Learnt Contrastive Concept Embeddings for Sign Recognition

Ryan Wong, Necati Cihan Camgoz, Richard Bowden

In natural language processing (NLP) of spoken languages, word embeddings have been shown to be a useful method to encode the meaning of words. Sign languages are visual languages, which require sign embeddings to capture the visual and linguistic semantics of sign. Unlike many common approaches to Sign Recognition, we focus on explicitly creating sign embeddings that bridge the gap between sign language and spoken language. We propose a learning framework to derive LCC (Learnt Contrastive Concept) embeddings for sign language, a weakly supervised contrastive approach to learning sign embeddings. We train a vocabulary of embeddings that are based on the linguistic labels for sign video. Additionally, we develop a conceptual similarity loss which is able to utilise word embeddings from NLP methods to create sign embeddings that have better sign language to spoken language correspondence. These learnt representations allow the model to automatically localise the sign in time. Our approach achieves state-of-the-art keypoint-based sign recognition performance on the WLASL and BOBSL datasets.