

論文内容要旨

報告番号	甲 先 第 408 号	氏 名	馮 鐸
学位論文題目	Research on Facial Expressions Recognition based on Deep Learning Methods (深層学習に基づく顔表情認識に関する研究)		
<p>内容要旨</p> <p>Facial expression recognition (FER) is a process of automatically recognizing and inferring the performance of human emotional states on the face through artificial intelligence technology. As the most important part of recognizing human emotion, FER technology crosses and integrates physiology, psychology, image processing, machine vision, pattern recognition, and other research fields. It has received extensive attention in the fields of human-computer interaction, information security, robotics, automation, medical care, communication technology, autonomous driving, etc. Although decades of research work on FER have been carried out, in actual situations, realizing accurate and effective FER is still a challenging problem. In recent years, with the success of deep learning technology in various fields, more and more deep neural networks are used to learn the discriminative representation of automatic FER. This thesis studies FER by combining traditional machine learning methods and constructing efficient deep model architecture. The main contributions of this thesis are summarized as follows:</p> <p>(1) This thesis first reviewed and summarized the currently widely used methods and existing problems in FER. After fully understanding the limitations of the traditional handcrafted features, a multi-stream neural network model combining the manually extracted LBP-TOP features and the deep learning model is proposed to recognize the dynamic process of facial expression changes. In the multi-stream neural network proposed in this thesis, to recognize dynamic facial expressions, the cascaded CNN-RNN model is used to extract the features of the input facial expression image sequence from space expand to time series. At the same time, the handcrafted features LBP-TOP is used to directly extract the spatiotemporal features of the image sequence, and then the CNN and RNN networks are used to process the spatiotemporal features. Finally, through the fusion of the two streams of features, and through experiments on the public database, it is proved that the handcrafted spatiotemporal features can effectively supplement the CNN-RNN model and improve the results of FER.</p> <p>(2) Application-oriented FER faces two challenges. One is the transition of FER from laboratory control to challenging in-the-wild conditions, and the other is the recent challenge of decentralizing deep network application technology to mobile platforms. Simply using larger and deeper neural network models for recognition</p>			

tasks can no longer cope with this problem. In recent years, FER has been proved to be more natural and effective from consecutive frames. The motivation of this thesis becomes to create a lightweight network that processes dynamic facial expression sequences. After studying the amount of calculation of the model architecture, the MobileNet series with a deep separable convolution architecture is chosen as the basic model of the CNN part and used GRU as the frame-to-sequence approach part to construct a lightweight CNN-RNN cascade network. The performance improvement is demonstrated by using the proposed technique on both the laboratory control and in-the-wild conditions databases.

(3) Through previous research, first the supplementary ability of handcrafted features extraction for deep learning methods is verified. Then the application of the lightweight depth model in FER is discussed. Further used the updated technology to combine the advantages of local binary convolution (LBC) and deep separable networks and proposed a new model architecture. Inspired by model pruning and SE optimization, using the feature that the convolution kernel parameters in LBC are not trainable, this thesis proposes a pruning method on depthwise LBC and SE optimization model architecture. Experiments were not only conducted on the general image classification database, but also on the in-the-wild conditions facial expression databases. The experimental results prove the effectiveness of our proposed model and pruning method.