ACM40960 Projects in Maths Modelling

Final Report

Dog and cat image recognition based on convolutional neural network

Yilu Zhao 20211696

Ruijia Sun 21203592

Abstract

This research presents a dog and cat image recognition problem, i.e., to determine whether the animal in each image is a cat or a dog. In the paper, deep learning approach is used to build a convolutional neural network to achieve binary classification. Convolutional neural network as a deep learning algorithm that performs well in computer vision. It takes an image as input, extracts, and learns features of the image, and classifies based on the learned characteristics. Then we train and evaluate the classifier on Kaggle's dog vs. cat data set.

Key words: Convolutional neural network, image classification, TensorFlow ,Keras library

1.Introduction

We know that even a child can tell what an animal is in a picture after learning. So can we similarly teach computers to recognize images. Maybe AI technology will bring us the answer, and the most powerful and potent artificial intelligence we make today is machine learning (Dick et al. 2019). Machine learning is a multi-disciplinary discipline used to study how computers simulate or implement human learning behaviors, which involves probability theory, statistics, approximation theory, convex analysis, algorithmic complexity theory, and other disciplines. Machine learning has become almost the first method of choice for developing practical software for computer vision, speech recognition, natural language processing, robot control, and other applications (Jordan et al. 2015). The most widely used machine learning methods are supervised learning methods. Deep networks have rapidly evolved in recent years as an area of supervised learning with significant advances and have had a major impact in areas such as computer vision and speech recognition, also achieving significant improvements in performance (Fang et al. 2019). The basic idea of deep neural networks is to obtain better feature robustness by constructing multilayer networks with multilayer representation of the target with a view to representing the abstract semantic information of the data through multiple layers of high-level features (Jordan et al. 2015). Among deep networks convolutional neural networks (CNN) are widely used as one of the methods in computer vision.

In this paper, we will build a picture recognition model based on CNN to implement cat and dog picture binary classification. So, what is a convolutional neural network (CNN), CNN is a kind of deep neural network with convolutional structure (Ráduly et al. 2018)

The convolutional structure can reduce the amount of memory occupied by the deep network with three key operations, one of which is local perceptual field, the other is weight sharing, and the third is pooling layer, which effectively reduces the number of parameters of the network and alleviates the overfitting problem of the model (Xiaoxiao et al. 2021).

The remainder of the paper is structured as follows: Section 2 presents the related work by the different researchers, while Section 3 is about problem statement. Section 4 details the experiment based on CNN. Section 5 provide the result of the study.

2. Related Work

We will briefly introduce problems are solved by different researchers on convolution neural networks and data augmentation. Because this paper focuses on convolution neural network classification and data augmentation to improve the performance of deep learning model for picture recognition.

CNN is a kind of deep neural network with convolutional structure that is well-known. A lot of research has been conducted on the image classification of cats versus dogs. Concerning " dog versus Cat" data set, a lot of good methods have been provided. P. Sermanet, achieved 98.9 percent on the "cat versus Dog" Kaggle rivalry in 2013, which is the most astounding accuracy in the past few years. However, various breeds resemble each other. Consequently, they need much finer-grained image order to deal with this problem. They employed cutting-edge deep learning techniques. Finally, the Dogs vs. Cats dataset had a few more requests for size information available for each classification: 12,500 versus 60, which solved the issue of multi-class characterization (Kumar et al. 2019). Geometric and appearance models of breeds using model-based data were created to address the breed recognizable proof issue (Sermanet et al. 2014). By this way, achieving a recognition rate of 67 percent. Consequently, these methods including segmentation, extension with noisy data and so on are progressive for image recognition.

In 2017 Howard et al. work on the Stanford Dogs dataset extended with noisy data learning of a CNN using the Mobile Net architecture as the main method to solve similarfine-grained image recognition problems.

3. Problem statement

This is a problem of binary classification of images of dogs and cats. The goal is to classify the input image into the correct class. In this study we will build a convolutional neural network based on the Keras package in TensorFlow. Furthermore, we'll use data augmentation techniques to reduce overfitting and increase model accuracy.

3.1 Data Exploration

The dataset named Dogs vs. Cats used in this study is from Kaggle website. The data is divided into a training set and a test set. Train dataset has 12500 dogs' images and cats' also. There are 12500 images in test dataset. The naming of photos in these two directories is different. The pictures in the training set have the label cat or dog in their names while the test set does not. The size, shape, and pixel values of the pictures in the dataset are not equal.

3.2 Algorithms and Techniques

The CNN model operates in two stages: extraction and classification of features. Feature extraction is the stage in which various filters and layers are applied to an image to extract information and features from it, and once completed, it is passed on to the next stage, classification, which classifies them based on the problem's target variable. CNN model involves the layer of input, layer of convolution, activation function, layer of collection and the layer of connection (Gu et al. 2018).

Data Augmentation is an effective means to overcome the shortage of training data by artificially extending the training dataset by generating more equivalent data from a limited amount of data (Wang et al. 2017). Rather than the artifact of the training images, we want the network to learn the important features that are invariant for the object classes. To realize this, various data augmentation techniques such as flip, rotation, cropping, and zoom are used, which are supervised data augmentation. In addition, unsupervised data augmentation is a more advanced approach than these supervised data augmentations (Wu et al. 2015).

TensorFlow is an end-to-end open-source machine learning platform that is widely used for the programmatic implementation of various machine learning algorithms.

Keras is an open-source artificial neural network library written in Python that can be used as a high-level application programming interface to TensorFlow for the design, debugging, evaluation, application, and visualization of deep learning models. Keras is the necessary library for our model.

4. Experiment

This article describes data preprocessing and convolutional neural network model building in this section. The initial CNN model is trained, and the model and data are optimized based on the results. In the optimization process we augment the data and adjust the model structure.

4.1 Dataset pre-processing

Loading the required datasets and libraries before the start of the experiment is necessary for the computer to work. Then use TensorFlow to execute the processing part.

The presented CNN learning methodology revolves around the Dogs vs. Cats dataset from Kaggle. In this experiment we selected a subset of this dataset for use. There are

1000 images in the training set, 500 images of cats and 500 images of dogs. The validation set and the test set each contain 400 images. The images in the datasets are all different.

Moreover, all images are not of the same size and shape, so to prepare the model we need to resize the images to the same size square. Here images are reshaped to 180x180 pixels and rescaling by 255.

4.2 Initial CNN model

In this part we specify a CNN model with four more convolutional layers, interleaved with four pooling layers, and then followed two fully connected layers.

The first convolution layer is set with 32 filters and a 3×3 kernel with default strides. The second convolution layer is set with 64 filters, with 3×3 kernels. The following two convolution layers are set with 128 filters, with 3×3 kernels. All max-pooling layers have a pool size of 2×2, thus halving width and height at every pass. In the fully connected layers, we add dropout.

We need to note that cat and dog classification is a binary problem, so the last dense layer of the network is a single unit activated using sigmoid and using binary cross entropy as loss function.

After model fitting, we plotted the curves of accuracy, validation accuracy, loss and validation loss and concluded from the line graph that the model was overfitted.

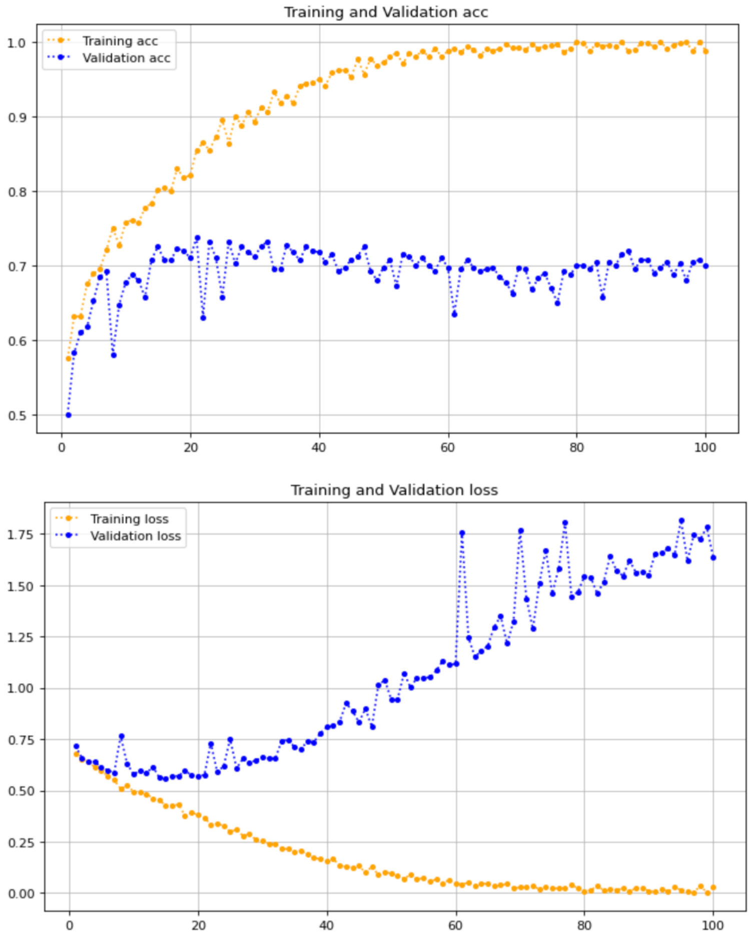

Fig.1 Accuracy and Loss for initial model

As the epochs increases, the training accuracy increases and approaches 100%, while the validation accuracy stays at 70%. The validation loss almost reaches the minimum after the 10th round and continues to rise within a certain number of rounds after that, while the training loss keeps decreasing and directly approaches 0.

4.3 Model improvement

In this section we improve the initial model. First, we perform data augmentation on the original data and then adjust the model parameters according to the fit. After this we use the optimized model to predict the test data.

4.3.1 Data augmentation

As a widely used solution to model overfitting, data augmentation generates more training data from the existing training sample, using a variety of random variations capable of generating plausible images to increase the data sample.

In this research, we randomly flip half of the images horizontally, followed by panning the images horizontally and vertically. We also randomly scale the images, randomly stagger transform, and randomly rotate them. The images after data augmentation with these methods are shown below.

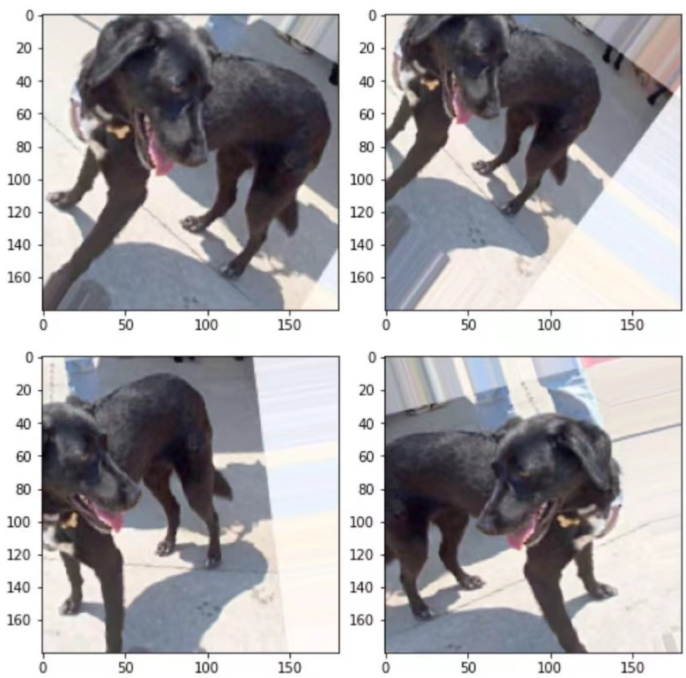

Fig.2 Image after data augmentation

4.3.2 Further model building

After experimentation and adjustment, we increased the number of fully connected layers from two to three and changed the dropout from 0.3 to 0.2.

5.Result

The models were trained for 100 epochs and basically reached convergence. Table I shows the Accuracy and table II shows the Loss from the model. Table I shows us the train accuracy is achieved 82.99% and the validation accuracy is also closed to 80%.

When the model is used to make predictions on the test data the prediction accuracy reaches 77.75%. This means that the probability of the model correctly categorizing a new set of pictures of cats and dogs is 77.75%, which is a relatively accurate result. Table II gives the information on the Loss: The training Loss decreases to 0.3841 while the validation and test Loss are both drop to approximately 0.45.

Train Accuracy	Validation Accuracy	Test Accuracy
82.99%	78.25%	77.75%

Table 1. Accuracy

Train Loss	Validation Loss	Test Loss
0.3841	0.4753	0.4548

Table 2. Loss

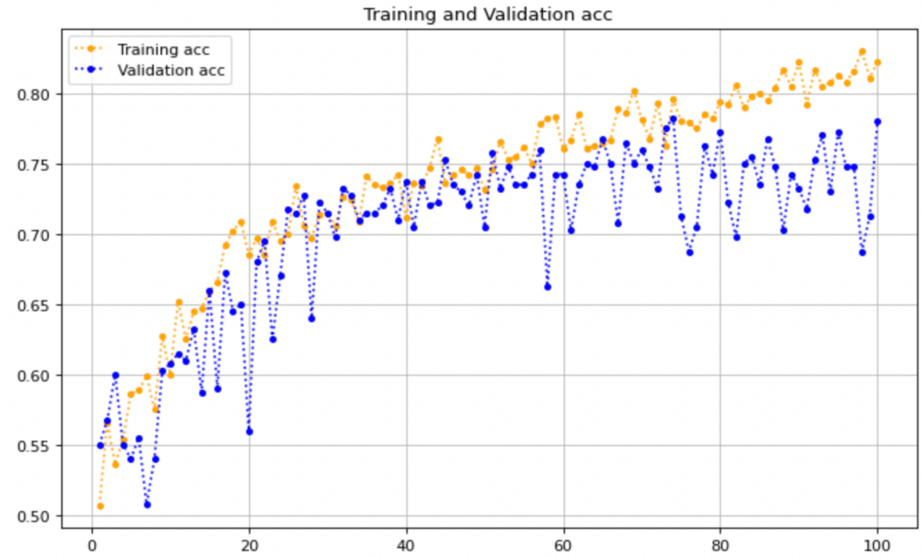

Fig.3 Training and Validation accuraccy for final model

Fig.3 provides the training and validation accuracy of the updated model for 100 epochs. We can observe that after adding data augmentation and improving the model structure, the model is no longer overfitted and the training curve follows the validation curve. After 80 epochs, the validation accuracy curve flattens out and reaches nearly 78%.

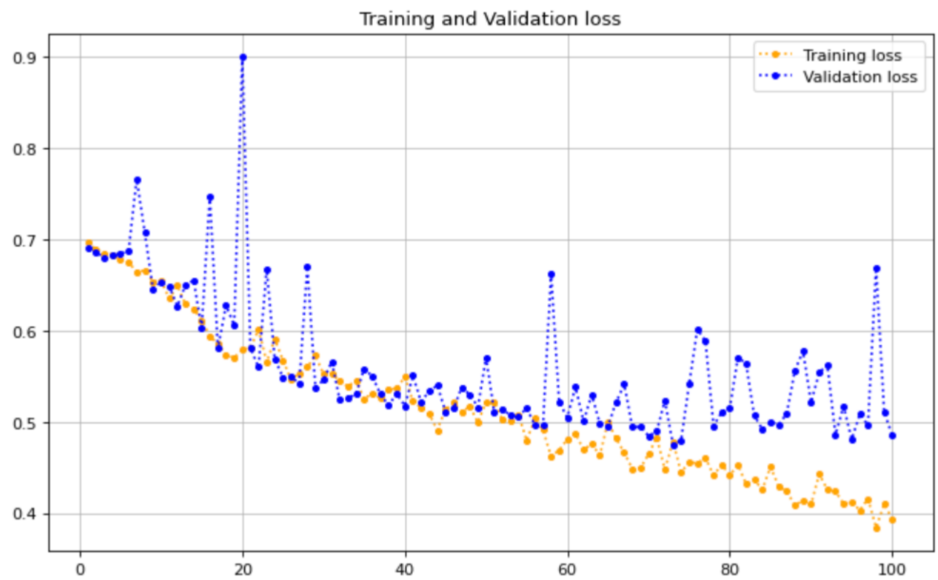

Fig.4 Training and Validation Loss for final model

Figure 4 details the training and validation Loss of the updated model for 100 epochs. Like the accuracy curve, the train Loss and validation Loss curves drop at almost the same rate during this process. This phenomenon also demonstrates that the overfitting situation has been largely alleviated.

Combining the information from these two images, we can notice that Accuracy and Loss change in opposite directions, i.e., the higher the Accuracy, the lower the Loss, and the more satisfying the results we get. Although the results of the classification model visualization tell us that the model accuracy is high and the overfitting situation is solved to a great extent, the validation curve of the model still plateaus before the training curve. This indicates that our model still has space to advance.

All In conclusion, we are satisfied with the results of the obtained cat and dog image recognition binary classification model, which has an accuracy of approximately 82.99% and a loss of approximately 0.3841. The results on the brand completely new test dataset are 77.75% accuracy and a loss of 0.4548. This model will help us to identify cats and dogs in a reliable way.

Conclusion

In the backdrop of the rapid development of deep neural networks in the field of computer vision, a convolutional neural network-based image recognition model is developed in this research. CNN were used to extract features from dog and cat images to achieve binary classification of dog and cat pictures. The initial classification model was adjusted step to step to reach the final version. In summary, the trained network performed well on the test dataset for image recognition and classification.

In future to increase the performance of model, we may use unsupervised data augmentation methods such as adversarial generative networks to improve model accuracy to reduce overfitting.

Appendix I Breakdown of tasks done by each student

Yilu Zhao:

Report Section1, 4, and 5, Conclusion and Abstract, Code implementation in experiment.

Ruijia Sun:

Report Section 2 and 3: Problem statement, Sort out literature and articles.

Reference

[1] Fang, W., Ding, Y., Zhang, F. and Sheng, V.S., 2019. DOG: A new background removal for object recognition from images. *Neurocomputing*, *361*, pp.85-91.

[2] Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), pp.255-260.

[3] Wu, R., Yan, S., Shan, Y., Dang, Q. and Sun, G., 2015. Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876, 7(8).

[4] Yangxiaoxiao, Z., 2021. Image modal analysis in art design and image recognition using AI techniques. Journal of Intelligent & Fuzzy Systems, 40(4), pp.6961-6971.

[5] Kumar, A. and Kumar, A., 2020, December. Dog breed classifier for facial recognition using convolutional neural networks. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 508-513). IEEE.

[6] Ráduly, Z., Sulyok, C., Vadászi, Z. and Zölde, A., 2018, September. Dog breed identification using deep learning. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000271-000276). IEEE.

[7] Kumar, R., Sharma, M., Dhawale, K. and Singal, G., 2019, December. Identification of dog breeds using deep learning. In 2019 IEEE 9th International Conference on Advanced Computing (IACC) (pp. 193-198). IEEE.

[8] Xie, G., Yang, K., Bai, Y., Shang, M., Rui, Y. and Lai, J., 2016, July. Improve dog recognition by mining more information from both click-through logs and pre-trained models. In 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-4). IEEE.

[9] Jajodia, T. and Garg, P., 2019. Image classification–cat and dog images. Image, 6(23), pp.570-572.

[10] Jones, H.G., 2020. What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. AoB Plants, 12(6), p. plaa052.

[11] O'Neal, A., Rodgers, B., Segler, J., Murthy, D., Lakuduva, N., Johnson, M. and Stephens, K., 2018, December. Training an emergency-response image classifier on signal data. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 751-756). IEEE.

[12] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T., 2018. Recent advances in convolutional neural networks. Pattern recognition, 77, pp.354-377.

[13] Sermanet, A. Frome, and E. Real, Attention for fine-grained categorization, CoRR, vol. abs/1412.7054, 2014.