2023 International Conference for Advancement in Technology (ICONAT) | 978-1-6654-7517-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONAT57137.2023.10080065

Dog Breed Classification using Inception-ResNet-V2

S. Manivannan[1]
Department of ECE,
Sri Sivasubramaniya Nadar College of Engineering (Autonomous),
Old Mahabalipuram Road, Kalavakkam,
Chennai-603110, Tamil Nadu, India.
individualvillas6@gmail.com

N. Venkateswaran[2]
Department of BME
Sri Sivasubramaniya Nadar College of Engineering (Autonomous),
Old Mahabalipuram Road, Kalavakkam,
Chennai-603110, Tamil Nadu, India.
venkateswarann@ssn.edu.in

***Abstract*—Dogs are one of the most faithful and loyal animals in the world.They are also the favourite pets for most of the pet lovers.Many feel relieved from stress and tension when they spent time with their pet dogs.So these special creatures are spread into various breeds across the world.It is very much essential to distinguish the breeds at many occasions.With the advent of development of artificial intelligence the methods to classify such large scale of breeds had become easier.This paper proposes a transfer learning based pretrained deep CNN architecture for classification of 120 breeds.The proposed model was trained on Stanford dogs dataset and the model achieved a training accuracy of 95.03% and a validation accuracy of 92.92% after training.The model performance and robustness had been inferred after testing with test images from internet.The network predicted correct breeds with a test accuracy of 88.92%.This paper provides an optimal solution for fine grained dog breed classification.**

Keywords—Dog breed classification, deep learning, artificial intelligence, Inception-ResNet-V2, CNN architecture

I. Introduction

Most of the people across the world own dog as their most favourite pet animal.One of the most important reason for having dogs as pets is the faithfulness that it has towards their owner.So people start to consider dog pets even as one of their family members.In today's world many people re becoming victims to frauds who sell similar dogs of different breeds for higher prices..The motivation for fine grained dog breed classification are high correlation between subordinated breeds and very large intra-class variation (such as different object poses)[2].More often than not, dog owners or veterinarians will need various images of dogs and preserve these images for later use, such as assisting in the search for lost canines.More often than not, dog owners or veterinarians will need various images of dogs for preserving these images for later use, such as assisting in the search for lost dogs and for other medical purposes[10].Artificial Intelligence and Machine Learning can provide a solution for classification of large variety of dog breeds.

II. Related Works

Initially, conventional image processing algorithms were applied for classification problems.Starting from Massinee Chanvichitkul et al.[8] had proposed a template matching technique and principal component analysis(PCA) to classify the dog breeds.This proposed approaches were based on conventional image processing algorithms.Initially the approach starts with coarse classification where face contours are considered as geometric feature of the dog.For comparing the two contours whether similar or not two measures euclidean distance and cosine distance are taken as metrics.Then the input image and the output from the coarse classifier is used as input for the PCA which is used for fine classification.For 35 breed classification the proposed architecture achieved an accuracy of 93%.The accuracy was considered to had improved by 16% for PCA classifier without coarse classifier.

Other types of descriptors and CNN based methods were later proposed . Punyanuch Borwarnginn et al.[1] had proposed a comparative study of different approaches for dog breed classification.The experiment is performed on Columbia dogs dataset containing 8351 dog images spread across 133 different breeds.The initial approaches proposed for dog breed classification are based on conventional image processing namely Local Binary pattern (LBP) and Histogram of Oriented Gradient (HOG).Following this approach a deep learning based pretrained CNN architecture is used as a classifier.The results from the experiment proved that prterained CNN architecture InceptionV3 achieved an accuracy of 96.75% outperformed the conventional HOG descriptor based approach which achieved an accuracy of 79.25%.Later, many algorithms and approaches were proposed for this fine grained classification with the parallel development of deep learning and computer vision. Minori Uno et al.[2] had a proposed a fusion based CNN architecture for fine grained classification task.The study uses Stanford dogs database for training and testing the robustness of the architectures.Considering the huge size of the dataset, transfer learning strategy is used for efficient training.With Alexnet and VGG16 as baseline models which are pretrained on the ImageNet dataset ,fusion of the two baseline CNN architectures is performed for better results.Fusion of architectures denotes the concatenation of the feature maps of the two architectures in a particular pattern.This approach proved to give a significant improvement in terms of accuracy.Results showed that through fusion of the different layers of the models ,the accuracy of the VGG16 architecture improved from 81.2% to 84.08%, thereby giving a 2.88% increase in accuracy score.

Experimentation on various deeper CNN were later proposed.Rakesh Kumar et al.[3] had proposed a comparative study of various deep learning based CNN architectures for dog breed identification. Resnet101, ResNet50, InceptionResNetV2, InceptionV3 are the transfer learning based deep learning models used in the study.Since the models mentioned above attained convergence within 10 epochs and considering the complexity of the networks the training is stopped at 10 epochs. ResNet101 ,ResNet50 ,InceptionResNetV2 , InceptionV3 performed with a training accuracy scores of 90.26% ,87.89% ,58.04% ,52.49% respectively.Results showed that ResNet101 was the best performing model with an accuracy of 91% and a total loss of 0.347.

III. Dataset and data Preprocessing

A. Dataset

Stanford Dogs dataset provided by TensorFlow is used for training and testing of the networks.This dataset consists of dogs of 120 different around the world.The dataset is built with images and annotations for fine grained image-classification.With a total 20,580 images the dataset is divided as 12,000 images used for training and 8580 images which are used for testing the performance of the trained network.Each of the 12,000 training images has bounding box annotations and class labels.

Fig. 1. Sample Images from the dataset

B. Data Preprocessing and Augmentation

1. Stanford dogs dataset is a standard dataset which is available readily in the TensorFlow backend.So the dataset is directly is loaded from the TensorFlow_dataset library.
2. Nextly, within the dataset the dog images are spread across various dimensions.This is found using Exploratory Data Analysis(EDA) where an histogram is plotted for the height and width distribution indicating the number of dog images spread across various height and width.
3. After analysing from the histogram plot ,it was seen that most of the training images had height and width closer to 300 and the InceptionResNetV2 architecture which is used for classification expects the input image to be in the shape of (299x299x3) .After these observations all the training images are resized uniformly between (299x299x3).
4. Following the scaling of images, pixel normalisation is carried out across the dog images.Since the training images are in RGB colour scale where the maximum pixel value is 255, each and every pixel from images are divided by 255. This bounds the pixels values between 0 and 1.
5. Since the model used for fine grained classification is a huge data hungry model , it is essential to increase the training dataset by data augmentation.This is done through operations such as Random flip, Random Rotation ,Random Zoom.The flip type chosen was horizontal, the factor for both random rotation and random zoom is 0.1.

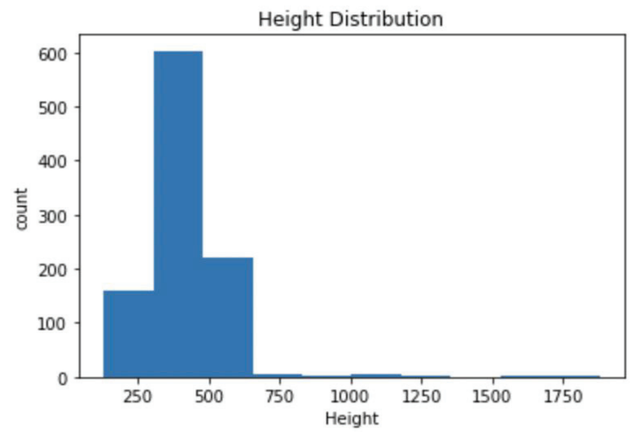


Fig. 2. Height distribution histogram

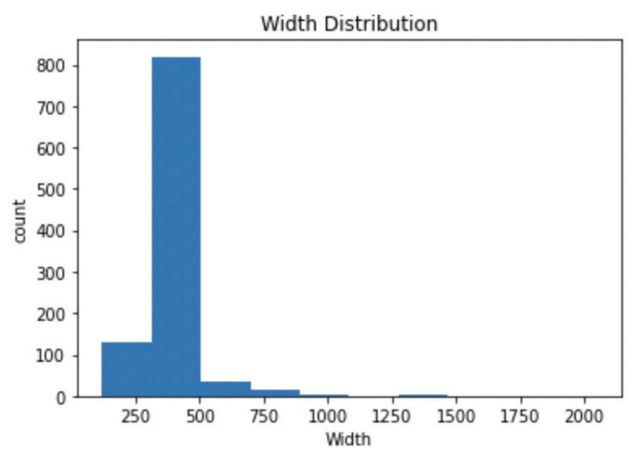


Fig. 3. Width distribution histogram

IV. Proposed CNN Architecture

Inception-ResNet-V2 is pretrained CNN architecture which is used for deep feature extraction and fine grained image classification.The architecture is pretrained on ImageNet dataset which consists of more than millions of images spread across hundred classes[4].Since it is difficult to train a custom deep CNN architecture from scratch for such a huge dataset, transfer learning approach is used to reduce training time and achieve earlier convergence.This model consists of 164 deep residual layers .The merits of this proposed architecture is that it combines the properties and advantages of both Inception and ResNet architectures. This model consists of both Inception and residual modules. Selecting residual blocks made it possible to eliminate the degradation issues and provide detailed features like location, tecture, colour ,size[5].Shortcut paths present in the residual block play a crucial role in feature retention and in improvement of the model performance[4] .The inception block, which was featured in Inception v3, is also greatly simplified and used in this architecture. [4].Inception module consists of multiple convolutions of different kernel sizes for extraction of local and general features[5].Blocks used in this architecture are Input, Stem, 5 x Inception-ResNet-A, Reduction-A, 10 x Inception-ResNet-B, Reduction-B, 5 x Inception-ResNet-C, Average Pooling, Dropout, and Softmax [6].This architecture outperformed its previous version Inception-ResNet-V1 in terms of performance and accuracy.

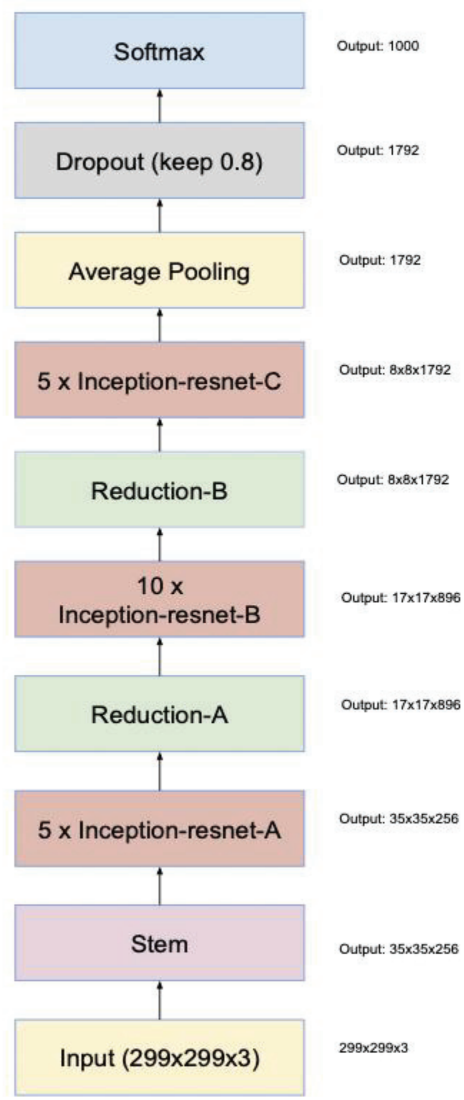


Fig. 4. Inception-ResNet-V2 architecture[9]

V. Configuration of Training and Experimental Process

The proposed architecture was trained in Google Colab environment. The classification model was trained in Python 3.9 environment , GPU of NVIDIA Tesla K80 with 12 GB RAM in deep learning framework of TensorFlow 2.9. After several experiments and analysis the training parameters for the network architecture is decided. The model is trained for 500 epochs with a batch size of 60 with ADAM as gradient descent algorithm at a learning rate of 0.00001.The images are shuffled before fitting and training the model.The loss function used for training is CategoricalCrossEntropy.

For training the Inception-ResNet-V2 pretrained CNN architecture the training data is spilt into training and validation data in the ratio of 8:2 respectively. The dataset used for training contained 20 images for per breed for 120 breeds. Performing 3 types of data augmentation on the dataset yielded upto 60 more images per breed. As the model is pretrained on the ImageNet dataset , the pertained layers are made non-trainable or frozen and the final fully connected dense layer is made trainable. In the last fully connected layer Softmax is used as the activation function since this classification is a multi class classification. The learning rate and the hyperparameters mentioned above are fine tuned based

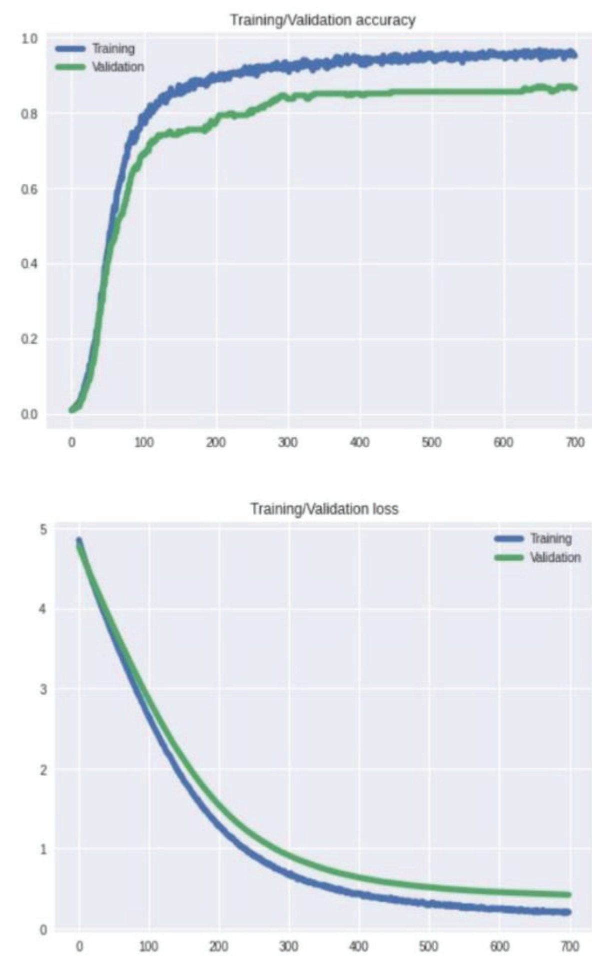


Fig. 5. Training accuracy and loss graphs

on several experimentation.The learning rate is kept very low (0.00001) so that the model learns slowly but accurately.In order to get the maximum efficiency from the architecture .The model was trained in Google colab free NVIDIA Tesla K80 GPU and it took around 4 hour 23 minutes completely train the model for 500 epochs.After training the model yielded a training accuracy of 95.03% and a validation accuracy of 92.92%.

VI. Simulation Results and conclusions

A. Simulation Results

After training and evaluating the model based on the accuracy and loss the model was tested with the individual images and the output was most of the times the correct prediction. The test image was a pug and the model predicted it as pug with an accuracy score of 94.1 % and this means that the model is working as expected.

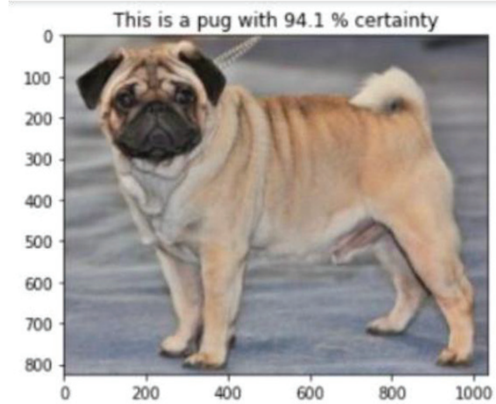
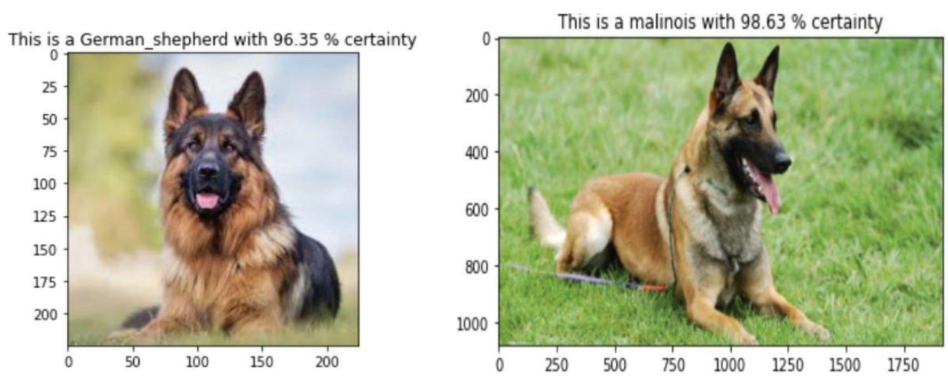


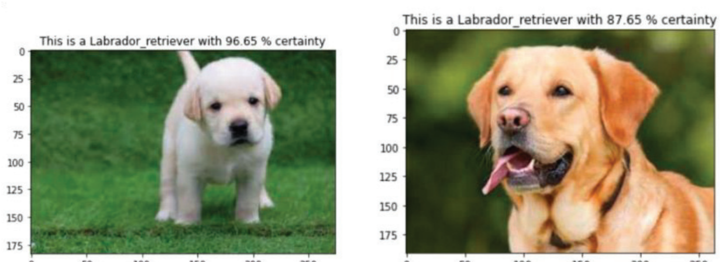
Fig. 6. Sample image prediction on test data

Interesting Results were observed during testing of the proposed architecture's efficiency

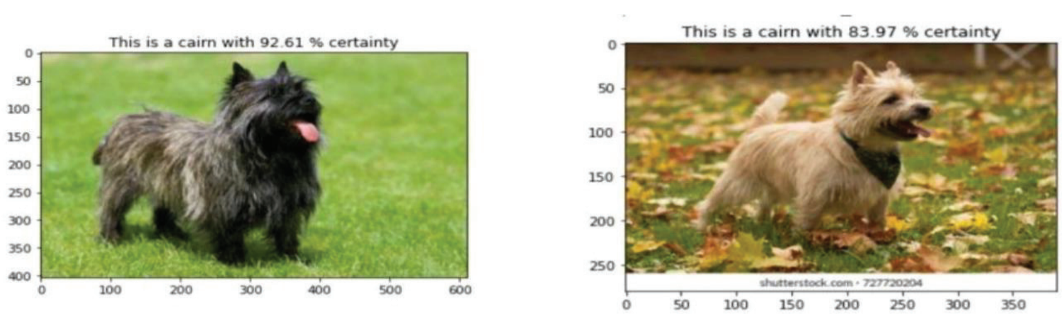
1.The model predicted accurately on the test images where the dogs appear similar to naked eyes but are actually different breeds



2.Again the model predicted accurately on a puppy and a fully grown Labrador Retriever



3.The model's robustness was observed when it predicted a black and a white Cairn accurately



B. Comparison with previous results

TABLE I.

Paper	Number of breeds	Accuracy
P.Borwarnnginn et al.[1]	20	96.75%
Sandra Varghese et al.[10]	35	93%
R.Kumar at al.[3]	120	91%
M.Uno et al.[2]	120	84.08%
Our proposed approach	**120**	**95.03%**

C. Conclusion and Future work

In this paper, a convolutional neural network(CNN) for fine grained dog breed classification has been proposed.This model is pretrained on ImageNet dataset thereby reducing the training convergence time.After training the model attained a train and validation accuracy scores of 95.03% and 92.92% respectively.While testing the model with test data it achieved a test accuracy score of 88.92% .In order to test the robustness and efficiency of the network , different classification situations on factors like dog's size, colour, breed and etc was fed to the classification model.The trained model didn't get confused on any of the above mentioned occasions. As a future work, the number of breeds used for classification can be increased from 120 to 180 and experiments can be conducted on various other CNN architectures like ResNet101 ,EfficientNet expecting better accuracy.

References

[1] P. Borwarnginn, K. Thongkanchorn, S. Kanchanapreechakorn and W. Kusakunniran, "Breakthrough Conventional Based Approach for Dog Breed Classification Using CNN with Transfer Learning," *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2019, pp. 1-5, doi: 10.1109/ICITEED.2019.8929955.

[2] M. Uno, X. -H. Han and Y. -W. Chen, "Comprehensive Study of Multiple CNNs Fusion for Fine-Grained Dog Breed Categorization," *2018 IEEE International Symposium on Multimedia (ISM)*, 2018, pp. 198-203, doi: 10.1109/ISM.2018.000-7.

[3] R. Kumar, M. Sharma, K. Dhawale and G. Singal, "Identification of Dog Breeds Using Deep Learning," *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, 2019, pp. 193-198, doi: 10.1109/IACC48062.2019.8971604.

[4] R. Badrahadipura, S. Q. Nur Septi, J. Fachrel, I. N. Yulita, A. A. Pravitasari and D. Agustian, "COVID-19 Detection In Chest X-Rays Using Inception Resnet-v2," *2021 International Conference on Artificial Intelligence and Big Data Analytics*, 2021, pp. 104-109, doi: 10.1109/ICAIBDA53487.2021.9689723.

[5] A. Thomas, H. P. M., P. P. and V. P. Gopi, "Moving Vehicle Candidate Recognition and Classification Using Inception-ResNet-v2," *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, pp. 467-472, doi: 10.1109/COMPSAC48688.2020.0-207.

[6] A. DEMİR and F. YILMAZ, "Inception-ResNet-v2 with Leakyrelu and Averagepooling for More Reliable and Accurate Classification of Chest X-ray Images," *2020 Medical Technologies Congress (TIPTEKNO)*, 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299232.

[7] A. Varshney, A. Katiyar, A. K. Singh and S. S. Chauhan, "Dog Breed Classification Using Deep Learning," *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498338.

[8] M. Chanvichitkul, P. Kumhom and K. Chamnongthai, "Face recognition based dog breed classification using coarse-to-fine concept and PCA," *2007 Asia-Pacific Conference on Communications*, 2007, pp. 25-29, doi: 10.1109/APCC.2007.4433495.

[9] arXiv:1602.07261

[10] Sandra Varghese; S Remya, "Dog Breed Classification Using CNN," in *Security Issues and Privacy Concerns in Industry 4.0 Applications* , Wiley, 2021, pp.195-205, doi: 10.1002/9781119776529.ch10.