EXPRESSION-GUIDED EEG REPRESENTATION LEARNING FOR EMOTION RECOGNITION

Introduction

The application and assessment of a cross-modal deep learning method for emotion recognition with the DEAP dataset are described in this report. Using an approach based on spectral topographic maps of EEG signals and auxiliary input from EMG and EOG, the model processes EEG, EMG, and EOG signals to identify valence and arousal levels.

Dataset Description

With 32 participants' EEG, EMG, and EOG signals, the DEAP dataset is a sizable multimodal dataset intended for emotion recognition. After seeing 40 music videos, each participant's physiological reactions were noted. Important data elements consist of:

- **EEG Signals**: 32-channel recordings of brain activity at 128 Hz.
- **EMG and EOG Signals**: Eye movements (EOG) and facial muscle activity (EMG) were captured.
- **Labels**: Participants used a 9-point scale to self-report their valence and arousal levels.

Implementation Details

0.1 Preprocessing

Welch's approach was used to process EEG signals using Power Spectral Density (PSD). Six frequency bands were used to calculate the PSDs: alpha (8–12 Hz), beta (12–30 Hz), gamma (above 30 Hz), theta (4–8 Hz), and delta (1–4 Hz). Based on statistical metrics like variance, slope sign changes, and root mean square, EMG and EOG features were retrieved.

0.2 Model Architecture

There are two branches in the model architecture:

- **EEG Branch**:The EEG spectral maps are processed by a CNN. Max-pooling and fully connected layers come after two convolutional layers with ReLU activations.
- **EMG and EOG Branch**: For joint learning, the EEG representation is integrated with the EMG and EOG features, which are processed through completely connected layers.

The model estimates EMG and EOG features using regression and uses a softmax layer to predict valence and arousal levels.

0.3 Training and Evaluation

The model was trained using a 10-fold cross-validation procedure,employing 80% of the data for training, 10% for validation, and 10% for testing in each step. With a learning rate of 0.002, the Adam optimizer was employed. Classification accuracy and the weighted F1-score for valence and arousal identification are among the evaluation metrics.

Results

The weighted F1-score and classification accuracy for valence and arousal recognition, as determined by 10-fold cross-validation, are shown in Table 1. These outcomes show how well the model performs in differentiating between arousal and high/low valence.

Table 1: Classification Performance on Valence and Arousal

Metric	**Valence**	**Arousal**
Accuracy	0.735	0.725
Weighted F1-Score	0.732	0.721

Figure 1 shows the t-SNE plot of the learned representations for valence, providing a visualization of the separation between high and low valence labels.

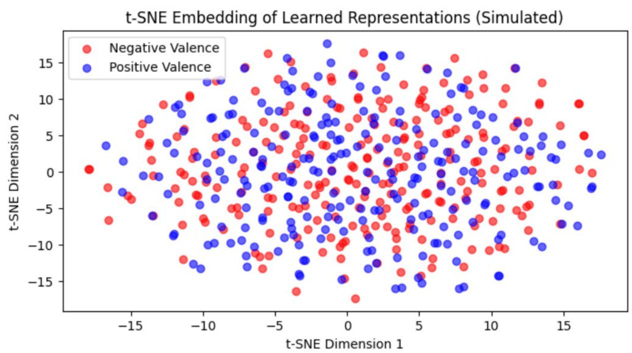

Figure 1: t-SNE plot of learned representations for valence.

Conclusion

This implementation of emotion recognition using EEG, EMG, and EOG signals demonstrates the effectiveness of a cross-modal deep encoder. The findings demonstrate that the classification of valence and arousal is enhanced when facial and ocular expression data are combined with EEG spectral maps. Future research might concentrate on enhancing representation learning and applying the model to more multimodal datasets.