Classifying points of interest in FAST ultrasound videos using neural networks

Ilan Gofman[a,b], Matthew Holden[b]
[a]Department of Computer Science, University of Toronto, Toronto, Canada
[b]School of Computer Science, Carleton University, Ottawa, Canada

Introduction: The Focused Assessment with Sonography in Trauma (FAST) is a Point-of-Care Ultrasound procedure used to identify possible trauma by examining areas for lacerations and free fluid. The assessment covers four regions of the body: left upper quadrant, right upper quadrant, cardiac, and pelvic. Each region has various points of interest (POIs) that are important for ultrasound operators to examine comprehensively to ensure an accurate diagnosis. Our goal is to develop an objective system that can identify which POIs have been scanned and provide feedback to the ultrasound operators regarding their scans. To accomplish this, we propose several Convolutional Neural Network (CNN) architectures for classifying POIs that have been scanned across regions. A POI classifier can be integrated within a real-time information system for the assessment that will notify the operator which POIs require further scanning to ensure that each patient is fully scanned.

Methods: The dataset contains 17 videos of scans completed of the 4 regions, where each region contains between 8 to 12 POIs. As multiple POIs can occur in the same video frame we use multi-label classification to predict the list of POIs that are present. We propose 2 architectures for this task, a single-frame architecture and an architecture for classifying video fragments. For the video classification, we split the original video into short snippets consisting of 5 to 20 frames, and classify all the POIs observed in the snippet. The single frame classification is done using the InceptionV3 model architecture proposed by Szegedy et al [1]. While the video classification model architecture is a modified Inflated 3D ConvNet (I3D) [2] that can be seen in Figure 1. It is a two-stream version of the InceptionV1 model architecture but with another dimension added to the convolutional layers to capture the temporal information. One stream uses the original ultrasound images to train, while the other stream uses the optical flow of the video frames. They are trained independently, and the predictions are then averaged. The test set consists of 3 videos, while the remaining 14 were used to train and tune the model through cross-validation.

Results: The performance of the model was measured using the full-label accuracy which measures the portion of predictions where all POIs are predicted correctly in a single label, as well as the micro F1-score. The best performing model was the InceptionV3 model that predicts POIs using single frames, having achieved a full-label accuracy of 88.1% with a micro F1-score of 0.63. The video classification achieved a full label accuracy of 63% with a micro F1-score of 0.34 when using 10 frames per video fragment. Although the results are worse when training using the I3D model, we have found that additional data augmentation improves the performance. By using techniques such as translations, mirroring, and temporal flipping, the micro F1-score improves by 0.07.

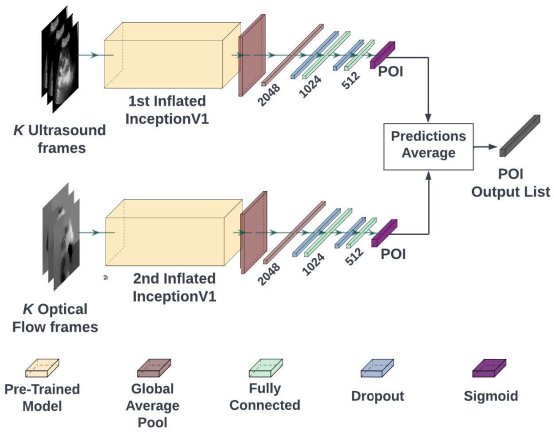

Figure 1: The modified I3D model architecture used for predicting POIs.

Conclusions: The single frame classification model using the InceptionV3 architecture performed the best. Given the small size of the dataset, applying data augmentation was critical to the performance of the model. To conclude, this work has shown CNNs can be used effectively to classify the POIs scanned during a FAST exam. By using a POI classifier, the models trained can provide objective feedback to the ultrasound operator on which areas have not been scanned in a particular region. This work can be expanded to provide real-time feedback as part of the FAST examination to ensure each patient is assessed comprehensively and all POIs have been scanned.

References

[1] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” Dec. 2015.

[2] J. Carreira and A. Zisserman, “Quo vadis, action recognition? a new model and the kinetics dataset,” May 2017.