DeMoAiR: Deep Learning based Multi-modal Airwriting Recognition for Smart Wearables

The ability to communicate stands out as a fundamental life skill for humans. As digital devices are rapidly evolving, the interaction between humans and these devices is also increasing. Conventionally, users engage with computers through peripheral devices such as keyboards, mouse, or touchscreen. However, the future of technology aspires to eliminate the dependence on these mediating devices, thereby relieving individuals from the burden of carrying additional physical tools. In pursuit of an alternative input method, this thesis explores the field of multi-modal airwriting recognition. The objective is to provide users with an innovative means of input and therefore contribute to the landscape of human-computer interaction. Airwriting is defined as the act of freely writing letters in the air through unrestricted finger movements, constituting a specialized form of gesture recognition. Such a system does not require users to learn a new set of gestures and also allows for a versatile range of words formed by concatenating letters. This broadens the spectrum of interaction capabilities for users. The process of airwriting necessitates coordination between the brain, forearm muscles, and the wrist to execute the desired movements. Consequently, recognition of airwriting may involve analyzing cues from each of these individual aspects, leading to a multi-modal approach. This approach integrates Electroencephalography (EEG) for brain signals, Electromyography (EMG) for muscle signals, and Inertial Measurement Unit (IMU) for wrist movement. Each modality provides a unique perspective on the same information. This thesis explores the feasibility of airwriting recognition utilizing diverse modalities.

The thesis begins with a focus on airwriting recognition using signals recorded from a wrist-worn IMU. A supervised contrastive loss based framework for airwriting recognition (SCLAiR) is presented. The framework employs a 2-stage classification method: in the first stage, supervised contrastive loss is applied, followed by the minimization of cross-entropy loss in the second stage. This approach enhances classification accuracy on both a publicly available dataset and a dataset recorded in the lab. Additionally, a real-time demonstration showcasing airwriting recognition is provided. Subsequently, image representation of time-series data (ImAiR) is also utilized for airwriting recognition. First, the time-series data is encoded to image representation and then deep learning-based models are utilized for identifying the written alphabets. Different techniques such as Self Similarity Matrix, Gramian Angular Field, and Markov Transition Field, are utilized for image encoding. Various standard model architectures for image classification are then used for classification.

The prospect of airwriting recognition using EMG signals recorded from forearm muscles is subsequently explored. The SurfMyoAiR dataset comprising EMG signals while writing English uppercase alphabets is constructed. Different time-domain features to construct EMG envelopes and different time-frequency image representations are explored to form the input to a deep learning model for airwriting recognition. Several different deep learning architectures are exploited for this task. Furthermore, a multi-loss minimization framework for EMG-based airwriting recognition (TripCEAiR) is proposed. The framework aims at learning a feature embedding vector that minimizes the triplet loss, while simultaneously

learning the parameters of a classifier head to recognize corresponding alphabets. The proposed method is validated on the SurfMyoAiR dataset. The effect of different variations of triplet loss, triplet mining strategies, and feature embedding dimension is also comprehensively investigated.

Leveraging EEG signals for airwriting recognition offers a promising alternative input method for Human-Computer Interaction and is explored as a part of this thesis. The NeuroAiR dataset comprising EEG signals recorded while writing English uppercase alphabets is first constructed. Various features are then explored in conjunction with different deep learning models to achieve accurate airwriting recognition. These features include processed EEG data, Independent Component Analysis (ICA) components, source-domain-based scout time series, and spherical and head harmonic decomposition-based features. Furthermore, the impact of different EEG frequency bands on system performance is comprehensively examined. This sets a strong baseline for future advancements and demonstrates the viability and utility of EEG-based airwriting recognition.