Dogs Emotion Recognition and Parameter Analysis Based on Efficientnet with Transfer Learning

Yinglin Xie
Faculty of Science and Technology, Beijing Normal University-Hong Kong Baptist University, United International College (UIC), Zhuhai, China

Keywords: Emotion Recognition in Dogs, EfficientNet, Fine-Tuning, Transfer Learning.

Abstract: Emotion recognition in dogs plays a crucial role in understanding their mental state and improving their welfare. This study aims to develop an efficient model for recognizing emotions in dogs' images using the Efficient Neural Network(EfficientNet) architecture. Specifically, this study uses a dataset of dog images labeled with different emotion categories. The dataset is preprocessed, while transfer learning techniques are used for training model. Fine-tuning of hyperparameters is performed to optimize the model. Finally, the trained model is evaluated on the testing set. This study is conducted on the dog-emotion-prediction dataset, achieving an accuracy of 78%. Experimental results demonstrate the effectiveness of using EfficientNet for this task. Recognizing emotions in dogs' images has practical implications in various domains, such as veterinary care, animal behavior analysis, and pet training. It can aid veterinarians in diagnosing and treating emotional distress in dogs, assist trainers in developing tailored behavior modification strategies, and promote a deeper understanding of dogs' emotions to strengthen the human-animal bond.

1 INTRODUCTION

Emotion recognition is the process of identifying and classifying emotions from verbal and non-verbal cues. Dogs, cats, and birds are favorite pets that have mirror neurons. Mirror neurons have been directly observed in humans for understanding and responding to emotions (Valentina et al 2019). Dogs are highly valued in a variety of roles such as therapy dogs, service animals, and even family pets due to their strong ability to understand human emotions and respond empathically. Recognizing and predicting dogs' emotions can provide valuable information for understanding their health. Also, a better understanding of dog emotions can help humans strengthen the dog-human bond. Studies have shown that dogs are able to extract and integrate bimodal sensory and emotional data to recognize emotions (Albuquerque et al 2016).

The analysis technology of human faces is relatively mature at present, while the emotion analysis technology of animals has not been significantly developed. In the realm of recognizing animal emotions, in order to study animal emotions and focus on animal emotion health, motion tracking and gesture recognition were used by researchers from 2014 to 2022 (Chen et al 2023). Animal emotion recognition from pictures is a research field that has received extensive attention in recent years. This method predicts an animal's emotional state by analyzing its image features and expressions. During the development process, the researchers mainly relied on machine learning techniques and computer vision to realize animal emotion recognition (Corujo et al 2021). The fundamental goal of the tests about dogs conducted by Franzoni and Boneh, which employed photographs to induce emotional states, was the identification of emotion through the animal's face (Tali et al 2022). By employing 23 body and facial areas as crucial locations, Ferres et al. were able to identify dog moods from body positions (Ferres 2022). To extract characters from the photos, such as color, texture, and form, the researchers employed computer vision algorithms. Deep learning models or conventional image processing methods can be used to extract these information.

Machine learning algorithms may be used to detect an animal's emotional condition efficiently in images. Some common examples of machine learning, such as support vector machines (SVM), convolutional neural networks (CNN), or random forests, are widely used. Currently, some progress has been made in research on recognizing animal emotions from images. Some

Xie, Y.
Dogs Emotion Recognition and Parameter Analysis Based on EfficientNet with Transfer Learning.
DOI: 10.5220/0012799600003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 240-245
ISBN: 978-989-758-705-4

studies have shown that the emotional state of animals can be predicted more accurately by analyzing their facial expressions and postures (Tsai et al 2020, Raman et al 2021 & Pramerdorfer and Kampel 2016). For example, a dog with upturned corners of its mouth may indicate happiness, while ears flattened backward may indicate fear.

The main objective of this study introduces the EfficientNet model for dog image emotion recognition task. Specifically, first, in the preprocessing stage, data set partitioning, traversal and storage operations are performed on dog images to realize the partition and storage of data sets. The second is to augment the training data, which can generate more training samples by randomly transforming and enhancing the input image. Third, the EfficientNet-B0 model is introduced to construct the recognition model. Transfer learning technique is introduced to enhance the generalization ability of the model. A pre-trained EfficientNet-B0 model is used to accelerate model convergence and improve accuracy by utilizing pre-trained weight parameters on large-scale image datasets. During training, the Adam optimizer and cross-entropy loss are applied to minimize the prediction error of this model. On the test dataset, the model shows good accuracy and Area Under the Curve(AUC) values. The results of this study embody that the proposed model is effective for dog emotion recognition. The research studied in this paper can help improve human understanding of dog emotions, promote animal welfare, training, and communication, and be used in a variety of areas such as intelligent assistive tools and entertainment applications.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The dataset used in this study, called Dog Emotions Prediction, is sourced from Kaggle (Dataset). The dataset comprises photographs of various dogs displaying a range of emotions. This dataset consists of color images that have been automatically scaled to include dog faces or full bodies. This dataset contains two main features: pixel values of images and sentiment labels. Dog images are classified into 4 categories according to emotion. These 4 categories are sad, happy, angry, and relax. The size of each image is 384 x 384 pixels. There are a total of 15921 examples in this dataset. In the data preprocessing part, the study divides the data set, and defines the relevant directory and category information, creating a storage directory, and finally moving the image accordingly after traversing each category. After the data set is divided, The public test set makes up 20%, whereas the training set makes up 80%. Fig. 1 shows some examples in the dataset.

Figure 1: Images from the Dog Emotions Prediction dataset (Original).

2.2 Proposed Approach

The major focus of this paper is dog emotion recognition with EfficientNet-B0. After preprocessing the data, advancing the training dataset aims to increase the capacity of the model for generalization. Then, the EfficientNet-B0 model is utilized to build a recognition model. The pre-trained weight parameters are used to promote model convergence and improve accuracy. The Adam optimizer and cross-entropy loss are employed to reduce the model's prediction error during training. The model shows good evaluation results on the test dataset. Finally, predictions are then made. In Fig. 2, the system's structure is shown below.

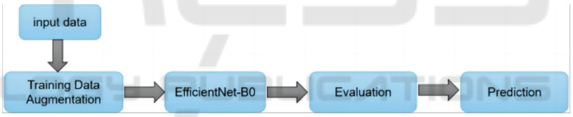

Figure 2: The pipeline of the system (Original).

2.2.1 EfficientNet-B0

EfficientNet-B0 is a variant in the EfficientNet model family. EfficientNet is a series of efficient and lightweight convolutional neural network architectures. Through the method of compound scaling, it reduces the parameter quantity and computational complexity while exactly maintaining high precision. The EfficientNet model family's design objective is to decrease the model's computational cost and number of parameters while retaining excellent accuracy. The difference between EfficientNet and EfficientNet-B0 is the size. EfficientNet is a model series that includes multiple models of different scales, such as EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, etc. These models differ in width, depth, and resolution, with larger models generally having higher accuracy and higher computational complexity.

EfficientNet-B0 adopts a comprehensive scaling method to improve model performance by expanding the depth, width, and resolution of the network. This comprehensive scaling method can effectively control the consumption of computing resources while increasing the model capacity. Depthwise Separable Convolution and Enhanced Inverted Residuals are adopted to reduce the amount of parameters and computational complexity, while improving the representation ability and feature extraction ability of the model. EfficientNet-B0 can find a reasonable balance between accuracy and computational efficiency by automatically searching and optimizing the network structure. The significance of using EfficientNet-B0 is that it can achieve efficient and accurate image classification tasks with limited computing resources. Compared with other more complex models, EfficientNet-B0 has a smaller model size and lower computational requirements, making it suitable for deployment in mobile devices and embedded systems. Therefore, in this experiment, the author uses the most basic EfficientNet-B0.

2.2.2 The Construction of Model.

The structure of this model consists of multiple repeated blocks, each block contains multiple convolutional layers, batch normalization, and activation functions. The whole network consists of fully connected layer, convolutional layer, pooling layer and global average pooling layer. The entire network starts with a convolutional layer, extracts and transforms features through the repetition of different blocks, and reduces the feature map size through pooling layers and global average pooling layers, and performs global pooling on features. Finally, the features are mapped to the corresponding classification results through a fully connected layer.

This study first creates a basic model of EfficientNetB0. The model uses the weights pre-trained on ImageNet, and the images' input size is (224, 224, 3), and the pooling method is set to None (that is, no pooling operation is performed), and the number of output categories is 4. At the same time, make all layers of the base model non-trainable. When building a model, a Sequential model object is first created. Then, the model structure is added layer by layer. Add EfficientNetB0 as the basic model, and then add following layers: Dropout layer (discard rate 0.5), Flatten layer, Batch Normalization layer, fully connected dense layer with He initializer (32 neurons). Add Batch Normalization layer again, Rectified Linear Unit (ReLU) activation function layer and with the output dense layer of the softmax activation function. Finally, a summary of the model is printed out, showing the layers of the model and their number of parameters. The layers of the model are represented in Fig. 3.

Layer
efficientnet-b0
dropout
flatten
batch normalization
dense
batch normalization
activation
dense

Figure 3: Layers of the model (Original).

2.2.3 Loss Function

Particularly in classification problems, cross-entropy loss function plays a pivotal role in many machine learning tasks, employed in this study, which optimizes model parameters by minimizing the cross-entropy of predicted probabilities and true labels. The following is the mathematical expression follows:

$$L_{(y,\hat{y})} = -\sum_{i=1}^{N} y_i \, log(\hat{y}_i) \quad (1)$$

Among the expression (1), $L_{(y,\hat{y})}$ represents the cross-entropy loss function. N is the total number of categories. y_i is the i-th category's value (0 or 1) in the true label, while $\hat{y}_i$ is the likelihood that the model correctly predicted the i-th category. The cross-entropy loss function gauges the difference between the probability distributions of the true label and the predicted label. It quantifies the difference between these distributions. When the model's prediction aligns perfectly with the true label, the cross-entropy loss reaches its minimum value of 0. However, as the dissimilarity between the predicted and true distributions increases, the cross-entropy loss also increases. Through minimizing the cross-entropy loss function, the model is able to continuously adjust the parameters when training the model, so that the output of the model is closer to the real label, thereby improving the classification performance of the model.

2.3 Implementation Details

This study uses a Graphics Processing Unit (GPU) T4 x2 configuration. During data augmentation, rescale

the input image pixel values to between 0 and 1. Specify the proportion that will be used for verification in the training data set as 20%. Set the angle range of the randomly rotated image to 5 degrees. Besides, the strength range of the shear transform, the scale of the random horizontal and vertical offset images are all set to 0.2. The probability of randomly flipping the image horizontally and the probability of randomly flipping the image vertically are set to half and the method of filling newly created pixels is set to 'nearest', which means fill using nearest neighbor interpolation. Resize the input image to (224, 224) and set the number of samples included in each batch to 64. The learning rate increases 0.50 times after the loss stops improving. A reduction in the learning rate is triggered when the validation metric does not improve for 20 consecutive epochs. The learning rate will be reduced to a minimum of 1e-10.

3 RESULTS AND DISCUSSION

In this study, EfficientNet-B0 is applied to train images to recognize emotions. This chapter evaluates and analyzes the performance of the generated model, mainly from four aspects: accuracy, loss, AUC, and precision. Each plot contains training and validation curves with corresponding titles and axis labels.

As shown in Fig. 4, it shows the change of accuracy rate. The most frequently employed metric for evaluating classification performance is accuracy. In general, accuracy refers to the ratio of quantities correctly classified and all quantities trained. Among 1-25 epochs, the accuracy rate of training gradually increases from 75% to 80%, basically showing a continuous upward trend. Validation accuracy consistently swings back and forth between 77 percent and 78 percent. It shows the history of the model loss value. With the optimization of the model, the loss value continues to drop from 1.30 to 1.05. The AUC is computed as the integral of the Receiver Operating Characteristic (ROC) curve and is typically greater than 0.5 but less than 1. For a classifier, the larger the value of AUC, the better the performance of the classifier. The value of AUC rose to 0.8 in the last training epoch and 0.77 in the last validation epoch. Precision is mainly used to measure how many cases that are predicted as positive examples according to the model are true positive examples. The range of accuracy is between 0 and 1. As the proportion of correct predictions among the samples predicted as positive through the model approaches 1, the model's performance improves. As shown Fig. 4-Fig. 7, at the end, the training precision continues to rise to 0.66. Validation precision fluctuates between 0.58 and 0.60.

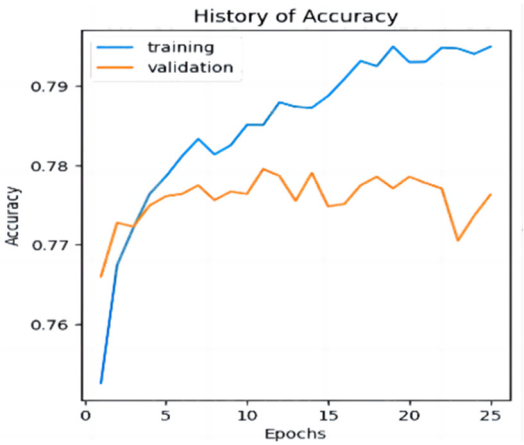

Figure 4: The curve of the value of the accuracy (Original).

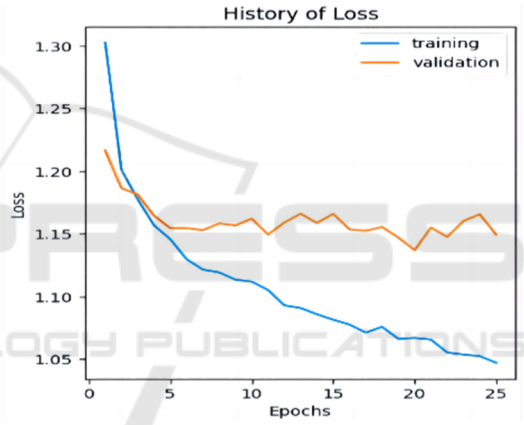

Figure 5: The curve of the value of the loss (Original).

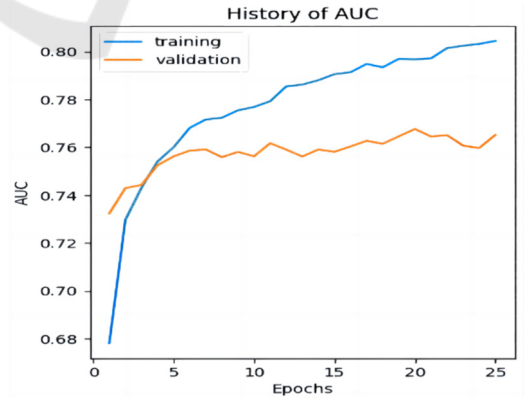

Figure 6: The curve of the value of the AUC (Original).

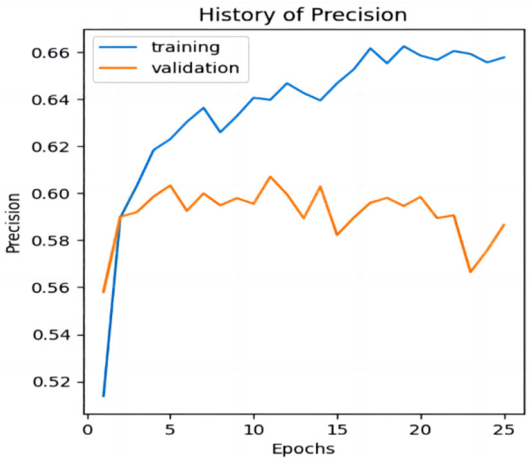

Figure 7: The curve of the value of the precision (Original).

In the early stages of training, the initial state of the model is poor, with relatively low accuracy, precision, and AUC. With the training of the model, the model gradually learns the characteristics and patterns of the data, and because the model adjusts the parameters through the optimization algorithm during the training process to adapt to the features of the data, the value of the model is continuously reduced. Loss function. This leads to a continuous upward trend in the individual performance of the training. Some of the fluctuations in validation values may be due to differences in data distribution or a small number of samples in the validation set.

As described in Tab. 1, respectively, the accuracy, precision, recall, and AUC of the model for the test are 0.78, 0.60, 0.37, and 0.77. Although the accuracy and precision are relatively high, the recall rate is low, indicating that the model's ability to identify true positive samples needs to be improved. At the same time, the value of AUC can also indicate that the model has good classification ability. Although, there is an overfitting phenomenon, the gradual improvement of various data during the training process of the model shows that the model has a certain learning ability, can adapt to the data, and extract pertinent features, while also enabling visualization of the training progress.

Table 1: The Results of Rest.

Test Loss	1.1416
Test Accuracy	0.7817
Test Precision	0.6032
Test Recall	0.3693
Test AUC	0.7683

4 CONCLUSION

This study aims to explore EfficientNet for dogs' emotion recognition, through a large amount of picture data. First, the database is preprocessed using partitioning, traversal, and storage operations to advance the generalization ability of the EfficientNet model. Then, the input image data is scaled, cut, and offset to enhance the training data. Second, the EfficientNet-B0 model is introduced to build the recognition model while using Adam optimizer and classification cross-entropy loss function to compile the model. The result demonstrates that EfficientNet can effectively capture important features related to emotions in dog images. The model exhibits robust performance across different dog breeds and varying emotional expressions. The findings of this study highlight the potential of EfficientNet as a reliable tool for emotion recognition in the domain of animal behavior analysis. This technology can help people better understand the expression of animal emotional information, so as to provide a more scientific and rational decision basis for animal health management and behavior intervention. In addition, this study provides useful ideas and methods for further exploring the application of EfficientNet model parameters and transfer learning strategies in emotion recognition. Further research could focus on expanding the dataset to include more diverse dog images, investigating transfer learning strategies, and exploring the generalization of the proposed approach to other animals.

REFERENCES

F. Valentina, M. Alfredo, B. Giulio, M. Francesco, “A Preliminary Work on Dog Emotion Recognition,” IEEE/WIC/ACM International Conference on Web Intelligence, vol. 19, 2019, pp. 91–96

N. Albuquerque, K. Guo, A. Wilkinson, C. Savalli, E. Otta, D. Mills, “Dogs recognize dog and human emotions,” Biol Lett, vol. 12, 2016, p. 20150883

H. Y. Chen, C. H. Lin, J. W. Lai, Y. K. Chan, “Convolutional Neural Network-Based Automated System for Dog Tracking and Emotion Recognition in Video Surveillance,” Applied Sciences, vol. 13, 2023, p. 4596

L. A. Corujo, E. Kieson, T. Schloesser, P. A. Gloor, “Emotion Recognition in Horses with Convolutional Neural Networks,” Future Internet, vol. 13, 2021, p.250

B. Tali, A. Shir, B. Annika, S. M. Daniel, R. Stefanie, F. Dror, Z. Anna, “A deep learning model for automatic classification of dog emotional states based on facial expressions,” arXiv, 2022, unpublished

K. Ferres, T. Schloesser, P. A. Gloor, “Predicting dog emotions based on posture analysis using deeplabcut,” Future Internet, vol. 14, 2022, p. 97

M. F. Tsai, P. C. Lin, Z. H. Huang, C. H. Lin, “Multiple feature dependency detection for deep learning technology—smart pet surveillance system implementation Electronics,” vol. 9, 2020 p. 1387

S. Raman, R. Maskeliūnas, R. Damaševičius, “Markerless dog pose recognition in the wild using ResNet deep learning model,” Computers, vol. 11, 2021, p. 2

C. Pramerdorfer, M. Kampel, “Facial expression recognition using convolutional neural networks: state of the art,” arXiv, 2016

Dataset https://www.kaggle.com/datasets/devzohaib/dog-emotions-prediction