DYNAMIC THRESHOLDING WITH SPATIO-TEMPORAL ATTENTION FOR WIFI GESTURE RECOGNITION

Author's: A. Sanjay[1], N. Vanjulavalli[2], K. Sujith[3]

Abstract

WiFi-based gesture recognition has emerged as a promising contactless human–computer interaction technology due to its device-free sensing capability, low deployment cost, and robustness under non-line-of-sight conditions. By leveraging Channel State Information (CSI) signals, WiFi systems can capture fine-grained spatio-temporal variations caused by human movements. However, practical WiFi gesture recognition systems face two critical challenges: severe class imbalance in labeled datasets and limited availability of annotated CSI samples. Traditional supervised learning approaches struggle under these constraints, often exhibiting biased predictions toward majority classes and reduced generalization performance.This research proposes an Imbalanced Semi-Supervised Learning framework using a Dynamic Threshold-Based Spatio-Temporal Attention Network (DT-STAN) to address these limitations. The proposed model integrates spatial attention to capture discriminative subcarrier–antenna correlations and temporal attention to identify salient motion dynamics across time. A bidirectional recurrent architecture encodes temporal dependencies in CSI sequences, while attention mechanisms enhance interpretability and feature discrimination.To mitigate data imbalance, class-aware sampling and weighted loss functions are incorporated into the supervised branch. For unlabeled data utilization, a semi-supervised pseudo-labeling strategy is introduced. Unlike fixed-confidence methods, the proposed dynamic threshold mechanism adaptively adjusts the pseudo-label acceptance criterion based on training progression, ensuring reliable label propagation during early stages while progressively leveraging more unlabeled samples as model confidence improves. An Exponential Moving Average (EMA) teacher model further stabilizes pseudo-label generation and reduces confirmation bias.The framework is evaluated using synthetically generated CSI datasets that simulate realistic WiFi gesture scenarios with long-tailed class distributions.