**FACIAL EXPRESSION RECOGNITION WITH SWIN TRANSFORMER**

Recognizing facial expressions in people is essential in many human-related systems, including the medical and health care industries. Facial expression recognition research has reached a mature stage because to deep learning's recent successes and the availability of a lot of annotated data.

to be used with audio-visual datasets in practical applications. In this article, we introduce the Swin transformer-based face expression technique for the Aff-Wild2 Expression dataset's audio-visual in-the-wild dataset. For the audio-visual videos, we specifically use a three-stream network (i.e., Visual stream, Temporal stream, and Audio stream) to combine the multi-modal data into face expression identification. The Aff-Wild2 dataset's experimental results demonstrate the viability of our suggested multi-modal techniques.

Recognition of human facial expressions has gained a lot of popularity recently, both in artificial intelligence (AI) research and in real-world settings like the healthcare and medical industries [1, 2]. Particularly, the impressive developments in deep learning and the accessibility of massive annotated datasets pave the door for facial expression detection applications in real-world settings. In response to this pattern, a sizable in-the-wild dataset of Aff-Wild2 is provided by the 3rd Affective Behaviour Analysis in-the-wild (ABAW 2022) competition by Kollias et al. in connection with CVPR 2022 [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. The Aff-Wild2 includes annotations for the three primary tasks of valence-arousal estimation, action unit (AU) identification, and frame estimation across 548 films and 2,813,201 frames.

In order to categorise eight facial expressions, we used multi-modal data with a three-stream model in this research, comprising cropped faces, multiple cropped faces, and audio from the Aff-Wild2 data set. It performed better than the baseline according to the recently announced Swin-transformer, and performance improvement was made using the suggested half-mix jittering augmentation.