ABSTRACT

The embodiments of the present disclosure relate to a method and apparatus for video matching, an electronic device and a product. The method comprises: on the basis of an understanding of content of an audio, generating a plurality of segment texts of the audio (202). The method comprises: on the basis of an understanding of multimodal content of a plurality of videos, generating a plurality of text descriptions for the plurality of videos (204). In addition, the method further comprises: on the basis of the plurality of segment texts of the audio and the plurality of text descriptions of the plurality of videos, determining a video matching with the at least one of the plurality of segment texts (206).