Deep Unsupervised Pre-trained Neural Network for Human Gesture Recognition

B.M.T.Kumarika,N.G.J.Dias[2]

Recognition of visual patterns for real world applications is a complex process that involves many issues. Varying and complex backgrounds, bad lighting environments, person independent gesture recognition and the computational costs are some of the issues in this process. Since human gestures are perceived through vision, it is a subject of visual pattern recognition. Hand gesture recognition is of higher interest for Human-Computer Interaction (HCI), due to its widespread applications in virtual reality, sign language recognition, robot control, medical industry and computer games. The main goal of the research is to propose a computationally efficient and accurate pattern recognition algorithm for HCI.

Deep learning attempts to model high-level abstractions (features) in data and build strong feature space for the recognition task. Neural network with five hidden layers was used and each layer can learn features at a different level of abstraction. However, training neural networks with multiple hidden layers was difficult in practice. At first, each hidden layer individually was trained in an unsupervised fashion using autoencoders. After training the first autoencoder, second autoencoder was trained in a similar way. The main difference is that features that were generated from the first autoencoder are used as the training data in the second autoencoder thus decreased the size of the hidden representation, so that the second autoencoder learns an even smaller representation of the input data. The original vectors in the training data had 101376 dimensions. After passing them through the first encoder, this was reduced to 10000 dimensions. After using the second encoder, this was reduced to 1000 dimensions. Likewise at the end, final layer was trained to classify 50 dimensional vectors into different image classes. The result for the deep neural network is improved by performing Backpropagation on the whole multilayer network.

Finally, we observed that average test classification error for traditional neural network with supervised learning algorithm is 3.6% while the error for pre-trained deep neural network is 1.4%. We can conclude that unsupervised pre-training adds robustness to a deep architecture and it proposes computationally efficient and accurate pattern recognition algorithms for HCI.

Key words: Deep neural network, Auto-encoders, Unsupervised pre-training, Hand gesture recognition.