Research on Facial Expression Recognition Technology Based on Convolutional-Neural-Network Structure

Junqi Guo, College of Information Science and Technology, Beijing Normal University, Beijing, China
Ke Shan, College of Information Science and Technology, Beijing Normal University, Beijing, China
Hao Wu, College of Information Science and Technology, Beijing Normal University, Beijing, China
Rongfang Bie, College of Information Science and Technology, Beijing Normal University, Beijing, China
Wenwan You, College of Information Science and Technology, Beijing Normal University, Beijing, China
Di Lu, College of Information Science and Technology, Beijing Normal University, Beijing, China

ABSTRACT

Human facial expressions change so subtly that recognition accuracy of most traditional approaches largely depend on feature extraction. In this article, the authors employ a deep convolutional neural network (CNN) to devise a facial expression recognition system to discover deeper feature representation of facial expression. The proposed system is composed of the input module, the pre-processing module, the recognition module and the output module. The authors introduce jaffe and ck+ to simulate and evaluate the performance under the influence of different factors (e.g. network structure, learning rate and pre-processing). The authors also examine the anti-noise property of the system with zero-mean gaussian white noise. In addition, they simulate the recognition accuracy on different expression pairs and discuss the confusion issue on similar expression recognition. Finally, they introduce the k-nearest neighbor (KNN) algorithm compared with CNN to make the results more convincing.

KEYWORDS

Affective Computing, Artificial Intelligence, Computer Vision, Convolutional Neural Network, Deep Learning, Facial Expression Recognition Research, Machine Learning

INTRODUCTION

Picard (2000) has predicted that Affective Computing would be an important direction for future artificial intelligence research. Human facial expression recognition is an important task for affective computing. The American psychologist Ekman and Friesen (1971) defined seven categories of basic facial expression, which are Happy, Sad, Angry, Fear, Surprise, Disgust and Neutral. Pentland and Mase (1991) held the first attempt to use optical flow method to determine the direction of movement of facial muscles. Then, they extracted the feature vectors to achieve four kinds of automatic expression recognition including Happy, Angry, Disgust, Surprise and reached the accuracy of nearly 80%.

DOI: 10.4018/IJSI.2018100108

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Hinton and Salakhutdinov (2006) published an article in “Science”, opening the door to a deep learning era. Hinton suggested that the neural network with multiple hidden layers had good ability for learning characteristics. It can improve the accuracy of prediction and classification by obtaining different degrees of abstract representation of the original data. So far, deep learning has achieved good performance in speech recognition, collaborative filtering, handwriting recognition, computer vision and many other fields (Chen & Lin, 2014).

The concept of Convolutional Neural Network (CNN) was presented by Yann LeCun (1989), where a neural network architecture was composed of two kinds of basic layers, called convolutional layers (C layers) and subsampling layers (S layers). However, many years after that, there was still no major breakthrough of CNN. One of the main reasons was that CNN could not get ideal results on large size images. But it was changed when Hinton and his students used a deeper Convolutional Neural Network to reach the optimal results in the world on ImageNet in 2012. Since then, more attention has been paid on CNN based image recognition.

In this paper, we present a method to achieve facial expression recognition based on a deep CNN. Firstly, we implement face detection by using Haar-like features and histogram equalization. Then we construct a four-layer CNN architecture, including two convolutional layers and two subsampling layers (C-S-C-S). Finally, a Softmax classifier is used for multi-classification. The structure of the paper is organized as follows: Section 2 introduces the whole system based on CNN, including the input module, the image pre-processing module, the recognition algorithm module and the output module. In Section 3, we simulate and evaluate the recognition performance of the proposed system under the influence of different factors such as network structure, learning rate and pre-processing. Finally, a conclusion is drawn.

**FACIAL EXPRESSION RECOGNITION SYSTEM BASED ON A CNN STRUCTURE**

In this section, we introduce the whole system based on a CNN structure and describe important details of all modules including face detection, image pre-processing and recognition algorithm.

**System Overview**

This part starts with the overall introduction of a CNN-based facial expression recognition system. System flow is showed in Figure 1.

First of all, the Input Module obtains the input image 2D data. We employ the Extended Cohn-Kanade Dataset (CK+) (Lucey, Cohn, Kanade, Saragih, Ambadar, & Matthews, 2010) and the Japanese Female Facial Expression Database (JAFFE) (Lyons, Akematsu, Kamachi, & Gyoba, 1998) for the simulation, which are both standard facial expression databases categorized into 7 kinds of expressions. The Pre-processing Module includes 2 steps: face detection and histogram equalization. Thus, we can get the main part of the human face and minimize the difference of lighting conditions in backgrounds. Recognition Module is based on a CNN structure and a Softmax multiple classifiers. The Output Module shows MSE convergence and calculates the recognition accuracy. If the recognition accuracy does not meet the requirement, it re-adjusts the network parameters and begins a new round of training until the accuracy is satisfying. Details of each module are described in the following part.

**Image Pre-Processing**

We employ the two standard facial expression databases mentioned above for simulation, which are both widely acknowledged by academia. JAFFE contains 213 images of 10 Japanese women, while CK+ covers 328 expression images coming from all over the world. Before the recognition, some pre-processing work need to be done firstly. In our image pre-processing procedure, we run a two-step process to reduce the interference in the original images, which are Face Detection and Histogram Equalization.
Using the Cloud for Testing NOT Adjunct to Development
www.igi-global.com/chapter/using-cloud-testing-not-adjunct/72233?camid=4v1a

Quality Metrics for Evaluating Data Provenance
www.igi-global.com/chapter/quality-metrics-evaluating-data-provenance/8245?camid=4v1a