Animal Image Classification with ANN, CNN and VGG16

Hongyang Zheng, Heng Zhou, Zhengqian Xu, Youyou Xie

November 28, 2019

Abstract

With the rapid development in deep learning and neural networks, more powerful tools are introduced to deal with image problem. In this paper, we use several methods to do animal classification. Our work begins with a brief introduction of object classification and related work. Then we introduce our dateset and how we preprocess and transform it. Next we use three different methods to train our data and build model. Finally, performance of each model is provided and compared, which we can use for further analysis and conclusion.

1 Introduction

Artificial Intelligence (AI) nowadays becomes a popular topic in our life and it seems to exist everywhere, like FaceId and robots. Here, we would like to specifically introduce one type of AI techniques - Computer Vision, which is defined as "a subset of mainstream artificial intelligence that deals with the science of making computers or machines visually enabled". Human use eyes to capture images while computers can do it in another way[4]. Thanks to the advantages of deep learning, this kind of technology is now widely used in object classification. For example, people can use it to identify different areas by the type of land usage. In addition, the method for image preprocessing and recognition have changed from traditional machine learning to Convolutional Neural Networks (CNN).

For our project, we will also focus on this field but concentrate on animal classification. We want to identify and classify animal species based on pictures taken by wildlife conservatories using different deep learning tools. We used different methods: ANN, CNN, VGG16 for classification and compared their performance.

2 Related work

One notable example is about how to use neural networks to o classify huge amount of images in the ImageNet. The ImageNet challenge has been traditionally tackled with image analysis algorithms such as SIFT with mitigated results until the late 90s[3]. In a long time, people did not make progress for it. However, a gap in performance has been brought by using neural networks. Inspired by Y. Lecun et al. (1998), a Convolutional Neural Network (CNN) called AlexNet achieved a top-5 error of 15.3% in the ImageNet 2012 Challenge, more than 10.8 percentage points lower than that of the runner up.

Since the 2012 milestone, researchers have tried to go deeper in the sequences of convolutional layers. In 2014, Karen Simonyan and Andrew Zisserman from the Oxford Vision Geometry Group (VGG) achieved top results for image classification and localization with their VGG model [2]. It is composed of sixteen convolutional layers, multiple max-pool layers and three final fully-connected layers. Since then, more people contribute to this field, which leads to its rapid development in the past few year.

3 Dataset

The dataset comes from Animal-10 dataset in Kaggle. It contains about 28,000 median quality animal images belonging to 10 categories. Limited by the computer speed, we only choose a subset of animal types, which are cows, sheep and horses, because they are very similar to each other.

After filtering files which are not JPEG format, we did image transformation and resized them into the shape of 100*100*3. Finally, there are 2672 horse images, 2039 cow images and 1559 sheep images. We randomly selected 80% of our data as the training data and 20% as the test data

4 Methods

We started with ANN as a benchmark with cross validation, then trained CNN with train-validation set to see how much the performance can be improved and finally applied pre-made transfer learning VGG16 to enhance the accuracy[1]. We played with different parameters for each model, kept the best result and compared their performances based on categorical cross entropy.

5 Results

5.1 ANN

We first trained an ANN with two layers as a benchmark. The first layer has 8 nodes and we used relu as activation function for it. The second layer has 3 nodes and we used softmax as activation function for multi-class prediction. The test accuracy is 45.14%.

Then We added one layer between the first and second layer, which has 4 nodes and took relu as activation function. The test accuracy is 44.42%.

5.2 CNN

We trained 3 CNN models with epochs at 8, 18 and 20. Results of CNN model with 8 epochs indicate underfitting, so we increased epochs to 20. However, results of CNN model with 20 epochs show that the highest validation accuracy occurred around 18 epochs. Therefore, we decided to use 18 epochs to train our final CNN model and the test accuracy is 71.53%.

5.3 VGG16

For VGG16, we also trained 2 models. First, we trained VGG16 with 8 epochs, dropout = 0.3, the test accuracy is 73.13%, which is close to the test accuracy by only using CNN. We next trained VGG16 with epoch at 10 and decreased the dropout to 0.2, it performs very well , with the test accuracy about 81%.

The plot of results for the final CNN model and VGG16 model is showed below.

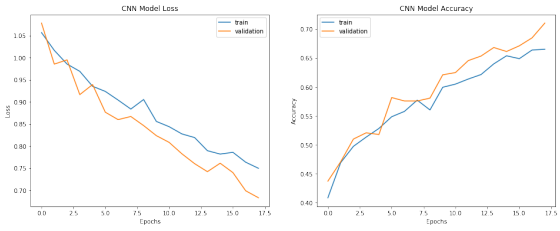

Figure 1: CNN Model Training Process

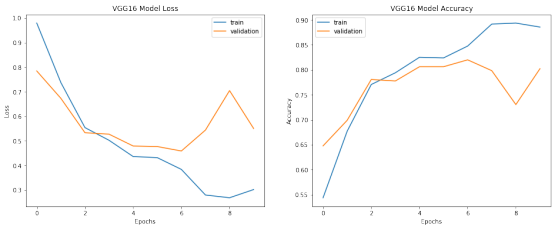

Figure 2: VGG16 Model Training Process

6 Discussion of Results

Table 1: The Result Table

Model	Test Loss	Test Accuracy(%)
ANN(2 layers)	-	44.41
ANN(3 layers)	-	45.13
CNN	0.67	71.53
VGG16	0.54	80.94

From the table, we can see that the accuracy for ANN model(2 layers) is not very high, so we improved it with three layers ANN model. The results show that increasing the number of layers does not improve the model performance much. Overall, the test accuracy is not high. It may because ANN could not deal with multi-class problem for image very well. Therefore we tried CNN model to solve this problem.

For the CNN model, we tried different training epochs and found it is an important factor to influence the model performance. Finally we chose a point with good balance between underfitting and overfitting and got a much higher accuracy than ANN accuracy. We can see that CNN is a more proper approach to deal with image classification problem than traditional ANN.

Our best performance came from VGG16, which agrees with our expectation since VGG16 embeds with lots of neural networks. However, it turns out the performance only increases by 9%, which is lower than our assumptions. After trying to tune epochs, dropout and activation function, the best test accuracy is still about 81%.

7 Conclusion

Our project starts from image input, then performs image transformation and modeling. It turns out both CNN and VGG16 perform well in image recognition compared to ANN. We finally get test accuracy of 81% using VGG16. Even though it is quite a good result for us, we believe we can do it better in the future if we try more transfer learning approaches with VGG and more parameters (i.e., epoch, dropout).

References

[1] Jason Brownlee. Animal classification with vgg16 cnn, 2019.

[2] Jason Brownlee. A gentle introduction to the imagenet challenge (ilsvrc), 2019.

[3] Arthur Ouaknine. Review of deep learning algorithms for image classification, 2018.

[4] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review, 2018. cite arxiv:1807.05511.