Hand Gesture Recognition with RCNN Optimized Vision Transformer

Bhavana Sharma *

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

Jeebananda Panda

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

Abstract:

The deep learning models and convolutional neural networks have been comprehensively used for different applications of gesture recognition. The vanishing gradient problem, which is a vital problem for precise feature extraction, is the reason why these models do not produce high-quality results for hand gesture recognition because of long-range dependencies. In order to solve this issue and to handle the vision-based tasks for classification and segmentation, respectively, an efficient model is proposed that combines the Vision Transformer (ViT) block with RCNN. This model is trained with three-fold cross validation on two publically available datasets i.e., Ego hand dataset and VIVA dataset. In this study, 20% of the data from both datasets are utilized for testing while 80% of the data are used for training. The other parameters like batch size are 16 and learning rate of 0.001 over the course of 100 epochs. The test accuracy obtained for Ego hand dataset and VIVA dataset are 96.04% and 98.39%, respectively.

Keywords:

Gesture recognition, segmentation, Feature extraction, RCNN, Vision Transformer.