Development and Evaluation of a Visual Attention Model with Python and Tensorflow

Oleg Yarin

A thesis presented for the degree of
Bachelor of Science

First supervisor: Prof. Dr. Christian Herta
Second supervisor: Dr. Vera Hollink

Applied computer science
HTW Berlin
Germany
31-05-2017

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Neural network approaches have received much attention in the last several years. It's becoming a popular choice for performing various tasks like speech and image recognition, object detections etc. as these methods have dramatically increased accuracy compared to traditional machine learning approaches. However, achieving high accuracy on recognition tasks is still computationally expensive and needs improvements in performance. This study will be a close resemblance of the recurrent neural network of visual attention which is able to extract necessary information from an image by looking at it in low resolution, and then adaptively select parts that are most relevant for a task [1].

The idea of visual attention was inspired by how human perception works. Humans do not perceive a visual scene as a whole but focus on parts of the scene that gives the most useful information to them. Humans are also capable of combining information from different parts of a picture. They then connect it to build a subjective knowledge of the picture (or sequence of pictures) [2]. Taking into account these properties, researchers from google Deepmind build a model which can be described as follows:

> Instead of processing an entire image or even bounding box at once, at each step, the model selects the next location to attend to; based on past information and the demands of the task.... The model is a recurrent neural network (RNN) which processes inputs sequentially, attending to different locations within the images (or video frames) one at a time, and incrementally combines information from these fixations to build up a dynamic internal representation of the scene or environment.[1]

One of the main advantage of this model, is that the computation required is controlled and is independent of the input image size. Deepmind's researchers evaluated their model on several image classification and dynamic visual control problems which showed a better performance when compared with convolution neural network[2].

The evidence from this study suggests application of this model on large scale object recognition as well as classification of sequence of images, which will be a great fit since the model's performance is not dependent on the size of an input object.

The main aim of this study is to extend the current knowledge of the work mentioned above and build a model which will be able to classify a set of images and develop appropriate prototype system since it can be useful in a variety of areas. However, the current work is limited by low-resolution images and mostly will concentrate on classifying a group of objects as this restriction will reduce complexity of the task and therefore reach better results on a task of classifying a group of images.

1.1 Motivation

This approach to classify a group of images has a potential to help with automated detection and classification of breast cancer metastases, which is the main concern of camelyon challenge [3]. Camelyon challenge is an inspiration for this work since pathologist's efforts along with the assistance of automated detection system will reduce significantly not only the workload of pathologists but the human error rate in diagnosis as well.

This work will be the first step in building software that will be capable of classifying whole-slide images of histological lymph node at the patient level. That is, bringing together estimations from multiple lymph node slides into a single outcome.

Digital pathology is a very attractive field for machine learning researchers since whole-slide images have a very high resolution and are typically about 200000 x 100000 pixels. To give you some sense of data, camelyon challenge provides data for 200 patients, where each patient has 5 different slides. It means that in total they release about 1000 slides and that is 55.88gb of uncompressed data [3].

It is quite clear that using CNN for this task is computationally very expensive. Applying model of visual attention promises to solve the issue of high-resolution pictures at a computational level. Therefore making an extensible piece of software, that will allow further improvements is also one of the main concerns of this work.

Chapter 2

Theory

2.1 Artificial Neural Networks

Why Neural Networks? Before going into what actually are artificial neural network, let's first try to face a question why do we need it in this paper. The problem that we give to our application to solve can be shortly summarized in the following statement: Given a group of images, find the patterns in this images that will change or influence on your belief more than other patterns that an image or a group of images belongs to a specific class. This is problem is known as pattern recognition problem or in our case visual pattern recognition problem[4]. The solve this problem it's required to develop ability for a machine to recognize patterns that will help to make a decision about a class. The obstacles that can appear by solving this problem can be more visible if we will try to write a conventional computer program, i.e. bunch of rules to identify these patterns. What seems to be easy for us, is really hard to describe algorithmically. In these system the computational steps are deterministic hence not flexible enough for the whole variety of input data[5].

Solving problem differently Artificial Neural Networks(and machine learning in general) are looking at the problem in a different way. They don't execute programs instructions, they don't have sequential semantic and normally are not deterministic. They acquire their "knowledge" by extracting patterns from raw data, which normally called training data(which normally is a set of tuple (input, label)) This approach also know as concept of statistical pattern recognition.[4] Artificial Neural networks have recently shown an excellent performance and accuracy at recognizing objects compared with other machine learning techniques[6].

What is Neural Network? Artificial Neural Network(ANN), often referred just as Neural Network(NN), in simply words is a computational model, which was inspired by how human/animal brain works. Artificial NN is modeled based on the neuronal structure in the brain's cortex. Though the inspiration was from the brain, it's indeed much much simpler than brain in terms of number of neurons that is used in ANN. [7] To understand how neural networks works it is crucial to understand first the way perceptron work, which is simple type of an artificial neuron. Given several binary inputs, perceptron is meant to produce a single binary output.

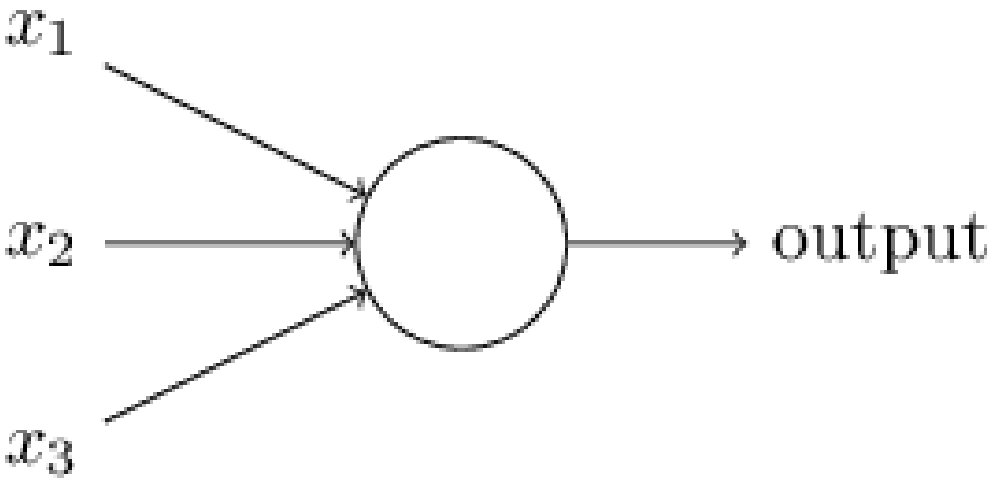

Figure 2.1: A perceptron.

In the figure **??** the perceptron has three inputs: x_1, x_2, x_3.
[8]

Perceptron

Basic concept of neural network There are a good variety of explain the example in terms of the elements layers, activation function. How output is produce.

- about learning weights/ parameters , about data. Ground truth.
- write shortly about how one updates parameters. About stoschastic gradient descent
- different activation functions - learning rate - a little bit about backpropagation.

Bibliography

[1] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent Models of Visual Attention," *Nips-2014*, pp. 1–9, jun 2014.

[2] P. Goldsborough, "A Tour of TensorFlow Proseminar Data Mining,"

[3] "CAMELYON17 - Home."

[4] C. M. Bishop, "Neural networks for pattern recognition," *Journal of the American Statistical Association*, vol. 92, p. 482, 1995.

[5] M. A. Nielsen, "Neural Networks and Deep Learning," 2015.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[8] F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Report (Cornell Aeronautical Laboratory), Spartan Books, 1962.