E3S Web of Conferences **390**, 03011 (2023) https://doi.org/10.1051/e3sconf/202339003011
AGRITECH-VIII 2023

Solving the problem of biodiversity analysis of bird detection and classification in the video stream of camera traps

Mikhail G. Dorrer[1*], and *Anna* E. Alekhina[1,2]

[1]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Ave., Krasnoyarsk, 660037, Russian Federation
[2]Siberian Federal University, 79 Svobodny pr., 660041 Krasnoyarsk, Russia

Abstract. The work is devoted to solving the problem of assessing the comparative efficiency of several common architectures of convolutional neural networks for monitoring birds in a natural environment. The problem was solved by detecting birds recorded by video traps installed on feeders in several regions of Panama by different architectures. Then a comparison was made between the recognition quality metrics – IoU and mAP, and based on the values of the metrics, a conclusion was made about the effectiveness of the architectures. Experiments have shown that the YOLO architecture of the Tiny version with comparative modules wins in the accuracy table. In the future, it is planned to improve the application of neural network architectures by finalizing the dataset with the involvement of expert bird watchers and open ornithological ontologies.

1 Introduction

Comparison of different neural network architectures for solving certain applications is currently a prevalent task among researchers in various fields of artificial intelligence. In such a comparison, rather valuable results can be obtained that allow developers of applied solutions to reduce the time and effort spent on selecting the optimal neural network architecture.

For example, compares how 3D convolutional networks and recurrent LSTM networks learn features in time-dependent frames. An urgent task is to compare the reliability of Caps Net with the reliability of a classical convolutional neural network and a fully convolutional network (FCN) in face recognition tasks. The authors of [1] propose using evolutionary methods to study the architectures of convolutional networks when creating neural networks for specific applications.

As an applied area for comparing the efficiency of convolutional neural network architectures, the problem of efficient detection and monitoring of bird species was chosen. This task is relevant in many branches of biology and ecology. Monitoring methods are divided into bird diversity study, spatial distribution study, and migration study. These methods can be subdivided into the aerial, sea, and land surveys depending on the location

* Corresponding author: dorrer_mg@sibsau.ru

and media used. Migration surveys can use trackers and radar for accurate and detailed data collection. As a rule, both point census and line-section methods can be used for monitoring, and bird species, number of birds, and place and time of observation should be recorded. The data collected as a result of bird monitoring can be used as a basis for choosing a site for a particular economic activity - deforestation, mining, construction of energy facilities [2]. At the same time, it becomes possible to comply with the conditions for minimizing harmful effects on nature, in particular, on the life of birds.

At the same time, manual methods of monitoring birds are highly laborious [3], which complicates their implementation. This problem makes it attractive to use automation tools for solving problems of monitoring the number and behavior of birds in nature.

This paper describes the use of YOLO v3 [4] to detect tropical birds of Panama in places of their feeding. The data from the surveillance cameras for the "Panama Fruit Feeders" project [5] are used. It was shown that thanks to the use of the YOLO v3 architecture, a confident recognition of the majority of birds found at feeders was achieved. Therefore, video cameras combined with a computer vision system based on a convolutional neural network of YOLO v3 architecture can effectively monitor birds in the wild.

2 Related works

The work [6] describes an automated solution for recognizing birds by their calls. Encouraging results were obtained - 96 - 100% correct recognition of two species of birds by recording their calls.

In [7], also using convolutional neural networks, the problem of identifying birds in an arbitrary image with an accuracy of 95.52% is solved.

The work [8] describes the development of a system for automatic recognition of harmful birds using video cameras of drones and the YOLO architecture.

Also, the DC-YOLO architecture is used in [9] to determine the number of birds around the power line. Experimental results show that the detection accuracy with this model reaches 86.31%.

Thus, having considered several solutions for using various architectures of artificial neural networks for recognizing birds and solving similar problems, we can talk about, at least, the attractiveness of the idea to apply the architecture of convolutional neural networks YOLO v3 for the task at hand.

3 Materials and methods

3.1 Materials

3.1.1 Test Zones

The project's camera [10] is installed on the territory of El Valle de Anton in the Cerro Gaital mountains (Figure 1) in Panama.

 https://doi.org/10.1051/e3sconf/202339003011

Fig. 1. Map of the location of the project cameras.

The feeder is located 12 meters from the house and is filled with food several times every day between 5 am and 7 pm. The project uses the Axis Q1785-LE camera. The video broadcast runs around the clock, is broadcast on the project's website and youtube channel [11], and is in the public domain. Also, on the project's youtube channel, there is a playlist with the moments of direct feeding of the birds. The total amount of time for the presence of various birds is approximately three and a half hours.

Using the bird breed classifier, the following birds were identified on video excerpts (Figure 2): Mexican jay Aphelocoma wollweberi (1), thick-billed euphonia Euphonia laniirostris (2), golden-billed aremon Arremon aurantiirostris (3), striped wood potter Thripadectes (4) , turquoise tanager honey plant Cyanerpes Cyaneus (5), brown-winged chachal Ortalis vetula (6), cream-bellied thrush Turdus amaurochalinus (7), Caroline melanerpes Melanerpes carolinus (8), collared arasari Pteroglossus torquatus (10 blue-gray tanager Thraupis episcopus (11), gray-necked forest shepherd Aramides cajanea (12), dark-faced reed tanager Mitrospingus cassinii (13), violet euphonia Euphonia violacea (14), black-breasted multi-colored jay Saltin cyanocorax aff 16), green sai Chlorophanes spiza (17), crimson tanager Crimson-backed tanager (18), Baltimore Oriole male (19), yellow-bellied egg painted tanager Ramphocelus icteronotus (20), banana songbird Coereba flaveola (21), brown-tailed amazilia Amazilia tzacatl (22), toco toucan Ramphastos toco (23), chestnut-headed oropendola Psarocolius wagleleagioneus mice (24), guardian Bird number 26 is currently unknown (26).

Fig. 2. Species of birds found and classified in Panama feeders.

3.1.2 Training sample

We created a dataset with 3548 images, where 80 percent are training data, and 20 percent are validation images. Twenty-six images are test images (Table 1). Unfortunately, the dataset is not uniform, as the appearance of some birds on the camera is inconsistent.

Table 1. Test images.

Dataset	Training images	Training segments	Validation images	Validation segments	Test images	Test segments	Images, total	Bird classification image dataset
Class 1	2839	426	708	85	26	1	3573	8129
Class 2		75		15		1		
Class 3		566		113		1		
Class 4		421		84		1		
Class 5		279		55		1		
Class 6		237		47		1		
Class 7		103		20		1		
Class 8		276		55		1		
Class 9		313		62		1		
Class 10		423		84		1		
Class 11		105		21		1		
Class 12		21		5		1		
Class 13		306		61		1		
Class 14		53		10		1		
Class 15		540		108		1		
Class 16		147		29		1		
Class 17		220		44		1		
Class 18		223		44		1		
Class 19		137		27		1		
Class 20		236		47		1		
Class 21		41		8		1		
Class 22		3		1		1		

Class 23		1122		224		1		
Class 24		482		96		1		
Class 25		2		1		1		
Total	2839	6757	708	1346	26	26	3573	8129

The dataset consisted of daytime photographs. All original images were in an additive color model (RGB, where R is red, G is green, and B - blue) with a size of 1280x720 pixels. This dataset was developed in a specially designed markup program for the YOLO standard models. An example of poultry sample marking is shown below (Figure 3).

Fig. 3. Example of markup of a sample of the class Pteroglossus torquatus.

Since the task of the project is to detect and classify the arriving bird, other animals that came to the feeding trough were not included in the training sample (Figure 4).

Fig. 4. Other animals appearing on camera traps.

During the classification of birds, it was decided to include birds with bright signs of sexual dimorphism in one class of birds. A similar decision was made with the adult chicks. Below is an example of a sample of a female and a male of the class Cyanerpes Cyaneus (Figure 5) and a group of birds of different ages of the class Ortalis vetula (Figure 6).

 https://doi.org/10.1051/e3sconf/202339003011

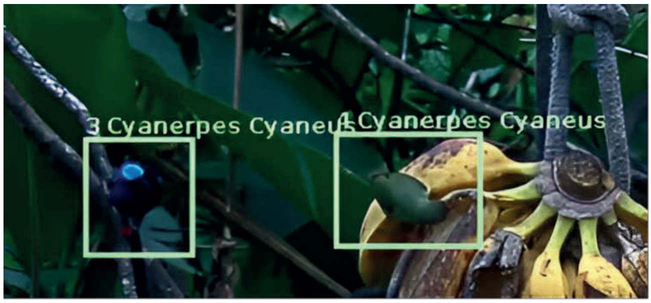

Fig. 5. Sexual dimorphism of birds of the same species.

Fig. 6. Individuals of birds of the same species, but of different ages.

4 Methods

4.1 Neural network architecture used to process the feeders' video stream

In this paper, we used various variants of convolutional neural networks belonging to the YOLO (You only look once) family of architectures. Its architecture is shown in Figure 7.

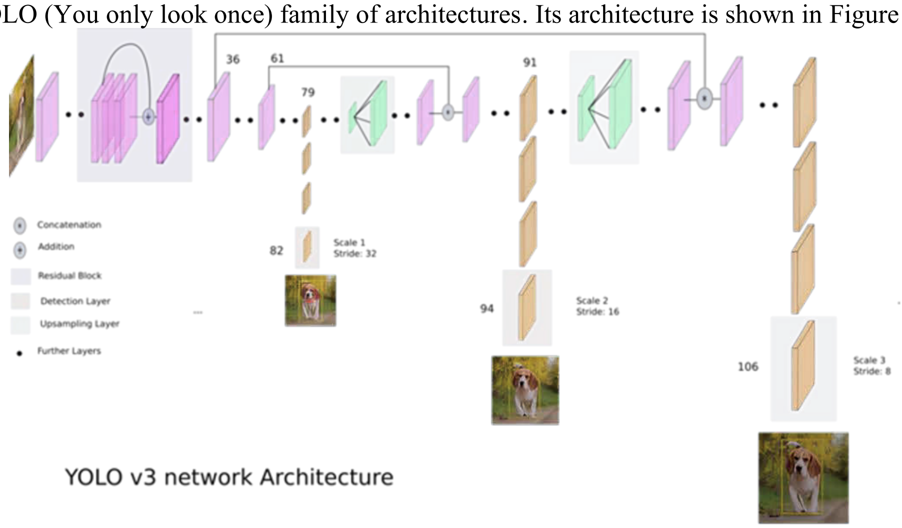

Fig. 7. YOLOv3 architecture diagram.

4.2 Analyzed architectures

This paper examines the ability to learn on a heterogeneous user dataset of various modifications of the YOLOv3 architecture. The base ("vanilla") version of the model is also included in the comparison.

The selected architectures differ in the following features:

– YOLOv3 spp uses "spatial pyramid join" to eliminate a fixed image size on the model input.
– YOLOv3 tiny pan3 CenterNet.
– YOLOv3-tiny 3L.
– YOLOv3-tiny comparison.
– YOLOv3 spp pan scale.

4.3 Experimental setup

The model was trained and tested on Ubuntu 16.04.6 LTS operating system with an NVIDIA GeForce GTX 1060 graphics processing unit (GPU) and CUDA 10.1 parallel computing platform. The input images have been scaled down to 416×416 pixels to fit the input layer of the training model. We used a global learning rate of 0.001, and four classes have 30,000 iterations for maximum packages. During training, standardized magnification methods were used, where decay 0.0005, saturation 1.5, exposure 1.5.

The training time for each YOLO v3 model took 30 hours.

4.4 Model accuracy metrics

To evaluate the performance of trained YOLO architectures in the bird detection problem, we used metrics of mean accuracy (mAP) and intersection by a union (IoU).

The $Precision$ metric measures the percentage of labels correctly recognized, and $Recall$ is part of the successful retrieval of matching labels. These criteria are used to calculate $F1Score$ and mAP to evaluate the performance of the model. $F1Score$ is calculated based on the accuracy and memorization of the test. mAP is the average of all grades or the area's location under the Precision-Recall curve. mAP is calculated in the range from 0 to 1.

5 Results

In Table 1, you can see the results of processing the video stream from feeders with various variants of the convolutional neural network architectures tested in this article.

Table 2. Metrics of trained YOLOv3-X configurations.

Model version	Precision	Recall	F1-score	IoU%	mAP%	TP	FP	FN	FPS
Yolov3	0.97	0.98	0.98	76.0	59.24	2559	79	50	30
Yolov3-spp	0.85	0.83	0.84	66.79	53.62	2153	366	456	15.6
Yolov3-tiny_pan3_CenterNet	0.90	0.98	0.94	71.34	59.67	2554	277	55	38.2

Yolov3-tiny_3l.	0.99	0.98	0.99	76.63	59.63	2559	26	50	98
Yolov3-tiny comparsion	0.99	0.99	0.99	80.21	59.78	2580	19	29	60
YOLOv3 spp pan scale	0.97	0.68	0.80	72.10	55.23	1763	50	846	12.8
YOLOv3 tiny pan	0.99	0.99	0.99	80.71	59.86	2590	17	19	54

Figure 8 shows the expert selection with the dotted line and the line with the results of the trained model. As can be seen from the results, seven birds were not identified in the test images. Thus, the recognition was successful, but the recognized classes suffer from a low IoU value.

Fig. 8. Results of detection and recognition of birds in the video stream for Yolov3.

As you can see from the accuracy table, the Yolo versions of teenies with comparative modules win. They have fewer false positives than other architecture configurations, and also the larger versions of the Yolov3 model also win in speed. Thus, these models can be used on weak computing architectures such as Raspberry pi 4 or Nvidia Jetson Nano and its following generations. Furthermore, this work shows that even with a bad and sparse dataset, it is possible to obtain good model training accuracy, such as precision and recall 0.99. In this situation, during the operation of the model, the domain of definition will suffer, but it is solved precisely by the refinement and saturation of the training sample.

6 Conclusion

The following stages of work to optimize the operation of the neural network include:

1. Classify the remaining unknown birds with the help of ornithology specialists
2. Supplement the neural network glossary, similar to the glossary compiled on the official website of the Allbirds project.

https://doi.org/10.1051/e3sconf/202339003011

3. Normalize the selection of existing classes using additional materials from open sources.

References

1. J. D. Homburg, M. Adams, M. Thies, T. Korthals, M. Hesse, U. Rückert, In: Advances in Computational Intelligence. IWANN 2019. Rojas I, Joya G, Catala A (Eds); Lecture Notes in Computer Science, 11507. (Cham: Springer, 2019) pp. 735-746
2. M. Hong, Wind Energy **10**, 31 (2019)
3. J. Seys, H. Offringa, J. Waeyenberge, P. Meire, E. Kuijken, Mar. Pollut. Bull. **44**, 322 (2002)
4. B.-C. Zhu, T.-H. Lin, Y.-C. Tsai, K.-W. Hsieh, F.-M. Fan, P.-K. Lei, Proc. Int. Disp. Work. **36** (2019)
5. J. Brown, S. Gripenberg, Y. Basset, O. Calderón, I. Simon, D. Fernández, M. Cedeno, M. Rivera, Proc. Entomol. Soc. Washingt. **122**, 12 (2020)
6. N. Priyadarshani, S. Marsland, I. Castro, A. Punchihewa, Plos One **11** e0146790. 10.1371/journal.pone.0146790. (2016)
7. F. Mashuk, Samsujjoha, A. Sattar, N. Sultana, Axioms **12** 818-822 (2021)
8. H. Kitakaze, R. Yoshihara, et al., Transactions of the Japan Society of Applied Engineering **8**, 10 (2020)
9. C. Zou, Y. Liang, Physicochemical aspects of studying clusters, nanostructures and nanomaterials **12** 222-231 (2020)
10. C. Arvind, V. Joshi, R. Charif, P. Jeganathan, V. V Robin, Oryx **57**, 55 (2023)
11. J. Houtz, R. Mady, J. Uehling, Ecol. Evol. **11**, (2020)