Abstract

Intro

Data

Experiment/Implementation

Conclusions

ABSTRACT

In this work, we applied recurrent neural network for image classification using glimpse action. Glimpse of images are now used because it reduces the total number of data needed for the image classification. Glimpse also ignore irrelevant part of the image. The overall idea of glimpses comes from human-eye perception and how humans can recognize objects by quick glimpses. Contours of images can give enough information for a human to accurately classify an object. We didn’t make use of convolution neural network for our glimpse classification because it is more computationally expensive and it scales linearly with the number of pixels. In our evaluation, we recorded a training accuracy of 80.05% on 47 Epochs and bets validation accuracy of 85.6% in 46 Epochs

1. Introduction

The idea of using glimpse action in image classification is from how the human eye works. It shows how the retina works by paying attention to a particular part of an object and blanking the rest out. If a person is looking at a building all other part of the building might be blurred and the eye only focus on the garden or any other part of the building.

Our model uses the idea such that our model only focuses on the interesting part of the object and use lower attention for the other part of the object. The model is then able to reduce the number of data needed in focusing on less interesting part of the image.

With the success story of convolution neural network in Image classification and object detection, images are now being classified seamlessly. Convolution neural network has also recorded very high accuracy in image classification. These successes come with some bottleneck which are the training with a large amount of data, running on multiple GPU.

These drawbacks have made researchers working on image classification to find a way in which the data that are needed for image classification can be reduced. The result is leveraging on how the human eye view objects. They noticed that the human eye does not process the entire object or a particular scene at once rather it focuses on a particular space to use the necessary object that is required at a particular time. Using this approach for our model will result in lower computational costs. It reduces the number of objects to be processed by the model and it also reduces the task complexity.

Our model uses a recurrent neural network that uses glimpses during the training process to quickly try and train the network. The network splits the set into training and validation sets

The recurrent model processes the input in a sequential way. It picks the different part one at a time for processing. It then combines the input to build an internal representation. A recurrent model was used because it has the ability to process the next image in a sequence. Our next object will always be the new part of the image that has the viewer’s attention.