# ACTL3143 Assignment: Image Classification of Australian Animals

William Li z5257749

## Problem Specification

The goal of the project is to create a model that can classify between a selection of Australian native animals. There are 8 animals in the dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| - Cockatoo  - Kookaburra | - Dingo  - Platypus | - Kangaroo  - Seadragon | - Koala  - Wombat |

Models will be assessed with (Top 1) categorical accuracy.

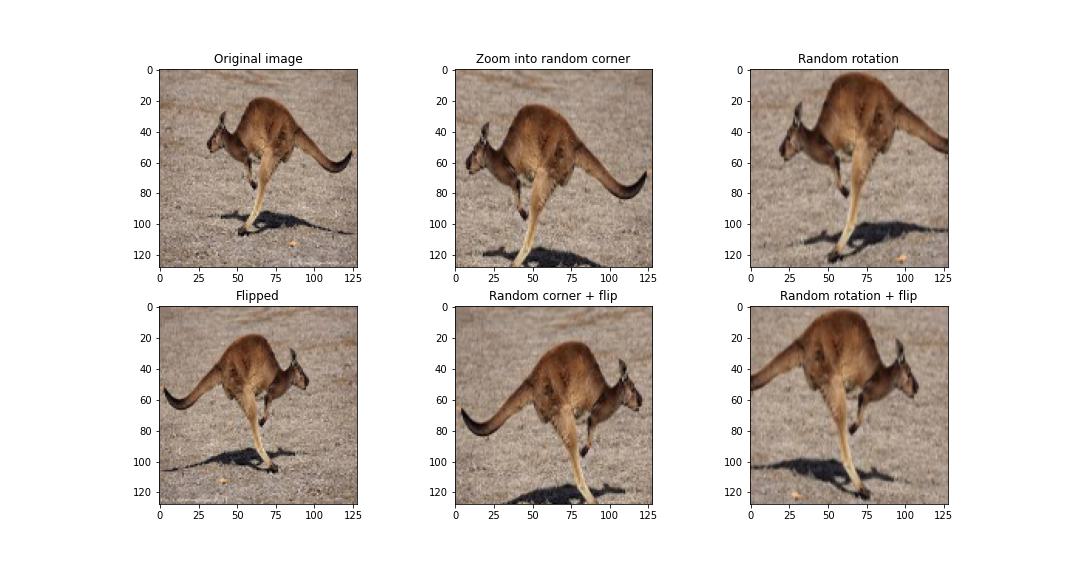
## Data collection and pre-processing

There are 200 images for each animal. The images were collected from Google Images with the help of the Image Downloader extension. They were selected manually to ensure accuracy and avoid duplicates (though the latter is not guaranteed as it is not possible to perfectly recall 200 images).

All images were rescaled to 128x128, as most machine learning techniques require all inputs to be the same size. The data was split into 4 sets. Their purpose and number of examples per class are:

* Training (96)
* Validation set 1 (32)
* Validation set 2 (32) and
* Test (40)

Prior to training the final model, validation set 1 is used for early stopping while validation set 2 is used for hyperparameter tuning. For the final model, validation set 1 was added to the training set while set 2 was used for early stopping. Reasoning for 2 validation sets is explained in Appendix 1.

The training set was augmented using zoom and rotation to get 6 times as many examples.

## Exploratory Data Analysis

Exploratory data analysis is not as crucial in image classification as in other problems. However, it is helpful to check that the colour channels are working as intended, which can be confirmed based on the photos below. For example, the red channel has low values (represented by dark colour on the “Red Channel”) for the background which is blueish-green, and high values (represented by light colour) for the seadragons coloured orange.

Calendar

Description automatically generated

## Simple Benchmark Model: Logistic Regression

Multi-class logistic regression was fit using the multinomial method. All images had to be scaled to 8x8 before fitting as this results in 8x8x3=192 inputs, which combined with an intercept results in 193 parameters. Having any more parameters than this would result in severe overfitting as logistic regression requires the number of observations to be larger than the number of features. The logistic regression model gave an accuracy of 46.88% in the validation set. It is not great but considerably higher than random guessing (which would give an accuracy of 12.50%).

A picture containing diagram

Description automatically generated

*Confusion matrix of logistic regression on validation set. It worked reasonably for some animals but really struggled to identify the koala, kookaburra, and platypus.*

## Simple neural network

The simple neural network is the most basic type of neural network. It uses dense layers where every neuron receives an input from all the neurons in the previous layer. The model resizes the image at the very start, in order to limit the input size to a reasonable scale.

## Convolutional neural network

Convolutional neural networks are one of the most common models for image classification. It makes use of convolution layers, which uses kernels that move across the previous layer to identify features within neighbouring pixels. The model also makes use of pooling layers, which groups information and reduces the dimensionality. After convolution and pooling, it is flattened into a single dimension and fed into a dense network.

The CNN architecture implemented in this report takes inspiration from classic models such as LeNet and AlexNet. The architecture with optimal hyperparameters the hyperparameter tuning process are shown below in Figure 3.1. Using the best hyperparameters, an accuracy of 71.09% was achieved on the validation set. The confusion matrix is included in Appendix # and analysis of the probability assigned to the correct label is in Appendix #.

*Figure 3.1: CNN architecture with best set of parameters found through grid search*

Shown below is the table of hyperparameters tuned using grid search. It shows the choices for each hyperparameter and the combination that produces the highest accuracy on the validation set. Accuracy values of the top combinations are included in Appendix #

|  |  |  |
| --- | --- | --- |
| Hyper-parameter | Choices | Optimal combination |
| Kernel size of 1st Conv | 5, 7, 9, 11 | 11 |
| Kernel size of 2nd Conv | 3, 5 | 3 |
| Pooling type | Max, Average | Max |
| Dropout rate | 0, 0.2, 0.4 | 0.4 |
| Activation | ReLu, tanh | tanh |

## Comparison of the 3 models on the validation set

|  |  |
| --- | --- |
| Model | Accuracy on  Validation set |
| Logistic Regression | 46.88% |
| Simple NN | 48.75% |
| CNN | 71.09% |

Deep learning is widely used in image classification tasks nowadays. This project showed that even a fairly straightforward CNN easily outperforms other models and achieves reasonable accuracy in the task of classifying photos of Australian animals.