World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:11, 2022

Audio-Visual Recognition Based on Effective Model and Distillation

Authors : Heng Yang, Tao Luo, Yakun Zhang, Kai Wang, Wei Qin, Liang Xie, Ye Yan, Erwei Yin

Abstract : Recent years have seen that audio-visual recognition has shown great potential in a strong noise environment. The existing method of audio-visual recognition has explored methods with ResNet and feature fusion. However, on the one hand, ResNet always occupies a large amount of memory resources, restricting the application in engineering. On the other hand, the feature merging also brings some interferences in a high noise environment. In order to solve the problems, we proposed an effective framework with bidirectional distillation. At first, in consideration of the good performance in extracting of features, we chose the light model, Efficientnet as our extractor of spatial features. Secondly, self-distillation was applied to learn more information from raw data. Finally, we proposed a bidirectional distillation in decision-level fusion. In more detail, our experimental results are based on a multi-model dataset from 24 volunteers. Eventually, the lipreading accuracy of our framework was increased by 2.3% compared with existing systems, and our framework made progress in audio-visual fusion in a high noise environment compared with the system of audio recognition without visual.

Keywords : lipreading, audio-visual, Efficientnet, distillation
Conference Title : ICILPISE 2022 : International Conference on Inductive Logic Programming, Information and Software Engineering
Conference Location : Istanbul, Türkiye
Conference Dates : November 29-30, 2022