Comparasion of ResNet50 and 3-layer Neural Network for Bird Identification

Wentai Ye
Department of Engineering Technology
KU Leuven
yewentai126@gmail.com

Joren Mangelschots
Department of Engineering Technology
KU Leuven
email

Abstract

The abstract is optional, depending on your available space. It should consist of 0.5 paragraph consisting of the motivation for your paper and a high-level explanation of the methodology you used/results obtained.

1 Introduction

In the realm of environmental conservation and biodiversity research, the development of a machine learning algorithm capable of accurately classifying bird species from images represents a groundbreaking advancement. This technological endeavor addresses a critical challenge with far-reaching implications in various fields.

The cornerstone of this project is its potential impact on conservation and biodiversity monitoring. Birds are often considered key indicators of an ecosystem's health. Therefore, the ability to automatically identify bird species is an invaluable asset for conservationists. This capability is particularly crucial in conservation areas, where monitoring the presence and diversity of bird species can yield vital information about the state of biodiversity and the effectiveness of ongoing conservation efforts.

Another significant aspect of this project is the efficiency and scalability it offers. Traditionally, the identification of bird species has been reliant on expert knowledge, a process that is both time-intensive and limited in scope. An automated system, on the other hand, can rapidly process a large volume of images, thereby facilitating the monitoring of more extensive areas and a greater variety of species than what is currently possible through manual methods.

Beyond its scientific and conservation applications, this tool also holds promise for public engagement and education. By integrating an automated bird identification system into apps and platforms used by bird watchers and nature enthusiasts, it can enhance their experience and knowledge. This, in turn, helps in raising public awareness about various bird species and broader conservation issues.

The input to our algorithm is an image. We then use a neural network to output a predicted bird species.

2 Related Work

You should find existing papers, group them into categories based on their approaches, and discuss their strengths and weaknesses, as well as how they are similar to and differ from your work.

Final report for Machine Learning (EA) (B-KUL-T4VML2)

In your opinion, which approaches were clever/good? What is the state- of-the-art? Do most people perform the task by hand?

You should aim to have at least 5 references in the related work. Include previous attempts by others at your problem, previous technical methods, or previous learning algorithms. Google Scholar is very useful for this: https://scholar.google.com/ (you can click cite and it generates MLA, APA, BibTeX, etc.)

Recent studies in bird species classification have explored various methodologies. Haobin et al. [2018] introduced a low-dimensional bilinear model that prioritizes computational efficiency by reducing feature dimensionality. Zhang et al. [2023] focused on improving the ShuffleNetV2 model, offering a lightweight yet effective solution suitable for resource-limited environments. Vo et al. [2023] employed a YOLO-based framework, emphasizing real-time processing capabilities. Zahra EL Bouni et al. [2021] utilized wavelet transform for feature extraction, showcasing effectiveness in handling varying textures and patterns. Lastly, Budiman et al. [2022] adopted a K-Nearest Neighbors approach, highlighting simplicity and ease of implementation.

3 Dataset and Features

1. Include a citation on where you obtained your dataset from. 2. Describe your dataset: how many training/validation/test examples do you have? 3. Depending on available space, show some examples from your dataset. Try to include examples of your data in the report (e.g. include an image, show a waveform, etc.). 4. What is the resolution of your images? 5. Is there any preprocessing you did? 6. What about normalization or standardization?

The dataset in this study encompasses 525 bird species, featuring 84,635 training images, along with 2,625 images each for testing and validation (5 per species). Rigorous dataset analysis tools were applied to eliminate duplicates and low-quality images, ensuring the dataset's integrity and preventing leakage across train, test, and validation sets.[1]

Each 224x224x3 color image in JPG format, predominantly featuring the bird, is meticulously categorized into 525 folders per species. Accompanying this is a comprehensive CSV file detailing file paths, labels, scientific names, dataset types, and class indices. The test and validation images, carefully chosen for their quality, enhance the dataset's robustness.

4 Methods

Describe your learning algorithms, proposed algorithm(s), or theoretical proof(s). Make sure to include relevant mathematical notation. For example, you can briefly include the SVM optimization objective/formula or say what the softmax function is. It is okay to use formulas from the lecture notes. For each algorithm, give a short description (1 paragraph) of how it works. Again, we are looking for your understanding of how these machine learning algorithms work. Although the teaching staff probably know the algorithms, future readers may not. Additionally, if you are using a niche or cutting-edge algorithm (e.g. long short-term memory, SURF features, or anything else not covered in the class), you may want to explain your algorithm using 1/2 paragraphs.

4.1 ResNet

ResNet, short for Residual Network, is a type of neural network architecture that was designed to enable the training of much deeper networks than was previously feasible. Introduced by Microsoft researchers in 2015, ResNet quickly became famous for its effectiveness, especially in image classification tasks. He et al. [2016] The ResNet architecture is composed of a series of residual blocks, which are composed of a series of convolutional layers.

[1]Laszewski, G. von. (2023). 100 Bird Species [Data set]. Kaggle. `https://www.kaggle.com/datasets/gpiosenka/100-bird-species`

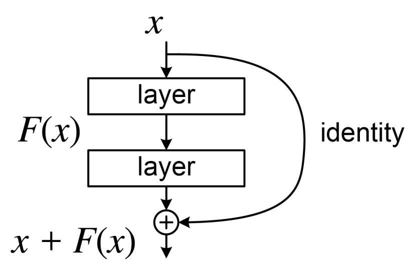

Figure 1: ResNet architecture

The ResNet (Residual Network) architecture, as illustrated in Figure 1, represents a significant advancement in deep neural networks. This architecture is primarily characterized by its residual blocks and skip connections. The structure of ResNet can be described as follows:

- **Residual Blocks:** Each residual block in ResNet is composed of several convolutional layers. These layers are designed to learn the residual functions with reference to the layer inputs. A typical residual block contains two or three convolutional layers.
- **Convolutional Layers:** Within each residual block, convolutional layers are used for feature extraction. Each layer applies a set of learnable filters to the input. The convolutional layers in ResNet typically use batch normalization and a ReLU activation function.
- **Skip Connections:** A distinctive feature of ResNet is the use of skip connections (or shortcut connections). These connections allow the input of a residual block to be directly added to its output. Mathematically, if $\mathbf{x}$ is the input and $\mathcal{F}(\mathbf{x})$ is the output of the residual block, the final output of the block is $\mathcal{F}(\mathbf{x}) + \mathbf{x}$.
- **Identity Function Learning:** Skip connections facilitate the learning of the identity function. This is crucial in deeper networks as it mitigates the problem of vanishing gradients, enabling the training of very deep networks.
- **Depth of Network:** Thanks to skip connections, ResNet architectures can be significantly deeper than traditional convolutional neural networks. Models with various depths have been proposed, such as ResNet-50, ResNet-101, and ResNet-152, indicating the number of layers in the network.

The overall architecture of ResNet enables the training of deep neural networks with improved performance on tasks like image classification, by effectively addressing the issues of vanishing gradients and degradation problem.

5 Results and Discussion

You should also give details about what (hyper)parameters you chose (e.g. why did you use X learning rate for gradient descent, what was your mini-batch size and why) and how you chose them. Did you do cross-validation, if so, how many folds? Before you list your results, make sure to list and explain what your primary metrics are: accuracy, precision, etc. Provide equations for the metrics if necessary. For results, you want to have a mixture of tables and plots. You should include a confusion matrix or AUC/AUPRC curves. Include performance metrics such as precision, recall, and accuracy. Include visualizations of results, heatmaps, examples of where your algorithm failed and a discussion of why certain algorithms failed or succeeded. In addition, explain whether you think you have overfit to your training set and what, if anything, you did to mitigate that. Make sure to discuss the figures/tables in your main text throughout this section. Your plots should include legends, axis labels, and have font sizes that are legible when printed.

Mention which part of the assignment code you used for which experiment.

6 Conclusion and Future Work

Summarize your report and reiterate key points. Which algorithms were the highest- performing? Why do you think that some algorithms worked better than others? For future work, if you had more time, more team members, or more computational resources, what would you explore?

7 Contributions

The contributions section is not included in the 6 page limit. This section should describe what each team member worked on and contributed to the project.

References

Ichsan Budiman, Diena Rauda Ramdania, Yana Aditia Gerhana, Alif Rakasha Pratama Putra, Nisairrizqy Nabilah Faizah, and Maisevli Harika. Classification of Bird Species using K-Nearest Neighbor Algorithm. In *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–5, September 2022. doi: 10.1109/CITSM56380.2022.9936012. URL `https://ieeexplore.ieee.org/document/9936012`. ISSN: 2770-159X.

Shi Haobin, Zhang Renyu, and Sun Gang. Fine-Grained Bird Classification Based on Low-Dimensional Bilinear Model. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 424–428, June 2018. doi: 10.1109/ICIVC.2018.8492897. URL `https://ieeexplore.ieee.org/document/8492897`.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. URL `http://ieeexplore.ieee.org/document/7780459/`.

Hoang-Tu Vo, Nhon Nguyen Thien, and Kheo Chau Mui. Bird Detection and Species Classification: Using YOLOv5 and Deep Transfer Learning Models. *International Journal of Advanced Computer Science and Applications*, 14(7), 2023. ISSN 2158107X. doi: 10.14569/IJACSA.2023.01407102. URL `https://www.proquest.com/docview/2858093373/abstract/E390BDC364874051PQ/1`. Place: West Yorkshire, United Kingdom Publisher: Science and Information (SAI) Organization Limited.

Fatima Zahra EL Bouni, Tareq El Hariri, Chaime Zouitni, Ilham Ben Bahva, Hafida El Aboui, and Aziza El Ouaazizi. Bird image recognition and classification using Watson visual recognition services from IBMCloud and Conventional Neural Network (CNN). In *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–6, June 2021. doi: 10.1109/ICECCE52056.2021.9514269. URL `https://ieeexplore.ieee.org/document/9514269`.

Liu-Lei Zhang, Ying Jiang, You-Peng Sun, Yuan Zhang, and Zheng Wang. Improvements Based on ShuffleNetV2 Model for Bird Identification. *IEEE Access*, 11:101823–101832, 2023. ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3314676. URL `https://ieeexplore.ieee.org/document/10247540/`.