Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20240302014
Volume 3, Issue2 (May-August 2024), PP: 105-109.
www.indjcst.com
ISSN No: 2583-5300

Bird Species Detection Using Deep Learning

S. Gopalakrishnan[1], Naveen B[2], Lochan Krishnan R[3], Nishanth S[4], Ashraf B[5]
[1]*Assistant Professor, Department of Information Technology, Er. Perumal Manimekalai College of Engineering, Hosur, Tamilnadu, India.*
[2,3,4,5] *Department of Information Technology, Er. Perumal Manimekalai College of Engineering, Hosur, Tamilnadu, India.*

To Cite this Article: *S. Gopalakrishnan[1], Naveen B[2], Lochan Krishnan R[3], Nishanth S[4], Ashraf B[5], "Bird Species Detection Using Deep Learning", Indian Journal of Computer Science and Technology, Volume 03, Issue 02 (May-August 2024), PP: 105-109.*

Abstract: *Bird species classification plays a crucial role in various ecological and conservation endeavors. Leveraging deep learning techniques has shown promise in automating this task, offering the potential for efficient and accurate species identification. This paper presents a comprehensive study on the application of the YOLOv5 architecture, a state-of-the-art object detection framework, for bird species classification from images. We explore the effectiveness of YOLOv5 in comparison to traditional convolutional neural networks (CNNs) for bird species identification. Our study involves extensive experimentation with different configurations of YOLOv5, including variations in model size, training strategies, and dataset preprocessing techniques. We evaluate the performance of the YOLOv5-based classification framework on diverse datasets, ranging from publicly available bird image datasets to custom datasets collected through field observations. Additionally, we compare the classification accuracy of YOLOv5 with other popular CNN architectures, such as ResNet and EfficientNet, to assess its efficacy in bird species recognition tasks. Furthermore, we investigate the transferability of pre-trained YOLOv5 models to different bird species datasets and examine the robustness of the models to variations in image quality, background clutter, and occlusions.Our experimental results demonstrate that YOLOv5 offers competitive performance in bird species classification tasks, achieving high accuracy and efficiency. We discuss the strengths and limitations of using YOLOv5 for this application and provide insights into potential avenues for future research.*

Key Words: *YOLO version 5, Object Detection, Deep Learning, Image-based Classification, Accuracy*.

I.INTRODUCTION

Bird species classification plays a pivotal role in ornithology, serving as the cornerstone of biodiversity conservation efforts, ecological research endeavors, and comprehensive environmental monitoring strategies. Traditionally, the identification of bird species has heavily relied upon the expertise of trained observers, a process susceptible to time constraints, subjectivity, and inherent human error. However, the advent of deep learning methodologies has heralded a new era in this field, offering automated solutions that promise to revolutionize the way we classify and understand avian diversity. Among these innovative techniques, convolutional neural networks (CNNs) have emerged as powerful tools in the realm of computer vision, exhibiting remarkable proficiency across a spectrum of tasks, ranging from image classification to object detection. Within this landscape, the You Only Look Once (YOLO) architecture, particularly its latest iteration, YOLOv5, has garnered significant attention for its unparalleled efficiency and accuracy in the domain of object detection. The allure of YOLOv5 lies in its ability to swiftly and accurately identify objects of interest within complex visual scenes, making it an enticing prospect for bird species classification endeavors.

This project embarks on a comprehensive exploration of the efficacy of the YOLOv5 architecture in the context of classifying bird species from images. By harnessing the robust capabilities of YOLOv5 in object detection, the overarching objective is to develop a sophisticated and streamlined framework capable of automating the intricate process of bird species identification. The multifaceted nature of this endeavor encompasses meticulous dataset curation, rigorous model training procedures, exhaustive evaluation protocols, and insightful analysis methodologies. Within the pages of this report, we endeavor to provide a comprehensive and detailed exposition of our approach, delving into the intricacies of our methodology, elucidating the nuances of our experimental setup, and presenting a thorough analysis of the results obtained. Central to our investigation is the comparative assessment of YOLOv5-based models against alternative deep learning architectures, allowing for a nuanced understanding of its performance metrics across a diverse array of bird species datasets. Furthermore, we meticulously scrutinize the robustness of these models in the face of environmental perturbations, image variations, and dataset idiosyncrasies, thereby providing valuable insights into their real-world applicability and reliability. Moreover, the implications of our findings extend far beyond the realms of academia, reverberating throughout the spheres of biodiversity conservation, ecosystem management, and wildlife preservation. By showcasing the unparalleled capabilities of YOLOv5 in the domain of bird species classification, this project serves as a testament to the transformative potential of deep learning methodologies in addressing pressing environmental challenges and advancing the frontiers of ecological research. Through our collective efforts, we strive to pave the way for the

development of innovative and effective tools that are poised to redefine the landscape of avian conservation and environmental stewardship.

II.RELATED WORK

The earlier approaches for the bird species identification used bird songs to identify the birds. The visual features i.e. SIFT (Scale invariant feature transform) [11] from bird images and acoustic features both were used in the classical Machine learning algorithms to train a standard support vector machine (SVM) for classification. The fine-grained visual cate- gorization methods [12] have shown great and better results of image classification and within computer vision research they have become a promising approach. For the generic object recognition numerous techniques have been applied [13]. A few strategies applied local part learning that uses deformable part models and region-CNN for object recognition [14], bounding box generation, and distinctive parts selection for image recognition.Some researches have centered on discriminative features based on the local traits of bird species [15], [17]. Simultaneous detection and segmentation are used to localize score detections effectively [16]. Pose-normalization and model ensembles [17] are also utilized to progress the execution of fine-grained detection by producing millions of key point sets through fully convolutional search. There are lots of strategies are available to distinguish the birds but they are costly and time consuming. There has been an increasing interest for automated acoustic monitoring of sound-emitting creatures, which may give reliable Information on the presence/ absence of target species and on the common biodiversity status of an area in recent a long time [18].

Xie et al. [10], investigated three types of time frequency representations (TFRs) such as Mel-spectrogram, harmonic component-based spectrogram, and percussive component based spectrogram of bird sounds to characterize the different acoustic components of birds to identify particular bird species. Stavros Ntalampiras et al. [18], statistically analysed the similarities in between the audio signal of the bird and music genres rather than looking at the bird's audio signal alone. And for that they utilized the transfer learning technology. Loris Nanni et al. [19], presented a combination of classifiers including AlexNet [20], GoogleNet [21], VGGNet [22], ResNet, and InceptionV3 to identify the bird species by processing its audio signal. The audio images such as spectrograms, ScatNet [23,24] scattering representations, and harmonic and percussion images are extracted from the bird's audio signal for the classification and prediction processes. In all these cases there might be a probability of occurring background noise such as environmental noise, sounds of insects while recording the bird's song. Thus it leads to the misclassification and reduces the probability to get accurate prediction results. The similarities existing among the songs of birds also provides inaccurate predictions about the species of the bird. Jiaohua Qin et al. [25], replaced fixed size images with appropriately large size images as input to convolutional neural network and few modules in that were replaced with an Inverted Residual Block module inorder to reduce the network parameters and computational cost. The convolution layer, Batch Normalization layer, ReLU activation function, Global Average Pooling, and Softmax were included to form this improved network. But they haven't specialized any animal dataset or any object dataset for the processes and focused mainly on general biological images. Plenty of techniques are available to classify the biological images into different categories and few of them are focused more on species identification. This proposed work also approached the species identification using convolutional neural network architecture.

III.DATA SET

For the bird image recognition, it is required to have a solid dataset on which the identification system can be trained, tested, and validated. Thus we used one of the fine- grained biological image classification dataset Caltech-UCSD Birds 200–2011[26]. Caltech and UCSD have gathered data to produce this particular dataset and it is the extended version of the CUB-200 dataset. Figure 1 shows the Caltech-UCSD Birds 200 datasets. The dataset contains 11,788 images of 200 different categories of bird species. The dataset was splitted into training set, testing set, and validation set. It is very important to keep the testing set completely separate from the training set since it needed to be sure that the classification model will perform well in the real-world scenarios. The pixel values are normalized in order to reduce the harshness, noise and disturbances in the images. Then it can be used for training the classification model. More than 60% of data allocated to the training set and rest of the data allocated to the testing set and validation set. Training set and validation set are randomly selected from dataset for the fine-tuning process.

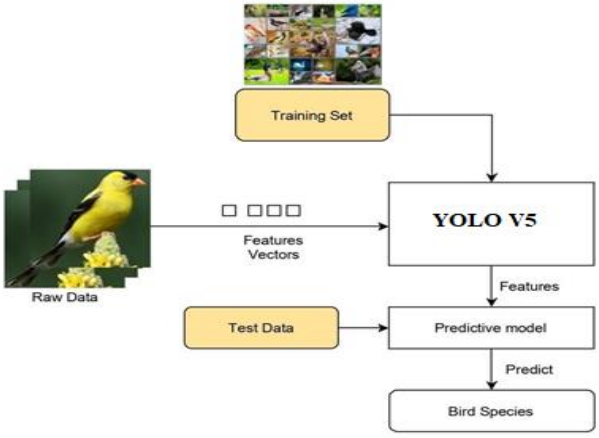

Figure 1 Architecture Diagram

IV. BIRD SPECIES IDENTIFICATION DEEP LEARNING FRAMEWORK

As of late, deep learning models have become the foremost well-known tool for artificial intelligence [27] and big data analysis. The rise of deep learning [28] algorithms has resulted in exceedingly complex cognitive tasks for computer vision and image recognition. The proposed deep learning model acquired to build this bird image classification system using the CNN framework is described as follows.

1. Building the YOLOV5

Building YOLOv5 involves designing an efficient object detection architecture optimized for real-time performance. YOLOv5 inherits the principles of the YOLO (You Only Look Once) series, focusing on speed and accuracy. It employs a single CNN to simultaneously predict bounding boxes and class probabilities. The architecture comprises convolutional layers, incorporating techniques like batch normalization (BN) for stable training and reduced training epochs. YOLOv5's architecture allows for flexibility in model size selection, with variations like YOLOv5s, YOLOv5m, catering to different performance and resource requirements. Training involves supervised learning on annotated datasets, optimizing parameters with gradient descent algorithms. Hyperparameter tuning and model evaluation ensure optimal performance. Inference with trained YOLOv5 models enables fast and accurate object detection in diverse applications. Overall, YOLOv5's building process emphasizes speed, accuracy, and efficiency in object detection tasks.

2. Skip Connection

In YOLOv5 architecture, skip connections serve as integral components to enhance model convergence and feature extraction. By offering an alternative pathway for gradients, skip connections provide additional routes for information flow during training, aiding in mitigating gradient vanishing issues commonly encountered in deep architectures. As the name suggests, these connections skip certain layers within the neural network, enabling the output of one layer to directly influence subsequent layers, fostering feature reusability and facilitating convergence. The strategic incorporation of skip connections among corresponding convolutional layers in YOLOv5, as depicted in the architectural design, aims to quantify the transition from general to specific features across network layers. Through weighted summation of feature maps from different layers, skip connections contribute to improved feature extraction and network stability, thus enhancing training efficiency and convergence. Overall, the utilization of skip connections in YOLOv5 underscores their significance in optimizing model performance and facilitating effective training processes The skip layer connections should improve feature extraction through weighted summation of corresponding layers as follows [1]:

$$G(X) = (1 - \alpha) F(X) + \alpha X \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where X is the input, G(X) is a linear combination of F(X) and X, F(X) is a function of input X, and α is a weight in the unit interval [0,1]. Result from the previous layer contributes less to overall performance than the layers preceding it if the weight α is greater than 0.5. And the result from the previous layer contributes more to the overall performance if the weight α is lesser than 0.5. By using the skip connections we can enable feature reusability and stabilize training and convergence and we can facilitate network training too.

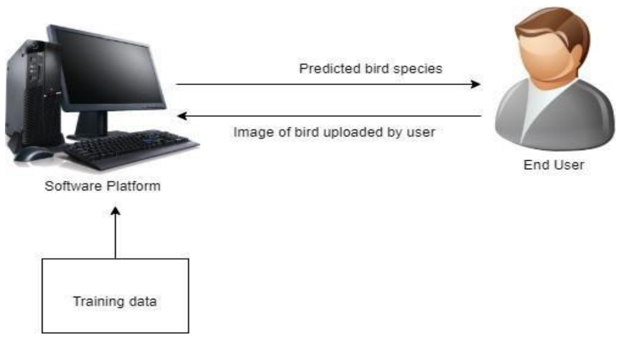

Figure 2 Client-server Architecture

3. Feature Extraction

In YOLOv5, feature extraction is fundamental for precise object detection, involving the extraction of relevant details from input images. The architecture encompasses multiple convolutional layers, systematically unraveling hierarchical features .Beginning with early layers, basic patterns like edges and corners are extracted, constituting low-level features. As the network progresses, subsequent layers build upon these features, discerning more intricate attributes such as object shapes and textures. Batch Normalization (BN) layers ensure training stability by standardizing inputs across mini-batches, while Rectified Linear Unit (ReLU) activation functions introduce non-linearity, aiding in pattern recognition.Pooling layers reduce spatial dimensions, capturing salient features while discarding redundant information, thus enhancing computational efficiency. Additionally, skip connections facilitate gradient flow and feature reuse across layers, enabling the transfer of features from early to later layers. Overall, feature extraction in YOLOv5 follows a hierarchical process through convolutional layers, batch normalization, activation functions, pooling layers, and skip connections. These extracted features serve as the foundation for subsequent object detection and classification, demonstrating the effectiveness of YOLOv5 in real-world applications normalization, activation functions, pooling layers, and skip connections. These extracted features serve as the foundation for subsequent object detection and classification, demonstrating the effectiveness of YOLOv5 in real-world applications.

4. System Implementation

The initial step involves accepting bird images in various formats, resolutions, and color spaces. Preprocessing ensures consistency and optimal performance by resizing images to a standardized resolution, converting them to RGB color space, and normalizing pixel values. Noise reduction techniques may also be applied to enhance image quality.YOLOv5, renowned for object detection, employs a deep architecture with multiple convolutional layers to extract intricate features from images. Analyzing images at different scales and abstraction levels, it identifies patterns, textures, shapes, and spatial relationships. As images traverse the network, features are refined, providing a comprehensive representation of detected objects, including birds. Extracted features are then inputted into YOLOv5's classification module, where advanced algorithms assign class labels to detected objects. Trained to recognize various bird species, the model categorizes birds based on distinct features. Classification involves computing probability scores for each class label using a softmax activation function, predicting the species with the highest probability.YOLOv5 generates output comprising bounding boxes around detected birds, along with class labels and confidence scores. These bounding boxes offer spatial localization information, indicating the position and size of each detected bird within the image. Class labels and confidence scores provide insights into the model's certainty, aiding users in understanding result reliability .Deployment typically employs a client-server architecture for accessibility and scalability. Clients, like web or mobile applications, send image requests to a server housing the YOLOv5 model. The server processes requests, executing model inference on input images, and returns detection outcomes to clients. This architecture facilitates remote interaction, enabling bird species identification from any location with internet access.

V.RESULTS AND DISCUSSION

This section explains the details about the experimental results of the system used to identify the bird species. The bird and non-bird images can be differentiated and predicted using this proposed deep learning framework. The bird species identification system sends an error notification to upload only the images that contains a bird, when non-bird images are uploaded to the identification system. 100 bird images were uploaded from a mobile phone for preliminary testing in order to validate the effectiveness and efficiency of the proposed bird species identification system and to filter non- bird images uploaded to the system automatically. For the differentiation and classification of the images as true bird images, the proposed model achieved 100 percentage accuracy. The bird detection results are shown in Table 1.

Table 1 Prediction results of images uploaded from an end-user device

Subjects	Predicted as Bird Image	Predicted as Non-bird Image
Image containing bird	100%	0%
Image containing non-bird	0%	%

Comparing the performance of YOLOv5 with other models like Support Vector Machine (SVM) involved training all three models on the same dataset under identical conditions. Each model was trained with a learning rate of 0.00001 over 100 epochs. For SVM, a linear kernel was utilized to address the high dimensionality of the feature space.Results indicated that the YOLOv5 model outperformed both the traditional CNN model without skip connections and the SVM model in terms of accuracy. The YOLOv5 model achieved notably higher accuracy compared to both the CNN model without skip connections and the SVM model. This finding underscores the efficacy and efficiency of YOLOv5 for bird species identification tasks.The development of an automatic deep learning model leveraging. YOLOv5 aimed to enhance efficiency and effectiveness in predicting different bird species. Through empirical studies, the YOLOv5 architecture was evaluated for its effectiveness in addressing challenges such as the vanishing gradient problem. While the focus was on predicting various bird species with increased efficiency, YOLOv5 demonstrated exceptional performance, achieving high accuracy in identifying uploaded bird images. This underscores the potential of YOLOv5, particularly in tasks requiring precise species identification, such as bird species classification

VI.CONCLUSION

Employing yolov5 for bird species classification using CPU resources has proven to be a highly effective approach, capable of handling a large number of species, such as the 200 species considered in this study. The yolov5 algorithm offers efficient object detection and classification, making it suitable for real-time applications even on CPU-based systems. By leveraging deep learning techniques, the model demonstrates robustness and accuracy in identifying diverse bird species from images. This advancement holds significant promise for biodiversity monitoring, conservation efforts, and ecological research, offering a scalable and accessible solution for automated species identification. As computational resources continue to improve, yolov5 stands as a powerful tool in the arsenal of techniques for bird species classification, paving the way for enhanced understanding and preservation of avian biodiversity

REFERENCES

1. *Yo-Ping Huang and Haobijam Basanta, "Bird Image Retrieval and Recognition Using a Deep Learning Platform", IEEE Access, Vol. 7, May 2019.*
2. *O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei- Fei, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211252, Dec. 2015.*
3. *H. Yao, S. Zhang, Y. Zhang, J. Li, and Q. Tian, "Coarse-to-ne description for ne-grained visual categorization," IEEE Trans. Image Process., vol. 25, no. 10, pp. 48584872, Oct. 2016.*
4. *F. Garcia, J. Cervantes, A. Lopez, and M. Alvarado, "Fruit classication by extracting color chromaticity, shape and texture features: Towards an application for supermarkets," IEEE Latin Amer. Trans., vol. 14, no. 7, pp. 34343443, Jul. 2016.*
5. *L. Zhu, J. Shen, H. Jin, L. Xie, and R. Zheng, "Landmark classication with hierarchical multi-modal exemplar feature," IEEE Trans. Multime- dia, vol. 17, no. 7, pp. 981993, Jul. 2015.*
6. *X. Liang, L. Lin, W. Yang, P. Luo, J. Huang, and S. Yan, "Clothes co- parsing via joint image segmentation and labeling with application to clothing retrieval," IEEE Trans. Multimedia, vol. 18, no. 6, pp. 11751186, Jun. 2016.*
7. *Y.-P. Huang, L. Sithole, and T.-T. Lee, "Structure from motion technique for scene detection using autonomous drone navigation," IEEE Trans. Syst., Man, Cybern., Syst., to be published.*
8. *C. McCool, I. Sa, F. Dayoub, C. Lehnert, T. Perez, and B. Upcroft, "Visual detection of occluded crop: For automated harvesting," in Proc. Int. Conf. Robot. Autom. (ICRA), Stockholm, Sweden, May 2016, pp. 25062512.*
9. *A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classication with deep convolutional neural networks," in Proc. 25th Int. Conf. Advance Neural Inf. Process. Syst., Lake Tahoe, NV, USA, Dec. 2012, pp. 10971105.*
10. *Jie Xie, Kai Hu, Mingying Zhu, Jinghu Yu, Qibing Zhu, "Investigation of Different CNN-Based Models for Improved Bird Sound Classification", IEEE Access, vol. 7, pp. 175353-175361, 2019.*
11. *Marini, A., Turatti, A. J., Britto, A. S., Koerich, A. L. (2015). Visual and acoustic identification of bird species. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015).*
12. *"Bird Species Categorization Using Pose Normalized Deep Convolu- tional Net" Steve Branson, Grant Van Horn, Serge Belongie ,Pietro Peron (2015).*
13. *Li Liu, W. Ouyang, X. Wang, P. Fieguth, X. Liu, and M. Pietika¨inen, "Deep learning for generic object detection: A survey," Sep. 2018, arXiv:1809.02165. [Online]. Available: https://arxiv.org/abs/1809.02165*
14. *K. Dhindsa, K. D. Gauder, K. A. Marszalek, B. Terpou, and S. Becker, "Progressive thresholding: Shaping and specicity in automated neurofeedback training," IEEE IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 12, pp. 22972305, Dec. 2018.*
15. *C.-Y. Lee, A. Bhardwaj,W. Di, V. Jagadeesh, and R. Piramuthu, "Region- based discriminative feature pooling for scene text recognition," in Proc. Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 40504057.*
16. *B. Hariharan, P. Arbela´ez, R. Girshick, and J. Malik, "Simultaneous detection and segmentation," in Proc. Eur. Conf. Comput. Vis., Jul. 2014, pp. 297312.*
17. *S. Branson, G. V. Horn, S. Belongie, and P. Perona, "Bird species categorization using pose normalized deep convolutional nets," in. Proc. Brit. Mach. Vis. Conf., Nottingham, U.K., Jun. 2014, pp. 114.*
18. *Stavros Ntalampiras, "Bird species identification via transfer learning from music genres", Ecological Informatics, Vol. 44, March 2018.*
19. *Loris Nanni, Yandre M. G. Costa, Rafael L. Aguiar, Rafael B. Mangolin, Sheryl Brahnam and Carlos N. Silla Jr., "Ensemble of convolutional neural networks to improve animal audio classification", EURASIP Journal on Audio, Speech, and Music Processing, Article number: 8(2020), May 2020.*
20. *A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.*
21. *C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,*
22. *V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, 2015.*
23. *K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint (2014). arXiv:1409.1556.*
24. *J. Bruna, S. Mallat, "Invariant scattering convolution networks", IEEE Trans. Pattern Anal. Mach. Intell. 35(8), 1872–1886 (2013).*
25. *L. Sifre, S. Mallat, "Combined scattering for rotation invariant texture analysis", ESANN, vol. 44, pp. 68–81, 2012.*
26. *Jiaohua Qin, Wenyan Pan, Xuyu Xiang, Yun Tan, Guimin Hou, "A biological image classification method based on improved CNN ", Ecological Informatics, Vol. 58, July 2020.*
27. *C. Wah et al. The Caltech-UCSD Birds-200-2011 Dataset. Tech. rep. CNS-TR-2011-001. California Institute of Technology, 2011.*
28. *J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, Q. V. Le, and A. Y. Ng, "Large scale distributed deep networks," in Proc. 25th Int. Conf. Adv. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, Dec. 2012, pp. 12231231.*
29. *L. Yang, A. M. MacEachren, P. Mitra, and T. Onorati, "Visually-enabled active deep learning for (Geo) text and image classication: A review," Int. J. Geo-Inf., vol. 7, no. 2, p. 65, Feb. 2018.*
30. *J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks," in Proc. Int. Conf. Advance Neural Inf. Process. Syst., Dec. 2014, pp. 33203328.*
31. *K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770778.*
32. *N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R- CNNs for ne-grained category detection," in Proc. Int. Conf. Eur. Conf. Comput. Vis., Cham, Switzerland, Jul. 2014, pp. 834849.*
33. *C. Huang, Z. He, G. Cao, and W. Cao, "Task-driven progressive part localization for ne-grained object recognition," IEEE Trans. Multimedia, vol. 18, no. 12, pp. 23722383, Dec. 2016.*