Comparison of Results: DenseNet-Based Method for Decoding Auditory Spatial Attention

October 22, 2024

Name: Shikhar Singh **Roll No.:** 210973 **Email:** shikhars21@iitk.ac.in

1 Introduction

The paper proposes a DenseNet-based method to improve spatial detection(ASAD) using EEG signals. We have used the KULeuven(KUL) dataset for this study. The model has a state-of-the-art accuracy of 94.3% for 1s data window. In this report, I compare the results obtained from my implementation and the results obtained from the paper.
We will be using the decoding accuracy for the comparison and model performance for different data window lengths and the effects of architectural variations.

2 Dataset Description

The dataset of KUL was used for the ASAD experiments. This has EEG recordings from 16 normal-hearing subjects, each recorded for 48 minutes. In the recording, EEG subjects had to focus on one of two simultaneously active competing speakers that were positioned at $\pm 90°$ along the azimuth direction corresponding to the left and right locations, respectively. The EEGs were recorded using a 64-channel BioSemi Active Two system.

3 Implementation Details

The implementation was done as follows:

- **Data Download**: Use 'download_data.py' to retrieve the KUL dataset.
- **Preprocessing**: Preprocess the raw EEG data using 'preprocess.m' in MATLAB; the generated files are saved in the '4_processed_data' folder.
- **Model Training**: Train and evaluate the model using 'main.py'.

Each epoch takes around **1.5-2** minutes.

4 Results Comparison

4.1 Decoding Accuracy

The paper presents the decoding accuracy of the DenseNet model in both subject-dependent and subject-independent conditions.
A subject-dependent condition is the one in which the model is tuned for each subject. A subject-independent condition is a condition in which the model is not tuned for a specific subject.

Table 1: Decoding Accuracy from the Paper (Subject-Dependent)

Window Length (s)	1	2	5	10
Decoding Accuracy (%)	94.3	95.9	96.6	96.8

As seen in Table 2, the accuracy achieved by my implementation for the 5-second and 10-second decision window is **96.9** and **96.7** compared to **96.6%** and **96.8%** in the paper. The differences could stem from variations in computational resources, dataset version and preprocessing environment.

Table 2: My Implementation Results (Subject-Dependent)

Window Length (s)	1	2	5	10
Decoding Accuracy (%)	94.3	95.9	96.9	96.7

4.2 Subject-Independent Performance

The paper also reports subject-independent performance, achieving decoding accuracy as shown in Table 3. My implementation results are shown in Table 4.

Table 3: Decoding Accuracy from the Paper (Subject-Independent)

Window Length (s)	1	2	5	10
Decoding Accuracy (%)	94.3	95.6	96.1	95.7

Table 4: My Implementation Results (Subject-Independent)

Window Length (s)	1	2	5	10
Decoding Accuracy (%)	94.6	95.5	95.8	95.8

4.3 Ablation Study

The paper includes an ablation study comparing CNN-baseline, CNN-3D, DenseNet-3D without bootstrapping, and DenseNet-3D with bootstrapping. The results from my ablation study are shown in Table 5.

Table 5: Ablation Study Results (1-second decision window)

Model	Paper Accuracy (%)	My Accuracy (%)
CNN-baseline	84.8	72.2
CNN-3D	88.6	88.5
DenseNet-3D (no bootstrapping)	91.8	91.6
DenseNet-3D (bootstrapping)	94.3	94.3

As seen in Table 5, the results of the CNN-baseline implemented in the paper and my implementation differ by a large margin because the ablation study of the CNN-baseline was hindered during the processing and training.

5 Discussion

The performance of my implementation is close to that of the original paper in most cases. However, certain discrepancies are noted, particularly with the 1-second decision window and in the ablation study of the CNN-baseline.

6 Conclusion

This report compares the results obtained from the paper and those achieved through my model implementation. The overall performance is consistent with the results reported in the original work.