DOG BREED IDENTIFICATION

Rijesh Shrestha

2018011897.rijesh@ug.sharda.ac.in

Parangat Narsingh Pradhan

2018009122.parangat@ug.sharda.ac.in

Department of Computer Science and Engineering

School of Engineering & Technology

Sharda University, Greater Noida-201310

Abstract

In this project, we attempted to build a classifier capable of identifying a dog's breed through a photograph. Many dogs of different breeds of dogs are very difficult to identify and classify even for humans. For training and evaluating the model we have used the Stanford Dogs dataset which contains images of 120 breeds of dogs from around the world. This dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization. It consists of 20580 images of dogs. Our system consists of three major stages: facial KeyPoint localization, facial normalization, and breed classification. This paper describes how we have used deep learning through convolutional neural networks and transfer learning for identifying dog breeds. The main reason why we have done this project is that different dogs need to be rescued and medicated based according to their breeds and also identifying dog breeds that are not suitable for family.

Index terms: Dog breed Identification, Stanford Dataset, facial KeyPoint localization, facial normalization, breed classification, deep learning, convolutional neural networks, transfer learning

1. Introduction:

1.1 Task Definition:

A lot of work has been done in the field of image recognition over the past few years. In this project a specific object recognition problem was examined. The goal of our project is to build a model capable of identifying to which breed a dog belongs to, based upon its image. To train and evaluate the model the Stanford dataset, containing 20580 real-world images of 120 dog breeds, was used.

There are several factors that make Dog Breed Identification challenging. Firstly, there are many breeds of dogs and there can be only subtle differences between some dog breeds. Furthermore, there is low inter-breed and high intra-breed variation; in other words, there are relatively few differences between breeds and relatively large differences within breeds, differing in size, shape, and color. In fact, dogs are both the most morphologically and genetically diverse species on Earth. It may also be difficult to identify breed even for people who have sufficient knowledge in this domain.

Therefore, identification of dog breeds provides an excellent domain for fine grained visual categorization experiments. Secondly, dog images are very rich in their variety, showing dogs of all shapes, sizes, and colors, under differing illumination, in innumerable poses, and in just about any location. The photos have different resolutions, backgrounds, and scales.

Models such as Convolutional Neural Networks (CNN) allow computers to automatically extract hierarchies of features from raw pixels. These techniques proved to be successful in a variety of visual analysis tasks. In this project, CNN - based approaches were applied to the problem of dog breed identification. A number of techniques for improving performance of the model were used. Next, we experimented with data augmentation and transfer learning - two techniques that can potentially achieve high performance on datasets. In this project it was shown that the models that use these two techniques can perform very well on a dog identification task.

1.2 Motivation:

This problem is not only challenging but also its solution is applicable to other fine-grained classification problems. For example, the methods used to solve this problem would also help identify breeds of cats and horses as well as species of birds and plants - or even models of cars. The great variety of dogs poses a significant problem to those who would be interested in acquiring a new canine companion. However, walking down the street or sitting somewhere, one might see a friendly, attractive dog and wonder of its breed. In many situations, it is impossible to ask an owner about the breed and in many cases, the owner themselves will be either unsure or incorrect in their assessment. It is also important for identifying the breeds of dog while rescuing them and also medicating them. Some breeds may also be highly aggressive and not suitable for family companion, but useful for police and army. Through this project we aim to identify the breeds of dogs for both human companion and for the purpose of rescuing dogs.

1.3 Dataset:

Having a good training dataset is a huge step towards the robust model. We have used the Stanford Dogs Dataset. Every image in the dataset is annotated with the breed of a dog displayed on it. The Stanford Dogs dataset contains images of 120 breeds of dogs from around the world. This dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization. It consists of 20,580 images of dogs. Convolutional neural network (CNN) is by all accounts the best machine learning model for image classification, but in this case, there are not enough training examples to train it. It would not be able to learn generic enough patterns off this dataset to classify different dog breeds. Most likely, it will just overfit to this small amount of training examples so that accuracy on the test set will be low.

Preview of images from Stanford Dogs Dataset

2. Related Work:

Plenty of previous work has been done in the field of fine-grained classification, and we used this literature to develop an understanding of the field. Likewise, there was a fair amount of research that has been done into part localization, which we heavily leverage in our project. However, we primarily focused on classification within species in our literature review, which bears closest resemblance to our problem.

A problem that is similar to dog breed identification, is the problem of Cats vs Dogs classification, and a lot of work has been done on it. J. Howell J. Elson, J. Douceur and J. Saul. used a classifier based on color features got 56.9% accuracy on the Asirra ”Dogs vs Cats” dataset. Golle et al achieved an accuracy of 82.7% using a Support Vector Machines classifier based on a combination of color and texture features.

However, the problem of dog breed identification has a number of major differences that make it significantly more challenging. Firstly, it is a multi-class classification problem, in which the algorithm would need to identify a dog breed out of 120 choices. Secondly, much more fine-grained photo categorization is required, because dogs of different breeds look more similar to each other. Tremendous progress has been made in fine-grained categorization with the advancement of deep CNNs. D. Steinkraus P. Y. Simard and J. C. Platt discusses data augmentation as an effective technique for improving the performance of CNNs when dealing with limited datasets. Sinno Jialin Pan Qiang Yang points out that deep learning models are by nature highly reprovable, and therefore a technique called transfer learning (transferring learned general knowledge from one, usually larger, problem to another one) is possible. It is emphasized that it is a very effective method for cases when only a small dataset is available for training.

2.1 Review:

Our dog breed identification algorithm also heavily relies on accurate facial detection as an initial step. We trained a Convolutional Neural Network (CNN) with a mean squared error loss function on all the facial Key Points. The output of the prototype was satisfying so we went through with this model. Once the facial KeyPoint locations have been predicted by the first phase of the breed identification pipeline, normalized sub images of a constant size are prepared for the second neural network. This step is extremely important for the performance of any subsequent analysis. First, the center of the face is estimated as the mode of the midpoint between the eyes and the nose. Next, the slope of the vector from the left eye to the right eye is calculated and the entire image is rotated so that it lies flat. Finally, the length of the segment between the eyes, is calculated and a box centered at the center of the face with side length four times the interocular distance is cropped from the image. This box is then scaled to a constant size to serve as input to the next phase of the pipeline, breed identification. Once facial Normalization is used on the input image. The image is run through the Dataset which we have uploaded and the CNN identifies the breed of the dog. A 2012 paper by Liu et. al attempted dog breed identification using a similar approach. Through this approach Liu et. Al's model is able to classify their test dataset with an accuracy of about 90%

2.2 Comparison:

Liu et. Al's results are certainly very impressive. However, significant effort is put into identifying the dogs face and then its key points. In fact, Liu et. al identify the failure of dog face detection as the primary bottleneck in their pipeline. In our project, we attempt to use a convolutional neural network (CNN) to assist with KeyPoint detection in dogs, namely identifying eyes, nose and ears. CNNs have seen success in identifying facial key points in humans, and we hope to apply this technique to dogs as well. By doing so, we eliminate dog face detection as a step in the process and the first step goes for facial KeyPoint localization. With our own approach, we hope to match the success of Liu et. Al.

2.3 Breed Identification:

Liu's breed identification algorithm produces impressive results via a pipeline backed by SVMs. For each different breeds, they train a one vs all SVM over grayscale SIFT descriptors centered at the predicted part locations and the midpoints of the lines connecting them as well as a quantized color histogram over the entire facial region. Locations of parts other than the eyes and nose are repeatedly estimated by transforming the closest matches for each breed onto the eyes and nose and merging the score of the SVM for each. Ultimately, this method achieves 67 percent first-guess accuracy and 93 percent top-ten accuracy on the breed identification problem, extremely robust results given the large number of classes, high intra-class variability, and often low inter-class variability inherent to the breed identification problem. They also compare this algorithm to other popular model types and find that it outperforms bag of words, multiple kernel learning, and locally constrained linear coding by a large margin.

3. Methodology:

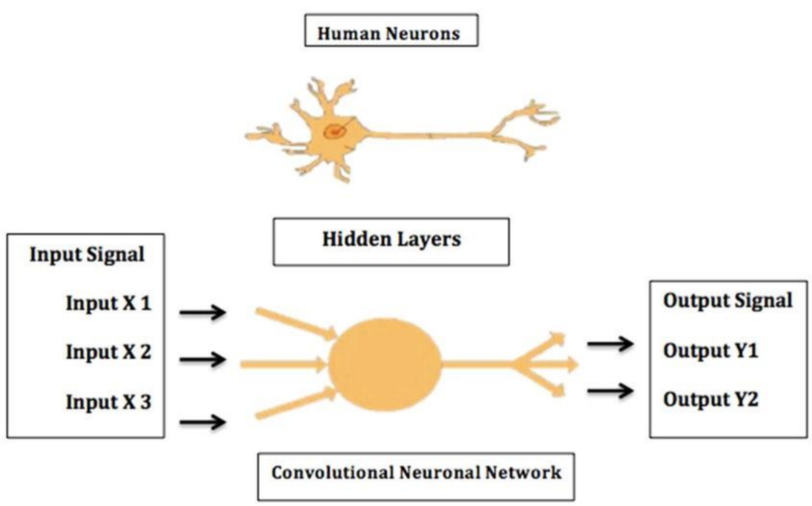
3.1 Definitions:

Deep Learning:

Deep learning is an Artificial Intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of Machine Learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. It is also known as deep neural learning or deep neural network.

Convolutional Neural Network (CNN):

A Convolutional Neural Network (CNN) is a type of artificial Neural Network used in image recognition and processing that is specially designed to process pixel data. It is a spatial neural network which is mostly used to recognize patterns in an image as an input. CNN imitates the human brain. Our brain is a vast collection of Neurons. Convolutional Neural Network consists of neurons, weights and biases. This network receives the input and uses the biases and weights to calculate the weighted sum of the input and passes this sum through an activation function and finally comes up with a prediction or an output.



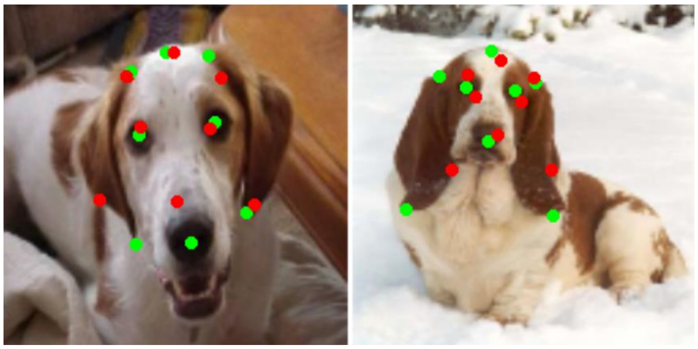
Real Neural Network VS Artificial Convolutional Neural Network

Transfer Learning:

In deep learning, transfer learning is a technique whereby a neural network model is first trained on a problem similar to the problem that is being solved. Transfer learning has the advantage of decreasing the training time for a learning model and can result in lower generalization error.

3.2 Facial KeyPoint Detection:

Our breed identification algorithm heavily relies on accurate facial detection as an initial step. To best identify dog breeds, the first step of analysis is KeyPoint detection. Dog face key points are defined and annotated in the Stanford Dog dataset as the right eye, left eye, nose, right ear tip, right ear base, head top, left ear base, and left ear tip. Thus, the goal of this part of the analysis pipeline is defined as follows: given an unseen image from the testing set, predict the key points of the dog face as close as possible to the ground truth points in terms of pixels. Convolutional neural networks have been shown to perform quite well on a variety of image detection and classification tasks, including human face detection. Once we had detected the facial key points, we used them to extract meaningful features about the image, which we would later use in classification.



Detecting the facial key points in the dog images

3.3 Facial Normalization:

Once the facial KeyPoint locations have been predicted by the first phase of the breed identification pipeline, normalized sub images of a constant size are prepared for the second neural network. This intermediate step is extremely important for the performance of any subsequent analysis. First, the center of the face is estimated as the mode of the midpoint between the eyes and the nose. Next, the slope of the vector from the left eye to the right eye is calculated and the entire image is rotated so that it lies flat. Finally, the interocular distance, the length of the segment between the eyes, is calculated and a box centered at the center of the face with side length four times the interocular distance is cropped from the image. This box is then scaled to a constant size to serve as input to the next phase of the pipeline, breed identification.

Examples of images before and after facial normalization. Concentrating on the face eliminates a huge amount of noise in the image and allows models to focus on a static, rigid object.

3.4 Dog Breed Identification:

Once facial Normalization is used on the input image. The image is run through the Dataset which we have uploaded and the CNN identifies the breed of the dog in the first ten tries with an accuracy of 55%. When the CNN is completely trained, we are aiming for an accuracy of about 70% to 80%

identifying breeds of dog through CNN

The problem of fine-grained image classification and breed identification is very difficult to begin with. Historically, computer vision researchers have struggled to effectively distinguish dogs from cats, so it is a fairly difficult task. There are three major issues with the breed identification problem as it pertains to this dataset: low inter-class variance, high intraclass variance, and pose variance.

4. Obstacles Faced:

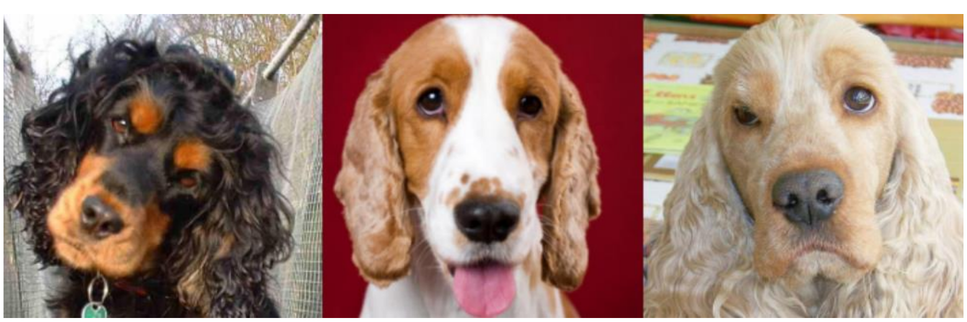
4.1 Interclass Variance:

Low inter-class variance is a phenomenon common within the animal domain. It is well known, for instance, that certain species of birds are more difficult to distinguish from one another than others due to their similarity in color. There are many breeds of dog which appear more or less identical to the untrained eye, yet to achieve high accuracy, a vision algorithm must be able to pick up on the minor details distinct to each class. In the figure below there are images of dog breeds which show the tendency of how similar dog breeds can be. Their similarity can even confuse people who have sufficient knowledge in this domain.

Images of Norwich terrier, Cairn terrier, and Australian terrier which display low inter-class variability.

4.2 Intraclass Variance:

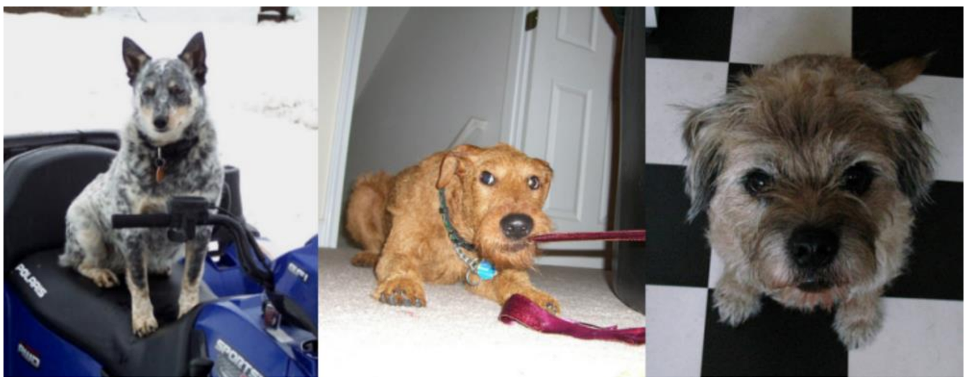
The Stanford Dog Dataset also suffers from high intra-class variance, in which a single breed of canine can take on a variety of appearances naturally. There are several breeds of dog which are best differentiated via the color of their coats, yet many can take on a range of colors and patterns from tawny through black, white and spotted. The breed identification model must make use of these contradictory signals in order to provide an intelligible estimate of the dog's true identity. The Figure below best illustrates example of a breed with high intraclass variance.



Three members of the English cocker-spaniel breed which demonstrate high intra-class variability.

4.3 Pose Variance:

Pose variance is a canonical problem within computer vision. In this specific case, it relates mainly to the variety of settings in which dogs have been pictured in the dataset. Canines are extremely deformable and pictured across a large number of tasks including at dog shows, in trucks, at home, and in people's handbags. This noisy atmosphere makes it extremely difficult for any vision algorithm to separate out meaningful signal for the dog's breed. In the figure, we have provided examples of this miscellany of subjects. This problem is arguably the most difficult and fundamental for dog breed identification.



Three dogs pictured in widely varying settings and poses.

5. Steps Taken:

The basic steps taken in our project are mentioned below:

- Dataset
- Imports
- Model
- Data Processing
- Training
- Testing

1. Dataset:

We first downloaded the dataset from Stanford's dog's dataset website and used it for the entirety of the project.

2. Import:

import all the packages needed for this task.

```
import os
import numpy as np
import pandas as pd
import cv2
from glob import glob
import tensorflow as tf
from tensorflow.keras.layers import *
from tensorflow.keras.applications import MobileNetV2
from tensorflow.keras.callbacks import *
from tensorflow.keras.optimizers import SGD, Adam
from keras import backend as K
from sklearn.model_selection import train_test_split
```

3. Model:

As we have already mentioned that this task is going to use Transfer Learning.So, let’s learn about Transfer learning. It is a technique of reusing a pre-trained model on a new task and here the task a multiclass classification.

```
def build_model(size, num_classes=120, trainable=False):
    inputs = Input((size, size, 3))
    backbone = MobileNetV2(
        input_tensor=inputs,
        include_top=False,
        weights="imagenet")
    backbone.trainable = trainable
    x = backbone.output
    x = GlobalAveragePooling2D()(x)
    x = Dropout(0.2)(x)
    x = Dense(1024, activation="relu")(x)
    x = Dense(num_classes, activation="softmax")(x)
    model = tf.keras.Model(inputs, x)
    return model
```

Here we use MobileNev2 as the pre-trained model. You can also use other models like Resnet50, VGG16, VGG19 and more available in TensorFlow. We add a few layers to pre-trained model to make it ready for the classification task.

```
def read_image(path, size):
     image = cv2.imread(path, cv2.IMREAD_COLOR)
     image = cv2.resize(image, (size, size))
     image = image / 255.0
     image = image.astype(np.float32)
     return image
```

4. Data Processing:

Now we will perform pre-processing on the dataset we are using for this task. In the process, we will first create a function that is going to read the image from the path and also resize it to the desired size. Then, we perform normalization to make the image pixels smaller by dividing them with the 255.

Now we will use TensorFlow tf.data functions to create the dataset pipeline for training.

The tf_parse function preprocesses the single instance of the complete dataset. It preprocesses a single image and its label and return it to the dataset function.

6. Conclusions & Future Work:

In the end, we concluded that deep learning models have a very great capability to surpass the human potential if the data provided is sufficient. Engineers and scientists are still working on the deep learning field because till now the exploration of deep learning is limited. In this project, we tried to tackle the dog breed identification problem using a very small dataset with only a few dozens of images per breed. First, a small convolutional neural network was built and trained on the dataset from scratch, and then the result was further improved by applying data augmentation to the training dataset. Overall, we consider our results to be a success given the high number of breeds in this fine-grained classification problem. We are able to effectively predict the correct breed over 50% of the time in one guess, a result that very few humans could match given the high variability both between and within the 120 different breeds contained in the dataset. Convolutional Neural Networks (CNN) takes a lot of time to train and Future work should further explore the potential of convolutional neural networks in dog breed identification. Given the success of our KeyPoint detection network, this is a promising technique for future projects. The problem of dog identification in CNN can also be used to identify the different breeds of cats, birds, horses and also different models of car. CNN shows promising results in image classification and never the less it is the future.

7. References:

1. Dog Breed Classifier Using Deep Learning

   Agrim Dogra and Harsh Massand

2. Dog Breed Classification Using Part Localization

   J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur

3. Dog Identification using Soft Biometrics and Neural Networks

   Kenneth Lai, Xinyuan Tu and Svetlana Yanushkevich

4. Dog breed classification via landmarks

   X. Wang, V. Ly, S. Sorensen, and C. Kambhamettu

5. Automatic Dog Breed Identification

   Dylan Rhodes

6. Using Convolutional Neural Networks to Classify Dog Breeds

   Hsu, David

7. Knowing Your Dog Breed: Identifying a Dog Breed with Deep Learning

   Punyanuch Borwarnginn, Worapan Kusakunniran, Sarattha Karnjanapreechakorn and Kittikhun Thongkanchorn

8. Dog Breed Identification

   Whitney LaRow, Brian Mittl and Vijay Singh

9. Dog Breed Identification Using Deep Learning

   Zalan Raduly, Csaba Sulyok, Zsolt Vadaszi and Attila Zolde

10. Dog Breed Identification

    Yerbol Aussat

11. image classification for dogs and cats

    K. Zhou, B. Liu and Y. Liu.

12. Dog Breed Identification with Fine tuning of pre-trained models

    B. Vijaya Kumar and K. Bhavya

13. Transfer learning for image classification of various dog breeds

Pratik Devikar

14. Transfer learning on convolutional neural networks for dog identification

X. Y. Tu, K. Lai and S. Yanushkevich

15. Dog breed classification using part localization

J. L. et al

16. captcha that exploits interest-aligned manual image categorization

J. Howell, J. Elson, J. Douceur and J. Saul.