Khidmat Report
WWF Bird Recognition Model

Ali Hamza
ah05084

Ali Haider Rizvi
ah05199

M. Usaid Rehman
mr04302

Contents

1 Introduction

1.1 Project Description

This project is aimed at creating and deploying a deep learning pipeline for WWF Pakistan to classify images of three different species of birds – namely: common myna, house crow, and the house sparrow. This project would serve as a proof-of-concept for a larger model that WWF can use to classify a larger number of birds using a mobile application.

1.2 About WWF Pakistan

The World Wildlife Fund for Nature (WWF) is an international non-governmental organization founded in 1961 that primarily works in the areas of wildlife conservation and preservation. WWF also works in the domains of food security, renewable energy, ocean and forest preservation, and many others[1].

WWF Pakistan is the Pakistan chapter of the World Wildlife Fund, working on similar problems as WWF International but in a more local context. WWF has more than five regional offices throughout the country, and has conducted several successful projects pertaining to freshwater conservation, water security, forests, climate change and others. What they are most known for, however, is wildlife conservation — the area that we have assisted WWF in.

1.3 Project Background & Context

Keeping track of biodiversity is a challenge in Pakistan due to a lack of resources, personnel and governmental support. Pakistan has a plethora of wildlife and a rich and diverse ecosystem. WWF Pakistan has helped in the conservation of several endangered species such as the Indus River Dolphin and the Snow Leopard — both of which are now experiencing a rise in numbers.

However, wildlife preservation and protection extends beyond just the conservation of endangered species. Maintaining biodiversity is extremely important for wildlife protection and to maintain a healhty ecosystem and environment. Unfortunately, a lack of biodiversity is a major crisis that is unfolding across the globe. In order to combat this, it is imperative that researches have updated data about species populations, habitats, interactions etc.

An important group of animals that contribute to biodiversity are birds. Therefore, it is vital that data pertaining to bird population is collected in Pakistan. To assist in this endeavor, WWF aims to enlist the help of citizen scientists who could help identify and locate birds and consequently update WWF's database.

Our project aims to implement a deep learning pipeline that would classify images of three different species of birds — house crow, common myna, and the house sparrow.

[1] https://wwf.panda.org/

The pipeline would take an image, preprocess it to a form that the model can work with, and then classify it using a deep neural network — telling us what species the bird in the image belongs to.

This pipeline serves as a proof-of-concept for a larger model that would be classifying hundreds of species of birds. This model will fit into a planned mobile application that citizen scientists can use to spot birds and then capture their image. Not only will the application classify the species of the bird but it will also update WWF's database, giving them access to more data about bird populations and their locations.

The development of a larger deep learning model and an application is beyond the scope of our *Khidmat* due to time restraints. We have succesfully implemented a computer vision pipeline for preprocessing and a deep neural network for classification for three species of birds.

1.4 Work Plan

1.4.1 Project Description

To create an algorithm that distinguishes between 3 birds that will act as the first stage in developing a larger ML model for the WWF app for recognizing birds.

1.4.2 Expected Deliverables

1. Dataset on all 3 birds (Common Myna, Housecrow, Sparrow)
2. Machine Learning Model (Python Notebook)
3. Validation and Training of Model
4. Demonstration of Final Model
5. Report on Model, including Code, Explanations, and Documentation.

1.4.3 Planned Schedule

- June Week 1: Images, Image Metadata, Relevance, Summary statistics, Scraped from Facebook, Instagram, Twitter, Datasets
- June Week 2: Labelled data for training and testing. [Data in the proper format, ready to be used for the model]
- July Week 2: Prototype ML Model
- July Week 3: Test run for the model using training data
- July Week 4: Validation of model

- August Week 1: Final training of reworked model
- August Week 2: Final Validation of reworked model
- Sept Week 2: Demonstration of Model
- Sept Week 4: Final Report and Delivery

2 Weekly Work Log

2.1 `Week 1:`

Item	Activity	Time	ID
1	Scraped images for house sparrow	2	05084
2	Scraped images for house sparrow	3	05084
3	Meeting with WWF	1	05084
4	Scraped images for house crow	2	05199
5	Scraped images for house crow	1	05199
6	Scraped images for house sparrow	2	05199
7	Meeting with WWF	1	05084
8	Scraped Common Myna Pictures	2	04302
9	Scraped Common Myna Pictures	2	04302
10	Scraped Common Myna Pictures	3	04302
11	Meeting with WWF	1	05084

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	6
ah05199	6
mr04302	8

2.2 Week 2:

Item	Activity	Time	ID
1	Analyzed the compiled dataset	2.5	050584
2	Wrote code for image preprocesing	5	05084
3	Internal Meeting	1	05084
4	Resized Images	2.5	05199
5	Coded Initial Model	3.5	05199
6	Training and Tuning	3.5	05199
7	Internal Meeting	1	05199
8	Research Neural Nets	1.5	04302
9	Coded Initial Model	3.5	04302
10	Researched Optimization Methods	1.5	04302
11	Internal Meeting	1	04302

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	8.5
ah05199	10
mr04302	7.5

2.3 Week 3:

Item	Activity	Time	ID
1	Worked on Model	2	ah05084
2	Training and tuning	2	ah05084
3	Worked on Report	2	ah05084
4	Explored ResNet for better performance	3	ah05199
5	Worked on Model	2	mr04302
6	Training and tuning	2	mr04302

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	6
ah05199	3
mr04302	4

2.4 `Week 4:`

Item	Activity	Time	ID
1	Shifted training to TPUs for faster training	2.5	05199
2	Training and Tuning	11.5	05199
3	Confusion Matrix/ Precision Recall	1.5	05199

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	0
ah05199	15.5
mr04302	0

2.5 Week 5:

Item	Activity	Time	ID
1	Scraped 40-50 images for all birds	3	05084
2	Worked on Intermediate Report	3	05084
3	Code Refactoring + Misc Google Colab work	4	05084
4	Worked on Neural Net + Github Syncing	2.5	05084
5	Worked on Intermediate Report	3	04302

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	12.5
ah05199	0
mr04302	3

2.6 `Week 6:`

Item	Activity	Time	ID
1	Training and Testing	4	05084
2	Worked on Intermediate Report	2	05084
3	Meeting with Dr. Sarah	1	05084
4	Meeting with Dr. Sarah	1	04302

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	7
ah05199	0
mr04302	1

2.7 Week 7:

Item	Activity	Time	ID
1	Cleaned Code	2	05084
2	Turned Notebook into Python Modules	3	05084

The total time spent on the Khidmat this week is as follows.

ID	Total Hours
ah05084	5
ah05199	0
mr04302	0

2.8 `Week 8:`

Item	**Activity**	**Time**	**ID**
1	Worked on Final Report	8	05084

The total time spent on the Khidmat this week is as follows.

ID	**Total Hours**
ah05084	8
ah05199	0
mr04302	0

3 Technical Overview

3.1 Resources

The resources that we used for this project can be categorized into two main categories: images, and technologies. Our first step was to acquire a set of images that would function as our dataset. We collected around 550 images for each of the three species of birds — the house crow, common myna, and the house sparrow. We used several different websites and online databases to find these images, we list these websites below:

1. https://search.macaulaylibrary.org/catalog?taxonCode=myna&mediaType=p&q=Common%20Myna
2. https://ebird.org/media/catalog?taxonCode=commyn&mediaType=p&sort=rating_rank_desc&q=Common%20Myna%20-%20Acridotheres%20tristis
3. https://ebird.org/media/catalog?taxonCode=houcro1&sort=rating_rank_desc&mediaType=p&regionCode=
4. https://search.macaulaylibrary.org/catalog?taxonCode=houcro1&mediaType=p&region=Pakistan%20(PK)&regionCode=PK&q=House%20Crow%20-%20Corvus%20splendens
5. https://www.kaggle.com/gpiosenka/100-bird-species
6. https://search.macaulaylibrary.org/catalog?taxonCode=houspa&mediaType=p&q=House%20Sparrow
7. https://ebird.org/media/catalog?taxonCode=houspa&mediaType=p&sort=rating_rank_desc&q=House%20Sparrow%20-%20Passer%20domesticus

All collected images can be viewed at https://drive.google.com/drive/folders/18k-roE_VJSB1dcrhvN1y_EosVF7Kb0dY?usp=sharing.

To preprocess our images, convert them into a usable dataset, and to create our deep learning model, we had to rely on several different preexisting tools. We list the tools that we used in this project below:

- `Python` — The primary programming language that we used for this project. Most of the other tools that we used are different libraries & modules for in Python.
- `OpenCV` — A computer vision module in Python to preprocess the images.
- `TensorFlow & Keras` — A Python library that we used to construct, and then ttrain and test our neural network.

- `Matplotlib` — A plotting library for Python, which we used to visualize our
- `Sci-kit learn` — A Python library for machine learning that we used to construct our training and testing set.
- `Google Colab` — We used Colab notebooks to write code and then run code using GPU-accelerated computation that is made available using Google Colaboratory.

3.2 Overview of CS Techniques Used

This project was an amalgamation of several different aspects of computer science techniques. The process can be summarized into 5 main steps:

1. Image Collection
2. Image Preprocessing
3. Construction of Neural Network
4. Training the Neural Network
5. Testing and Optimization

We briefly expand upon these steps in the following sections. For a more technical description, please view the documentation (LINK HERE).

3.2.1 Image Collection

We used the sources listed at 7 to collect images of the three species of birds. We downloaded approximately 650 images for each birds, which we then filtered down to around 550 to maintain uniform image quality. Based on our results, we believe the number of images that we have, i.e. the size of our dataset is adequate.

3.2.2 Image Preprocessing

The next step was to pre-process these images to convert them into a usable dataset. The first step is to create bounding boxes surrounding the birds within the images. Since we need images of equal dimensions, i.e. a square, we pad the images in order to convert it into a square.

Then we can scale the image up or down in order to convert it into a specific dimensions. For example, an image of dimension 500×500 can be scaled down to an image of dimension 50×50. Examples of this can be seen in Figure 1. Our entire data set consists of 50×50 images, numbering around 550 images per bird. Most of the inner working of image preprocessing is handled by the OpenCV Python library.

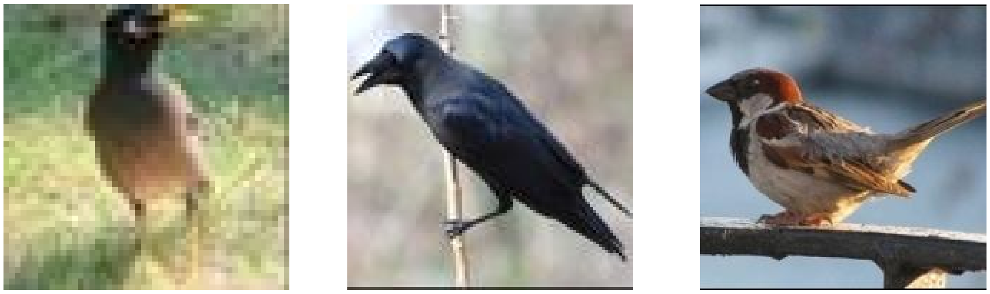

Figure 1: Pictures of birds from the data set after preprocessing.

3.2.3 Neural Network Construction

Artificial neural networks form the basis for solving most artificial intelligence problems. Image classification is one such problem. Neural networks represent connections between artificial neurons, also called units. These connections are represented as weights in the neural network. These weights are also called parameters and learned by the neural network during supervised learning.

There are several types of neural networks that are optimal for solving different intelligence problems. Image recognition/classification has been solved using several types of neural networks since the 1980s such as multilayer perceptrons. A newer method to solve image classification problems (and problems generally pertaining to computer vision) is the Convolutional Neural Network (CNN).

Unlike other neural networks, a CNN takes spatial information about the input image into consideration during learning, allowing it to generally be more accurate. Therefore, we decided to implement a CNN for this project. We first implemented a simplistic CNN using two-dimensional convolutional layers that preexist in TensorFlow. However, we found that our results were not as satisfactory as we hoped for them to be.

We then tried implementing the Residual Neural Network (ResNet), which is built upon the same idea as CNNs but significantly improves accuracy and performance. ResNet utilize skip connections or shortcuts between layers. Specifically, we leveraged TensorFlow's built-in ResNet50 implementation, a variant of ResNet that utilizes 48 convolutional layers, a MaxPool layer and an Average Pool layer. However, since ResNet50 is a fairly deep architecture which is more useful when implementing models for data sets that are quite large, with at least several thousand examples.

Therefore, we resorted to the CNN once again since it was more suited to our data set. For more technical information on the neural network construction, please see ¡LINK¿.

3.2.4 Training the Neural Network

To train neural networks, there are several different methods that optimize the cost function. The most famous of these is the stochastic gradient descent (SGD). SGD is a stochastic approximation of the simpler gradient descent optimization. Gradient descent approximates the gradient for the entire training set at each iteration. However, stochastic gradient descent approximates the gradient for one example in the training set per iteration — making it much faster.

A newer method called minibatch stochastic gradient finds the middle ground between gradient descent and SGD. It approximates the gradient in each iteration using a randomly selected minibatch of training examples. These minibatches are also effective for parellel computation using TPUs.

We used the Adaptive Moment Estimation or *Adam* optimization algorithm. Adam is an effective algorithm for problems with sparse gradients such as computer vision problems. This optimization algorithm relies on three hyperparameters: α, β_1, β_2, and ϵ. Adam is very effective in optimizing problems that contains a large number of trainable parameters. The value we set for α was 0.001 and the β values were set as default, i.e. 0.9 and 0.999. The loss function for our model was the cross-entropy loss function which is quite popular in image classification tasks.

3.3 Results

To train our neural network, we ran the optimization algorithm on it for 110 epochs, and we kept the minibatch size as 64, in order to not overburden the TPU. At the end of training, our accuracy reached approximately 86% and our loss value was reduced to almost 0.67 as shown in Figure 2.

Other than just the accuracy and loss metrics, we used the confusion matrix to analyze the performance of our model. The confusion matrix is a standard for judging the performance of a model in a classification problem. Our confusion matrix looked like this:

$$\begin{bmatrix} 117 & 15 & 9 \\ 13 & 134 & 6 \\ 14 & 14 & 128 \end{bmatrix}$$

Lastly, we also calculated the accuracy, precision, recall, and F1 scores. Our accuracy score was almost 0.8222, precision was 0.84209, recall was 0.842039, and F1 score was 0.8417. Therefore, considering the small data set, our model did do reasonably well in terms of performance.

4 Conclusion

Our *Khidmat* Project is the starting step for a larger model that should be able to identify many more species of birds commonly found in Pakistan. This will clearly make the

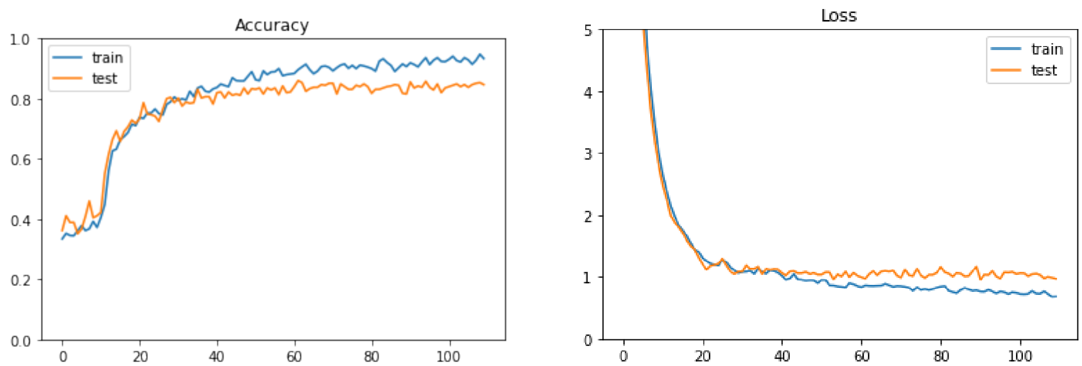

Figure 2: Plots for the accuracy and loss resulting from training the CNN.

data set much larger. One of the primary challenges we faced was that our data set was fairly small. An improvement that we would suggest would be to use a much larger data set. This data set can be accumulated using web scraping. Ideally, there should be several thousand images for each birds. Not only does this have the potential of improving accuracy but it also allows for better, more complex architectures to be used such as DenseNet and ResNet. We would also recommend tweaking the existing model more and adjusting hyperparameters in order to find a better configuration that optimizes the model the most.

This project was a great learning experience for us and we are thankful to WWF Pakistan, especially Mr. Jamshed Iqbal for letting us be a part of the great work that WWF Pakistan does for animal conservation. We are also thankful to our internal supervisor, Dr. Sarah Hasnain, for letting us know of this opportunity and for supporting us throughout the project.

A Documentation

- `load_dataset(size)` This function has been designed to take in the size of the images being used for training, each image being a square. The function will then return a list of images, and labels for training. The dataset variable returns a list of paths for the images, and a the labels variable returns list of labels for those images.

- `read_image(dataset)`

- This takes the dataset variable as an argument and returns a list of images that have been turned where each image is represented by a list of RGB values for each pixel in the image.

- `create_model()`

Khidmat Completion Form

To be completed by the external supervisor.

Please use the space below to provide any comments you may have on the students' performance, the Khidmat program, or any other feedback you want to share with Habib University's Khidmat committee. We can also be reached at khidmat@sse.habib.edu.pk.

I hereby certify that I supervised XXX and XXX for the Khidmat described in this report. Furthermore, that I have read and agree with the weekly updates included in this report. My signature below marks the successful completion of the work undertaken for the Khidmat.

______________________ & ______________________

Name and signature Location and date