Capstone Proposal

Dog Breed Classifier (CNN)

1. Domain Background

Image Classification task mostly has its roots in ImageNet Classification Challenge. In 2010, Fei Li proposed to join PASCAL VOC challenge team (which worked on dataset of 20 classes with 20 000 images in total) to create a competition in which researchers from all around the world would test their algorithms on the dataset in order to push forward Computer Vision development. The resulting dataset called "ILSVRC" consists now of 22 000 classes with over 14 000 000 of images with weight of about 2 Terabytes of data.[i] Recognising images by machines is crucial for the future of Artificial Intelligence and Robotics.

Deep Learning, a subset of Machine Learning focusing on Artificial Neural Networks, has been recently used in many fields: from Computer Vision, through Speech Recognition, to Bioinformatics and NLP problems. The Deep Learning architecture used for Image Classification is "Convolutional Neural Networks" – an algorithm which consists of two parts: Feature Learning and Classification. An input image which is represented by a matrix of pixels (in case of coloured images each cell of the matrix is RGB value) consist of Spatial and Temporal dependencies which may be captured using appropriate filters. The features can then be used in Classification part which ends usually in a probability distribution of image's label certainty.[ii]

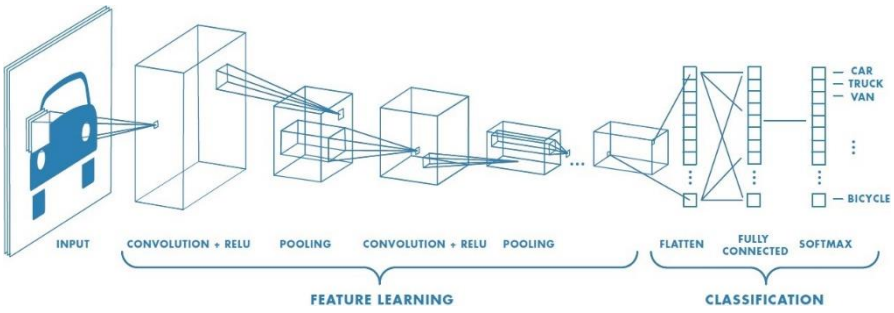

Figure 1: an example of a typical CNN architecture. Source: medium.com

2. Problem Statement

The problem consists of developing an algorithm that takes supplied image as an input and if a dog is detected in the image it provides its breed, if a human is detected the output of the algorithm should be a label of the dog breed that the human is most resembling.

3. Datasets and Inputs

   The datasets are provided by Udacity. There are two datasets:
   - dog dataset (weight 1.08GB): contains 133 classes of different dog breeds, the classes have different number of samples ranging from ~20 to ~60 images per class in training dataset. There are 3 sets provided: training (6 680 images), test (836 images) and validation (835 images)
   - human dataset (weight GB): contains images of 5750 humans, each human has 1-8 representing images. There is just one set, not split for test/train/validation.
   The two sets will be used during training to decide if the new input is a human (and what dog breed it resembles) or a dog, if it is a dog then what specific breed it is.

4. Solution Statement

   Solution to the problem is creating Convolutional Neural Network architectures that will take as an input images from dog and human datasets and learn to specify the dog's breed (or in case human – resembling dog breed). In order to achieve that several steps must be made: firstly, the datasets must be imported, loaded and pre-processed. Then we need to teach the model to detect humans, then dogs. The project's notebook uses two approaches in creating CNN (from scratch where we build the architecture by ourselves and using transfer learning which is a method of using pre-trained models and changing its parameters to suit our task). Then the output must be shown to the user in a neat, easy to understand way.

5. Benchmark Model

   State-of-the-art models for Image Classification like GoogLeNet, Resnet or NASNet achieve 95%+ of accuracy in the classification tasks. However, they were trained on a huge Image Net dataset for thousands of GPU hours. In this project, the dataset is much smaller, so the benchmark model accuracy that should be possible to achieve is around 80%.

6. Evaluation Metrics

   Evaluation is rather simple. The test set can be used to measure the model's loss using for example Categorical-Cross-Entropy algorithm for multi-label classification. It calculates the difference between two (or more) probability distributions. Measuring the model's loss function and its accuracy we can easily determine whether the implemented architecture is good or bad.

7. Project Design

   As specified in point 4 (Solution Statement) – there are several steps needed to achieve the goal. Firstly, the data must be loaded and appropriately prepared (i.e. data augmentation, resizing so all the input is the same), then following the given notebook's instructions we can create appropriate CNNs for image classification. One

of the CNNs made from scratch can be based on one of the winners of ImageNet for example ZFNet which won the challenge in 2013 – it is simple, shallow and relatively fast to compute CNN network. There is no possibility to implement more complex CNNs like GoogLeNet or Resnet due to its incredibly high computational power needed. After implementing the architecture, it will be trained and compared to pre-trained VGG model. At the end, appropriate output of the given image and its classification should be shown in easy-to-read manner.

---

[i] D., Gershgorn, "The data that transformed AI research—and possibly the world", qz.com, Available: https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world.

[ii] J., Brownlee, "Deep Learning & Artificial Neural Networks", machinelearningmastery.com, Available: https://machinelearningmastery.com/what-is-deep-learning/