Gustavo Grinsteins
Date Submitted: 07/28/2022

Convolutional Neural Networks: Dog Breed Classifier

This capstone project consists of using state-of-the-art Convolutional Neural Networks (CNNs) to provide an estimate of a dog's breed given an image. Furthermore, if the algorithm is supplied an image of a human, the code will identify the image as a human and proceed to provide a resembling dog breed.

Domain Background: Computer Vision and Convolutional Neural Networks

Humans can easily navigate and understand the three dimensional world. We can easily recognize shape, colors, translucency, and other object features. After this recognition, we can combine this information instantaneously to classify a plethora of objects and people. The field of study that aims to simulates this human ability using computers is called Computer Vision.

Computer vision first started in the 1970s with universities attempting to add a visual input to machines in order to achieve intelligent behavior. These tasks proved to be highly complex and let to the development of the mathematical and technological infrastructures in the following years. As a result, computer vision has made great strides in the last two decades. 3D models of an environment captured by several overlapping images can be reconstructed and algorithms can delineate people and objects in photographs (Szeliski, 2022). However, toddlers have an easier time explaining images in details than computers. Computer vision is a challenging task because it seeks to recover the rich complexity of the visual world in just one image.

The dramatic increase in computational power during the last decades led to the implementation of deep CNNs as the algorithm of choice for image recognition and semantic segmentation tasks (Szeliski, 2022). In recent years, focus has been shifted to modify CNN architectures to improve image classification accuracy. These improvements have been showcased in competitions like the ImageNet Large-Scale Visual Recognition Challenge and have resulted in high image recognition accuracy scores (Simonyan and Zisserman 2015).

Problem Statement: Dog Breed Classification

This project seeks to create an algorithm that correctly labels an image of a dog with the corresponding breed. If the algorithm is supplied an image of a human, the code will identify the image as a human and proceed to provide a resembling dog breed. The project will use publicly available datasets for reproducibility. Image recognition accuracy scores will be used to evaluate the model.

Datasets and Inputs: Dog and Human image datasets

The project will make use of two datasets provided by Udacity:

- Dog Images (DogImages.zip)

  - Total of 13,233 labeled images
  - Image folders partitioned for training, validating, and testing

- Labeled Faces in the Wild (lfw.zip)
  - Total of 8,351 labeled images
  - Image folders partitioned for training, validating, and testing

The input to the model will consists of the processed images according to the CNN architecture that is used.

Solution Statement

The project solution consists of creating two CNNs: one from scratch and one using transfer learning. The CNNs need to achieve a minimum of 10% (From scratch) and 60% (transfer learning) accuracy in recognizing images from the dog data set. The higher accuracy CNN will be used to implement an algorithm for classifying a dog breed from an image input.

Benchmark Model: VGG-Nets

The benchmark model used for the project will be VGG-Nets architectures (Simonyan and Zisserman 2015). These are state-of-the-art architectures that have won top prizes in image recognition competitions. Simonyan and Zisserman have demonstrated that this architecture generalizes well using other datasets. Furthermore, VGG-Nets are publicly available to facilitate further research.

Evaluation Metrics: Image Classification Accuracy

Image classification accuracy will be used as the metric to evaluated the algorithms. The algorithms predicted label will be compared to the ground truth label in the Dog image dataset. The test algorithm will sum the number of correct predictions in the testing partition and divide this over the total partition size to obtain accuracy results.

Project Design

The workflow of this project can be separated into the following sections: dataset loading, image processing, development of CNNs, and final dog classification algorithm testing.

The dataset loading will consists of loading the datasets making sure to take advantage of the conveniently labeled and partitioned (training, validating, and testing) images.

Following proper loading, the images will be processed using the open source Pytorch computer vision modules and following VGG-Nets processing guidelines. Data exploration will be performed by assessing methods for detecting humans in images as well as exploring algorithms for detecting a dog in an image using a pretrained VGG-16 CNN (Simonyan and Zisserman 2015).

After the data exploration and processing is complete a CNN will be developed from scratch. The architecture for this CNN will resemble a VGG-11 (Simonyan and Zisserman 2015). Since this model is a simpler architecture in relation to other VGG-Nets it will reduce complexity during training. Following training, this CNN will be tested using the testing partition and accuracy results will be outputted (need to be > 10%).

Given the low accuracy from the CNN created from scratch due to the relatively small dataset, lack of computing power, and time, the accuracy score of the model will be increased by implementing transfer learning on a pre-trained VGG-19 (Simonyan and Zisserman 2015) CNN. This CNN has more complexity than other VGG-Nets but has the advantage of recognizing a wide range of features from images given its deep network architecture. The transfer learning will consist of substituting the last classification layer with a layer that outputs the right number of label predictions from the provided dataset. Following training, the transfer learning CNN will be tested on the testing partition and accuracy results will be outputted (need to be > 60%).

A classification algorithm will be employed using the transfer learning CNN in order to provide user-friendly feedback regarding the classification of the provided dog images. A smaller subset of images will be used to test the robustness of the created model and the algorithm logic.

References

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations.*

Szeliski, R. (2022). *Computer Vision: Algorithms and Applications.* Springer Cham.

Udacity. (2022). *Dog Image Dataset.* Retrieved from https://www.udacity.com/: https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/dogImages.zip

Udacity. (2022). *Labeled Faces in the Wild Dataset.* Retrieved from https://www.udacity.com/: https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/lfw.zip