Kangaroo and Wallaby Detection

1 Project selection

There is a list of projects available for selection. They are mushroom edibility classification, apple detection and kangaroo and wallaby detection. Kangaroo and wallaby detection is selected for this project. Difficulty is the main concern for project selection.

For the mushroom edibility classification project, the main difficulty is there are too many species for mushroom, different mushrooms can have different appearances when it comes to edibility. In order to build a model that can classify the edibility of most mushroom species, a huge dataset is required, and it is time-consuming. Therefore, mushroom edibility classification project was not selected.

For the apple detection, the main difficulty is the similarity of the three required categories. The apple detection requires detection and classification for different types of apples - royal gala, pink lady, and kanzi. These categories look similar and cannot classify by the Author easily. This will affect the process of data annotation. Therefore, the apple detection project was not selected.

For the kangaroo and wallaby detection, the main difficulty is whether the required number of images – 1400 images, enough for develop an accurate model. On the other hand, there are only two categories required for the project, so the dataset does not necessarily need to be huge. Moreover, the differences of the categories are obvious which is helpful for data annotation efficiency. Therefore, Kangaroo and wallaby detection project are selected.

2 Real-life benefits

Kangaroo is one of the symbols of Australia. When it comes to Australian wildlife, kangaroo is of one the most popular animals. Developing a model that can detect and classify kangaroo and wallaby are beneficial for Australian wildlife advertisement. There are lots of people do not know much about different categories of kangaroo. Developing the model can help those people to classify kangaroo and wallaby, and be familiar with Australian wildlife.

3 Data collection

In order to train a model, a dataset is essential. This project is to detect and classify kangaroos and wallabies from images. Therefore, images for kangaroos and wallabies are required. This project has collected 956 images for kangaroo and 903 images for wallaby. Half of the images are collected from Google image search using Python script (Figure 1). The Python script helps users to download images from Google images search for the keywords.

```python
#importing the library
from google_images_download import google_images_download

#class instantiation
response = google_images_download.googleimagesdownload()

#creating list of arguments
kangaroo = {"keywords":"Kangaroo","limit":1000,"size": "large","format":"jpg","chromedriver":"./chromedriver.exe"}

#passing the arguments to the function
response.download(kangaroo)

#creating list of arguments
wallaby = {"keywords":"Wallaby","limit":1000,"size": "large","format":"jpg","chromedriver":"./chromedriver.exe"}

#passing the arguments to the function
response.download(wallaby)
```

Figure 1. Python script for downloading kangaroo and wallaby images from Google image search

Another half of the images are collected using a Chrome extension called Image downloader (Figure 2) from the internet such as Pinterest and Unsplash. Image downloader let users download all the images on a webpage.

Figure 2. The Chrome extension called image downloader

4 Data annotation

In order to let a machine learn from the dataset, data annotation is required. In the dataset, 727 kangaroo images and 727 wallaby images are labelled. The tool that this project used for data annotation is called LabelImg (Figure 3). It is a graphical image annotation tool that can help users label images quickly and export the annotations as XML files in different formats such as PASCAL VOC format and YOLO format.

Figure 3. The graphical image annotation tool called LabelImg

5 Splitting data

The dataset is split into 3 different sets – train, validation, and test set. 70% of the images (1301 images) are split into train set. 20% of the images (372 images) are split to validation set. 10% of the images (186 images) are split to test set.

6 Data pre-processing

In order to increase the training efficiency, image transformations are applied to the dataset. The images are auto-orient and resized – stretched to 416*416.

7 Augmentation

In order to let a model learn from new training samples, augmentations are applied to training set. The images are flipped horizontally and vertically and sheared ±15° horizontally and ±15° vertically. The brightness of the images is adjected between -25% and +25%. The images are also blurred up to 2px and cropped from 0% minimum zoon and 20% maximum zoom.

8 Model selection for model development

Faster R-CNN, YOLOv5 and RetinaNet are selected for model development. When it comes to object detection, YOLOv5 and Faster R-CNN are popular. Faster R-CNN has high performance and YOLOv5 is extremely fast. RetinaNet also has good performance in object detection.

Transfer learning technique is used in this project, which means the selected model need to have a pre-trained model. For Faster R-CNN, pre-trained model R50-DC5 is selected. For YOLOv5, pre-trained model yolov5s is selected. For RetinaNet, pre-trained model R50 with 1x learning rate schedule is selected.

9 Model analysis

The model developed using the Faster R-CNN pre-trained model (R50-DC5) has around 96% accuracy (Figure 4).

Figure 4. Accuracy of the model developed using the Faster R-CNN pre-trained model (R50-DC5)

The model developed using the YOLOv5 pre-trained model (yolov5s) has around 77% precision, around 86% recall and around 0.018 loss (Figure 5).

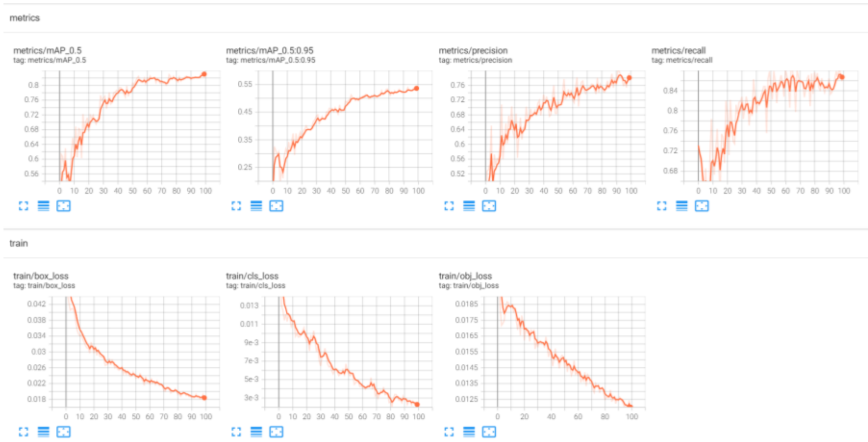

Figure 5. Training process of the model developed using the YOLOv5 pre-trained model (yolov5s)

The model developed using the RetinaNet pre-trained model (R50 with 1x learning rate schedule) has around 0.3 total loss (Figure 6).

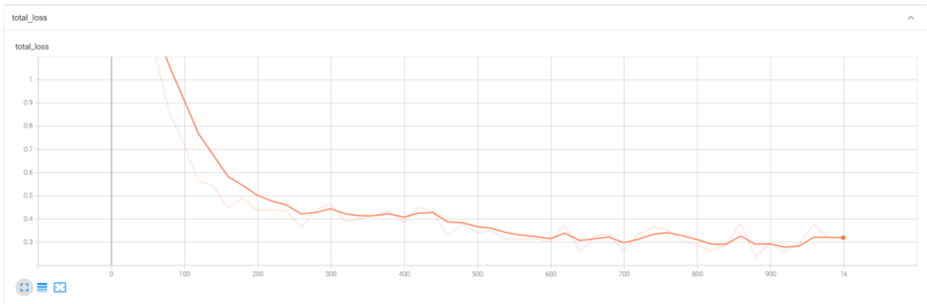

Figure 6. Total loss of the model developed using the RetinaNet pre-trained model (R50 with 1x learning rate schedule)

10 Final model selection

Accuracy is the main factor when it comes to model selection. The better accuracy a model has, the better performance it has. The model developed using the Faster R-CNN pre-trained model (R50-DC5) has around 96% accuracy (Figure 4). Therefore, the model developed using the Faster R-CNN pre-trained model (R50-DC5) is selected in this project.

11 Real-life usage of selected model

The selected model has great performance (around 96% accuracy) so it can be used to advertise Australian wildlife. For example, a mobile app with kangaroo and wallaby detection and classification can be developed to help tourist to classify kangaroo and wallaby when they travel to Australia or visit the Australian national park. Moreover, the Australian government can use the model to develop an application to help counting the total number of kangaroo and wallaby in Australia. Besides, the model can be used to classify kangaroo and wallaby in videos such as YouTube video to help people to learn more about kangaroo and wallaby.

12 Advantages of transfer learning

The main benefit of transfer learning is efficiency. It is because a model can be built using a pre-trained model, a model development for model training is not necessary. Therefore, transfer learning can save a lot of time.

Transfer learning also allows a model development with a small dataset due to a pre-trained model.

13 Disadvantages of transfer learning

Negative transfer is the biggest problem of transfer learning. The only way that transfer learning works is that when there is a similar initial and target problem. If a pre-trained model is not for object detection but is used for object detection model development, the model will perform worse.

For transfer learning, there is not much customization such as layers cannot be removed due to using pre-trained model. That means the performance relies on the pre-trained model and cannot have much improvement.

14 Possible improvement of the final model

Now the selected model cannot classify albino wallaby as wallaby because there is no albino wallaby annotation in the dataset. The dataset can include albino wallaby to solve the problem.

Now the model can only detect and classify kangaroo and wallaby. In order to turn the model into a commercial product, more animals such as Tasmania devil and wombat can be added to the dataset to train the model to detect and classify more animals so as to advertise wildlife.

15 Other possible approaches

When it comes to object detection, SSD is also a popular model. This project can consider finding an SSD pre-trained model for model development.

This project can also consider having one model development that does not involve transfer learning so as to avoid the limitation of pre-trained models and have more customization during the model development such as layer removal in order to increase the model performance.