Smartphone Software Retina

Slide 1

Good afternoon, I am presenting my research project, which I completed at the University of Glasgow during my masters.

It is entitled Smartphone Software Retina

Slide 2

Currently many deep learning architectures use the raw image data as input into neural networks for tasks such as image classification.

These networks have millions of parameters and take days to weeks to train and require large datasets with data augmentation to ensure the model does not overfit.

So we proposed a biologically inspired image transformation approach to reduce the image size while increasing their invariance to scale and rotation changes.

This would be useful for robotic vision systems for detecting points of interesting and classifying objects in a scene efficiently.

Slide 3

Mammalian vision systems have evolved over thousands of years to reach where it is at now.

Currently when we focus on an object, the object that is in focus has lots of details and our peripheral vision is blurred.

This is due to the arrangements and size of the receptive fields, known as the retina tessellations.

The retina tessellations are used to generate a back projected image, which will see later on.

Secondly the signal from each eyeball is split into two halves which are projected separately onto primary visual cortex using the receptive fields.

The image undergoes a translation to a complex logarithmic mapping, which potentially contribute towards scale invariance in biological vision systems.

Slide 4

These two image transformations, form basis of my project and it was part of a pilot study with the main focus on the initial stages of data acquisition and the image transformation.

The objectives of the project was to create a live preview of the retinal image transforms of each frame captured by the smartphone video camera. As well as to apply a gaze control mechanism to find points of interest in the images and save the results.

Slide 5

Here is the approach I took to create the mobile application, where on start up it underwent a preparation stage then during the live preview performed the retina sampling to create the cortical image and back-projected image.

To find new focal points I developed a gaze control mechanism.

And finally a method to save the relevant image data.

Slide 6

Here is an example of the resultant retinal image transform.

The left is the original image and right, represents both the back-projected image and the cortical image will the focal point in the center of the image.

This shows the difference in scale of the two generated images.which represent the same information, where the cortical image on the extreme right would be used as input into the neural network.

Slide 7

Here is a closer look at the two transformed images.

The back projected image is mainly used as a validation technique to check which part of the image is in focus.

As you can see in this example, the stems are visually detailed while the peripherals are slightly blurred.

Slide 8

The approach I used for the gaze control mechanism was a SIFT keypoint detector to find focal points within the cortical image space and applied a heuristic to balance between exploration and fixation.

On the right you can you can see the gaze control mechanism in action.

Slide 9

Since this project was mainly a optimising and application development.

The approach was evaluated based on the computational time on the mobile device which was an iPhone 5C.

Slide 10

One of the main points of optimisation was the computational time for preparation stage of the retinal transform.

Based on the existing pseudocode available, I found that the application was unusable without preprocessing of certain computations offline.

I found points where it could be preprocessed offline and loaded into memory at startup.

By preprocessing some of the computation I was able to reduce the preparation stage from around 70 seconds to under 5 seconds on the mobile device.

Slide 11

Here we achieved under half a second to generate each image transform.

In particular the cortical image took on average a quarter of a second on a iPhone 5C.

Although this was not realtime of 30 fps there was concurrent works in investigation of using parallel computing to optimise the image transformation.

Slide 12

This is how my final iOS application looked with options to select the different image transformation modes, and an option to select at a live video preview or image stored on the device.

Slide 13

So that was the conclusion of my 2-3 months masters project which mainly involved image processing and data acquisition which I completed a few years ago.

But for this presentation I decided to extend on this and apply the knowledge of machine learning to analyse what would have been the next stages of the project.

Slide 14

Can we train existing image classification deep learning architectures with the cortical images as input and achieve similar results to the original image dataset?

In order to ensure a balanced and fair results, I used the following constraints:

Slide 15

Here we have an example image of one of the classes which is an English Terrier.

The left is the original image and the middle is the cortical transformed image.

Slide 16

So what I found was that the cortical transformed dataset did not perform as well as the original image dataset.

Here is a confusion matrix of where the cortical image model performed poorly on the class “french horn”.

Slide 17

One area I believe that should be focused on in machine learning, is ensure the model is interpretable, this is especially important in life critical environments such as medical image analysis.

So I dug into the reasons why the model predicted certain classes using a method called Grad-cam.
As you can see the original images could identify the region of the classes relatively well but for the cortical image approach, the heat map regions generally either covered the whole image or was around points on the edges.

Slide 18

So this leads me to believe that current neural network architectures do not seem suitable for training on the cortical images but instead a custom input approach must be used to handle the black surrounding regions which are not part of the cortical image.

Additionally, another step preprocessing step of finding a focal point within the image must be first employed before the transform instead of just using the image centers.

Slide 19

So that was a brief overview of the full process required for retinal transformed images which involved image processing, image acquisition and image analysis.