# Reindeer Recognition and Counting System Based on Aerial Images and Convolutional Neural Networks

Vladimir Mikhailov
St. Petersburg Federal Research Centre
Russian Academy of Sciences
Saint Petersburg, Russia
mwwcari@gmail. com

Vladislav Sobolevskii
St. Petersburg Federal Research Centre
Russian Academy of Sciences
Saint Petersburg, Russia
arguzd@yandex. ru

***Abstract.* Described animal recognition and counting system based on convolutional neural networks with MRCNN architecture. Initial training of the network is performed using a basic array of MS COCO images, and additional training is performed using an array of aerial photographs of reindeer herds. A web-interface of the system has been developed. The error of counting reindeer in the image from the verification sample is about 13%.**

***Key words:* reindeer, recognition, convolutional neural networks, aerial imagery**

## I. INTRODUCTION

The development of an automatic system for recognition and counting the number of reindeers was driven by the following reasons. The currently used methodology for counting the number of wild reindeer in tundra populations (Taimyr, Yakutia, Chukotka reindeer, and migrating herds of reindeer from Canada and Alaska) is based on the ecological characteristics of the species, consist in the fact that in hot weather, during the flight of blood-sucking insects, reindeer gather in herds of many thousands in a limited area in the northern part of their summer range (subzones of arctic and typical tundra) [1, 2]. The herds in the aggregations are photographed and the number of animals in them is counted directly, "by head".

The advantage of this approach over purely approximation-based population estimation methods [3] is the significantly greater accuracy of the results, as the vast majority of animals in the population (up to 90%) are counted directly from herd photographs and only a small number are estimated by area-based approximation. However, manual processing of the survey results for large populations takes about three months, whereas for the ecologically based management of population dynamics, the non-depletion of biological resources of the species and the determination of the norms of commercial reindeer harvesting, it is desirable to have the population data in the second half of August, that is, 10-15 days after the end of the aerial count.

The task was therefore to automate the processing of aerial photographs in order to reduce the time it takes to obtain aerial survey results.

## II. SELECTING A CALCULATION MODEL

The technology of convolutional neural networks (CNN) is adopted as the intellectual basis of the recognition system. This class of architectures is a highly specialized tool, suitable primarily for images and other data that can be represented in matrix form. As images store all information as two-dimensional matrices (i.e., as pixels), it is necessary to consider not only values from the neurons themselves but also values from a group of nearest neurons when working with images. To this end, besides neurons there is another type of elements in convolutional layers of CNN that apply certain linear operations to all input data of each neuron of the layer - the convolution core. The convolutional kernel is a grid that "slides" across the image (or previous layer convolutional layer) and looks for patterns and patterns in the data. If it finds a part of the image that matches a kernel pattern, it passes a large positive value to the current layer's computational neuron. If there is no match, the kernel will pass a small value or zero.

Because the convolution kernel is applied to every position in the image, the convolution layer of CNN is extremely effective in image processing tasks because features or patterns in the images can appear anywhere in these images. That is, CNN is able to analyze context-dependent data.

The Mask Regions with Convolution Neural Networks (MRCNN) architecture [4] was chosen for this task. This architecture is a subset of the classical CNN. Due to the complexity of the architecture, it more successfully copes with the tasks of semantic and object segmentation of images. It is a modification of the existing architecture Fast Region-based Convolution Neural Network (FRCNN), in which was added a module responsible for recognition and generation of object masks.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

FRCNN is an CNN that searches the image for objects and then additionally classifies the found object. The output of this model is the bounding rectangles localizing each object in the image and the class label of the found object with a confidence score.

In terms of operating logic, MRCNN, the first stage of operation is the same as that of a conventional FRCNN. It consists of simultaneously running two enabled artificial neural networks: a mainline network (ResNet, VGG, Inception or similar) and a regional positioning network. These networks process each image received at the MRCNN input and provide an output of a three-dimensional array - an array of suggested regions. This array contains the coordinates of regions in the input image that contain an object.

In the second stage of the work, conventional FRCNN predicts the bounding rectangle coordinates and feature classes for each of the proposed regions obtained in the first stage. Each proposed region can have a different size, but since convolutional layers in CNN always require a fixed size vector for prediction, this step also scales the regions found. The size of the regions is scaled using either the RoI algorithm or the RoIAlign method.

MRCNN is in turn an extended version of FRCNN, augmented with a branch to predict segmentation masks for each area of interest. The second phase of MRCNN already uses only RoIAlign, which helps to preserve the original spatial coordinates that are offset when RoI is used. This is then necessary so that the RoIAlign output can be combined with the data from the first phase, and a mask can be generated for each RoIAlign response using the Mask Head module (which in turn is also implemented using convolutional layers). Such masks are a two-dimensional matrix which, for each pixel within a region's boundaries, determines whether or not that pixel belongs to the object in question.

This approach allows the boundaries of the object being searched to be defined more precisely. Ideally, the MRCNN can accurately calculate all the pixels in the image that represent the object being searched for.

## III. THE MODEL CREATED

Since there is insufficient data for training in the chosen application domain, it was decided to perform the main training on a dataset that includes images of other animals. This was necessary for the MRCNN to learn to recognize animals as a class of objects. And for the specific task, this MRCNN was already further trained on a specific dataset, including aerial images of reindeer herds. The specifics of the task are that the herds are photographed from different distances, in different landscapes, under different light conditions, the animals on the pictures have different colors and can be at different angles to the camera, can overlap each other. These peculiarities of aerial photos create additional difficulties in solving the task of reindeer identification. Therefore, the presented task is non-trivial and the application of CNNs trained on common datasets is not possible.

The main training dataset for MRCNN was the MS COCO (Microsoft Common Objects in Context) image array [5]. This array is one of the largest datasets used to date for training machine learning models to solve detection and segmentation problems. The dataset consists of 328 thousand images. All images are marked up and formed into training samples. Therefore, using this dataset for basic training of MRCNN allows all the basic concepts of different object classes, including animals, to be specified for CNN. However, images of reindeer are not part of MS COCO and MRCNN by default is not able to distinguish them from a number of other animals (sheep, gazelles, cows, horses). Therefore, in order to recognize reindeer, the MRCNN needs to be further trained using aerial photo arrays with images of these animals.

An input dataset containing a training sample of 100 aerial images of herds with all animals tagged and a test sample of 30 original herd images was prepared for training the CNN. The model was trained on 20 epochs, with 60 training steps per epoch, with a training rate of 0.0058 and a detection miss threshold of 0.7. On the test data set, the trained model recognized an average of 82% of the deer correctly. The deer recognition accuracy can be improved by retraining the CNN on the extended training dataset.

So far, a software package with a web interface https://regionview.ru/ai/ has been created for the developed network, and the program itself has been deploy for limited use.

To use the system, users must upload a JPEG (.jpg) or GIF (.gif) image to their computer.

The system interface contains a set of window forms that provide:

- downloading of aerial images from the user's computer for processing,
- running a program to recognition and count the reindeer in the pictures,
- presentation of the results of the program work, with an image on the screen showing the reindeer images recognized by the system and the total number of animals counted,
- downloading the results to the user's computer.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

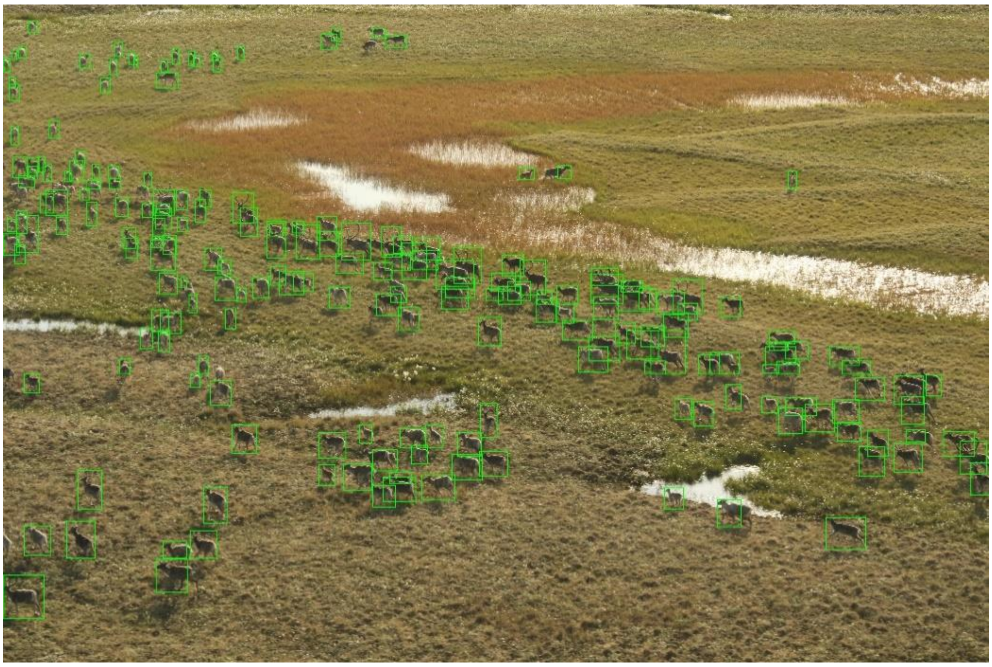

Fig. 1. Program-marked aerial photo of a herd of wild reindeer

After viewing the marked image, if the user is not satisfied with the accuracy of the software package, they can continue to process the image manually in any graphics editor that supports the. jpg file extension.

As an example, Figure 1 shows the result of the software, a raw image with automatically recognized and tagged animals. The figure shows that MRCNN can work with images that are noisy with background objects - puddles, lakes, bumps, etc. None of the background objects were mistaken for a reindeer. Also, it is noticeable that the MRCNN well with herds in which the reindeer gather in very dense groups and partially overlap each other. The recognition error was about 10%.

## IV. CONCLUSION

In general, verification of the system on an independent sample of aerial images showed that MRCNN can work with images that are noisy with background objects - puddles, lakes, hillocks, polygons, etc. The network successfully distinguishes individual animals in dense groups, animals at different angles and at different distances from the camera. The recognition error of the deer in the image was about 17%.

## ACKNOWLEDGEMENTS

The work was carried out at the St. Petersburg Federal Research Centre of the Russian Academy of Sciences (SPb FRC RAS) with the support of the World-Wide Fund for Nature - Contract №BBF003107, RFBR grant №19-37-90112 and Budget Subject №0073- 2019-0004.

## REFERENCES

[1] V. A. Zyryanov, B. M. Pavlov, G. D. Yakushkin, Ecological basis of counting the number of game animals in the tundra zone of Taimyr. Problems of hunting economy of Krasnoyarsk Region, Krasnoyarsk, 1971, P. 70-72.

[2] L. A. Kolpashchikov, B. M. Pavlov, V. V. Mikhailov, Methodology of aerial counting and determination of polling rates of Taimyr wild reindeer population: methodological recommendations, Saint-Petersburg, 1999, 25 p.

[3] N. G. Chelintsev, Mathematical basis of animal recording, Moskva, 2000, 431 p.

[4] K. He Mask R-CNN, G. Gkioxari, P. Dollar, R. Mask Girshick, Computer Vision and Patter Recognition, Cornell University, 2017.

[5] Common Objects in Context, https://cocodataset.org/.

[6] Automated Animal Identification Using Deep Learning Techniques, Proc. of the Nat. Acad. of Sciences of the USA, 2018.

[7] P. Ganesh, K. Volle, T. F. Burks, S. S. Mehta, Deep Orange: Mask R-CNN-based Orange Detection and Segmentation, IFAC-PapersOnLine, 2019, vol. 52, issue 30, pp. 70-75.

[8] G. Zhao, J. Hu, W. Xiao, J. Zou A mask R-CNN based method for inspecting cable brackets in aircraft, Chinese Journal of Aeronautics, 2020.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.