

Method and System for Automatic Speech Recognition Correction

Abstract

In a described embodiment, candidate text transcriptions associated with captured speech information from an automatic speech recognition (ASR) system are processed and image data and reference transcripts associated with a plurality of image frames of a dataset of videos are extracted. The extracted image data is further filtered based on a similarity calculation between the image data and the reference transcripts to obtain a filtered dataset. A machine learning model is used to output corrected candidate text transcriptions based on the filtered dataset and the received candidate text transcriptions.

10 Figure 1