Bird Species Identification Using Yolact Classifier

Sofia K. Pillai[1], M. M. Raghuwanshi[2] and Snehalata Dongre[2]

[1,3]*Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India*

[2]*G H Raisoni College of Engineering & Management, Pune, India*

ABSTRACT

Capturing a perfect shot for an ornithologist is a challenging task when a bird is distant or the image captured is blurred or not recognizable due to motion, geographical, or weather phenomenon. Researchers from MIT have developed a novel semantic segmentation model based on public segmentation benchmarks achieving a state-of-art performance on datasets being efficient and accurate compared to other pre-existing models. In this paper, we have proposed a novel segmentation model based on the YOLACT classifier for real time instant segmentation of bird species and classifying them according to their class. This model overcomes the development and classification challenges faced using other pre-existing classification models. The model is trained using a Caltech-UCSD dataset containing more than 11,788 images classified under 200 species categories. The model creates pseudo masks at a rate of 34 fps from features extracted and predicts the class of the species combining all the pseudo masks and comparing the features from the train data.

KEY WORDS: BIRD CLASSIFICATION, DEEP LEARNING, YOLACT CLASSIFIER.

INTRODUCTION

Growing species of birds, many times can confuse to specify which class it belongs to. An ornithologist should be aware of the physical and geographical properties of the birds to recognize, but as there are thousands of species one may find it difficult to recognize it in different environments. To deal with these problems there are already many feature extraction and classification architectures like ResNet, YOLO, Inception, etc, they have several drawbacks. While taking pictures of the bird it should be stationary and also the weather should be suitable for example hail, sleet, fog, snow, and rain can give blur and hazy pictures as well as when the bird is not stable and is flying it makes a challenging task to capture a perfect shot for the specification of the bird.

To capture a perfect shot there is a need for a pre-trained architecture, where the focus is much more accurate than the others and it should capture the various colors and their variations in beak, eyes, feathers clause, etc. The problems of previously architectures take a lot of time to be trained and may have some issues with the specifications identification. YOLACT is a classifier consisting of 101 ResNet layers that are responsible for perfect object detection. The higher the epoch more the accuracy but if the training times longs too much it may defect the model. YOLACT consists of strong focus, less training time, and higher accuracy.

Semantic segmentation of LIDAR achieves a performance above benchmark that is significantly faster and more efficient than existing methods for semantic analysis of bird species. A group of researchers has developed a new instance segmentation method that works in real-time. Identifying the bird species is a complicated task that often results in ambiguous labels sometimes even professional bird watchers fail to identify the bird in the image. This results in pushing the limits of the visual ability of humans and computers and is a difficult problem.

Although different bird species have the same parts they dramatically differ in shape, size, and appearance. Due to extreme variation in pose, background, and variation in lighting intraclass variance is high. Bird watching is an art of studying, observing, and researching birds. Birdwatchers are those who enjoy observing birds and are passionate about it and others involved in scientific study and research of birds are termed as ornithologists. Bird-

Biosc Biotech Res Comm P-ISSN: 0974-6455 E-ISSN: 2321-4007

Identifiers and Pagination
Year: 2021 Vol: 14 No (9) Special Issue
Pages: 108-114

Article Information
Received: 04th May 2021
Accepted after revision: 09th July 2021

DOI: *http://dx.doi.org/10.21786/bbrc/14.9.23*

identification should be accurate and is an important aspect of bird watching. In India, not many software are available for the identification of bird species thus making the identification process tough for the bird-watchers. Feature extraction obtained from an image with Transfer learning and pre-trained algorithms.

The Mask R-CNN algorithm builds on the previous Faster R - CNN, enabling the network to not only performs object detection but pixel-wise instance segmentation as well. Unlike polygonal segmentation devised specifically to detect a defined object of interest, full semantic segmentation provides a complete understanding of every pixel of the scene in the image. By applying an instance segmentation method and geometry projection of the LIDAR points in world coordinates we can generate the semantic segmentation of each object in terms of LIDAR points. Object instance segmentation in 3D point clouds is formulated by considering all the input LIDAR points X-1 from Y-1 to Geo points. The semantic segmentation of each object is created by applying a combination of two methods: semantic and non-semantic segmentation. In the case of the image scan, we formulate the semantics of objects in terms of their position in the scene, taking into account the position of LiDAR points and their distance from each other and other objects.

Literature Survey: Automated model identification of bird species has been developed by the researchers from their audio segments. Using the signal process and machine learning techniques they tend to resolve the bird's species identification problem. Initially, audio features are extracted from the audio fragments then according to a traditional machine learning scenario, where labeled information of earlier identified bird songs are used to style a choice procedure that's wont to predict the species of a new bird song the problem is Experiments are performed in a dataset of recorded songs of bird species which emerge in a specific region. The experiments are results compared based on the performance obtained in different scenarios, enclosing the complete audio signals, as recorded in the field, and short audio segments (pulses) acquired from the signals by a split procedure. The impact of the number of classes (bird species) in the identification accuracy is evaluated. While on the contrary, most approaches apply within reach to neighbor matching or decision trees using extracted templates for each bird species, our attracts upon recent advances within the domain of deep learning strategies and from speech recognition, on the largest publicly available dataset we train a 3D-CNN.

Researchers have used a 3D-CNN with one dense layer and five convolution layer. Every convolution layer firstly uses a rectify activation function followed by a max-pooling layer. The signal including bird songs or calls were audible and a noise part where no bird is calling or singing. It includes background noise. Spectrogram (Short Time Fourier Transform) of both parts is computed and splits each spectrogram into two equally sized chunks. Around 3 seconds each chunk can be seen as the spectrogram. As a special testing sample for our neural network, we can use every chunk from the signal part.

However, a large number of object categories and occlusion from nearby objects in complex environments pose great challenges in urban Coupling instances and semantic segmentation. iMerit enrichment teams identify the pixels in images as belonging to a class and identify what instances of that class they belong to. Instance segmentation models are a little more complicated to evaluate whereas semantic segmentation models output a single segmentation mask instance segmentation models produce a collection of local segmentation masks describing each object detected in the image. The way YOLACT addresses the problem of instance segmentation is by breaking the task into two smaller tasks that run in parallel: generating a dictionary of prototype masks .The function of this model is to map the input LiDAR to a space that reflects the objects in the scene like semantic segmentation or object detection networks . LiDAR points within the ground truth 3D bounding box are utilized as the supervising signal for 3D instance segmentation to filter out background points, the point cloud data is segmented to determine the class of objects using the points network.

Evidence theory adopts a formalism that can effectively counter the incoherence of multiple segmentation pieces: instance segmentations that give each example of a particular object in the image a unique label. Depending on the boundary field used, semantic segmentation can distinguish between individual instances of an object or only regions with more meaningful segmentation. The network prints a list of all instances of an object in the image with the same label and label for each instance of that object.Mask (RCNN 9.0) has taken over the instance, but is integrated differently than the Mask R - CNN algorithm, which builds on the previous Faster - R CNN, to enable the network not only a more efficient and accurate image recognition but also a semantic task. While polygonal segmentation is specifically designed to recognize the defined objects, full semantic segmentation provides a complete understanding of each pixel in the scene of the image. In the case of the image scan, the segmentation mask predicted by DCNN is projected to infer the position of each object in an image, such as a bird, tree, or object with a name.

The iMerit Enhancement Team identified each pixel in the image as belonging to a class and identified the instance of the class to which it belongs. The instance segmentation model is a bit more complicated, as it needs to be evaluated and evaluated in real-time, which creates a local segmentation mask that describes all objects discovered in an image. This is a big challenge, but the way YOLACT addresses the issue of instance segmentation is to break the task down into two small tasks that run in parallel: creating a dictionary of prototype masks and predicting a series of linear combinations of coefficients for each instance. First, the 3D LiDAR data is transformed and fed into the instance segmentation model to obtain a predicted instance mask for each class. The function of the model is to assign the input LiDAR to a space that reflects the objects in the scene through semantic segmentation in an object recognition network. These points are used as background points, provided by the three-dimensional model for the 2D suggestion region and the background point for 3D instance segmentation. The class of the object used by the Point segment network is determined by semantic segmentation in the 3D LiDAR data and by a binary classification of the LiDAR points.

Researchers want to explore how they can use deep learning and deep neural networks (DNN) to solve the problem of image classification in many classes. The data set includes 200 bird species, so there are over 200 different classes of output that we want to predict by training our model. What makes the problem difficult is that the order can vary, includes a very large vocabulary of input symbols, and must be chosen by 12 December 2019. Faced with this problem, transfer learning techniques are generally used to strengthen the capacity of deep neural networks. Important thing that makes deep learning compatible for NLP is the ability to select functions from a large number of different datasets. A text collector is a great example of a project in which we build a deep neural network using natural language processing. We take a text sequence as input, learn its structure, and then summarize it as we do the structure.

This is often referred to as a large-scale fine grain, but identifying bird species can be difficult. The proposed system will enable a deep neural network to mimic a pathologist's perception and acquire knowledge related to pathology. This should be possible for medium-level datasets by imitating the deep neural network in the way it is perceived and by acquiring knowledge about pathology. This has proved to be a very difficult task, and it can challenge even the most advanced deep learning techniques, such as neural networks. Sources: 5, 7, 12 The classification of multiple markings is a type of classification in which an object can be divided into more than one class. Ornithology experts carry out bird identification based on classifications proposed by Linnaeus, such as the classification of bird species.

In image classification by transfer learning, one usually takes a layer of the Convolutional Neural Network (CNN) from a pre-trained model and adds a final layer. The output of the feature Extraction Network can then be fed into a classifier of choice, which does not necessarily have to be machine learning to perform the class prediction. If you are interested in how to train your classification or object recognition model, there is a great article about Deep Learning in Computer Vision, which Python calls Deep Learning for Computer Vision. In image classification through transfer learning, the pre-trained network is usually supplemented and used as a feature extraction network, removing the last layers of classifiers and using the extracted feature output as input, the class identifier choice may be an ensemble based on a decision tree or a random number generator.

Some approaches show excellent results in the transfer of learning outcomes, which means that the image classification that the model generates can be applied to other AI systems. In deep learning, a revolutionary neural network (CNN) or CNN is a type of machine learning in which a model learns to perform classification tasks directly on an image, video, text, or sound. Neural networks such as CNN are deep neural networks that are mainly used for the analysis of visual images. Programming ResNet models is easy because they allow the creation of networks that are often used in many transfer tasks - learning tasks related to image classification, object localization, segmentation, etc.

Convolutional neural network models are developed for image classification, in which the model learns to display two-dimensional input internally through a process called feature learning. For image classification, models of conventional neural networks were developed that learn the internal representations of two-dimensional input through the process of "feature learning." For Bild-Kleinanzeigen, evolution-oriented models of the Neural Network were developed, for example in the form of convolutional, a type of machine learning in which a model learns the internal representations for two-dimensional input via the processes known as "feature learning. Transfer learning refers to the process of reusing a large, pre-trained model by recycling the parameters it has learned in a new model with other learning parameters. In the case of image classifieds, this means, for example, using a trained neural network with the same parameters as the original model. Transfer learning can also be applied to other types of machine learning, such as image classification.

Design Methodology: We have used the Caltech-UCSD Bird detection dataset containing over 11,788 images mostly of North American Birds categorized into 200 bird species. Every image in the dataset is labeled with one bounding box and 312 binary attributes. The classes from the dataset are much imbalanced containing an unequal amount of images. The images are of high resolution ranging from image resolution 800X600 to 4000X6000. The downloaded dataset is then divided into training and testing categories in a ratio of 80:20. The Pre-processing stage is being carried out on the dataset before using the images to train the model. All the images were scaled and re-sized into the same ratio using CV2 libraries. To reduce the noise, harshness, and disturbances in the image, the pixel values are normalized and later used for training the model. The model focuses and extracts features such as color, pattern or shape of a particular part highlighting a particular area. The training dataset contains over 9,430 images and the testing dataset contains over 2,358 images. The model is trained using several features as per the Table 1. such as size, wing_shape, body_color, back_pattern, wing_color, eye_color, head_pattern, bill_pattern, wing_shape, etc.

Architectural Overview: The input images are fed into a fully convolutional neural network backbone. The CNN backbone is a combination of ResNet-101 which is a long 101 layer network combined with a feature pyramid network. ResNet-101 is a long neural network that passes the image through tons of layers. It pulls out the result at a certain point before it's fully processed and then it passes it into the feature pyramid network which is a simplified processed version of images at different scales. It is difficult to train deep neural networks progressively; the ResNet-101 architecture is capable of extracting features from birds. Low resolution maps are generated via CNN backbone identifying rough locations.

Table 1. Multi-valued bird attributes

Attribute	Values
Size	3-5 in, 5-9 in, 9-16 in, 16-32 in, 32-64 in
Wing_Shape	long-wing, pointed-wing, broad-wing, tapered-wing
Body_Color	brown, gray, pink, white, green, red, blue
Back_Pattern	spotted, striped, multi-colored, lined
Wing_color	Black, buff, gray, white, pink, yellow, brown
Eye_Color	Black, buff, gray, white, pink, yellow, brown
Head_Pattern	Malar, stripped, spotted, eyebrow, crapped, unique pattern
Bill_Pattern	cone, all purpose, hooked, dagger
Wing_shape	Long, short, tapered, pointed, round, broad

Figure 1: YOLACT evaluation results on Caltech-UCSD Dataset

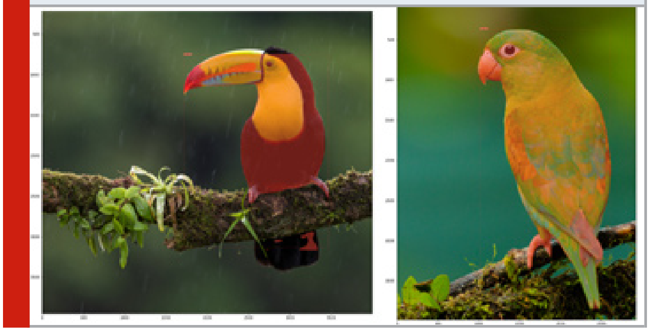


Figure 2: System flow for YOLACT Bird classification model

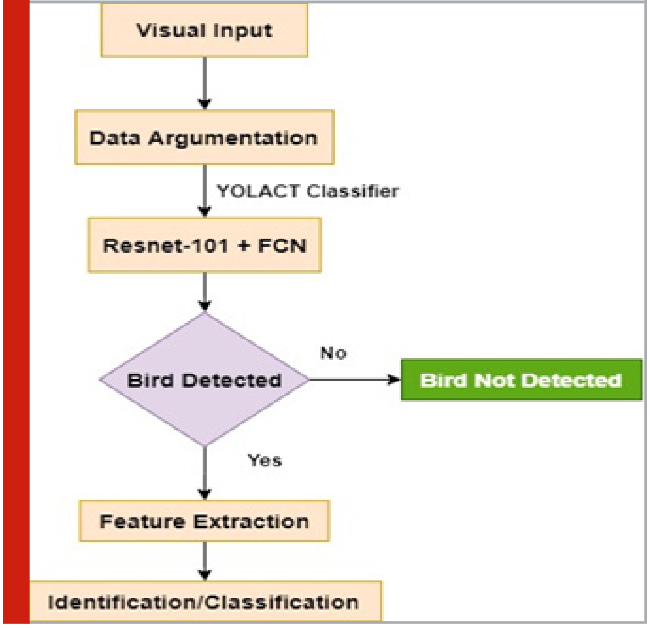
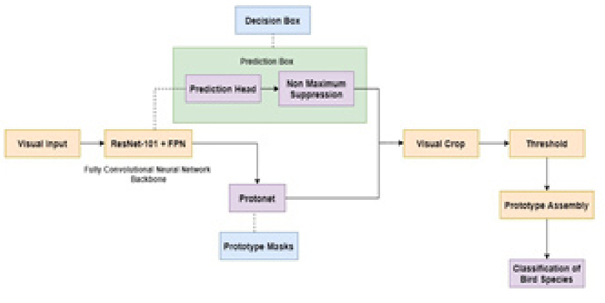


Figure 3: YOLACT Architecture for Bird Detection Model



The prediction head looks over the scales and makes predictions according to the pseudo mask scales. Non Maximum Suppression (NMS) filters the mask scales based on the features extracted and trained dataset to predict the model correctly. Protonet makes the prototype masks and tries to predict the mask then combine those in a way that makes pseudo masks as per the features extracted and compiled using convolutional neural network models and then it combines them to make an intelligent mask decision. The intelligent mask decision is a correct prediction about the species made according to the trained dataset and their characteristics.

YOLACT Architecture: YOLACT architecture is a fully CNN model for segmentation at real-time for bird species greater than 30 frames per second that have a state-of-art performance and results above benchmarks using Caltech-UCSD-200 dataset executed on Google Collaboratory using a single Titan XP GPU which is significantly considered to be faster and state-of-the-art approaches and architectures. The initial branch of the model uses a fully convolutional neural network to produce pseudo prototype masks of the size of images using several features extracted from the input image. Later, the model adds an extra head to the branch for predicting vector masks from prototype space that represents the instances for bird detection.

Finally, the model linearly and partially combines the work of the first and second branches and constructs a prototype mask that survives non-maximum suppression. After creating a pseudo mask following parallel steps the model carries out the assembly step. For each instance, matrix multiplication and cropping operation is being carried out by a simple linear combination with the predicted bounding boxes. Cropping the instances reduces the network load and suppresses the disturbance factor outside the bounding box but still keeps a look at leakages on other instances of the same class of the bounding box. The instance segmentation divides the architecture into two parts. The first part generates the prototype masks and another part produces mask coefficients as per the instances. The feature pyramid network develops prototype masks that openly benefit from semantic segmentation in advances.

Data Argumentation: The Caltech-UCSD dataset contains bird images that can be fairly categorized according to the size into two types as 800 X 600 pixels and 4000 X 6000 pixels. The size of the image was approximately calculated and pre-processed to 600 X 600 pixels as it is difficult to process images with different several parameters. We applied cropping techniques on the dataset and obtained a total of 11,788 images. To make bird species detection robust, it is very important in the aspects of deep learning to ensure the diversity of data. We augmented all the images from the dataset and randomly selected images for dividing the dataset for training and testing purposes. Initially, we considered the center of a bird image and added it to the center of another image and produced sub-images for the whole dataset. Secondly, all the images were 50% cropped and flipped in vertical and horizontal manner. Similarly, the RGB values and probability were re-adjusted.

RESULTS AND DISCUSSION

We have used Caltech-UCSD-200 dataset for instance segmentation using standard metrics. We have divided the dataset into training and testing categories with a ratio of 80:20. The pre-processing stage is being carried out on the dataset before using the images to train the model. All the images were scaled and re-sized into the same ratio using CV2 libraries. To reduce the noise, harshness, and disturbances in the image, the pixel values are normalized and later used for training the model. The model focuses and extracts features such as color, pattern or shape of a particular part highlighting a particular area. The training dataset contains over 9,430 images and the testing dataset contains over 2,358 images. We have trained the model with a batch size 8 on Google colaboratory with a single Titan XP GPU using ImageNet based pre-trained weights. We have normalized the pre-trained batch and kept unfrozen and haven't added any extra convolutional pooling layer as the batch size used is enough for batch normalization.

Table 2. Prototype Results

k	AP	FPS	Time
9	27.2	34.1	31.0
17	28.3	33.8	31.4
*33	28.6	33.6	31.8
65	28.4	32.5	32.7j
129	28.9	32.3	32.5
257	28.1	30.0	30.2

Table 3. Accelerated Baselines

Method	AP	FPS	Time
FCIS w/o Mask Voting	28.9	21.4	106.5
Mask R-CNN (550 x 550)	33.6	10.7	74.8
FC-Mask	14.4	26.7	39.6
YOLACT-550 (Ours)	30.1	34.0	31.2

The model carries out up-to 1, 50,000 iterations and divides the training rate at 50,000 and 1, 00,000 for Pascal dataset and for bigger objects multiplied the anchor scale with 4/3. Training the model is a time consuming process depending on the GPU configuration. We have compared the performance of YOLACT model with other pre-existing architectures on MS COCO dataset setting up benchmarks and using state-of-the-art methods considering the main parameters as speed, train time and test time. As compared to the fully convolutional instance aware semantic segmentation and region based convolutional neural network, the YOLACT architecture has comparatively less noise and follows a boundary.

The map with 5.6 worse overall intersection over the union threshold at 95% threshold our model has a 1.6 AP with R-CNN. This indicates that re-polling results decrease the mask quality. As compared to the R-CNN model architecture, our model produces more temporal stability on visual masks by not applying temporal smoothing even when objects are stationary. As our masks produced by our model are of higher quality and have good stability as our model is a one stage shot detection model. Two stage shot detector methods are mostly dependent on features extracted at the initial stage. The prototypes are not much affected if the model predicts boxes across different frames giving more stable masks.

The speeds are calculated based on the experimental results on Titan XP GPU. YOLACT offers the fastest instance segmentation method on Caltech-UCSD-200 at a speed of 4.0x segmentation performance based on competitive instances as compared to other performances of existing architectures. The YOLACT architecture performs 50% more accurately than the R-CNN model with a threshold of 9.6 arithmetic progressions and

79% intersection over union threshold. There exists dissimilarity for instance between the results of Caltech-UCSD -200 with arithmetic progressive values as 7.5 and 7.6 respectively. The model clearly specifies that it has been built for speed. Hence we compare the no test time argumentation results with our model.

Table 4. Comparative analysis of different object detection architecture and their Intersection over Union (IoU) values

Architecture	Feature Extractor	Interference Time	AP	
			Intersection over Union: 0.3	Intersection over Union :0.5
YOLACT	ResNet-101	97	96.54	81.12
	ResNet-50	83	95.48	80.36
Faster R-CNN	Inception V2	89	95.65	80.45
	ResNet-50	76	92.45	74.87
YOLO V2	DarkNet-19	24	86.12	67.96
	Tiny YOLO	41	92.32	55.85
YOLO V3	MobileNet-V2	35	91.85	59.52
	DarkNet-53	22	89.25	57.41

From the accuracy results, it can be evaluated that the YOLACT model with ResNet-101 as a feature extractor has a good mean error value of 4.2% which is higher than the values for YOLACT ResNet-50 module and other subsequent values. The IoU can be elaborated as the ratio between the union and intersection of the detected box and ground truth box. These values are usually used to determine the precision and accuracy of the model architecture as the bird is identified and classified into the correct class or not. The values of AP depend on IoU values and change accordingly. Basically, The values for AP are calculated with Intersection over union threshold over 0.3, or by changing the threshold values from 0.3 to 0.5. The Table IV shows results for comparative analysis between YOLACT and other preexisting models for their intersection over Union (IoU) values and concludes that YOLACT model has the highest AP values among all other pre-existing models. On the basis of performance results for bird species detection and identification using Caltech-UCSD-200 dataset can be considered fairly satisfactory for AP values ranging from 85% to 97% during testing performed on different models.

CONCLUSION

In this paper, the model is developed to classify and identify bird species according to their feature characteristics extracted using YOLACT classifier. As per the results obtained the YOLACT architecture has best performance and accuracy compared to other architectures which balances the precision between training and accuracy values. The model extracts the features from the visual images as per the trained data and makes pseudo masks for the visuals plotting them on scales. The model uses the ResNet-101 backbone architecture combined with a feature pyramid network for extracting the bird features. The architectural model combines the pseudo masks and scales produced and predict the bird species. These areas can be considered to have more scope for real time classification using YOLACT according to the accuracy and prediction rate and its capability to produce good masks and speed. This model overcomes the development and classification challenges faced using other pre-existing classification models. The model is trained using a Caltech-UCSD dataset containing more than 11,788 images classified under 200 species categories. The model creates pseudo masks at a rate of 34 fps from features extracted and predicts the class of the species combining all the pseudo masks and comparing the features from the trained data.

REFERENCES

Bolya, Daniel & Zhou, Chong & Xiao, Fanyi & Lee, Yong Jae. (2019). YOLACT: Real-Time Instance Segmentation. 9156-9165. 10.1109/ICCV.2019.00925.

Bolya, Daniel & Zhou, Chong & Xiao, Fanyi & Lee, Yong Jae. (2020). YOLACT++: Better Real-time Instance Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2020.3014297.

C. Tang, Y. Feng, X. Yang, C. Zheng and Y. Zhou, ”The Object Detection Based on Deep Learning,” 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, 2017, pp. 723-728, doi: 10.1109/ICISCE.2017.156.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2011-001, 2011.

Choi, Se-Jun & Ki, Kyong. (2019). Nocturnal Birds Detection and Ecological Characteristics through Bioacoustic Monitoring. Korean Journal of Environment and Ecology. 33. 636-644. 10.13047/KJEE.2019.33.6.636.

Hong, Hyungi & Chung, Mokdong. (2020). Performance Improvement on Object Detection for the Specific

Domain Object Detecting. 10.1007/978-981-13-9341-9_26.

Islam, Shazzadul & Khan, Sabit & Abedin, Md & Habibullah, Khan & Das, Amit. (2019). Bird Species Classification from an Image Using VGG-16 Network. 38-42. 10.1145/3348445.3348480.

Jo, Jeongjin & Park, Junwon & Han, Jinyoung & Lee, Minsun & Smith, Anthony. (2019). Dynamic Bird Detection Using Image Processing and Neural Network. 210-214. 10.1109/RITAPP.2019.8932891.

Jo, Jeongjin & Park, Junwon & Han, Jinyoung & Lee, Minsun & Smith, Anthony. (2019). Dynamic Bird Detection Using Image Processing and Neural Network. 210-214. 10.1109/RITAPP.2019.8932891.

K. M. Ragib, R. T. Shithi, S. A. Haq, M. Hasan, K. M. Sakib and T. Farah, "PakhiChini: Automatic Bird Species Identification Using Deep Learning," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, United Kingdom, 2020, pp. 1-6, doi: 10.1109/WorldS450073.2020.9210259.

Nadimpalli, Uma & Price, Randy & Bomma, Pallavi. (2006). A Comparison of Image Processing Techniques for Bird Recognition. Biotechnology progress. 22. 9-13. 10.1021/bp0500922.

Prof. Pralhad Gavali , Ms. Prachi Abhijeet Mhetre , Ms. Neha Chandrakhant Patil , Ms. Nikita Suresh Bamane, Ms. Harshal Dipak Buva, 2019, Bird Species Identification using Deep Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 04 (April – 2019)

R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 2023-2027, doi: 10.1109/TENCON.2018.8650517.

S. Lee, M. Lee, H. Jeon and A. Smith, "Bird Detection in Agriculture Environment using Image Processing and Neural Network," 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), Paris, France, 2019, pp. 1658-1663, doi: 10.1109/CoDIT.2019.8820331.

Sankupellay, Mangalam & Konovalov, Dmitry. (2018). Bird Call Recognition using Deep Convolutional Neural Network, ResNet-50. 10.13140/RG.2.2.31865.31847.

Sharma, Kartik & Thakur, Nileshsingh. (2017). A review and an approach for object detection in images. International Journal of Computational Vision and Robotics. 7. 196. 10.1504/IJCVR.2017.081234.

W. Zhiqiang and L. Jun, "A review of object detection based on convolutional neural network," 2017 36th Chinese Control Conference (CCC), Dalian, 2017, pp. 11104-11109, doi: 10.23919/ChiCC.2017.8029130.

X. Zhou, W. Gong, W. Fu and F. Du, "Application of deep learning in object detection," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 631-634, doi: 10.1109/ICIS.2017.7960069.

Zhu, Bo-Cheng & Lin, Tzung-Han & Tsai, Yao-Chuan & Hsieh, Kuang-Wen & Fan, Fuh-Min & Lei, Perng-Kwei. (2019). Outdoor Wild Bird Detection based on YOLO algorithm. Proceedings of the International Display Workshops. 36. 10.36463/idw.2019.0036.

Zou, Cong & Liang, Yong-quan. (2020). Bird Detection on Transmission Lines Based on DC-YOLO Model. 10.1007/978-3-030-46931-3_21.