Dog Breed Identification
A Convolutional Neural Network Approach via Transfer Learning

Cheng Zhong, Ze Zheng, Ye Zhang
Graduate School of Arts and Sciences
Georgetown University

***Abstract*— To determine breed characteristics in dogs is difficult due to the fact that many dogs come in similar shapes and sizes. This project aims to identify different dogs from different breeds. Convolutional Neural Networks (CNN) is the main method in species classification. A baseline model utilized a random forest method is set up to compare the performance of the CNN model. Due to the fact that the dataset contains roughly 10000 images which led to a small number of images per breed, we utilize data augmentation and transfer learning methods to avoid the overfitting on the training set.**

I. INTRODUCTION

Dog breed identification has been an interesting topic for machine learning in recent years. However, the accuracy of the prediction has always been a hot topic. In this project, a dog breed identification problem will be examined through a convolutional neural network approach via transfer learning. The goal of this project is to build a model that is capable of a dog's breed information through a single dog picture.

There are several key points that make this project challenging. First of all, there exists lots of different category of dogs, more importantly, some of the dogs are highly identical, for example, it is even very difficult for a human to distinguish silky Terrier and Yorkshire Terrier. The second challenge is about the diversity of the data. In order to detect the dog breed as accurate as possible, the data has to be as large as possible. Each breed of dog should have enough number of pictures to train the model. In order to overcome these challenges, several methods including Transfer Learning and Data augmentation was applied to build the model.

II. DATASET

The dataset was downloaded directly from Kaggle. The dataset is originally from a strictly canine subset of ImageNet website, which is composed of 120 different breeds with 10222 images in total. The dataset is separated into a training set, test set and a completely new dataset are created as a validation dataset. The dataset is very imbalanced since some of the breeds have lots of pictures like Scottish deerhound and some of the breeds only has about 20 pictures. Thus, a data augmentation is needed for this dataset.

III. RELATED WORK

Some typical machine learning approach was completed by scholars. For example, the approach based on Support Vector Machine (SVM) was applied. SVM is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis.

In 2012, Liu et. al proposed a model based on SVM regressor [1]. Their paper used an SVM regressor using greyscale SIFT descriptors as features to isolate the face of the dog. According to their paper, the result accuracy can reach up to 90%. Also, in 2012, they proposed a novel approach to fine-grained image classification by

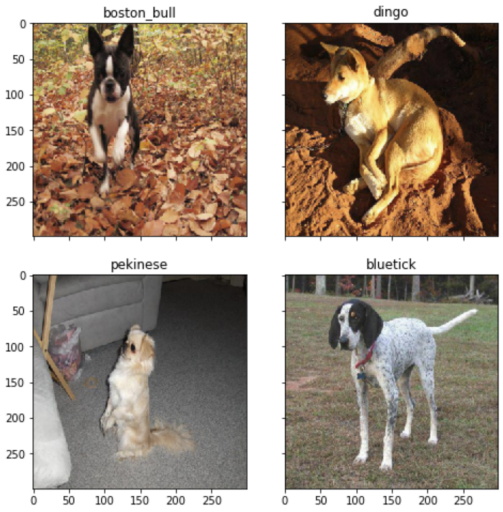

Fig. 1. Sample Images from the Dataset

using part localization [1]. They take dog breed identification as a test case. According to their paper, by extracting corresponding parts, they made the classification. Their approach features a hierarchy of parts and breed-specific part location. The prediction accuracy for this method is up to 67%.

In this project, rather than train a convolutional neural network from scratch, we decide to use a transfer learning method that uses several pre-trained models. This method will also avoid some overfitting problem that caused by the low volume of data.

IV. METHODS

A. Data Augmentation

Since we only have few examples, our number one concern should be overfitting. Overfitting happens when a model exposed to too few examples learns patterns that do not generalize to new data, i.e. when the model starts using irrelevant features for making predictions [2]. Hence, a data augmentation is needed. Since we are choosing Keras as our framework, we are using the ImageGenerator function in Keras to do the data augmentation. Data augmentation includes things like randomly rotating the image, zooming in, adding color filter etc. The method is only applied to the training dataset but not test and validation set. Here in our project, the original picture size is 256x256. By flipping the picture, crop the picture to 224x224, scale the picture, and add some color filter to the training picture. Our training set size was three times larger than our original training set.

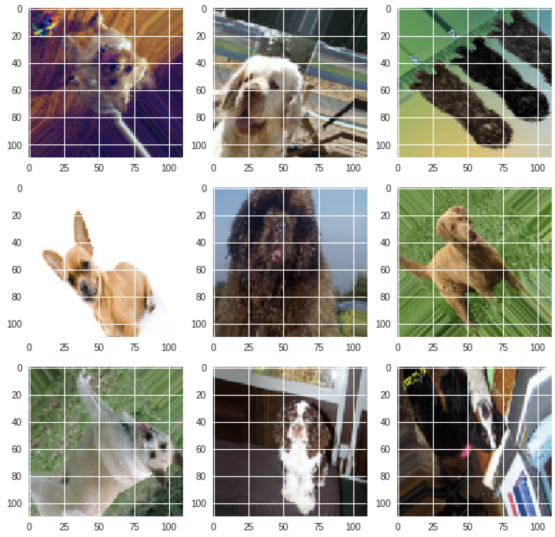

Fig. 2. Sample Image after Data Augmentation [4]

B. Baseline Model

Since the dataset is a little bit unbalanced, the baseline model here we choose is a random forest model with bagging. We decided to use a bag of words model for our baseline because it is a simple model that is frequently used for object classification. The bag of words model ignores our detected facial key points and instead just finds a visual vocabulary based on the cluster centers of SIFT descriptors obtained from the training set of images. Because the bag of words models performs better classification for objects with very high inter-class variability, we did not expect it to work very well for our fine-grained classification problem [3]. Before applying the baseline model, we need to extract the image features first. Initially, we used SIFT which is Scale-Invariant Feature Transform to extract all key descriptors of each image, then stacking all the descriptors together as the input of the k-means clustering from the Pythons Scipy library. Then, by applying the bag

of words method, we got the features of the data. Thus, a classifier like the random forest with bagging will perform well in reducing variance.

C. Transfer Learning

The transfer learning method is applied to avoid overfitting. Transfer learning method can be helpful to reduce computational time and deal with the problems caused by the small dataset. We compared the result from the following models:

- VGG16: A widely used CNN model proposed by K.Simonyan and A.Zisserman
- VGG 19: VGG 19 is a network with 19 layers used by Visual Geometry Group at ImageNet ILSVRC-2014.
- ResNet is the deep residual network. The ResNet-50 model has 50 layers.
- Xception: An extension of the inception architecture which replaces the standard inception modules with depth-wise separable convolutions.

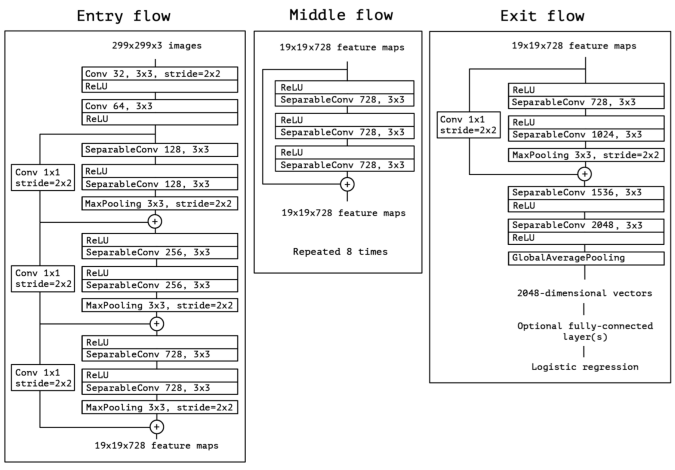

Fig. 3. Architecture of Xception Model

For these four types of pre-trained model, we used the weights from ImageNet and freezes the weights for all the layers except the last layer. Also, we added a top layer including a dense layer with size 1024 and a dropout layer with a drop rate of 0.7.

V. RESULTS

A. Baseline Model Performance

The baseline model result we got is not good. The random forest model performed on the test data got a 0.0488 accuracy. The reason might be dimensional of the data is too high since we have over 100 classes here.

B. Pre-trained Model Performance

TABLE I

LOSS RESULT TABLE

Model	Train Loss	Test Loss
VGG 16	0.3455	0.3087
VGG 19	0.3168	0.3864
ResNet-50	0.2785	0.2795
Xception	0.2560	0.2666

TABLE II

ACCURACY RESULT TABLE

Model	Train Accuracy	Test Accuracy
VGG 16	0.8052	0.8026
VGG 19	0.8108	0.7943
ResNet-50	0.8317	0.8266
Xception	0.8502	0.8450

VI. DISCUSSION OF RESULTS

A. Loss Analysis

In this project, we compared four different pre-trained model which is VGG16, VGG19, ResNet-50, Xception. The final performance was measured by softmax and categorical cross-entropy in terms of the training and testing loss. Cross-entropy loss measures the performance of a classification model that outputs a probability between 0 and 1. From the above table, we can see that the Xception model got the best performance with both smallest training loss and test loss. The performance of the rest model is followed by ResNet50, VGG 16, and VGG 19.

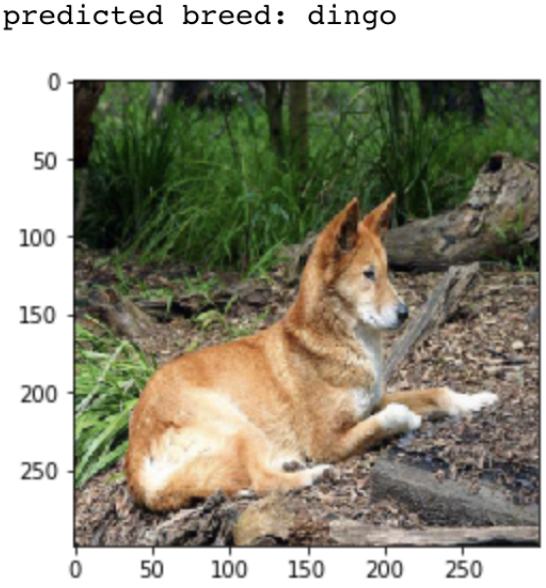

Fig. 4. Sample Output of Prediction Image

dingo	0.974487
dhole	0.011417
kelpie	0.003814
basenji	0.001791
redbone	0.000578
eskimo_dog	0.000574
malinois	0.000321
appenzeller	0.000249
cardigan	0.000242
boxer	0.000227
schipperke	0.000225
welsh_springer_spaniel	0.000175
siberian_husky	0.000172
golden_retriever	0.000172
chow	0.000147
pembroke	0.000146
soft-coated_wheaten_terrier	0.000145
saluki	0.000133
african_hunting_dog	0.000130
norwich_terrier	0.000129
cocker_spaniel	0.000118
norwegian_elkhound	0.000115
scotch_terrier	0.000114
ibizan_hound	0.000099
sealyham_terrier	0.000098
irish_setter	0.000096
tibetan_mastiff	0.000094
bouvier_des_flandres	0.000093
basset	0.000091
rottweiler	0.000090

Fig. 5. Sample Output of the Probability of Prediction

B. Accuracy Analysis

The above results show the accuracy for a different model on train data and test data. As the Xception got the best result which has an 85.02% accuracy on train set and 84.5% accuracy on the test set. The performance of the rest model is followed by ResNet50, VGG 16, and VGG 19.

C. Overall Analysis

From the above result, we can see that Xception is no doubt the best model and that makes sense. As the newest and more efficient model, Xception performs better in both measurement criteria and training time. Another finding from the result is that the VGG 16 performs better than VGG 19. It is obviously with more convolutional layers, the VGG 19 model got some overfitting issue thus it got a smaller accuracy and greater loss on the test set.

VII. CONCLUSIONS

In this project, we tried to solve the dog breed identification problem using a small dataset with only several images per breed. First of all, a random forest model with bagging was built as a baseline model, the model does not perform well and got a 4.88% accuracy on the test dataset. The reason leads to this is because of the high dimensionality of the data.

Next, a data augmentation and transfer learning were applied to the dataset. By applying flipping, crop, scale, and color filter to data, we got a more robust data. The Xception pre-trained transfer learning model achieved an 84.5% accuracy on the test set which is the highest accuracy among all the transfer learning models. Following models ResNet-50, VGG 16, and VGG 19 achieved accuracy 82.66%, 80.26%, 79.43% on test data respectively.

The results show that although the dog breed identification seems like a challenging task with 120 dog breeds, the CNN based transfer learning model with data augmentation is very powerful and has a decent performance on the dataset.

VIII. FUTURE WORK

Since the size of the dataset is small, overfitting can be a continuing concern. There are many methods can be applied to avoid model overfitting. Approaches like adding more images to the dataset or a more aggressive data augmentation might help solve this problem. The customized architecture that we used only have one flatten layer and on the hidden layer. For the future works, the performance of the pre-trained model can be improved by adding few more layers such as add a more aggressive dropout layer might also help the model with less chance to overfit and more accurate. Also, ensembling and combing pre-trained models would also give a better result.

REFERENCES

[1] Liu J., Kanazawa A., Jacobs D., Belhumeur P. (2012) Dog Breed Classification Using Part Localization. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7572. Springer, Berlin, Heidelberg

[2] Chollet, Francois. The Keras Blog. The Keras Blog ATOM, 5 June 2016,blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html.

[3] LaRow, Whitney, et al. Dog Breed Identification. Stanford University, 2016, web.stanford.edu/class/cs231a/prev_projects_2016/output%20(1).pdf.

[4] Chollet, Francois. Xception: Deep Learning with Depthwise Separ -able Convolutions. Google, Inc., 7 Oct. 2016,arxiv.org/abs/1610.02357