Boundary Based Anomaly Detection in Deep Neural Networks

Ahiakpor, Priscilla (School: Lincoln Community School)

Gupta, Ridaan (School: Lincoln Community School)

Purpose When artificial neural networks are deployed in real world situations, they often show flaws in their handling of data outside the training set. This kind of data is called out of distribution data (OOD). In some applications such as medical imaging or self-driving cars, this can pose danger as it may give high confidence prediction in situations they haven’t been properly trained to handle. Our research seeks to find new methods of distinguishing OOD samples so as to improve their reliability. Method Our method attempts to take advantage of clusters in feature space of the neural network to form boundaries that detect any OOD data that tries to cross and prevents them from crossing. Results The performance of this method was evaluated against a benchmark for OOD detection and it attained an accuracy of 90% across all experiments. Conclusion The results obtained from the tests proved that it was suitable for OOD detection. Future research may seek to explore this method further.