

Multimodal Data Fusion Techniques in Audiovisual Speech Recognition

Authors : Hadeer M. Sayed, Hesham E. El Deeb, Shereen A. Taie

Abstract : In the big data era, we are facing a diversity of datasets from different sources in different domains that describe a single life event. These datasets consist of multiple modalities, each of which has a different representation, distribution, scale, and density. Multimodal fusion is the concept of integrating information from multiple modalities in a joint representation with the goal of predicting an outcome through a classification task or regression task. In this paper, multimodal fusion techniques are classified into two main classes: model-agnostic techniques and model-based approaches. It provides a comprehensive study of recent research in each class and outlines the benefits and limitations of each of them. Furthermore, the audiovisual speech recognition task is expressed as a case study of multimodal data fusion approaches, and the open issues through the limitations of the current studies are presented. This paper can be considered a powerful guide for interested researchers in the field of multimodal data fusion and audiovisual speech recognition particularly.

Keywords : multimodal data, data fusion, audio-visual speech recognition, neural networks

Conference Title : ICCCN 2022 : International Conference on Communications and Computer Networks

Conference Location : Jeddah, Saudi Arabia

Conference Dates : November 14-15, 2022