SAMUEL OVERINGTON

Questionnaire

1. Why do you want to work at MMT? (50 words max)

2. Describe a novel computer vision model you developed (link to any relevant online materials - articles, papers, deployed applications)

3. What date would you want to start (or if currently in work, how long is your notice period)?

What was the most complex problem/challenge you have worked on? Why was it so challenging? What was the outcome?

During my time on the Applied ML use case team at arm, I was the lead researcher for creating a proof of concept computer vision model for dynamic hand gesture recognition, aimed at being able to control device using an interface of simple gestures (raise volume, change channel / track … etc), with the hardware constraints of IOT / embedded / home entertainment systems.

This involved reading through research papers exploring this and similar use cases (like language models for sign language), and comparing models and datasets and ultimately choosing one that best fit our purpose (dynamic gesture recognition). Models initially explored included visual transformers, hand landmark point-cloud + RNN, and finally decided upon 3D CNN based model to customize, along with a class based video dataset for hand gestures with multiple modes (depth, RGB, IR, composite).

One of the issues that arose during the project: the 3D CNN model architecture was chosen (MoViNets) as we did not have a working model in our model zoo for operator and hardware testing using 3d CNNs, and also it had limited support of TensorFlow operators and layers (eg AveragePooling3D), so we implement custom layers to approximating output, and raised tickets with our internal tooling team, who fixed and raised with TensorFlow core.

For the final outcome, several "flavours" of the model were released into our internal model zoo with trained weights on the dynamic gesture dataset. For the larger flavours, the model achieved comparable accuracy to published results using the same dataset and different models - with our model footprint, which was used to exercise the model on an embedded arm processor, by running previously unsupported math operators and other. The model was also used to create a show-and-tell for arm partners.

Describe an example of what state of the art/novel approaches have you taken?

When implementing gesture recognition model above and training on the dataset, the accuracy for our model was significantly below the accuracy level of published results on the same dataset with different model architectures. I had tried training from scratch, and transfer learning from pre-trained weights, and fine tuning on our target dataset - all with minimal improvements. I then researched novel methods of increasing accuracy, and came across learning rate schedulers and subsequently added ReduceOnPlateau to our model. After training, I was able to improve our validation accuracy by a further 15 points, taking us to within 10 points below published values for our model which was one quarter of the size.

Which direction do you feel that the field of Deep Learning & Computer Vision is heading? (this is all your opinion so just say what you think – e.g. having 1 big model to solve all tasks like DeepMind Perceiver)

Metaverse

Innovation in computer vision and deep learning models are being lead in industry as well as R&D by many developments brought on by work being done in trying to bring the metaverse to life - both in AR as well as VR. Companies researching and building products aimed at users having social or entertainment experiences and interactions are having wider reaching impacts in areas such as construction and healthcare. eg for AR: Improvements in model abilities to map landmarks in 3d space and improve methods of automating the process of stitching them together to create a single scene. This is impacting the technology used in for example construction, where CV models can extract useful information about a 3d reconstructed construction work-site, or in a similar way; auto insurance companies are using CV models being able to extract damage information from a 3d reconstruction of a damaged car.

Explainable AI using Bayesian techniques in healthcare

One of the problems with using deep learning models in healthcare, is not understanding how a model has come to the decision about something. For most models, the inference output doesn't tell the end user (eg a medical professional) how confident the model is of its own output. This can be problematic as real world inputs may have biases that were not seen during training data, and the model has no way of representing this uncertainty. It also causes issues on a model not generalising well on new data.

One way that researchers are trying to tackle this, is by exploring approaches of Bayesian uncertainty in deep learning. Specifically within computer vision tasks, where a model is created or modified to output the model uncertainty of a prediction, which in turn would give medical practitioners the ability to have explainable behaviour of model outputs, and give them the tools to understand when to add human intervention in the predictions.

Have you had any experience developing your own datasets?

During my internship with Deimos Space UK, I created a proof of concept ML model and dataset for image recognition-detecting different types of weeds mixed with crops producing a heat-map for live feed images, to be used on hardware mounted cameras on a tractor / drone, and trying to detect unwanted plant varietals.

For the dataset, I was given raw images in categorical folders, categorised by weed type, and sometimes multiple types of different plants in same image). Using CVAT I hand labeled a portion of images from each category the segmentation outlines of each class of vegetation, and then trained the model on this subset of this dataset, and then used it to create semi labeled data on the remaining images.