

Dog Breed Identification with Fine tuning of Pre-trained models

B. Vijaya Kumar, K. Bhavya

Abstract: Dog Breed identification is a specific application of Convolutional Neural Networks. Though the classification of Images by Convolutional Neural Network serves to be efficient method, still it has few drawbacks. Convolutional Neural Networks requires a large amount of images as training data and basic time for training the data and to achieve higher accuracy on the classification. To overcome this substantial time we use Transfer Learning. In computer vision, transfer learning refers to the use of a pre-trained models to train the CNN. By Transfer learning, a pre-trained model is trained to provide solution to classification problem which is similar to the classification problem we have. In this project we are using various pre-trained models like VGG16, Xception, InceptionV3 to train over 1400 images covering 120 breeds out of which 16 breeds of dogs were used as classes for training and obtain bottleneck features from these pre-trained models. Finally, Logistic Regression a multi-class classifier is used to identify the breed of the dog from the images and obtained 91%, 94%,95% validation accuracy for these different pre-trained models VGG16, Xception, InceptionV3.

Keywords: Image Classification, Transfer Learning, Convolutional Neural Network, vgg16 , Xception, Inception-V3.

I. INTRODUCTION

In modern science neural network is a network composed of artificial neurons or nodes. Artificial Neural Networks are used for solving Artificial Intelligence problems. Deep learning is growing fastly as a key instrument in the applications of artificial intelligence. For example in the field of computer vision such as natural language processing and speech recognition. These fields produce results that are remarkable, and leads to the growing interest in the area of deep learning. One area where deep learning serves excellently is image classification. The most frequently used deep learning method for image classification is Convolutional Neural Networks. CNN are similar to artificial neural networks which have learnable weights and biases[1]. The difference between artificial neural networks and CNN is that in CNN's filters will process over the whole image and are the effective methods for image recognition and classification problems.

A Convolutional Neural network is a deep learning algorithm which takes in an input image and assigns importance to various features in the image and will be able to differentiate one from the other. CNN learns from the image directly there

by eliminating the manual feature extraction, which makes them to serve as an excellent feature extractor by allowing end-to-end learning of features from image data in raw form for classification. The main power of CNNs present in its deep architecture that makes CNNs powerful.

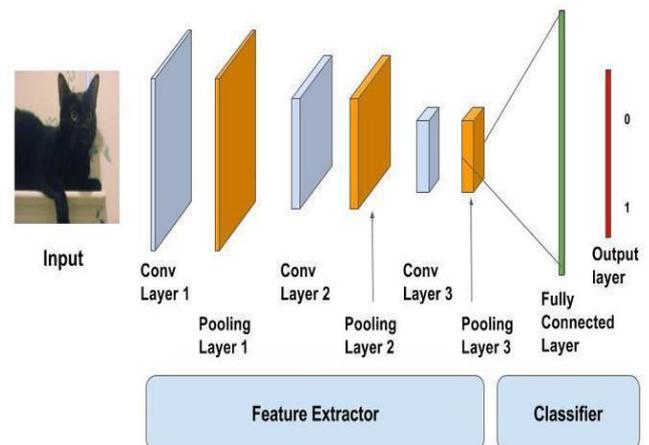


Fig 1. Convolutional Neural Network Model

In general many methods of machine learning will work under a common assumption that both the test data and train data are used from the same distribution and feature space. But if the distribution changes, then the model should be rebuilt from scratch using the newly collected data. But in many of the real world application it is expensive or impossible to rebuild these type of models.[6] Training deep CNNs from scratch requires huge computational and memory resources such as expensive GPUs. If these resources are not available, the training will be too time-consuming. So we use Transfer learning to overcome this problem.

Transfer learning refers to the use of pre-trained models. These pre-trained models are trained by feeding an image to the input, and at each layer it performs some computations on the data until it provides a output a label and classification accuracy.

In this paper, we firstly explains the concept of Convolutional Neural Network Model and the pre-trained model is described, then this dataset is used for the research and the process of image classification is done.

II. RELATED WORK AND BACKGROUND

Image recognition using deep learning performs better than human vision on many tasks, as humans can't recognize the



Revised Version Manuscript Received on 16 September, 2019.

Dr. B.Vijaya Kumar, Professor & Head of Department, Vidya Jyothi Institute of Technology (VJIT), Hyderabad, India. vijiyasree.b@gmail.com

K.Bhavya, PG Scholar CSE, Vidya Jyothi Institute of Technology (VJIT), Hyderabad, India. kodumuribhavya@gmail.com

breed of the dog by seeing it, we need to be thankful to Convolutional Neural Networks for providing a solution to this classification problem.

Hasbi Ash Shiddieqy, Farkhad Ihsan Hariadi, Trio Adiono “Implementation of Deep-Learning based Image Classification on Single Board Computer”, In this paper, a deep-learning algorithm based on convolutional neural-network is implemented using python and tflearn for image classification, in which two different structures of CNN are used, namely with two and five layers and It conclude that the CNN with higher layer performs classification process with much higher accuracy.[2]

Karan Chauhan , Shrwan Ram “Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensorflow and Keras”,In this paper firstly step image dataset is prepared, and then defined parameters for image classification to python, and then created CNN with two convolutional layers, then select different combination of activation functions and classifiers for comparison purpose. In next steps, we fit the created CNN to image dataset to Train, Test the system with training and test datasets respectively. Finally, obtain the accuracy for different CNN structures and compare these accuracies for performance measurement, and then get the resultant CNN structure.[9]

In many of the situations the owners of the dogs themselves will be unable to correctly identify the dog breed if it is a rare breed. So it might proved difficult to identify one’s ideal companion without a great deal of experience and research.

Interestingly, Zeiler and Fergus [3] use an ImageNet [It contains more than 14 million images, which are hand labelled with the presence/absence of 21,000+ categories. The images are mostly centered and the data-set is considered less challenging in terms of clutter and occlusion.]-trained CNN model frozen and train a new classifier, softmax on top of it using the training samples of the target data-set and report impressive results.

III. RESEARCH METHODOLOGY

A. General Overview

Traditional classification methods use hand engineered and hand-crafted features or contains large amount of prior knowledge and contains multiple pipelines for processing. So the feature extraction plays vital role than classification, which makes classification less effective. In order to overcome this traditional classification problem pre-trained CNN’s are used, as their feature embedding is significantly appreciated. In this paper the pre-trained CNN models are trained over 1000 images and then a multi-class classifier is applied on the top of the features that are obtained by using these pre-trained model. Thus, finally obtain the classification accuracy.

The flow diagram represents the methodology used for classification in fig.2. Each block of the flow diagram clearly shows the processing steps. Using this methodology, we obtain the validation accuracies obtained by these different pre-trained models.

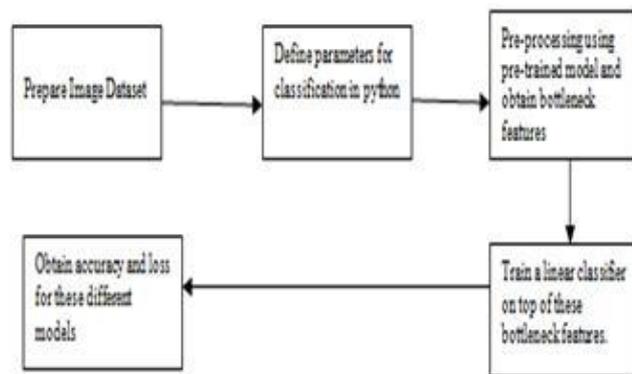


Fig 2. Proposed Methodology

B. Transfer learning

In general, training a deep Convolutional Neural Network from scratch is not easily possible. Because the dataset of required size and depth of the Neural Network is found rarely. So, we use Transfer learning, so that by using a pre-trained model which is trained over a large dataset can be used as a feature extractor for the task.

Transfer Learning is a vital component in deep Convolutional Neural Networks and provides solution to these problems. Transfer Learning in computer vision refers to the use of pre-trained model. A pre-trained model is a model that is trained on a large benchmark dataset to solve a problem to the one similar to our problem.

Transfer learning mostly depend on two factors namely the size of the new dataset and the similarity of the dataset to the original dataset. So, before adopting transfer learning we need to look upon few scenarios. If the new dataset has smaller size and same content compared to that of the original dataset then overfitting arises there by reducing the effectiveness of the CNN. If the new dataset have large size and similar content compared to that of the original data, then we can try to fine-tune the model through the full network.

C. Fine Tuning Pre-trained CNN model

In this paper, we use Imagenet as a basic source domain for the pre-trained models for our dog breed identification task. Similarly, some papers also show that better deep models can be learned based on CNN transfer learning from ImageNet than to the other small datasets .All the pre-trained CNN models that are trained are based on ImageNet dataset and are available in the websites.[7]

For our project, in order to find the effect of depth of the network, we used pre-trained CNN models that have different depth. The different pre-trained CNN models include VggNet, GoogLeNet and their variants.

D. VGG-16

Vgg16 is a CNN model that was proposed by K.Simonya and A.Zisserman from the university of Oxford to compete in 2014 ILSVRC[4]. This model has improvement over its previous model ALEXNET. It contains 16 convolutional layers, with a max pooling layer after each set of two or more convolutional layers, two Fc layers and a final softmax



output layer. It contains multiple 3X3 filters one after the other.[5]

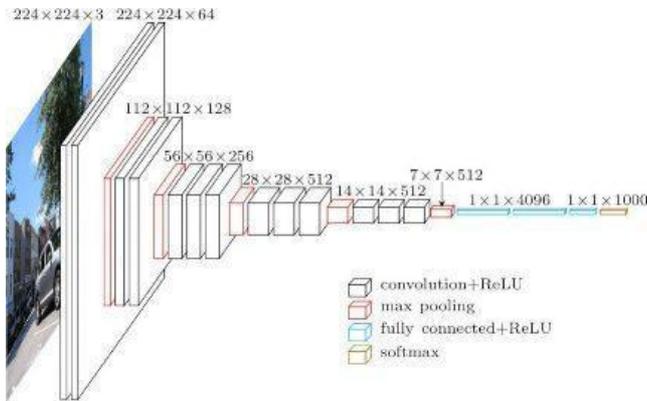


Fig 3. VGG16 Model

E. Xception

The Xception architecture is a linear stack of depth wise separable convolution layers with residual connections as shown in fig.4. This arrangement makes the architecture to define and modify easily. It takes only few lines of code using a highlevel library such as Keras or TensorFlow not like other architectures such as VGG-16, Inception V3 which are very complex to define.

F. Inception V3

InceptionV3 is a variant of GoogleNet that was proposed by Szegedy based on the original paper “Rethinking the Inception Architecture for computer vision”. This model consists of both symmetric and asymmetric building blocks, including convolutions, pooling, max pooling, concats, dropouts and fully connected layers as shown in fig.4. This network is designed to reduce computational cost meanwhile improving the classification accuracy.

For this model the ImageNet dataset consists of 1,333,167 images which are divided into training and evaluation datasets that contains 1,281,167 and 50,000 images respectively.

This model perfectly utilizes the computing resources that are present inside the network.

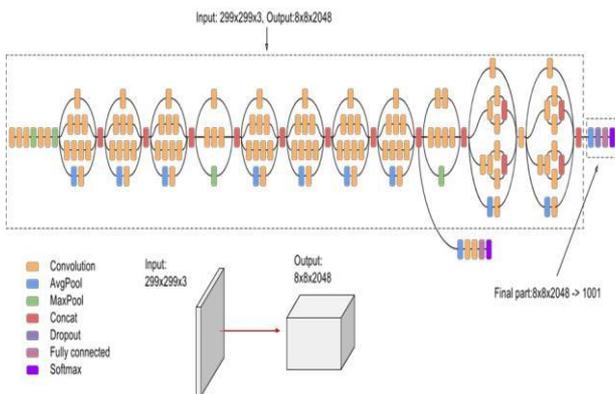


Fig 4. Inception V3 Model

G. Classifier-Logistic Regression

The pre-trained model already contains many layers that are stacked on top of each other. These layers are valuable as they

help to classify most images. Here we are intended to train only the last layer as the previous layers are already in a trained state.

The last layer that is before the final output layer which actually performs the classification is known as bottleneck features as shown in fig.5. We apply multinomial Logistic Regression on top of these bottleneck features to obtain the accuracy score and validation loss.

Logistic Regression is a binary classifier that is commonly used to estimate the probability that an instance belongs to a particular class. If the estimated probability is greater than 50% then the model predicts that the instance belongs to that class or it predicts that it does not belong to that class. logistic regression works even when two class of samples cannot be linearly separated thus generating a chance of higher rate of correct classification. The Logistic Regression model can be generalized to support multiple classes directly without training and combining multiple binary classifiers called Softmax Regression or Multinomial Logistic Regression.

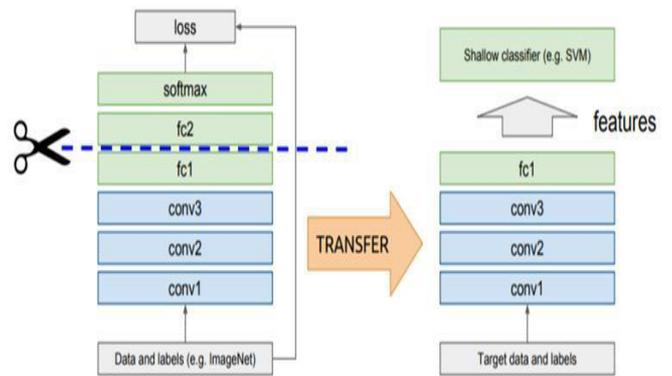


Fig 5. Fine Tuning

The Logistic Regression model can be generalized to support multiple classes directly without training and combining multiple binary classifiers called Softmax Regression or Multinomial Logistic Regression.

IV. RESULTS

The experimental results are tabulated in Table 1 with respect to the dataset taken from internet and performing Transfer learning by using Pre-trained models over a dataset containing 1400 images of 120 breeds dogs to obtain bottleneck features and finally a Multinomial linear classifier is applied on top of these features provides different accuracy and loss, these results produced by this are better than the results that are produced by retraining InceptionV3 model model over 275 dog images belonging to 11 different categories[8].

Table 1. Experimental Results of different Pre-trained Models

<i>Model</i>	<i>Accuracy</i>	<i>Loss</i>
VGG-16	91%	0.3
Inception V3	94%	0.06
Xception	93%	0.08

V. CONCLUSION

This paper describes the fine tuning of bottleneck features extracted through transfer learning by using multiple pre-trained deep CNN's and produced better results compared to traditional classification methods. In transfer learning we are mainly concerned with the training of fully connected layers as they act as feature extractors, moreover training the last Convolutional layer is worth exploring. It shows that Convolutional Neural Networks with transfer learning provides better solution for different image classification and identification problems. It also proves that the deeper the network the higher will be the classification accuracy. In future combining different pre-trained models can reduce the mis-classifications done by one model with the other model so that better results can be produced.

REFERENCES

- [1] Zalan Raduly,Csaba Sulyok,Zsolt Vadaszi,Attila,Zolde " Dog Breed Identification Using Deep Learning",SISY 2018 ,IEEE 16th International Symposium on Intelligent Systems and Informatics • September 13-15, 2018, Subotica, Serbia
- [2] Hasbi Ash Shiddieqy, Farkhad Ihsan Hariadi, Trio Adiono "Implementation of Deep-Learning based Image Classification on Single Board Computer". IEEE, 8253319, 11 January 2018 Yogyakarta, Indonesia.
- [3] Zeiler, M.D., Fergus, R.: „Visualizing and understanding convolutional networks“, 2014, pp. 818–833
- [4] Szegedy, C., Liu, W., Jia, Y., et al.: „Going deeper with convolutions“. Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
- [5]Thangarajah Akilan,Qingming jonathan,HuiZhang,"Effect of fusing features from multiple DCNN architectures in image classification" research article, [IET Image Processing](#) (Volume: 12 , [Issue: 7](#) , 7 2018), [Institution of Engineering and Technology](#).
- [6]Bihong Zhang,Lie Xie, YOugen,Yuaan,Huaiping Ming, "Deep neural network derived bottleneck features for accurate audio classification, 2016 IEEE International Conference on Multimedia & Expo Workshops, Seattle, WA, USA
- [7] Bahrapour, S., Nasrabadi, N.M., Ray, A., et al.: „Multimodal task-driven dictionary learning for image classification“, IEEE Trans. Image Process., 2016, 25, pp. 24–38.
- [8]pratik Devikar "Transfer learning for image classification of various dog breeds,IJARCET,volume 05,issue 2,Dec-2016.

- [9] Karan Chauhan , Shrawan Ram "Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensorflow and Keras",SJIF:5.71,FEB-2018.

I. AUTHORS PROFILE



Dr.B. Vijayakumar is Professor in Computer Science & Engineering in Vidya Jyothi Institute of Technology (VJIT) ,Hyderabad . He has about 23 years of Academic and Industry experience. He is a Life Member of CSI, ISTE, NESA, ISCA. He has more than 50 publications in the field of Image Processing, Digital Watermarking, IoT and Cloud Computing.



K.Bhavya is student of M.Tech Computer Science and Engineering (CSE)working under the guidance of Dr B.Vijayakumar in Vidya Jyothi Institute of Technology and , Hyderabad