, the U-Net model training and detection can be implemented fully automated online. It does not require much manual intervention like traditional semi-automated algorithms or expert-knowledge dependent object-based classification approaches (Kellenberger, Marcos, & Tuia, 2018c; Xue et al., 2017; Yang et al., 2014). It is thus proved that this U-Net deep learning model is more advanced in accurate and rapid animal detection and counting from very-high-resolution satellite imagery at larger scales.

The advantage of this wildebeest detection model is also reflected in the precise prediction of the location of the wildebeest individuals. The vectorized wildebeest predictions extracted from the map allow straightforward population counting. This counting method is more accurate and simple compared to research that estimates animal populations by performing regression (Gonçalves et al., 2020). It also benefits the wildebeest-level accuracy assessment, which is more intuitive than pixel-level assessment and consistent to commonly used accuracy metrics (precision, recall, and f-score) in object detection tasks. Moreover, the pixel-by-pixel classification of the U-Net model enables localization of the animal on the image, which makes it possible to analyze the spatial distribution of wildebeest like understanding the animal migration pattern, and further explore the response of animal movement to other environmental or human factors such as food resources (S. A. R. Mduma, Sinclair, & Hilborn, 1999). The automated wildebeest detection method can efficiently provide animal location data for such ecological research, which previously required laborious manual labelling work (Hughey et al., 2021). Therefore, the approach in this research will constitute a robust basis for wildlife ecological research and wildlife location and population monitoring that entails fine-scale spatial distribution data.

In this study, the performance of this wildebeest detection model is highly dependent on these following foundations: the quality and quantity of the training dataset, the weighted loss function, and the utilization of the ensemble model. First, in a heterogeneous environment like the savannahs, it is extremely important to select adequate non-target samples of background objects that cause misclassifications, such as termite mounds, riverbank shadows, dwarf shrub, the shadow of vehicles. The model needs to “see” the non-animal pixels so it can learn to avoid the possible large amounts of false positives. In short, the training dataset should include the variations of not only the detection targets but also the background landscape as completely as possible until the model achieves a good generalization ability. Another key issue is the imbalance of the target objects and background. The model tends to classify all the pixels into non-wildebeest pixels to achieve high overall accuracy. But by using weighted loss functions, such as Tversky

loss and Focal loss, we are able to force the model to emphasize learning the targets when the target pixels occupy a very small proportion (Lin et al., 2017; Salehi et al., 2017). In this wildebeest detection study where the wildebeest pixels are less than 1%, adopting the Tversky loss function increased the accuracy (F1-score) from 0.69 to 0.87 after weight parameter tuning. Moreover, the animal detection model performance is sensitive to randomness during the process of model training. The result slightly varies when running the model training procedure with exactly the same configurations. The utilization of the ensemble model can overcome this issue and improve the stability and robustness of the model. The ensemble model minimizes the chance of selecting one single model with poor performance on certain landscapes and achieves a better balance between false positives and false negatives.

There are still several limitations to the wildebeest detection model. This study is restricted within the 140 km² Mara Triangle area and the wildebeest detection at a larger scale is not explored yet due to the lack of data. Besides, the uncertainty of the model still exists in several steps: batch splitting in model training, weight initialization, and k-fold cross-validation splitting. The uncertainty has been reduced by the U-Net-based ensemble model approach; however, it is still inevitable to get slightly different results when running the model multiple times, especially the Precision and Recall. But the fluctuations of both metrics are within an acceptable range, and the F1-score hardly changes (Figure 21). The model is also sensitive to some parameters, including the searching distance, the threshold, and number k in the K-means clustering method for data vectorization. Sensitivity analysis is required to achieve optimal performance as presented in this research.

4.2. Spatial and temporal transferability of the wildebeest detection model

The transferability test results suggest the potential of the wildebeest detection model to transfer over space and time. Normally transfer learning needs more training data to adapt to the new variations of the new study image, however, this is not possible for this new temporal image due to the scarcity of potential wildebeest training samples. The spatial transferability cannot be quantitatively evaluated either as there is a limited number of wildebeests on the highland. Despite these limitations, it is clear that the model can still locate the animal clusters even though not all the individual wildebeest are correctly detected. This indicates its potential to apply to more spatially and temporally different satellite images and achieve better generalization ability when more training data are available.

Future study could upscale the model to a larger area (i.e., the whole 25000 km² Serengeti-Mara ecosystem) to build a massive wildebeest detection dataset to count the full population of wildebeest, and monitor the population fluctuations by leveraging multi-temporal satellite images in this area (Thirgood et al., 2004). It would be revolutionary for large-scale wildlife monitoring and management practices, especially when comparing to traditional ground surveys, like the Mara wildlife count in Mara ecosystems that involved 22 vehicle counting teams and 3 aircraft observing teams with 84 people (Reid et al., 2003). It is promising to establish a comprehensive wildlife spatial distribution database built on very high-resolution satellite images coupled with this advanced deep learning animal detection technique.

The unexpected detection of cattle on the highland points out the possibility of detecting other small animal targets that have similar characteristics to the wildebeest. This also poses a challenge to the model toward discriminating cattle from wildebeest since the resolution of satellite imagery is very limited. Multiple species that have similar characteristics with indistinct features cannot be differentiated by this model, and it is beyond the scope of this animal detection model in the current stage.

Regarding this limitation, it could be argued that expert knowledge can assist in refining the spatial and temporal boundaries of the data. If the local people and ecologists know that the dominant animal species is wildebeest within this specific area at a specific time, we can confidently say that the detection and counting results are still close to reality.

4.3. The role of the near-infrared band in the detection of wildebeest

The results demonstrate that adding the near-infrared band to the RGB combination in the GeoEye-1 satellite image does not significantly improve the wildebeest detection accuracy. This is the opposite of our hypothesis, where the NIR band is more sensitive to vegetation and thus should be useful to discriminate vegetation from the wildebeest (Knipling, 1970). In fact, in this research, the number of pixels for each wildebeest is very limited (less than 9). Among these 9 pixels, some are the shadow of the animal and background land, which makes it more complicated to distinguish animals from vegetation. According to our preliminary results where the problem of false positives was not yet solved, it was also observed that the confusion is mainly attributed to the shadow pixels of the vegetation rather than the “true” vegetation pixels. But, simply adding the NIR band may not add the presumed variations to the shadow pixels and thus does not contribute to distinguishing the wildebeests from the vegetation. Another reason could be associated with the classifier. The key to convolutional neural networks is the extraction of spatial features such as low-level edges and high-level shapes by computing the receptive field. This finding indicates that the RGB band combination provides good enough information on top of the spatial information for animal detection. Therefore, the wildebeest detection model can be applied to the commercial satellite imagery products of different levels, and the cost of data acquirement can be reduced since the NIR band is not necessary.

However, for animal detection on higher-resolution images (such as wildebeest on aerial photographs), the vegetation and animals will show different characteristics from those on satellite images, and more details that can help separate them will be available. Hence, the distinguishment of vegetation and animals could benefit from adding an extra NIR band. It still needs further research to investigate the role of the near-infrared band with various data and environment settings.

4.4. Applicability of the wildebeest detection model to other animal species

This study built a U-Net-based deep learning model for detecting and counting wildebeest from GeoEye-1 satellite images and proved the feasibility of detecting small animals with high accuracy. Nevertheless, to deploy the model on detecting other small-size feature indistinct animal species, there are still some potential challenges.

The first issue is data imbalance caused by the low proportion of animal samples. As discussed before, a large amount of training data representing all the data variations is required by the U-Net deep learning model. In this research, there is a large number of wildebeests migrating in this area, thereby allowing adequate dataset sampling for model training. But for other species or wildebeest itself with a small population on the image shot, the limited sample size may pose a challenge to the model. It can be expected that animal with a small population size also contains a lower level of variations of the characteristics across the background, so a smaller training dataset may not hinder the model performance. However, it is still not investigated that how the model will behave when the whole animal population on one satellite image is

limited. Besides, to deal with the data imbalance, the parameter in the weighted loss function also needs to be re-tuned to adapt to the new proportion of animal and non-animal pixels in the dataset.

The second potential challenge is posed by the image resolution. This model has been only applied to the animal with a size of 4-9 pixels in total on the satellite image. Therefore, we recommend deploying the model to detect animals with a similar or even larger size on the remote sensing image, such as around 1-meter-long animal on WorldView-3 satellite image (31-cm resolution), or at least 1.2-meter-long animal on GeoEye-1 satellite image (43-cm resolution). If the animal is smaller than this size, then a higher image resolution could be necessary to generate reasonable results. When the image resolution is restricted, then the small-size animal may not be able to be distinguished from the image. Future research could investigate the model performance on smaller-size (2×2 pixels) animal detection and counting using the U-Net-based ensemble model.

The data resolution also affects the confidence of ground truth labelling. It is important to note that the wildebeest labelling in this study is conducted using a temporally different satellite image as a reference, and the final labels are the agreement of three observers, thus improving the confidence of the ground truth quality when the features of animals are not distinct on the image and the contrast to the background is low. Under the condition of small animal size with a low image resolution (less than 6 pixels in length), even visual interpretations by humans can be contradictory and bring uncertainty to the model (Bowler et al., 2020b). Therefore, it would be crucial to use expert knowledge to refine the ground truth labels of the animal targets.

Considering animal detection on images with higher resolution, such as aerial images, the U-Net model as the widely used image segmentation method in computer vision, can be expected to have good performance as well, especially when the higher resolution introduces more clear detailed features. But thus far, the more commonly used method in animal detection from high-resolution imagery is object-based deep learning architectures, such as YOLO, Mask R-CNN (Torney et al., 2019; Xu et al., 2020). There is one study using U-Net and Inception-v4 regression network for livestock detection on aerial images, but the background environment is rather homogeneous (Han et al., 2019). In future work, it would be useful to compare the U-Net model and other object-detection models in terms of the performance on aerial images with the heterogeneous environment, and thus build a solid work pipeline for animal detection.

5. CONCLUSIONS

In this study, we tested the ability of the U-Net deep convolutional neural networks for detecting and counting the number of migrating wildebeest in the Mara Triangle from GeoEye-1 satellite imagery. To determine the role of the near-infrared band on the detection accuracy of wildebeest, the model performance using RGB and RGB+NIR band combination were compared using a 5×2 cross-validation paired t-test. The transferability of the U-Net model over space and time was evaluated by applying it to a spatially different area with varied landscapes and a temporally different GeoEye-1 satellite image, respectively. Based on the results, the following conclusions can be drawn:

- 1) The U-Net model is capable of detecting small animal targets (2-4 pixels in length) with indistinct features on sub-meter level resolution images, and it achieves a high generalization accuracy (F1-score) of 0.87 on the independent testing dataset distributed on a large-scale area (area of 140 km²).
- 2) Adding the near-infrared band to the RGB combination in the GeoEye-1 satellite image does not significantly improve the wildebeest detection accuracy.
- 3) The transferability test results suggest the potential of the wildebeest detection model to transfer over space and time. The wildebeest detection model can rapidly locate the animal clusters on the spatially and temporally different satellite images.

In conclusion, this study built an effective and efficient U-Net deep convolutional neural network model for accurate and rapid wildebeest detection and counting from GeoEye-1 satellite imagery. The model also demonstrated its potential to be upscaled for monitoring and counting the whole wildebeest population on the Serengeti-Mara ecosystem when more high-resolution satellite images are available. It would be promising to establish a comprehensive wildlife spatial-temporal distribution database built on multi-temporal very-high-resolution satellite images coupled with this deep learning animal detection technique, thereby allowing for efficient monitoring of wildebeest population dynamics in the whole Serengeti-Mara ecosystems. In addition, the wildebeest locating approach in this research enables straightforward and accurate animal population counting. The spatially precise wildebeest location dataset produced by this model will constitute a robust basis for a wide range of fine-scale spatial distribution-related research, such as animal migration pattern analysis.

LIST OF REFERENCES

- Allen, L., Holland, K. K., Holland, H., Tome, S., Nabaala, M., Seno, S., & Nampushi, J. (2019). Expanding Staff Voice in Protected Area Management Effectiveness Assessments within Kenya's Maasai Mara National Reserve. *Environmental Management*, 63(1), 46–59. <https://doi.org/10.1007/s00267-018-1122-6>
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Barber-Meyer, S. M., Kooyman, G. L., & Ponganis, P. J. (2007). Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology*, 30(12), 1565–1570. <https://doi.org/10.1007/s00300-007-0317-8>
- Ben-Shahar, R., & Coe, M. J. (1992). The relationships between soil factors, grass nutrients and the foraging behaviour of wildebeest and zebra. *Oecologia*, 90(3), 422–428. <https://doi.org/10.1007/BF00317701>
- Betts, M. G., Mitchell, D., Diamond, A. W., & Bety, J. (2007). Uneven Rates of Landscape Change as a Source of Bias in Roadside Wildlife Surveys. *Journal of Wildlife Management*. <https://doi.org/10.2193/2006-004>
- Bowler, E., Fretwell, P. T., French, G., & Mackiewicz, M. (2020). Using deep learning to count albatrosses from space: Assessing results in light of ground truth uncertainty. *Remote Sensing*. <https://doi.org/10.3390/rs12122026>
- Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: A survey and categorisation. *Information Fusion*, 6(1), 5–20. <https://doi.org/10.1016/j.inffus.2004.04.004>
- Christiansen, P. (2004). Body size in proboscideans, with notes on elephant metabolism. *Zoological Journal of the Linnean Society*, 140(4), 523–549. <https://doi.org/10.1111/j.1096-3642.2004.00113.x>
- Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *The Journal of Machine Learning Research*, 7, 1–30.
- Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 10(7), 1895–1923. <https://doi.org/10.1162/089976698300017197>
- Duporge, I., Isupova, O., Reece, S., Macdonald, D. W., & Wang, T. (2020). Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*. <https://doi.org/10.1002/rse2.195>
- Eikelboom, J. A. J., Wind, J., van de Ven, E., Kenana, L. M., Schroder, B., de Knegt, H. J., ... Prins, H. H. T. (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, 10(11), 1875–1887. <https://doi.org/10.1111/2041-210X.13277>
- Eva, H., & Lambin, E. F. (2000). Fires and land-cover change in the tropics: A remote sensing analysis at the landscape scale. *Journal of Biogeography*, 27(3), 765–776. <https://doi.org/10.1046/j.1365-2699.2000.00441.x>
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2015.169>
- Goebel, M. E., Perryman, W. L., Hinke, J. T., Krause, D. J., Hann, N. A., Gardner, S., & LeRoi, D. J. (2015). A small unmanned aerial system for estimating abundance and size of Antarctic predators. *Polar Biology*. <https://doi.org/10.1007/s00300-014-1625-4>
- Gonçalves, B. C., Spitzbart, B., & Lynch, H. J. (2020). SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, 239, 111617. <https://doi.org/10.1016/j.rse.2019.111617>
- Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-50795-9>
- Han, L., Tao, P., & Martin, R. R. (2019). Livestock detection in aerial images using a fully convolutional network. *Computational Visual Media*, 5(2), 221–228. <https://doi.org/10.1007/s41095-019-0132-5>
- Hansen, L. K., & Salamon, P. (1990). Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993–1001. <https://doi.org/10.1109/34.58871>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer*

- Vision*, 1026–1034.
- Hodgson, J. C., Baylis, S. M., Mott, R., Herrod, A., & Clarke, R. H. (2016). Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports*. <https://doi.org/10.1038/srep22574>
- Holdo, R. M., Holt, R. D., Sinclair, A. R. E., Godley, B. J., & Thirgood, S. (2013). Migration impacts on communities and ecosystems: empirical evidence and theoretical insights. In *Animal Migration* (pp. 130–143). <https://doi.org/10.1093/acprof:oso/9780199568994.003.0009>
- Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T., & Robinson, A. (2018). How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12973>
- Hopcraft, J. G. C., Sinclair, A. R. E., Holdo, R. M., Mwangomo, E., Mduma, S., Thirgood, S., ... Olf, H. (2015). Why are wildebeest the most abundant herbivore in the Serengeti ecosystem? In *Serengeti IV: Sustaining Biodiversity in a Coupled Human-Natural System* (pp. 4–8). University of Chicago Press.
- Hughey, L. F., Shoemaker, K. T., Stewart, K. M., McCauley, D. J., & Cushman, J. H. (2021). Effects of human-altered landscapes on a reintroduced ungulate: Patterns of habitat selection at the rangeland-wildland interface. *Biological Conservation*, 257, 109086. <https://doi.org/10.1016/j.biocon.2021.109086>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015, 1*, 448–456. Retrieved from <https://arxiv.org/abs/1502.03167v3>
- IPBES. (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In S. Díaz, J. Settele, E. S. B. E.S., H. T. Ngo, M. Guèze, J. Agard, ... C. N. Zayas (Eds.), *IPBES secretariat*. Bonn, Germany.
- Kellenberger, B., Marcos, D., & Tuia, D. (2018a). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, 216(December 2017), 139–153. <https://doi.org/10.1016/j.rse.2018.06.028>
- Kellenberger, B., Marcos, D., & Tuia, D. (2018b). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2018.06.028>
- Kellenberger, B., Marcos, D., & Tuia, D. (2018c). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, 216(June), 139–153. <https://doi.org/10.1016/j.rse.2018.06.028>
- Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Retrieved from <https://arxiv.org/abs/1412.6980v9>
- Kiszka, J. J., Mourier, J., Gastrich, K., & Heithaus, M. R. (2016). Using unmanned aerial vehicles (UAVs) to investigate shark and ray densities in a shallow coral lagoon. *Marine Ecology Progress Series*. <https://doi.org/10.3354/meps11945>
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1(3), 155–159. [https://doi.org/10.1016/S0034-4257\(70\)80021-9](https://doi.org/10.1016/S0034-4257(70)80021-9)
- LaRue, M. A., Rotella, J. J., Garrott, R. A., Siniff, D. B., Ainley, D. G., Stauffer, G. E., ... Morin, P. J. (2011). Satellite imagery can be used to detect variation in abundance of Weddell seals (*Leptonychotes weddellii*) in Erebus Bay, Antarctica. *Polar Biology*, 34(11), 1727–1737. <https://doi.org/10.1007/s00300-011-1023-0>
- Larue, M. A., Stapleton, S., Porter, C., Atkinson, S., Atwood, T., Dyck, M., & Lecomte, N. (2015). Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution. *Wildlife Society Bulletin*. <https://doi.org/10.1002/wsb.596>
- Leslie, M., Perkins-Taylor, C., Durban, J., Moore, M., Miller, C., Chanarat, P., ... Apprill, A. (2020). Body size data collected non-invasively from drone images indicate a morphologically distinct Chilean blue whale (*Balaenoptera musculus*) taxon. *Endangered Species Research*, 43, 291–304. <https://doi.org/10.3354/esr01066>
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal Loss for Dense Object Detection*.
- Longmore, S. N., Collins, R. P., Pfeifer, S., Fox, S. E., Mulero-Pázmány, M., Bezombes, F., ... Wich, S. A. (2017). Adapting astronomical source detection software to help detect animals in thermal images obtained by unmanned aerial systems. *International Journal of Remote Sensing*, 38(8–10), 2623–2638. <https://doi.org/10.1080/01431161.2017.1280639>
- Mduma, H., Musyoki, C., Maliti Kyale, D., Nindi, S., Hamza, K., Ndeti, R., & Machoke M, Kimutai D,

- Muteti D, Maloba M, Bakari S, K. E. (2014). *Aerial Total Count of Elephants and Buffaloes in the Serengeti-Mara Ecosystem*.
- Mduma, S. A. R., Sinclair, A. R. E., & Hilborn, R. (1999). Food regulates the Serengeti wildebeest: A 40-year record. *Journal of Animal Ecology*, 68(6), 1101–1122. <https://doi.org/10.1046/j.1365-2656.1999.00352.x>
- New biodiversity targets cannot afford to fail. (2020). *Nature*, 578(February), 337–338.
- Ogutu, J. O., Owen-Smith, N., Piepho, H. P., & Said, M. Y. (2011). Continuing wildlife population declines and range contraction in the Mara region of Kenya during 1977–2009. *Journal of Zoology*, 285(2), 99–109. <https://doi.org/10.1111/j.1469-7998.2011.00818.x>
- Onyeanusi, A. E. (1989). Large herbivore grass offtake in Masai Mara National Reserve: implications for the Serengeti-Mara migrants. *Journal of Arid Environments*, 16(2), 203–209. [https://doi.org/10.1016/s0140-1963\(18\)31027-9](https://doi.org/10.1016/s0140-1963(18)31027-9)
- Owen-Smith, N., & Mills, M. G. L. (2008). Predator-prey size relationships in an African large-mammal food web. *Journal of Animal Ecology*, 77(1), 173–183. <https://doi.org/10.1111/j.1365-2656.2007.01314.x>
- Peng, J., Wang, D., Liao, X., & Shao, Q. (2020). Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169(October), 364–376. <https://doi.org/10.1016/j.isprsjprs.2020.08.026>
- Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., & Ye, H. (2020). Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 364–376. <https://doi.org/10.1016/j.isprsjprs.2020.08.026>
- Pham, M. T., Courtrai, L., Friguier, C., Lefèvre, S., & Baussard, A. (2020). YOLO-fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sensing*, 12(15), 1–26. <https://doi.org/10.3390/RS12152501>
- Polikar, R. (2012). Ensemble Learning. In *Ensemble Machine Learning* (pp. 1–34). https://doi.org/10.1007/978-1-4419-9326-7_1
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2016.91>
- Reid, R. S., Rainy, M., Ogutu, J., Kruska, R. L., Kimani, K., Nyabenge, M., ... Lamprey, R. (2003). *People, wildlife and livestock in the Mara Ecosystem: the Mara Count 2002*. Retrieved from <https://www.researchgate.net/publication/266852551>
- Ronneberger, O., Fischer, P., & Brox, T. (2015a). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-24574-4_28
- Ronneberger, O., Fischer, P., & Brox, T. (2015b). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-24574-4_28
- Sahu, R. (2019). Detecting and Counting Small Animal Species Using Drone Imagery by Applying Deep Learning. In *Visual Object Tracking with Deep Neural Networks* (pp. 1–14). <https://doi.org/10.5772/intechopen.88437>
- Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-67389-9_44
- Secretariat of the Convention on Biological Diversity. (2020). Global Biodiversity Outlook 5: Summary for Policymakers. *Secretariat of the Convention on Biological Diversity*, 19. Retrieved from <https://www.cbd.int/gbo/gbo5/publication/gbo-5-spm-en.pdf>
- Serneels, S., & Lambin, E. F. (2002). Impact of land-use changes on the wildebeest migration in the northern part of the Serengeti-Mara ecosystem. *Journal of Biogeography*, 28(3), 391–407. <https://doi.org/10.1046/j.1365-2699.2001.00557.x>
- Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. In *Journal of Machine Learning Research* (Vol. 15). Retrieved from <http://jmlr.org/papers/v15/srivastava14a.html>
- Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C., & Atwood, T. (2014). Polar bears from space: Assessing satellite imagery as a tool to track arctic wildlife. *PLoS ONE*.

- <https://doi.org/10.1371/journal.pone.0101513>
- Subalusky, A. L., Dutton, C. L., Rosi, E. J., & Post, D. M. (2017). Annual mass drownings of the Serengeti wildebeest migration influence nutrient cycling and storage in the Mara River. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(29), 7647–7652. <https://doi.org/10.1073/pnas.1614778114>
- Talbot, L., & Talbot, M. (1963). Wildlife Monographs. In *The Wildebeest in Western Masailand, East Africa* (Vol. 12, pp. 1689–1699).
- Thirgood, S., Mosser, A., Tham, S., Hopcraft, G., Mwangomo, E., Mlengeya, T., ... Borner, M. (2004). Can parks protect migratory ungulates? The case of the Serengeti wildebeest. *Animal Conservation*, *7*(2), 113–120. <https://doi.org/10.1017/S1367943004001404>
- Torney, C. J., Lloyd-Jones, D. J., Chevallier, M., Moyer, D. C., Maliti, H. T., Mwita, M., ... Hopcraft, G. C. (2019). A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, *10*(6), 779–787. <https://doi.org/10.1111/2041-210X.13165>
- Wang, D., Shao, Q., & Yue, H. (2019). Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): A review. *Remote Sensing*, *11*(11). <https://doi.org/10.3390/rs11111308>
- Wishitemi, B. E. L., Momanyi, S. O., Ombati, B. G., & Okello, M. M. (2015). The link between poverty, environment and ecotourism development in areas adjacent to Maasai Mara and Amboseli protected areas, Kenya. *Tourism Management Perspectives*, *16*, 306–317. <https://doi.org/10.1016/j.tmp.2015.07.003>
- WWF. (2020). Living Planet Report 2020 - Bending the curve of biodiversity loss. In R. E. A. Almond, M. Grooten, & T. Petersen (Eds.), *WWF*. WWF, Gland, Switzerland.
- Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., ... Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, *171*. <https://doi.org/10.1016/j.compag.2020.105300>
- Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Sun, Z., & Li, C. (2020). Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *International Journal of Remote Sensing*, *41*(21), 8121–8142. <https://doi.org/10.1080/01431161.2020.1734245>
- Xue, Y., Wang, T., & Skidmore, A. K. (2017). Automatic counting of large mammals from very high resolution panchromatic satellite imagery. *Remote Sensing*, *9*(9), 1–16. <https://doi.org/10.3390/rs9090878>
- Yang, Z., Wang, T., Skidmore, A. K., De Leeuw, J., Said, M. Y., & Freer, J. (2014). Spotting East African mammals in open savannah from space. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0115989>