Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/30/20230066

Animal detection and classification from camera trap images using residual neural networks

B. Bizu[1], Sathishkumar V. E.[2,4], T. Kumaravel[1], Hari Prasath P.[1], Dharanish K. C.[1], A. S. Arun Prabu[1], N. Krishnamoorthy[3]

[1]Department of Computer Science and Engineering, Kongu Engineering College, Erode, Tamil Nadu, India
[2]Department of Software Engineering, Jeonbuk National University, Jeonju-si, Republic of Korea
[3]Department of Software Systems and Engineering, Vellore Institute of Technology, Vellore campus, Vellore, Tamil Nadu, India

[4]sathish@jbnu.ac.kr

Abstract. Using camera traps is common in animal studies. The camera is often activated when the movement is detected to prevent recording when nothing happens. It includes a collection of images of wildlife from Tanzania's Serengeti National Park. Deep Learning is built on an understanding of the composition and it is the working of behaviour like CPU of the computer. Deep learning model is mainly working as the basic principle of neural networks to analyse any inputs like data or images and videos and make better accurate with predicted value with less loss percentage. With current systems, wind and sunshine may potentially move the plants and start recording, leading to a large number of blank images. Researchers will manually eliminate them from the study, which is a hard way of classification by manually and very much wastage of time. When there is a lot of data accessible, the system has all it needs to train itself. Deep residual neural networks, such as ResNet50, which are very helpful for object detection of many image data and make more viable to the conservation of wildlife are used in this proposed system. It aids in determining if the provided picture data is of an animal or not with better prediction, as well as training on a useful dataset like Serengeti2, where camera trap image collection yields accuracy of 94.64% with better prediction of tested data with greater precision and recall value.

Keywords: deep learning, object detection, camera trapping, and animal identification.

1. Introduction

Camera traps are broadly utilized in natural life reviews and biodiversity checking. Occasionally, a more number of data instances like images or videos are generated, depending on its triggering mechanism. The use of residual neural networks in identifying the animal as object in given input image by camera setup in Tanzania forest image dataset like seregenti2 has been used in, which has the potential to significantly speed up analysis processes and reduce manual labours. The proposed algorithm is RESNET50. Surveys using camera traps may teach ecologists and wildlife conservationists a lot about the distribution of species richness, animal behavior, population density, community dynamics, and other topics. Camera traps are frequently used in monitoring biodiversity and managing animals due to

DOI: 10.54254/2755-2721/30/20230066

their low impact, excellent concealment, and 24-hour continuous operation. When animals pass by, a camera which is fixed in many places of the serengenti forest may trigger image clips for monitoring the movement of animals which is used for species classification. However, complex environments, such as animal crossings, pose a threat to camera traps as well resulting in false triggers and frequently resulting in a large number of images or videos devoid of wildlife.

2. Related work

Neil A. Gilbert et al. (2021) [2] argue that in order to make good decisions, precise details on species distributions, or the incidence and abundance patterns, is necessary for wildlife management. In the past, managers have depended on harvest records to monitor animal populations with poor resolution but large spatial scales. Wildlife movement and behavioral ecology parameters have been monitored using camera-trapping techniques, according to Pablo Palencia et al. (2021) [1]. However, when evaluating movement patterns to estimate the speed ratio must be taken into account in the formulation; otherwise, the outcomes would be skewed. For instance, some populations of wildlife exhibit patterns of migration. For example, Foraging or travelling across habitat patches are two behaviors that only occur in certain wildlife populations, and it will take more research to incorporate these behaviors into the estimation of movement parameters. According to David U. Hooper et al. (2012) [3], there is growing evidence that extinctions alter key processes that are crucial to the resilience and productivity of the ecosystems on Earth. It is unknown how these effects stack up against other environmental changes that have direct consequences for ecosystem function and biodiversity loss. However, future extinctions of species will hasten changes in ecological functions. Small rodents are an essential indication for understanding the influence that the rapidly changing winter climate has on Arctic tundra ecosystems, according to Eeva M. Soininen et al. (2021) [5]. However, throughout the lengthy Arctic winter, conventional trap monitoring of mouse populations has been impeded by snow cover and harsh environmental conditions.

3. Materials and methods

3.1. Datasets

The dataset contains features extracted from the Kaggle image dataset to predict the class of animals. I've used the data form Serengeti Dataset to get train and test model. The data comprises of camera trap images from Serengeti National Park in Tanzania. About 75% percent of images are blank. There are >1TB of images labeled by numerous volunteers in Zooniverse. The dataset consists of images from 4 locations: 3 of them are in the train set and 1 is in the validation set. Each location is represented by images from one roll (images taken to the next battery change). The images were labeled with species names, which I turned into non blank class. Blank class (no animals present) was remained. I've chosen the data so that the dataset will be balanced. I revised the dataset to ensure the labels are valid. About 40 problematic images were excluded (animals are very far or hardly visible). Testing dataset was used which is public for usage and it contains 112 images and it can be obtained from Kaggle site. They are indeed very valuable for the model, but they could be misleading for small datasets. Besides that I aimed to train proof of concept model, which can be a subject of further experiments. All features represent the class of animals. The underlying method image analysis using RESNET50 as our classification technique. The images which we used are 2381 images and especially 2269 images for training and 112 belongs to testing data which contains two classes which is blank and non blank representing animal as non blank and object with no animal image as blank and it is used for training the model like many deep learning models for better prediction with better accuracy.

DOI: 10.54254/2755-2721/30/20230066

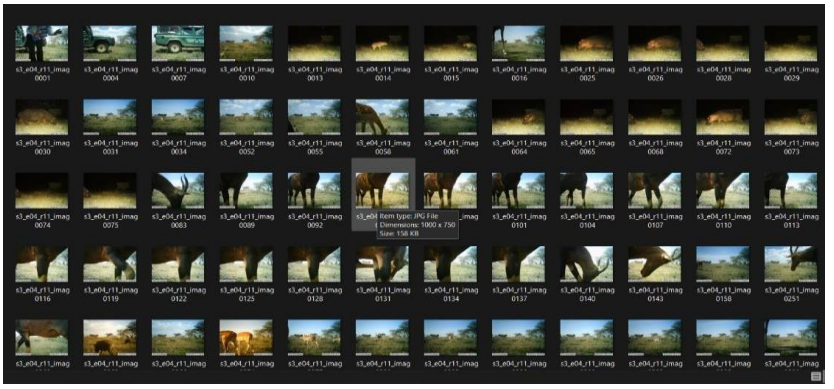

Figure 1. collection of sample datatset.

3.2. Deep learning

The main component of the machine learning subfield known as "deep learning" is a three- or more-layer neural network. With the help of these neural networks, it can "learn" from a lot of data and try to imitate how the human brain functions, though they are far from matching its capabilities. The accuracy of predictions made by a neural network with only one layer can be improved by adding additional hidden layers. Numerous artificial intelligence (AI) programmes and services enhance automation by physical and analytical tasks due to deep learning, without human involvement. Voice-activated television remotes, digital assistants, and fraud detection for credit cards are all examples of everyday products and services based on deep learning technology.

3.2.1. Fully connected neural network. The Feed forward with all connections. The most fundamental neural network applications all employ neural networks as their standard network design. It is made up of a number of seamlessly interwoven layers. Fully connected refers to the state in which every neurons in the layer above is linked with every neuron in layer below. Additionally, feed forward refers to the fact that neurons in any previous layer are never coupled to neurons in a layer above them. In a neural network, each neuron has an activation function that modifies its output in response to its input. Every input dimension affects every output dimension. The last layers of the neural network are typically the fully connected layers.

3.2.2. Cascaded residual convolutional neural network. A particular deep neural network architectures called a convolutional neural network (CNN) is made for specialized applications like image categorization. The plan of neurons in the visual cortex of the creature mind served as the model for CNNs. An input layer is a component of a CNN. However, for fundamental image processing, this input generally consists of a two-dimensional array of neuron which represent the picture's pixel values. It also has output layer, which generally consists of collection of one-dimensional output neuron. Convolution layers with sparse connections are combined in CNN to process the input images. They likewise incorporate down sampling levels called pooling layers to additionally limit the quantity of neurons expected in the network's succeeding layers.

3.2.3. Resnet-50's distinguishing features. The architecture of ResNet-50 is based on the concept shown above, with one significant exception. The bottleneck building block is used in the 50-layer ResNet model. A bottleneck residual block, sometimes referred to as a "bottleneck", employs 11 convolutions to cut down on the number of parameters and matrix multiplications. This makes each layer's training significantly faster. Instead of using a stack of two levels, it employs three layers. Figure 2 illustrates the architecture diagram of ResNet 50 with above mentioned layers.

DOI: 10.54254/2755-2721/30/20230066

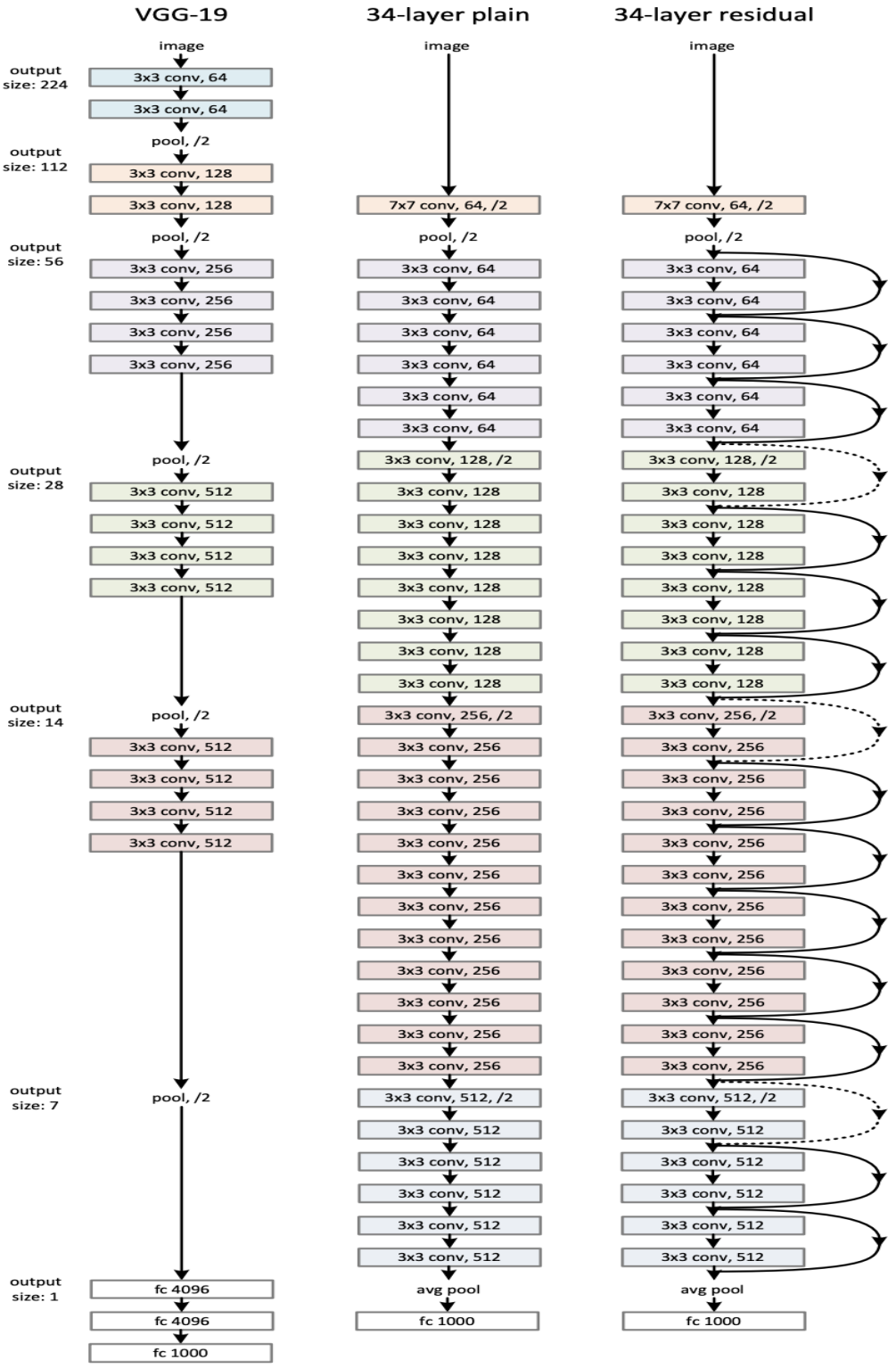

Figure 2. Resnet 50 architecture diagram.

4. Proposed models

In our project we are using the TensorFlow GPU as the requirements are satisfied we have we need a processor based on the certain modules or project is mainly based on the training and testing data set of

DOI: 10.54254/2755-2721/30/20230066

the wildlife detection where the image size will be fixed and the proposed algorithm of Resnet 50 is used this results the best output than the previous existing model which was used with yolo v5 algorithm, Fully Connected Neural Network, Residual Convoluted neural network compare to that our model will show the better result with high accuracy and better construction.

4.1. Training the dataset

Initially, we will use ResNet-50 to identify the training model, and the input shape will be determined by the image size and total number of trainable layers. Developing a model that generates the convolutional layer of the pulling layer, in addition to a few other models, will be utilized for the categorical cross entropy that the Adam optimizer used to improve accuracy. As can be seen, the accuracy will primarily be the parameter that we will use to display the results.

4.2. Testing the dataset

In testing, we have separated the few part of images from the original dataset for testing with two labels as Non blank and blank images. After trained the model , we have tested using two class labels for predicting the results of out trained model and collected the results using those images and plot that images with actual class as animal or Non blank with predicted results of correct classification .In our model testing places the major role to achieve the correct experimental identification by RESNET 50.As we have used this testing dataset is completely free for usage and it remains public and fetched from kaggle site.

4.3. Model execution

The training set and validation data, which correspond to the test set for epoch values, will be calculated when the model runs. A deep convolutional neural network with 50 layers is called ResNet-50, where the highest accuracy can be plotted for each value that will be calculated for validation loss, training loss, and training accuracy. It will load a pretrained version of the network trained on the ImageNet database more than one million images. Images, including many animals, can be classified into one of 1,000 object categories by the pretrained network. Thus, the organization has gotten rich element portrayals for a scope of pictures. Figure 3 [6] illustrates the flow diagram of the model.

DOI: 10.54254/2755-2721/30/20230066

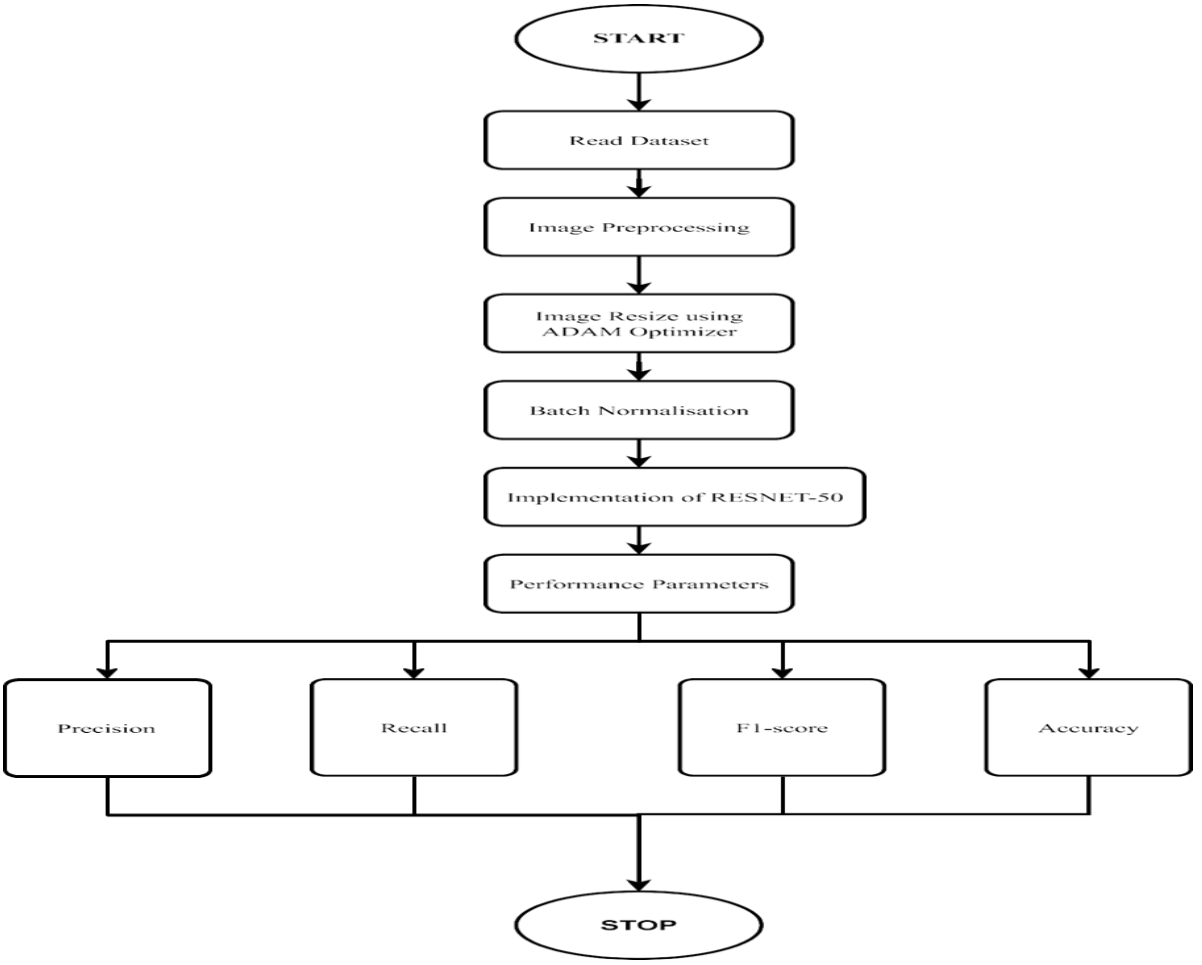

Figure 3. Flow diagram.

5. Result and discussion

In our classification model there are 2381 images that we have used for our paper and which were 2269 instances of images used for training the algorithm, 112 instances of images which are used for testing pretrained saved network for better prediction. The accuracy and loss for the YOLO V5 Model's training and validation datasets are shown in Table 1. With a learning rate of 0.01 across 25 epochs, this model was trained. Table 1 displays the YOLO V5 Model's accuracy and the loss for the training and the validation datasets. With a learning rate of roughly 0.01, this model performed 25 iterations of training, yielding a training accuracy of 93.18%. The table demonstrates that when training accuracy increases, the validation loss will rise relative to the first epoch.

Table 1. YOLO V5 model-performance.

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	1.7775	0.3370	1.1633	0.4817
5	0.6377	0.7316	0.6344	0.7429
10	0.3709	0.8349	0.4430	0.8138
15	0.2837	0.8725	0.3663	0.8586
20	0.1825	0.9281	0.3658	0.8663
25	0.1385	0.9318	0.3077	0.8912

DOI: 10.54254/2755-2721/30/20230066

Table 2 displays how well the Resnet-50 model performed. The model's accuracy was tested using the test dataset, and it came out at 88.36%. With 25 epochs and a learning rate for 25 epochs is 0.01, the RESNET50 model fared better in contrast, achieving a training accuracy of 93.58%. On the test dataset, the test accuracy for the model was 94.64%.

Table 2. Resnet50 -Performance.

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	1.872	0.6809	0.6390	0.7569
5	0.1931	0.9185	0.3484	0.8692
10	0.0905	0.9210	0.3785	0.8823
15	0.0336	0.9351	0.3461	0.9366
20	0.0378	0.9358	0.3211	0.9464

Table 3 displays the InceptionV3 model's performance. It has a training accuracy of 91.20% and was trained for 25 epochs with a learning rate for respective epochs is 0.01. The model's testing precision was 88.93%.

Table 3. Inception v3-performance.

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.986	0.633	0.510	0.784
5	0.246	0.800	0.301	0.888
10	0.097	0.861	0.294	0.904
15	0.041	0.886	0.358	0.908
20	0.041	0.912	0.367	0.903

Table 4 displays the comparison of model's performance between Inception V3 and RESNET 50 .It has testing accuracy of both image classification model.

Table 4. Increase in performance comparison.

Models evaluted	Accuracy (%)
Inceptionv3	92.02
Resnet 50	94.64

6. Conclusion and future work

The model that was utilized to train the dataset reveals that our model offers superior accuracy, recall, f1 score, and support compared to the earlier models that were applied to the current system. Consequently, 94.64% accuracy is the maximum level of accuracy. In comparison to the ResNet50 model, which has been trained for greater accuracy which is illustrated in figure 4 [8], earlier models like RCNN, ResNet34, and Fully Connected Neural Networks with HRNet32 performed very low among all other algorithms. Profound learning provides biologists with a ton of commitment as biology joins the universe of large information. Regardless of whether it is trying for the calculation to accomplish 100 percent precision, the innovation will rapidly help scientists and actually remove data from gigantic measures of information while likewise lessening the manual recognizable proof exertion. In the future, ecological research and conservation will benefit even more from deep interdisciplinary collaboration. Interdisciplinary collaboration will help to advance technical innovation in ecological

DOI: 10.54254/2755-2721/30/20230066

study and conservation in the future. In order to categorize animals as belonging to certain species for wildlife conservation and beneficial for ecologist study and discovering the extinction of species in that particular gathered dataset in the future, the dataset size may be increased by utilizing additional animal photographs. It can be obtained as more number of species for further classification of extinction of particular species and by our training algorithm we can identify the animal in any number of images and used for prediction of extinction species with particular classes of animal in future.

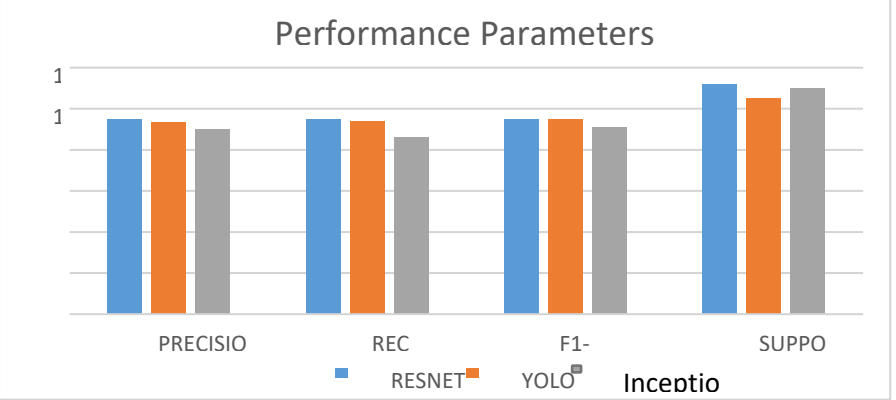

Figure 4. Performance analysis.

References

[1] Palencia, P.; Fernández-López, J.; Vicente, J.; Acevedo, P. Innovations in movement and behavioural ecology from camera traps: Day range as model parameter. Methods Ecol. Evol. 2021, 12, 1201–1212.

[2] Gilbert, N.A.; Pease, B.S.; AnhaltDepies, C.M.; Clare, J.D.; Stenglein, J.L.; Townsend, P.A.; Van Deelen, T.R.; Zuckerberg, B. Integrating harvest and camera trap data in species distribution models. Biol. Conserv. 2021, 258, 109147.

[3] Hooper, D.U.; Adair, E.C.; Cardinale, B.J.; Byrnes, J.E.; Hungate, B.A.; Matulich, K.L.; Gonzalez, A.; Duffy, J.E.; Gamfeldt, L.; O'Connor, M.I. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 2012, 486, 105–108.

[4] Almond, R.E.; Grooten, M.; Peterson, T. Living Planet Report 2020-Bending the Curve of Biodiversity Loss; World Wildlife Fund: Washington, DC, USA, 2020.

[5] Mölle, J.P.; Kleiven, E.F.; Ims, R.A.; Soininen, E.M. Using subnivean camera traps to study Arctic small mammal community dynamics during winter. Arct. Sci. 2021, 8, 183– 199.

[6] Anderson, C.B. Biodiversity monitoring, earth observations and the ecology of scale. Ecol. Lett. 2018, 21, 1572– 1585.