ARTICLE

Recognition of Bird Species of Yunnan Based on Improved ResNet18

Wei Yang[1,2,*] **and Ivy Kim D. Machica**[1]

[1]College of Information and Computing, University of Southeastern Philippines, Davao City, 8000, Philippines
[2]College of Big Data, Baoshan University, Baoshan, 678000, China
*Corresponding Author: Wei Yang. Email: wei04207@usep.edu.ph

ABSTRACT

Birds play a crucial role in maintaining ecological balance, making bird recognition technology a hot research topic. Traditional recognition methods have not achieved high accuracy in bird identification. This paper proposes an improved ResNet18 model to enhance the recognition rate of local bird species in Yunnan. First, a dataset containing five species of local birds in Yunnan was established: *C. amherstiae*, *T. caboti*, *Syrmaticus humiae*, *Polyplectron bicalcaratum*, and *Pucrasia macrolopha*. The improved ResNet18 model was then used to identify these species. This method replaces traditional convolution with depth wise separable convolution and introduces an SE (Squeeze and Excitation) module to improve the model's efficiency and accuracy. Compared to the traditional ResNet18 model, this improved model excels in implementing a wild bird classification solution, significantly reducing computational overhead and accelerating model training using low-power, lightweight hardware. Experimental analysis shows that the improved ResNet18 model achieved an accuracy of 98.57%, compared to 98.26% for the traditional Residual Network 18 layers (ResNet18) model.

1 Introduction

In the pursuit of long-term sustainable development, ecological environment protection must be considered a critical priority, with dedicated efforts focused on research, exploration, and practical applications. Accurate identification of bird species plays a crucial role in understanding their distribution, behavior, and population dynamics. Traditional methods for bird identification heavily rely on manual observation and field expertise, which are time-consuming and prone to subjective biases. This limitation impedes the efficient development and accurate identification of large-scale bird data. Artificial intelligence (AI) [1], particularly machine learning [2], and deep learning algorithms [3] employing convolutional neural networks (CNNs) [4], have emerged as groundbreaking technologies in recent decades. These technologies offer promising solutions for rapid and highly accurate image recognition, including bird identification [5]. As a result, they have the potential to significantly improve the efficiency of ecological studies [6].

Yunnan Province, situated in Southwestern China, is celebrated for its abundant biodiversity and serves as a habitat for a diverse array of bird species. The province's diverse topography [7], encompassing mountains, forests, wetlands, and rivers, offers habitats for numerous bird species. Yunnan is home to one-third of China's bird species [8], making its wild bird resources valuable for birdwatching and economic purposes. Despite the rich and varied diversity of wild bird species in Yunnan, the abundance of each species varies [9].

According to statistics, among the wild bird species in Yunnan, there are 136 rare species, accounting for 16% of the recorded species [8]; 336 uncommon species, making up 39.6% [8]; 363 common species, comprising 42.8% [10]; and 13 dominant species, accounting for 1.5% [11]. This indicates that while Yunnan has a wide variety of wild bird species, the populations of most species are relatively small, with low densities. The survival conditions for many species are fragile, and any disturbance or fragmentation of their habitats can easily lead to endangerment or extinction, especially for rare species [11].

According to literature, in China, the frogmouth family (Podargidae) and the Asian fairy-bluebird family (Irenidae) are found only in certain regions of Yunnan. Among the 110 species primarily found in India and Malaysia, such as the black-necked stork, black hornbill, and red-necked crake, they are only recorded in Yunnan, accounting for 13% of bird species in Yunnan [11]. Tropical bird families such as the jacamar family (Galbulidae), hornbill family (Bucerotidae), broadbill family (Eurylaimidae), pittas (Pittidae), leafbirds (Chloropseidae), flowerpeckers (Dicaeidae), woodpecker family (Picidae), bunting subfamily (Emberizinae), waxbill family (Estrildidae), and sunbird family (Nectariniidae) are primarily distributed in Yunnan [11].

In 2012, with the introduction of AlexNet (Alex Convolutional Neural Network) [12], deep learning models achieved significant success in general image recognition. Donahue et al. [13] attempted to apply CNNs to fine-grained image recognition, demonstrating the robust generalization capability of CNN features through experiments and introducing the concept of DeCAF (Deep Convolutional Activation Feature). This method analyzed CNN models trained on the ImageNet [14] dataset, using the output of AlexNet's first fully connected layer as image features. They discovered that CNN-extracted features contain stronger semantic information and higher discriminability compared to handcrafted features. DeCAF bridged the gap between CNNs and fine-grained image recognition, marking a significant shift in bird image recognition algorithms towards CNN features.

Although deep convolutional neural networks currently represent the most effective approach for bird recognition, they encounter challenges such as uncertain target sizes and difficulty in accurately extracting bird features amidst complex backgrounds. Therefore, this paper proposes a bird recognition method based on an improved ResNet18 (Residual Network 18 layers). The method automatically extracts bird features using neural networks and integrates a channel attention mechanism to broaden the scope of feature extraction. It replaces traditional convolution with depth-wise separable convolution to reduce model size and optimizes hyperparameters and training mechanisms to enhance the ResNet18 neural network model. This approach achieves high-quality bird recognition, thereby advancing the conservation efforts for bird biodiversity in Yunnan Province.

2 Related Work

Since deep CNNs achieved great success in general image recognition, many scholars have begun applying CNNs to various image recognition tasks. Common deep learning models such as AlexNet, ResNet, and VGG16 have been widely adopted across different industries.

Researchers Egbueze et al. [15] proposed utilizing the convolutional layers of the ResNet-15 model to extract basic features from images. Their experimental results demonstrated that this method could accurately identify weather conditions in images with higher accuracy, faster speed, and a smaller network model size.

Researchers Huo et al. [16] applied the deep mutual learning algorithm to traffic sign recognition. During the model training process, two ResNet-20 models were employed to learn from each other. The results indicated that this approach led to improved performance for each sub-network compared to training them individually.

For the rock image classification problem, Liu et al. [17] utilized the ResNet-50 residual network and validated it with their dataset. Experimental results showed significant enhancements in average recognition accuracy and reduction in loss.

Since Donahue et al. [13] successfully applied CNNs to bird recognition, many algorithms for bird image recognition have adopted CNN features. Researchers Wang et al. [18] demonstrated the identification of habitat elements using four end-to-end trained deep convolutional neural network (DCNN) models based on images. All proposed models achieved good results. The ResNet152-based model achieved the highest validation accuracy (95.52%), while the AlexNet-based model had the lowest test accuracy (89.48%). This study provided baseline results for the newly introduced database, addressing the issue of lacking robust public datasets for automatic habitat element recognition in bird images, thus enabling future benchmarking and evaluation.

Researchers Shazzadul et al. [19] classified different bird species in Bangladesh using bird images. They employed the VGGNet-16 (Visual Geometry Group 16-layer network) model to extract features from the images and applied various machine learning algorithms to classify bird species, achieving an accuracy of 89% using SVM and kernel methods.

In another study, researchers Biswas et al. [20] proposed a method for recognizing native birds in Bangladesh using transfer learning. Their research involved using six different CNN architectures: DenseNet201, InceptionResNetV2, MobileNetV2, ResNet50, ResNet152V2, and Xception.

Experiments have demonstrated that significant increases in the depth of convolutional neural networks do not always lead to improved performance [21]. In traditional neural networks, merely adding more layers can result in issues such as gradient vanishing or exploding. To address these challenges, deep residual networks (ResNet) introduced residual modules, with ResNet152 achieving first place in the ILSVRC 2015 classification task [21].

The receptive field is a fundamental concept in CNNs, defined as the area of input image pixels covered when each pixel in the feature map of a CNN layer is mapped back to the input image. Generally, as the number of network layers increases, the receptive field of each pixel in the feature map grows larger. A smaller receptive field is sensitive to detailed image information, while a larger one has a stronger ability to capture global information. However, it is challenging for a single layer in a convolutional network to effectively extract both detailed and global information simultaneously, often leading to feature maps that fail to accurately represent the relationship between image details and the entire scene.

In the context of wild bird images, where the positions and sizes of birds may vary significantly, feature maps may inadvertently capture complex background information rather than focusing on the birds' distinctive features. Therefore, this paper introduces a channel attention mechanism to enable the network to encompass a broader range of feature extraction. Additionally, it replaces traditional convolution with depth-wise separable convolution, aiming to enhance the accuracy of bird recognition.

3 Materials and Methods

3.1 Dataset Source

In this study, creating a dataset of local bird species in Yunnan is a crucial first step. Given the wide variety of bird species, totaling more than 400, we created the Gaoligong Mountain Bird Image Dataset (GMBID) based on the bird list from Reference [22] (in Chinese), as illustrated in Fig. 1. This dataset comprises images collected from 5 bird species, with 200 images per species, resulting in a total of 2000 images.

Figure 1: Sample of GMBID (Gaoligong mountain bird image dataset)

Collecting bird images from Gaoligong Mountain was the foundational task in establishing the dataset. Researchers conducted surveys to identify five bird species present on Gaoligong Mountain and recorded their names: *C. amherstiae*, *T. caboti*, *Syrmaticus humiae*, *Polyplectron bicalcaratum*, and *Pucrasia macrolopha*, as shown in Fig. 2a–e.

(a) C. amherstiae (b) T. caboti (c) Pucrasia.macrolopha

(d) Syrmaticus.humia (e) Polyplectron.bicalcaratum

Figure 2: Five bird species

Fig. 2a depicts *C. amherstiae* [23], primarily found in Yunnan, Sichuan, and Guizhou provinces of China. It features a red patch behind the head, a pale blue eye ring, yellow to red coloration on the back, a white belly, and a long tail with black and white stripes. Fig. 2b shows *T. caboti* [24], distributed in Yunnan, Zhejiang, and other regions of China. It has a brown body densely patterned with black, brown-yellow, and white fine lines. Fig. 2c shows *Pucrasia macrolopha* [25], characterized by a black beak, a dark metallic green head, long crest feathers, white patches on both sides of the neck, a gray-brown back with numerous "V"-shaped black stripes, white feather shafts resembling willow leaves in the middle of the stripes, and nearly gray tail feathers. Fig. 2d portrays *Syrmaticus humiae* [26], found in Yunnan and Guangxi regions of China. It has long white tail feathers with black or brown stripes, two white patches on the wings, bluish secondary feathers, a white lower back and waist with black scale-like markings, a purple gloss on the head and neck, and a red face. Fig. 2e displays *Polyplectron bicalcaratum* [27], mainly found in Yunnan. The male has black-brown feathers adorned with almost

pure white fine dots and stripes, and a brilliant eye-shaped pattern with purple or emerald metallic luster on the upper back, wings, and tail feathers.

For ease of recording during the training process, assigned short names to these 5 bird species. Table 1 lists these short names along with their corresponding conservation statuses. *C. amherstiae*, C.a, National First-Level Protected Animal. *T. caboti*, T.c, National First-Level Protected Animal. *Syrmaticus humiae*, S.h, National First-Level Protected Animal. **Polyplectron bicalcaratum**, P.b, National First-Level Protected Animal. *Pucrasia macrolopha*, P.m, National Second-Level Protected Animal.

Table 1: Research bird species names

Serial number	Species	Short name	Protection level
1	*C. amherstiae*	C.a	Second-level
2	*T. caboti*	T.c	First-level
3	*Syrmaticus humiae*	S.h	Second-level
4	*Polyplectron bicalcaratum*	P.b	First-level
5	*Pucrasia macrolopha*	P.m	Second-level

3.2 Preprocessing Techniques

This process primarily involves preparing the data to maximize the model's performance after training. All images in the dataset must be in the same format to facilitate easier model training. For image pre-processing, the image dimensions were standardized to 224×224 pixels. Batches for training were created, each containing 32 images. This approach is essential because training in batches allows for the comparison of expected results with actual results to calculate the error, which is then used to update the algorithm and improve the model. All images used are RGB images, so their dimensions are (224, 224, 3).

In deep learning approaches, an algorithm is systematically managed with a multi-layered neural network, where a large number of training samples are convolved within several hidden layers between the input and output layers. This process helps the network learn complex features, enhancing its ability to predict or classify sophisticated data. However, limited data and constrained resources can lead to overfitting, causing significant data features to be missed and affecting classification accuracy. To counter overfitting and address data imbalance, various data augmentation techniques were employed for the images of different bird species based on their features. Each species' images were augmented ten times, including the original images. Techniques used included 45° rotation, Gaussian noise addition, horizontal/vertical flipping, contrast enhancement, image sharpening, zoom range [0.7, 1.3], and affine transformations [28].

3.3 Model Training Including the Architecture

3.3.1 ResNet18 Model

To enhance the recognition accuracy of the model, it is common practice to deepen the network layers to extract higher-level features from images. However, as the number of layers increases, issues such as gradient vanishing, gradient explosion, and model degradation may arise, causing deep networks to underperform compared to shallower ones. To mitigate these challenges, Microsoft

Research introduced the deep residual neural network, ResNet, in 2016 [21]. Its key innovation lies in constructing ultra-deep network structures by introducing residual modules and leveraging Batch Normalization (BN) layers to expedite training. The residual modules address the problem of degradation, while BN layers mitigate issues related to gradient vanishing and explosion in neural networks. The structure of ResNet18 [29] is detailed in Table 2. The core component of ResNet, the residual module, is depicted in Fig. 3.

Table 2: Structure of ResNet18

| Layer name	ResNet18
Input layer	$224 \times 224 \times 3$
Convolutional layer	64 convolutional filters with size of 7×7, stride with 2
Down sampling layer	3×3 max pooling, stride with 2
Residual module	$\begin{vmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{vmatrix} \times 2$
Residual module	$\begin{vmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{vmatrix} \times 2$
Residual module	$\begin{vmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{vmatrix} \times 2$
Residual module	$\begin{vmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{vmatrix} \times 2$
Down sampling layer	7×7 global average pooling
Output layer	Fully connected layer with 1000 units
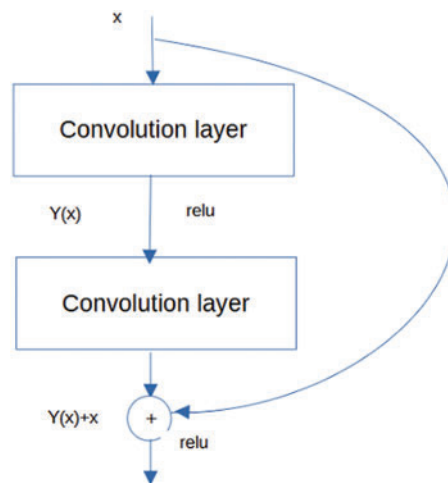
Figure 3: Structure of residual module

3.3.2 Improved ResNet18 Model

Depth-wise Separable Convolution [30] is an algorithm that improves standard convolution calculations. By separating the spatial and channel (depth) dimensions. This approach reduces the number of parameters required for convolution computations and has been demonstrated in several studies to enhance the efficiency of convolutional kernel parameters.

In paper [31], the authors combined depth-wise separable convolution with residual blocks to construct a convolutional layer for segmentation recognition, achieving precise real-time identification of beets and weeds.

Another paper [32] proposed a lightweight residual network model named DSC-Res14, based on depth-wise separable convolution. Initially, this model utilizes a three-dimensional convolutional layer to extract spectral and spatial features from hyperspectral images, which have been dimensionally reduced using principal component analysis. Subsequently, three scales of three-dimensional depth-wise separable convolutional residual layers are introduced to extract deep semantic features from the images. This approach reduces the number of training parameters and enhances the network's capability to express high-dimensional, multi-scale spatial feature information. The model achieved an accuracy of 99.46% on a standard dataset.

During model deployment, a large model size can lead to excessively long system runtimes. Therefore, light-weighting the model becomes necessary. This paper introduces depth-wise separable convolution processing [33] for the 3 × 3 convolutional kernels in the residual modules. Depth-wise separable convolution comprises two main components: depth-wise convolution [34] and pointwise convolution [35], illustrated in Fig. 4.
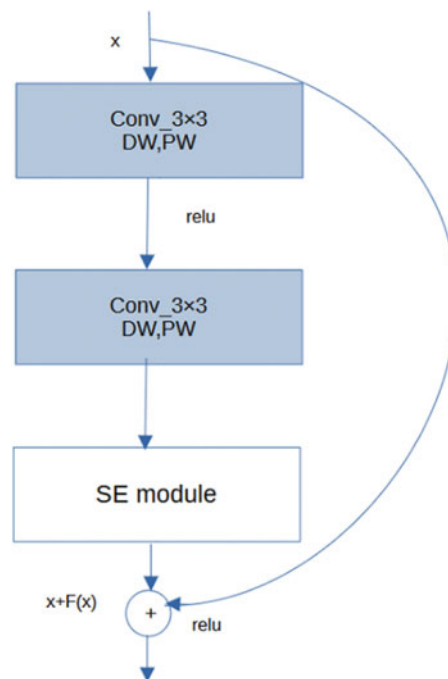

Figure 4: Improved residual module

Fig. 5 depicts the process where the feature matrix, initially of dimensions $n \times n \times k \times 1$, undergoes depth-wise convolution. Here, the number of depth-wise convolution groups equals the number of channels in the feature matrix, dividing the feature matrix into k groups. A 3×3 convolutional kernel is then applied to each group of feature matrices. Post convolution, the output dimensions are merged into a matrix of size k. Subsequently, a conventional 1×1 convolutional kernel is applied to the resultant feature matrix, where t represents the desired number of channels. The final output dimensions are $c \times c \times t \times 1$ for the feature matrix. This method effectively reduces the number of parameters.
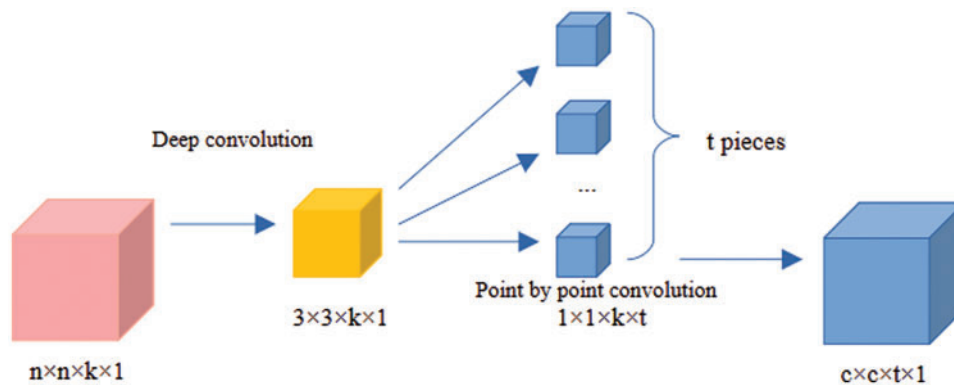
Figure 5: 3×3 convolution kernel is deeply separable

To enhance the efficiency and accuracy of the model, a channel feature extraction network SE module is integrated into the residual structure. As depicted in Fig. 6, a feature matrix with dimensions $H \times W \times C$ undergoes initial global average pooling, computing the average value for each channel, resulting in a feature matrix of dimensions $1 \times 1 \times C$. Subsequently, it proceeds through a fully connected layer with C/R channels, where W, H, and C denote the width, height, and number of channels of the matrix, respectively, and R is set to 16. The ReLU activation function is applied to its output, enabling the attention mechanism to learn non-linear mappings. Another fully connected layer aligns the output channels with the input feature matrix, producing a feature matrix of dimensions $1 \times 1 \times C$. Finally, after passing through the Sigmoid function, the output yields a weight value ranging from 0 to 1, representing the channel feature of the input matrix. These weight values are then multiplied with the corresponding channels of the input feature matrix, enabling the neural network to selectively emphasize different channels, thereby improving recognition accuracy.

For the 3×3 convolutional kernels in the residual modules, enhancements are implemented by integrating depth-wise separable convolution and incorporating a channel feature extraction mechanism. The modified residual module is illustrated in Fig. 6, where the input feature matrix undergoes two layers of depth-wise separable convolution, followed by the inclusion of an SE module for channel feature extraction.
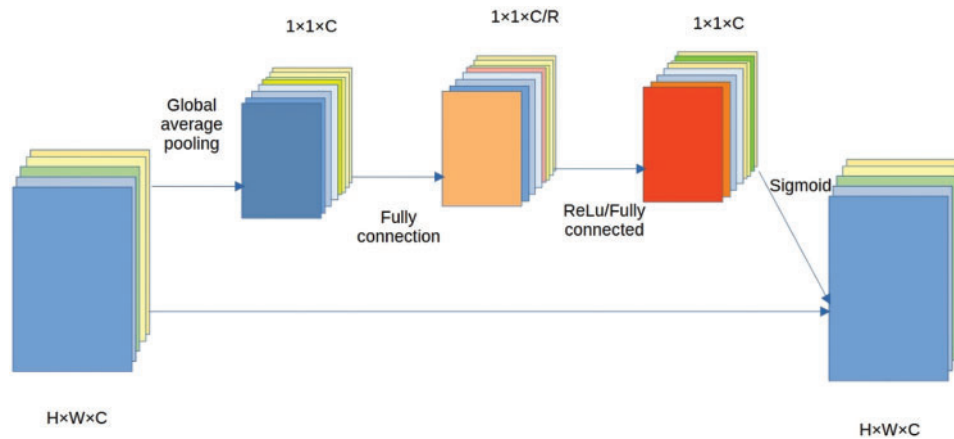
Figure 6: Channel feature extraction

3.4 Evaluation Metrics

The Birds dataset comprises 5 categories, necessitating robust metrics for evaluating the performance of different pre-trained models in multi-class classification tasks. These metrics include ROC AUC (Receiver Operating Characteristic Area Under the Curve) atrix indices such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were computed. TP represents the number of samples correctly classified as positive, TN indicates samples correctly classified as negative, FP denotes samples incorrectly classified as positive when they are actually negative, and FN indicates samples incorrectly classified as negative when they are actually positive.

4 Results

This research utilized the configurations detailed in Table 3, which are particularly suited for lightweight and resource-limited scenarios. This setup proves advantageous in environments constrained by resources, offering robust performance for small-scale deep learning tasks and straightforward data processing. While not tailored for large-scale or complex deep learning tasks, this configuration effectively meets the requirements for simple experiments and learning processes, offering a cost-effective solution. The selection of experimental environments should align closely with specific tasks and research objectives. In scenarios with limited resources, budget constraints, or modest performance requirements, this configuration represents a sensible choice that adequately fulfills basic deep learning needs. For more demanding tasks or extensive datasets, a more powerful hardware configuration may be necessary.

Table 3: Experimental configuration

Requirement	Configuration
Operating system	Windows 10
CPU	Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz 2.30 GHz
Memory	8.00 GB
Deep learning framework	PyTorch
Graphics card	Intel UHD Graphics with 3 GB of memory
Language	Python

In the species classification experiment, the objective was to accurately classify five bird species using both traditional ResNet and an improved method. To ascertain the superiority of the improved ResNet18 network, traditional ResNet18 was employed as a control with identical training parameters. The experimental results are presented in Table 4. Analysis of the table reveals that the verification accuracy of the improved ResNet18 network improved by nearly 0.31% compared to traditional ResNet18, achieving a classification accuracy of 98.57%. Moreover, its training speed showed noticeable enhancement. Upon comprehensive evaluation, the improved ResNet18 network demonstrates superior performance in identifying and classifying Yunnan bird species datasets.

Table 4: Results obtained in species classification

Network	Train_Acc	Val_Acc	Val_Loss
ResNet18	98.26	94.16	0.25
Improved ResNet18	98.57	96.7	0.17

Fig. 7a depicts the classification results of these species, where some were misclassified as other species. For instance, *C. amherstiae* (C.a) was misclassified as *Syrmaticus humiae* (S.h), and *T. caboti* (T.c) was misclassified as *Pucrasia macrolopha* (P.m). In contrast, the improved ResNet showed superior performance across all metrics. Fig. 7b illustrates the classification results of these species using transfer learning, demonstrating high accuracy.

Fig. 8 shows the improved ResNet18 training and test curve. It shows the accuracy of improved ResNet on the training and test sets. At the 5th learning cycle, the classification accuracy value (ACC) in the figure reaches relative stability, with reduced oscillation and a stable improvement in curve convergence speed. The learning curve is slightly noisy, and it needs to be smoothed later to reduce the interference of noise.

(a)

Figure 7: (Continued)

(b)

Figure 7: Species classification examples with (a) traditional ResNet and (b) improved ResNet: Top row represents prediction classifications; Bottom row represents true classifications; Red box represents the wrong prediction
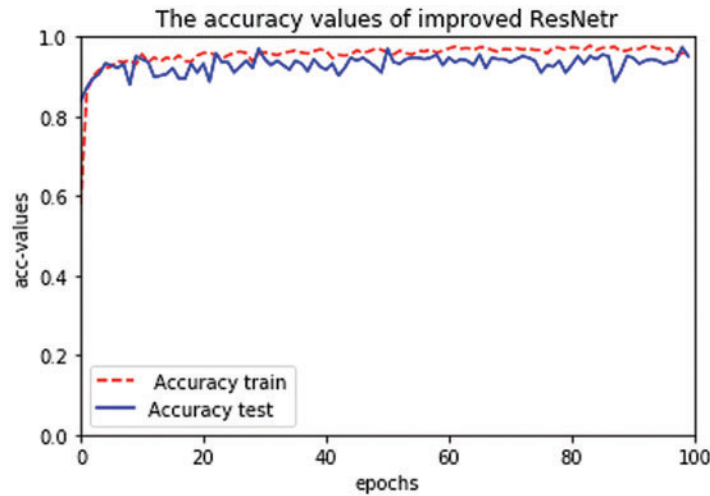
Figure 8: The accuracy of improved ResNet

5 Conclusion

This work introduces a new dataset designed to classify Gaoligong Mountain birds into five species across diverse situations. A benchmark solution for the bird classification problem using the ResNet18 approach is also provided. This dataset addresses the challenge of fine-grained classification by including sightings of birds in Yunnan Province. The aim is to support the environmental preservation of these birds and raise awareness and care for Yunnan's biodiversity.

The result achieved an accuracy of 98.57%, demonstrating that the improved ResNet18 can effectively classify the dataset. By providing this benchmark solution, it is hoped that other researchers can use it to train models for fine-grained classification problems. This result also indicates that there is still room for improvement in applying computer vision models to unbalanced and fine-grained datasets.

The improved ResNet18 model faces challenges such as difficulty in optimizing hyper-parameters and limited generalizability. While the enhancements may perform well on specific datasets or applications, their applicability and performance improvements in other scenarios may be limited.

As a final conclusion, this work aims to stimulate further research in the field while highlighting the biodiversity of Yunnan. For future work, the plan is to enhance this dataset by increasing noise levels and the number of classes and samples per species. Additionally, new challenges will be created, such as classifying birds by gender and age. The exploration of other deep learning models to improve classification performance, especially at the species level, is also intended.

grateful to Ivy Kim D. Machica for her profound influence on my academic career and her selfless dedication.

Author Contributions: The authors confirm their contribution to the paper as follows: paper writing: Wei Yang was responsible for drafting the entire manuscript and integrating feedback from co-authors and supervisors during the writing process. Paper review and improvement: Ivy Kim D. Machica undertook a detailed review and improvement of the manuscript. She provided valuable academic advice, and made extensive revisions to the structure and content of the paper, ensuring its accuracy and logical flow. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All datasets used in this study are available upon request via email at wei04207@usep.edu.ph.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. T. K. Ng, M. Lee, R. J. Y. Tan, X. Hu, J. S. Downie and S. K. W. Chu, *A Review of AI Teaching and Learning from 2000 to 2020*. Springer US, 2023.

[2] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989. doi: 10.1162/neco.1989.1.4.541.

[3] J. Niemi and J. T. Tanttu, "Deep learning case study for automatic bird identification," *Appl. Sci.*, vol. 8, no. 11, pp. 1–15, 2018. doi: 10.3390/app8112089.

[4] J. Gómez-Gó Mez, E. Vidaña-Vila, and X. Sevillano, "Western Mediterranean wetlands bird species classification: Evaluating small-footprint deep learning approaches on a new annotated dataset," 2022, *arXiv:2207.05393*.

[5] V. Omkarini, G. Krishna, and M. Professor, "Automated bird species identification using neural networks," *Ann. Rom. Soc. Cell Biol.*, vol. 25, no. 6, pp. 5402–5407, 2021.

[6] Y. P. Huang and H. Basanta, "Recognition of endemic bird species using deep learning models," *IEEE Access*, vol. 9, pp. 102975–102984, 2021. doi: 10.1109/ACCESS.2021.3098532.

[7] H. P. Li, "Analysis of ecological conservation measures in Gaoligong mountain national nature reserve," (in Chinese), *Contemp. Horticul.*, vol. 22, no. 43, pp. 168–169, 2020.

[8] J. Liu, "One-third of China's birds in Yunnan province," (in Chinese), *Xinhua Net*, p. 10, 2008.

[9] L. Dan *et al.*, "Diversity and vertical distribution of the birds in middle Gaoligong mountain Yunnan," (in Chinese), *Sichuan J. Zool.*, vol. 34, no. 6, pp. 930–940, 2015.

[10] S. L. Ma, L. X. Han, D. Y. Lan, W. Z. Ji, and R. B. Harris, "Faunal resources of the Gaoligongshan region of Yunnan, China: Diverse and threatened," *Environ. Conserv.*, vol. 22, pp. 250–258, 2009.

[11] Y. L. Zhao, Y. C. Li, and H. Chen, "Yunnan wild bird image automatic recognition system," (in Chinese), *Comput. Appl. Res.*, vol. 37, no. s1, pp. 423–425, 2020.

[12] K. Alex, S. Ilya, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Handb. Approx. Algorithms Metaheuristics*, vol. 60, no. 6, pp. 84–90, 2017.

[13] J. Donahue, Y. Jia, O. Vinyals, and J. Hoffman, "DeCAF: A deep convolutional activation feature for generic visual recognition," *Int. Conf. Mach. Learn.*, vol. 32, no. 1, pp. 647–655, 2014.

[14] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and F. F. Li, "ImageNet: A large-scale hierarchical image database," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 20, no. 11, pp. 1221–1227, 2009. doi: 10.1109/CVPR.2009.5206848.

[15] P. U. Egbueze and Z. Wang, "Weather recognition based on still images using deep learning neural network with ResNet-15," *ACM Int. Conf. Proc. Ser.*, vol. 1, pp. 8–13, 2022. doi: 10.1145/3556677.3556688.

[16] T. J. Huo, J. Q. Fan, X. Li, H. Chen, B. Z. Gao and X. S. Li, "Traffic sign recognition based on ResNet-20 and deep mutual learning," *Proc.-2020 Chinese Autom. Congr., CAC 2020*, pp. 4770–4774, 2020, Art. no. 61790564. doi: 10.1109/CAC51589.2020.

[17] J. Z. Liu, W. Y. Du, C. Zhou, and Z. Q. Qin, "Rock image intelligent classification and recognition based on ResNet-50 model," *J. Phys. Conf. Ser.*, vol. 2076, no. 1, pp. 2–7, 2021. doi: 10.1088/1742-6596/2076/1/012011.

[18] Z. J. Wang, J. N. Wang, C. T. Lin, Y. Han, Z. S. Wang and L. Q. Ji, "Identifying habitat elements from bird images using deep convolutional neural networks," *Animals*, vol. 11, no. 5, 2021, Art. no. 1263. doi: 10.3390/ani11051263.

[19] I. Shazzadul, K. S. I. Ali, A. M. Minhazul, H. K. Mohammad, and D. A. Kumar, "Bird species classification from an image using VGG-16 network," *ACM Int. Conf. Pro. Ser.*, vol. 4, no. 6, pp. 38–42, 2019. doi: 10.1145/3348445.3348480.

[20] A. A. Biswas, M. M. Rahman, A. Rajbongshi, and A. Majumder, "Recognition of local birds using different CNN architectures with transfer learning," *2021 Int. Conf. Comput. Commun. Inform., ICCCI 2021*, vol. 1549, pp. 1–6, 2021. doi: 10.1109/ICCCI50826.2021.9402686.

[21] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," *Indian J. Chem.-Sect. B Org. Med. Chem.*, vol. 45, no. 8, pp. 1951–1954, 2016. doi: 10.1109/CVPR.2016.90.

[22] X. Q. Su, "Research on the ecological environment status of avian species in gaoligong mountain nature reserve," (in Chinese), *Environ. Develop.*, vol. 30, no. 5, pp. 2–3, 2018.

[23] Y. Hu, B. Q. Li, D. Liang, X. Q. Li, L. X. Liu and J. W. Yang, "Effect of anthropogenic disturbance on chrysolophus amherstiae activity," *Biodivers. Sci.*, vol. 30, no. 8, pp. 1–14, 2022. doi: 10.17520/biods.2021484.

[24] J. P. Zhang and G. M. Zheng, "The studies of the population number and structure of *T. caboti*," *J. Ecol.*, vol. 1990, no. 4, pp. 291–297, 2020.

[25] M. Z. Shabbir *et al.*, "Complete genome sequencing of a velogenic viscerotropic avian paramyxovirus 1 isolated from pucrasia macrolopha in Lahore, Pakistan," *J. Virol.*, vol. 86, no. 24, pp. 13828–13829, 2012. doi: 10.1128/JVI.02626-12.

[26] X. Q. Xu *et al.*, "Characterization of highly polymorphic microsatellite markers for the lophophorus lhuysii using illumina miseqsequencing," *Mol. Biol. Rep.*, vol. 50, no. 4, pp. 3903–3908, 2023. doi: 10.1007/s11033-022-08151-0.

[27] Z. Shi, H. Yang, and D. Yu, "Polyplectron bicalcaratum population in Xishuangbanna national nature reserve," *J. West China For. Sci.*, vol. 47, no. 6, pp. 67–72, 2018.

[28] M. Kim and H. J. Bae, "Data augmentation techniques for deep learning-based medical image analyses," *J. Korean Soc. Radiol.*, vol. 81, no. 6, pp. 1290–1304, 2020. doi: 10.3348/jksr.2020.0158.

[29] M. Längkvist, L. Karlsson, and A. Loutfi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," *Pattern Recognit. Lett.*, vol. 42, no. 1, pp. 11–24, 2014.

[30] F. Chollet and I. Google, "Xception: Deep learning with depthwise separable convolutions," *SAE Int. J. Mater. Manuf.*, vol. 7, no. 3, pp. 560–566, 2017. doi: 10.1109/CVPR.2017.195.

[31] S. Jun, W. J. Tan, X. H. Wu, J. F. Shen, B. Lu and C. X. Dai, "Real-time identification of beets and weeds in complex backgrounds using a multi-channel depthwise separable convolution model," (in Chinese), *Trans. Chinese Soc. Agric. Eng.*, vol. 35, no. 12, p. 7, 2019.

[32] R. J. Cheng, Y. Yang, L. W. Li, Y. T. Wang, and J. Y. Wang, "Lightweight residual network for hyperspectral image classification based on depthwise separable convolution," *Acta Opt. Sin.*, vol. 43, no. 12, pp. 311–320, 2023.

[33] K. Zhang, K. Cheng, J. J. Li, and Y. Y. Peng, "A channel pruning algorithm based on depth-wise separable convolution unit," *IEEE Access*, vol. 7, pp. 173294–173309, 2019. doi: 10.1109/ACCESS.2019.2956976.

[34] Q. Han *et al.*, "On the connection between local attention and dynamic depth-wise convolution," *ICLR 2022-10th Int. Conf. Learn. Rep.*, pp. 1–25, 2022.

[35] W. An, C. Xu, and J. Kong, "Plant growth stage recognition method based on improved ResNet18 and implementation of intelligent plant supplemental lighting," *J. Crops*, vol. 23, no. 8, pp. 1–14, 2024.