Private Eye

An ANN Based Computer Vision API

Theodoros Nikopoulos-Exintaris
Aberystwyth University Computer Science
thn2@aber.ac.uk

Introduction

This project is concerned with the use of artificial neural networks and specifically deep convolutional neural networks for image classification. Its general aim is to investigate the performance of CNNs compared to other computer vision and types of ANN for this task as well as produce a simple wrapper API for interacting with the finished neural network which will be used to produce demonstration applications.

The ultimate goal of the project is to demonstrate the use of CNNs for image classification and hopefully multi-label classification and to then employ this technology to develop applications that will use such a technology.

Main Objectives

1. Decide on an ANN backend to use.
2. Develop a suitable ANN structure to train on the CIFAR-10 and MINST datasets to classify images.
3. Compare performance of different ANN structures for this problem
4. Create a suitable wrapper/API to for applications to interact with the ANN.
5. Develop simple web interface to submit images for classification to demonstrate the ANN working.
6. Extend to different demo applications and image classification problems.

Materials and Methods

1 Platform

After investigating several tools and libraries like TensorFlow, we have decided to settle with the Keras Library with the Theono backend running on an nVidia GPU using CUDA.

This is largely because TensorFlow is currently lacking in critical platform support and is considered as of now an alpha release.

Keras was chosen because it is widely used and provides a very simple programming interface which greatly simplifies the prototyping of neural networks and training as well as performance measurement.

Theano is supported in Keras as a backend and is another widely used backend. It was chosen because of its speed and ease of use.

2 Datasets

We are using the MNIST and CIFAR-10 datasets to train our neural network these consist of several thousand small, under 40px images with associated labels which will be used for classification.

2.1 MNIST

This dataset is a reduced subset of the NIST handwritten digit database and is made up of 60,000 hand written digits for training examples and 10,000 testing examples. The dimensions of each image is 28x28px.

These are used as a proof of concept for the neural network and a successful model using this dataset should be able to recognise a wide variety of normalised hand-written digits.

2.2 CIFAR-10

This dataset is a smaller set of the CIFAR-100 dataset. It consists of 60,000 32x32 images organised into 10 classes hence its name. Each class is 6,000 images big and in addition contains 1,000 testing images per class for a total of 10,000 test images.

The CIFAR-100 dataset on the other hand contains the same number of images except we now have 100 classes of 600 training images and 20 superclasses for those 100 classes. Each class also includes 100 test images.

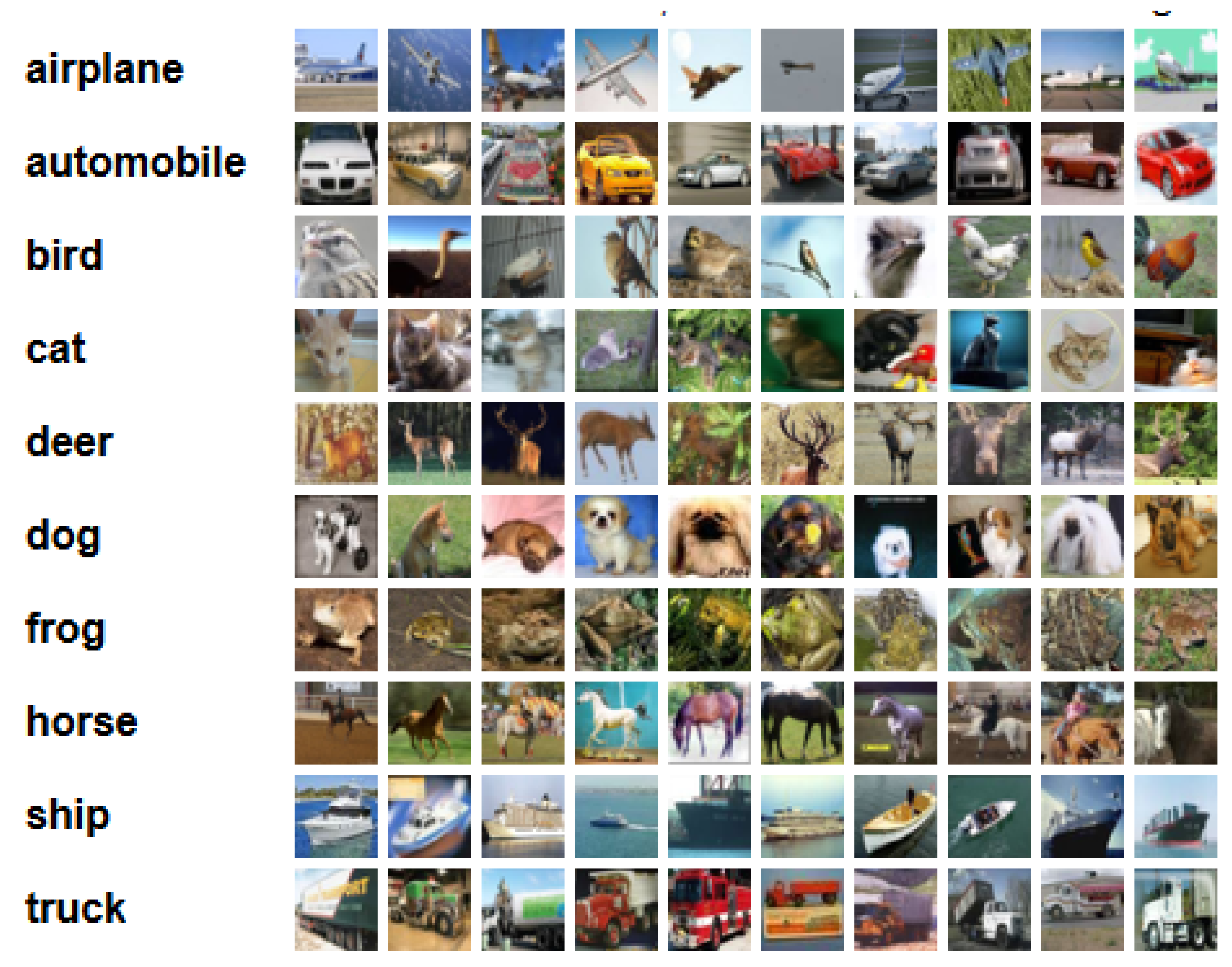

Figure 1: Example of CIFAR-10 images taken from CIFAR website

2.3 Network Structure

The prime candidate for this task is a convolutional neural network. These types of networks have been found to work quite well with computer vision problems so our investigation has been largely oriented towards those.

Their structures tend to be quite simple and most if not all networks use a very similar application of an input layer followed by stripes of a convolution layer followed by a subsampling or pooling layer which is repeated until the desired network depth is reached. These are then mapped to a set of outputs which give use our ANN classification results.

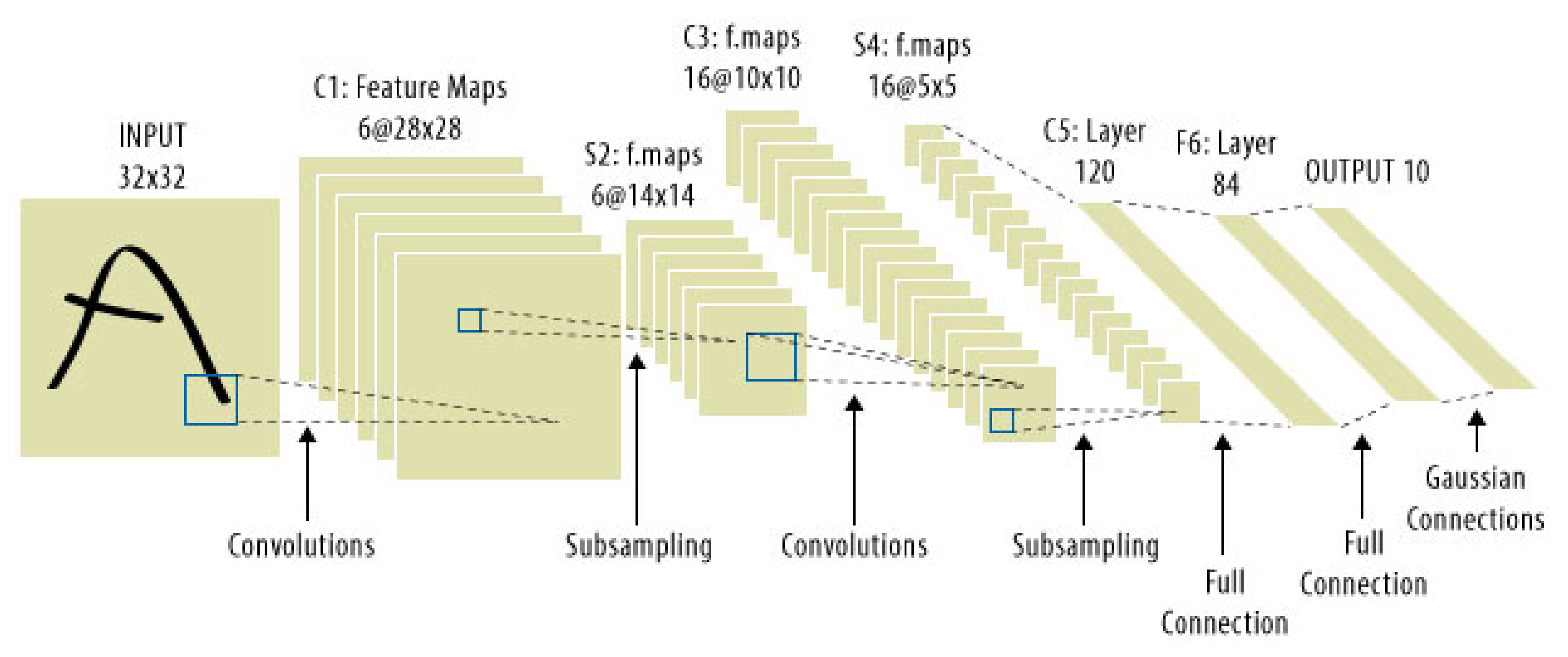

Figure 2: Example of a common CNN design. See acknowledgements

Convolution layers divide the visual field into smaller more manageable partitions while subsampling layers will merge inputs down. This is inspired by biological processes and on the assumption that interesting features tend to be very localised in the visual field so one only needs to consider parts of an image that are close together.

The main variables is the choice of weights, activation functions, convolution patterns and number of layers. These many variations will be tested throughout this project as well as other factors

2.4 Testing Methods

To evaluate the performance of our neural networks we will be using provided sets of test images which will then provide us with an accuracy for each experiment as well as appropriate tables. We will also try to benchmark our solutions with other people using the same data which is easy to do since CIFAR-10 is used to host competitions.

To The Future And Beyond

By May I hope to achieve good performance on both CIFAR-10 and CIFAR-100 datasets which is meant to prove the viability of such a project. Once good ANN performance is achieved two ideas for demonstration apps include a site where images can be submitted to and classified. As well as perhaps advanced web crawlers that use ANN output to decide whether pictures are worth keeping or not based on a search query.

After May, I hope to move this project to my own dataset where I aim to build similar applications but on a larger scale and with images of a different nature. Ideally I would like to use this to classify large datasets of obtained images as well as equip existing datamining projects to be more 'inteligent' perhaps even implementing quality predictive functions. The site itself can be expanded to have user feedback evolving the project to an 'online' trained neural network.

Acknowledgements

This Poster was produced in LaTeXusing the a0poster class:
http : //www.latextemplates.com/template/a0poster − portrait − poster
Figure 2 is taken from:
http : //systemdesign.altera.com/can − you − see − using − convolutional − neural − networks/