Bird Species Identification Using YOLO Neural Network

Yawei Li

Aircraft Control and Information Engineering, Beihang University, Beijing, 102206, China
Leeyawei3@gmail.com

Abstract. Ecological conservation efforts increasingly rely on biodiversity indicators, with bird populations serving as critical sentinels of ecosystem health. Bird conservation is fundamental to maintaining biodiversity and health. Thus, bird species monitoring constitutes a crucial part of conservation efforts. Because of the situation that traditional bird identification methods have low efficiency and the insufficient focus of existing deep learning models on fine-grained species recognition, this study applies the version 11 of You Only Look Once (YOLO) neural network for bird species identification. Experiments employ two distinct datasets with varying background complexities and target quantities to simulate ideal and realistic scenarios. Under ideal conditions, the model maintains 100% TOP5 accuracy. In simulated realistic scenarios, the model achieves 76.1% TOP5 accuracy. This research demonstrates that YOLOv11 exhibits high efficiency and strong generalization capabilities for bird species identification tasks, providing a feasible technical solution for ecological monitoring. The solution is a robust, automated technical solution that significantly enhances the feasibility and scalability of large-scale bird monitoring programs, ultimately strengthening data-driven decision-making in ecological conservation and habitat management practices.

Keywords: YOLO, Computer Vision, Bird Species.

1 Introduction

In natural ecosystems, birds serve as an important component due to their species and special behaviors. So bird species play an irreplaceable role in maintaining ecological balance and protecting biodiversity [1]. Consequently, monitoring and identifying wild bird populations for targeted conservation is critical. However, traditional bird identification methods primarily rely on manual observation and documentation. This approach consumes substantial human resources and time, while struggling to meet long-term or large-scale monitoring demands [2].

With advancements in computer vision, bird monitoring has entered an automated processing phase. Early bird image detection methods distinguished species by manually extracting features [3,4]. Such human-driven key points and feature extraction demonstrated limited performance in complex scenarios [5]. In recent years, deep learning progress in computer vision has introduced new approaches for bird species identification [6]. Nevertheless, most current deep learning methods focus

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_14

solely on bird presence detection rather than species recognition. For instance, Jang et al. implemented single-species detection using Faster Regions with Convolutional Neural Network features (R-CNN) [7]. Thus, bird species identification remains an open research challenge.

Deep learning recognition models are broadly categorized into one-stage and two-stage architectures. Among these, two-stage models evolved from the original R-CNN through iterative improvements into Faster R-CNN [8]. Compared to one-stage models, two-stage models exhibit inferior real-time performance. Conversely, one-stage models like the You Only Look Once (YOLO) series offer stronger real-time capabilities, better aligning with biological monitoring requirements in agricultural and forestry environments [9].

This study adopts the YOLOv11 model, utilizing the publicly available dataset "BIRDS 20 SPECIES-IMAGE CLASSIFICATION" for training and ideal condition experiments [10]. The "200k images of birds (iNaturalist)" dataset is employed for simulated realistic scenario testing [11]. The research on bird species identification via the YOLOv11 neural network provides valuable insights for avian monitoring and conservation efforts.

2 Methods

2.1 Data Source and Statement

The BIRDS 20 SPECIES-IMAGE CLASSIFICATION dataset comprises 20 bird species. Each category contains approximately 160 images, resulting in 3,308 total images. Five images per category are allocated to the validation set, while the remaining images form the training set. This yields 100 validation images and 3,208 training images. Additionally, most images in this dataset feature a single dominant subject occupying over 50% of the frame against simple backgrounds, representing ideal conditions.

Conversely, the "200k images of birds (iNaturalist)" dataset includes 1,486 bird species. Each species averages 140 images without predefined test splits. Numerous images contain multiple birds. Besides, birds occupy less than 50% of the frame amid complex backgrounds, effectively simulating real-world observational scenarios.

For the cross-dataset consistency, this research utilizes 10 species overlapping between both datasets since the remaining 10 species in BIRDS 20 SPECIES are absent from iNaturalist. By combining the 10 overlapping species from iNaturalist with images of the non-overlapping 10 species from BIRDS 20 SPECIES, a simulated realistic validation set of 20 species is constructed, and it includes 1450 images.

2.2 Methodology Introduction

The YOLO series represents highly efficient object detection and recognition models. Innovatively, the YOLO algorithm approaches object detection as a regression problem and accomplishes detection through an end-to-end framework.

This study employs the YOLOv11 model, which comprises three core components: a backbone network, a neck network, and a head detector [12]. The backbone network

extracts multi-scale features from input images. Subsequently, the neck network fuses and refines these multiscale features. Finally, the head detector performs the ultimate detection and recognition tasks.

For detection tasks, the head detector contains sub-networks responsible for classification, localization, and instance segmentation. However, in the present recognition study, the head detector excludes localization and segmentation tasks, focusing on the classification sub-network. The network architecture diagram of YOLOv11 is illustrated in Figure 1.

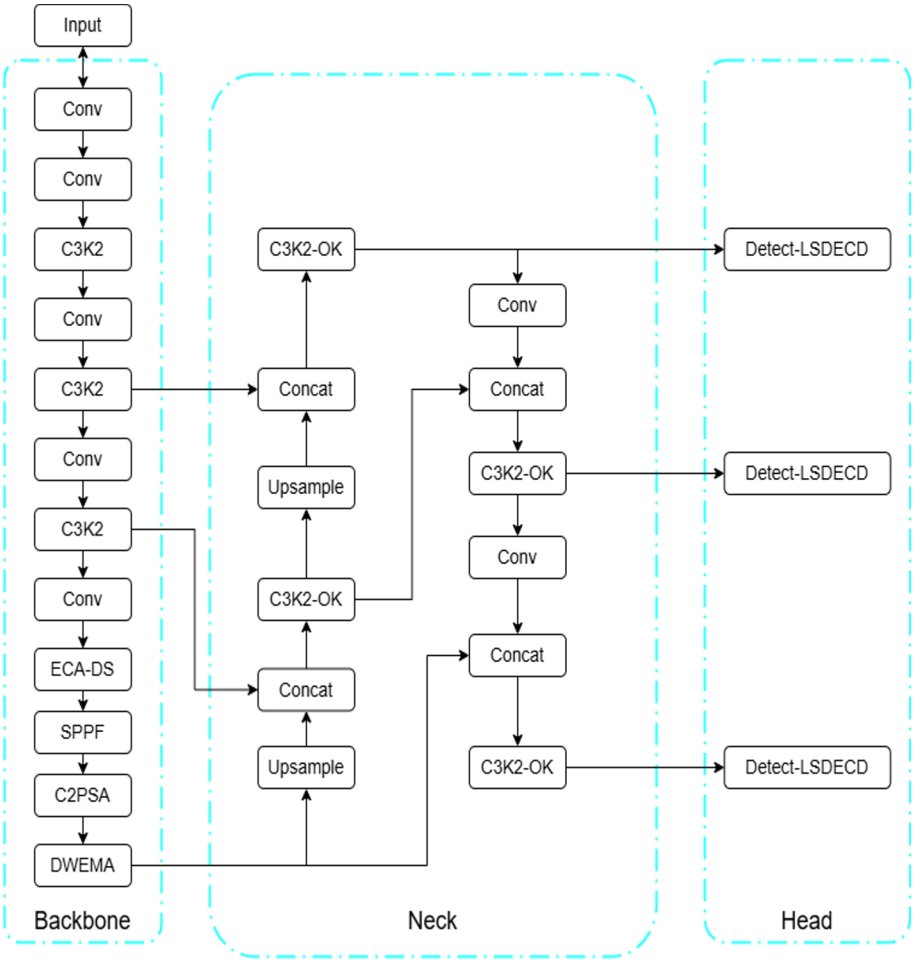

Fig. 1. The network architecture diagram of YOLOv11[12]

3 Results and Discussion

The primary evaluation metrics in this experiment are TOP5 accuracy and the normalized confusion matrix. TOP5 accuracy measures the classification correctness rate of bird species predictions in images. Meanwhile, the normalized confusion matrix analyzes accuracy variations across different bird species.

First, YOLOv11 is trained using the BIRDS 20 SPECIES-IMAGE CLASSIFICATION dataset to develop a classification model for bird species identification. Subsequently, both qualitative and quantitative experiments are conducted to validate the model's classification capability and generalization performance under ideal and simulated realistic conditions.

The initial experiments utilize the validation set partitioned from the BIRDS 20 SPECIES dataset. This validation set contains 100 images, with 5 images per species. All images feature a single dominant subject occupying over 50% of the frame, representing an ideal scenario validation set. Figure 2 displays a part of the recognition results from this validation set.

Fig. 2. recognition results from the ideal validation set (Picture credit: Original)

Figure 2 demonstrates the model's high performance on the ideal condition validation set. Notably, it effectively classifies bird species in input images. According to experimental results, all classifications are correct in this validation set. Moreover, the normalized confusion matrix for these classification outcomes appears below.

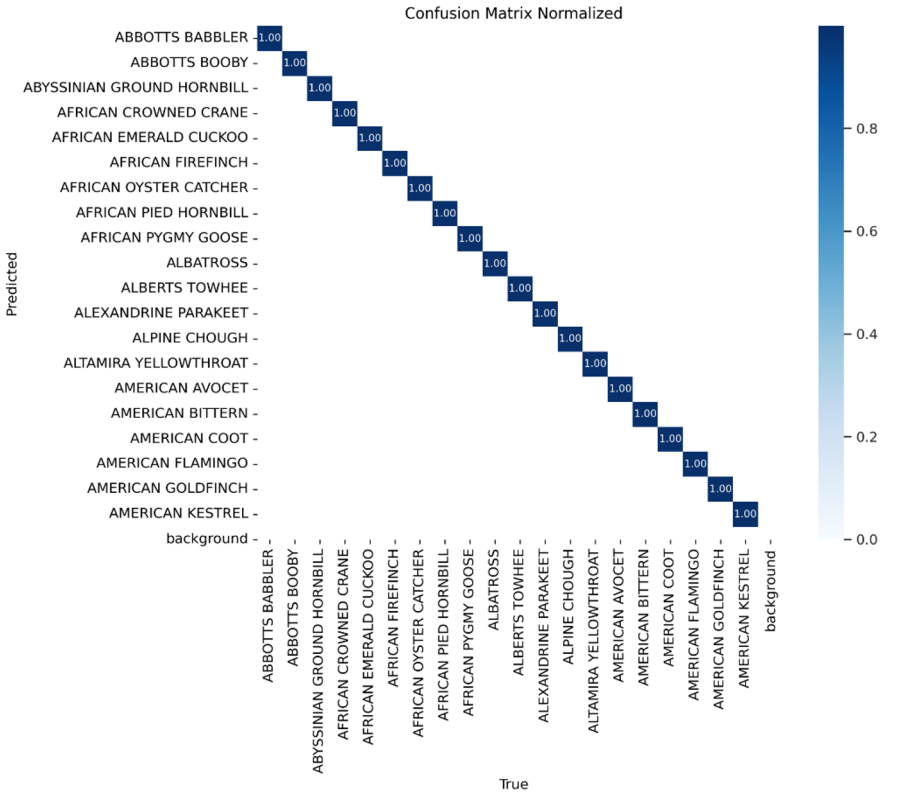

Fig. 3. normalized confusion matrix of the ideal validation set (Picture credit: Original).

Figure 3 indicates good classification performance on the ideal validation set with five images per species. All images are correctly classified. Consequently, the model achieves precision, recall, and F1 scores of 1.0 under ideal conditions. Furthermore, training progress metrics including train loss, validation loss, TOP5 accuracy, and TOP1 accuracy appear in Figure 4.

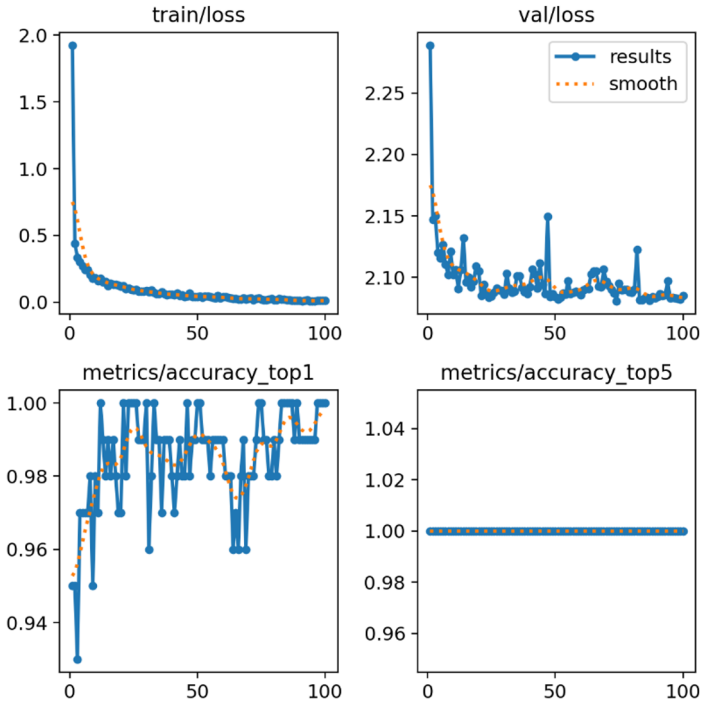

Fig. 4. The diagrams of model training progress metrics (Picture credit: Original)

Figure 4 demonstrates normal decreasing loss values, indicating stable training and model convergence. During later training stages, TOP1 accuracy stabilizes at high levels while TOP5 accuracy maintains 100% correctness.

Comprehensive analysis of Figures 2， Figure3 and Figure 4 confirmed excellent model performance under ideal conditions. To evaluate generalization capability, this study subsequently constructs a new validation set using challenging data from the iNaturalist 2021 Birds dataset, which contains multi-subject images and complex backgrounds for simulating real-world bird observation scenarios. Specifically, ten bird species overlapping between iNaturalist 2021 and the BIRDS 20 SPECIES supplement the original validation set. Consequently, the simulated validation set integrates ten species from iNaturalist 2021 (complex scenarios) and ten species from BIRDS 20 SPECIES (ideal-case images), totaling 1,450 images. The normalized confusion matrix for this validation set appears in Figure 5.

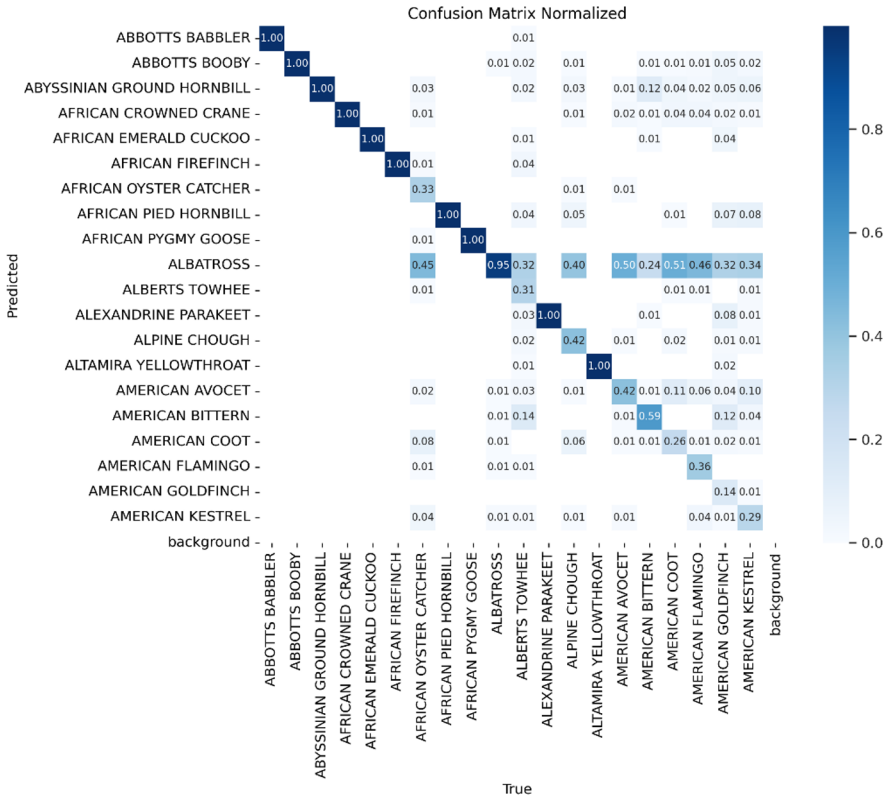

Fig. 5. Normalized confusion matrix of simulated realistic validation set (Picture credit: Original)

Figure 5 reveals a primary misclassification issue in the simulated realistic validation set, which is the frequent erroneous assignment of other bird species to the Albatross category. This occurs because the iNaturalist 2021 Birds dataset contains images with multiple birds and complex backgrounds, whereas the training set predominantly features single-subject images against simple backgrounds. Nonetheless, the model achieves 76.1% TOP5 accuracy, demonstrating robust generalization capability. Relevant experimental metrics are summarized in Table 1.

Table 1. Results of the experimental

validation set	precision	recall	F1	TOP5_ACCURACY
20 SPECIES	1.0	1.0	1.0	1.0
20 SPECIES+ iNaturalist	0.704	0.556	0.621	0.761

Qualitative and quantitative analyses in this section demonstrate the model's strong performance for bird species identification tasks. Under ideal conditions, it achieves high accuracy rates. While maintaining notable generalization capability in complex

simulated realistic scenarios. Overall, the model demonstrates practical utility in avian recognition applications.

4 Limitations and Future Prospects

Early research using neural networks for bird identification primarily focused on distinguishing birds from other objects. This approach addressed scenarios like airports requiring bird avoidance. Recent studies increasingly emphasize bird species classification, yet recognition accuracy requires improvement, especially for visually similar species. Future development directions include four key areas: lightweight models, multimodal systems, weakly-supervised learning, and large-scale modeling.

First, lightweight bird recognition models aim to enhance efficiency for real-time field monitoring. Researchers achieve this through knowledge distillation, transfer learning, and specialized hardware like Application-Specific Integrated Circuit (ASIC) or Field-Programmable Gate Array (FPGA) to boost training efficiency and inference speed [13].

Similarly, multimodal integration represents another critical direction. Incorporating bird vocalizations, motion videos, and even scent data can significantly improve system robustness. Future work may build multimodal systems leveraging diverse information sources to increase accuracy [14].

Conversely, weakly-supervised learning addresses the challenge of limited high-quality datasets. These methods enable model training with scarce or imperfectly labeled bird data.

Alternatively, large-scale modeling pursues opposing objectives to lightweight approaches yet remains essential. Constructing complex models trained on massive datasets could achieve high-precision recognition across diverse species [15].

Future optimizations for this study should prioritize diversifying training data. Consequently, this enhancement will address bird recognition challenges in complex multi-subject scenarios.

5 Conclusion

Bird monitoring is an essential component of avian conservation. This study successfully applies the YOLOv11 neural network to bird species identification tasks. In the ideal condition validation set, the model achieves 100% classification accuracy and an F1 score of 1.0, validating its high efficiency in single subject scenarios with simple backgrounds.

To further assess practical utility and generalization capability, a simulated real-world validation set is constructed, incorporating 1,450 bird images from both iNaturalist and BIRDS 20 SPECIES datasets. This set contains a significant proportion of multi-subject images and complex backgrounds. Results show 76.1% TOP5 accuracy and a 0.621 F1 score, confirming the model's moderate generalization ability despite misclassification issues particularly for albatross species.

These errors primarily stem from disparities in background complexity and bird count per image between training and test data. Nevertheless, this work offers a viable technical pathway for automated bird monitoring, potentially enhancing ecological conservation efficiency. Future research should prioritize enhancing the model's adaptability to multi-subject and complex environments, achievable through diversifying training data to improve robustness.

References

1. Das, S., Pradhan, B., Shit, P.K., et al.: Assessment of wetland ecosystem health using the pressure-state-response (PSR) model: A case study of Mursidabad district of West Bengal (India). Sustainability 12(15), 5932 (2020)
2. Fu, Y., Mao, B., Fang, X., et al.: Investigation and diversity analysis of bird resources in the Huanggang section of the Yangtze River main stream. Environmental Science Research 38(1), 68–77 (2025)
3. Xie, J., Zhu, M.: Handcrafted features and late fusion with deep learning for bird sound classification. Ecological Informatics 52, 74–81 (2019)
4. Kahl, S., Clapp, M., Hopping, W.A., et al.: Overview of BirdCLEF 2020: Bird sound recognition in complex acoustic environments. In: CLEF 2020 – Conference and Labs of the Evaluation Forum, pp. 262–270. CEUR-WS.org, Thessaloniki, Greece (2020)
5. Kumar, R., Kumar, A., Bhavsar, A.: Bird region detection in images with multi-scale HOG features and SVM scoring. In: Proceedings of 2nd International Conference on Computer Vision & Image Processing, pp. 353–364. Springer, Singapore (2018)
6. Hong, S.-J., Han, Y., Kim, S.-Y., et al.: Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. Sensors (Basel, Switzerland) 19(7) (2019)
7. Jang, W., Kim, T., Nam, U., et al.: Image segmentation and identification of parrot using Faster R-CNN. In: Proceedings of the ICNCT 11, 91–92 (2019)
8. Mao, R., Zhang, Y., Wang, Z., et al.: Identification of wheat stripe rust and dwarf virus using improved Faster-RCNN. Transactions of the Chinese Society of Agricultural Engineering 38(17), 176–185 (2022)
9. Wang, A., Chen, H., Liu, L., et al.: YOLOv10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems 37, 107984–108011 (2025)
10. Umair shah Pirzada, BIRDS 20 SPECIES- IMAGE CLASSIFICATION. Retrieved from https://www.kaggle.com/datasets/umairshahpirzada/birds-20-species-image-classificationn, (2023).
11. Sharan Sajiv Menon, 200k images of birds (iNaturalist). Retrieved from https://www.kaggle.com/datasets/sharansmenon/inatbirds100k, (2021).
12. Peng, L., Zhao, B.: Surface defect detection model for shaft hoistway based on MELE-YOLOv11n. Journal of Hubei Minzu University (Natural Science Edition), 1–6 (2025)
13. Wu, Z., Zhang, Y., Wang, X., et al.: Algorithm for detecting surface defects in wind turbines based on a lightweight YOLO model. Scientific Reports 14(1), 24558 (2024)
14. Zheng, X., Zheng, W., Xu, C.: A multi-modal fusion YOLO network for traffic detection. Computational Intelligence 40(2) (2023)
15. Xiong, Y., Jiang, Z., Li, Y., et al.: MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detection. Infrared Physics & Technology 118 (2021)