Title: Attention mechanism-based CNN for facial expression recognition

Article preview

Neurocomputing
Volume 411, 21 October 2020, Pages 340-350
Attention mechanism-based CNN for facial expression recognition
Author links open overlay panelJing Li a, Kan Jin a, Dalin Zhou b, Naoyuki Kubota c, Zhaojie Ju b

https://doi.org/10.1016/j.neucom.2020.06.014Get rights and content
Abstract
Facial expression recognition is a hot research topic and can be applied in many computer vision fields, such as human?computer interaction, affective computing and so on. In this paper, we propose a novel end-to-end network with attention mechanism for automatic facial expression recognition. The new network architecture consists of four parts, i.e., the feature extraction module, the attention module, the reconstruction module and the classification module. The LBP features extract image texture information and then catch the small movements of the faces, which can improve the network performance. Attention mechanism can make the neural network pay more attention to useful features. We combine LBP features and attention mechanism to enhance the attention model to obtain better results. In addition, we collected and labelled a new facial expression dataset of seven expressions from 35 subjects aged from 20 to 25. For each subject, we captured both RGB images and depth images with a Microsoft Kinect sensor. For each image type, there are 245 image sequences, each of which contains 110 images, resulting in 26,950 images in total. We apply the newly proposed method to our own dataset and four representative expression datasets, i.e., JAFFE, CK+, FER2013 and Oulu-CASIA. The experimental results demonstrate the feasibility and effectiveness of the proposed method.
Introduction

Facial expression is one of the most direct signals to express inner feelings in people's daily communication. The physical or mental state of a person at one time can be obtained by analyzing facial expressions. Therefore, facial expression recognition is of great significance in autopilot, human?computer interaction, medical treatment and other fields related to facial expression, and has gradually become a more and more important research direction. In machine learning, a variety of facial expression recognition algorithms have been proposed. Due to the complexity, diversity, occlusion, lighting and other challenges in facial expression recognition, the recognition accuracy in practical applications is still unsatisfactory.

In this paper, our goal is to design a recognition model that can automatically and accurately recognize different expressions in various types of images. Generally, the process of facial expression recognition consists of the following steps: i) pre-processing of the facial expression data; ii) feature extraction of facial expressions; and iii) classification of facial expressions. The process is depicted in Fig. 1. We usually consider two kinds of features, namely, facial features and face model features. The facial features are specific points on the face, like eyes, mouth, and eyebrows; the face model features are the features used to model the face. Therefore, there are several ways for facial representation, like using the whole face to get the holistic representation, using specific points for local representation, and combining different points to get a hybrid approach. The final step is to define some set of categories to which the expression belongs.

When dealing with expression recognition as a classification problem, traditional methods often use hand-crafted features such as Local Binary Patterns (LBP) and traditional machine learning algorithms such as Support Vector Machine (SVM) to classify. These methods may work well on datasets collected under laboratory conditions, but with the introduction of more challenging expression datasets in uncontrollable environments (e.g., FER2013), they cannot effectively achieve this task. Fortunately, deep learning has made a breakthrough in convenience and effectiveness since it has been used to deal with the image classification problem.

The attention mechanism has been widely used in various computer vision tasks such as saliency detection [15], crowd counting [16] and facial expression recognition [38]. The operation can select the most useful features for classification by learning an intermediate attention map and then applying element-wise product on attention maps and source feature maps to weight the importance of different features. For the task of facial expression recognition, the features that are useful for recognition are mainly in some key parts such as eyes, nose and mouth. The attention mechanism increases the weights of these key features and helps improve the expression recognition results.

In this paper, we design a novel Convolutional Neural Network with an attention model for recognizing facial expressions. In [43], it showed that using LBP features is better than using HOG and Gabor features because LBP can achieve rotation invariance and grey-scale invariance and thus is suitable for extracting texture features at different scales and can solve the imbalance of displacement, rotation angles and illumination conditions in facial images. In addition, LBP features can reflect fine facial changes in skin textures like wrinkles and furrows, which shows the changes of expressions. In [38], an end-to-end network with an attention model was presented for facial expression recognition. The attention module makes the network focus more on useful features which are vital for expression recognition by increasing the weights of these features. This makes the network recognize expressions more efficiently. Inspired by [38], [43], we combine LBP features with an attention model for facial expression recognition. Embedding the attention model into the network allows the network to pay different attention and weight to different parts of the input data. This can make the neural network pay more attention to useful features, which is vital to expression recognition. Furthermore, we combine LBP features with convolution features to improve our recognition results. The proposed method has been tested on five facial expression datasets, which are CK+ [11], JAFFE [13], FER2013 [39], Oulu-CASIA [25], and our self-collected Nanchang University Facial Expression (NCUFE).

To verify the effectiveness of our algorithm, we collected a new facial expression dataset called NCUFE. The dataset consists of seven expressions (i.e., anger, disgust, fear, happiness, sadness, surprise and neutral). We collected these facial expression images from 35 graduate students (6 females and 29 males) by a Microsoft Kinect sensor for acquiring both RGB images and depth images. The sample images are shown in Fig. 2. For each student, the size of these two types of images is 1280*1024 and 512*424, respectively. For each image type, there are 245 image sequences, each of which contains 110 images, resulting in 26,950 images in total. During the process of image capture, the students sat on a chair in front of the Kinect and faced the camera. The distance between the face and the Kinect was about 100 cm. We asked each student to look the expression examples printed on some pieces of paper and then make seven expressions.

The main contributions of this paper are as follows:

* 1)

We introduce a novel facial expression recognition method with attention mechanism. Not only raw images, but also LBP features are added to the attention layers of the network. LBP features contain texture information and can reflect fine facial changes in skin textures, which can help distinguish expressions with subtle difference.

* 2)

We collected and labelled a new dataset named Nanchang University Facial Expression (NCUFE) for facial expression recognition. The dataset includes 490 image sequence collected from 35 subjects labeled with seven facial expressions (i.e., anger, disgust, fear, happiness, sadness, surprise and neutral). For each subject, we captured both RGB images and depth images.

* 3)

We implement substantial experiments on five different datasets, as shown in Fig. 3. There are not only datasets collected under laboratory conditions, such as CK+, JAFFE, Oulu-CASIA and NCUFE, but also those collected in real world like FER2013. We also compare the model performance with some state-of-the-art expression recognition algorithms, and the results show that our model is superior.

The remainder of this paper is as follows. In Section 2, we introduce the related works in expression recognition and the existing algorithms. In Section 3, we describe the proposed method in detail. We describe our experimental process and results on different datasets and compare them with the results of the state-of-the-art algorithms in Section 4. We give a conclusion of the whole paper in Section 5.

Section snippets

Related works

Traditional facial feature extraction algorithms can be separated into two categories: 1) geometric-based methods, such as Active Appearance Models (AAM) [17]; and 2) appearance-based methods, such as LBP [9] and Gabor Wavelet Representation [21]. After feature description, the features are fed into a

classifier, such as SVM [22] and K-nearest Neighbors (KNN) [24], for recognizing different facial expressions. Therefore, the performance of the classifier depends to a large extent on the quality
Network architecture
In this section, we introduce our newly proposed convolutional neural network with attention mechanism for automatically recognizing facial expressions. The new network consists of four parts, i.e., the feature extraction module, the attention module, the reconstruction module and the classification module. The architecture starts from the feature extraction module composed of two separate CNN processing streams: one is for raw images and the other is for LBP feature maps. Our model uses pure
Experimental results
In this work, we design a novel deep Convolutional Neural Network with an attention model to automatically recognizing facial expressions. Except for the famous facial expression datasets such as CK+ [11], JAFFE [13], Oulu-CASIA [25] and FER2013 [39], we also evaluate our proposed method on our self-collected dataset NCUFE. Because CK+, JAFFE, Oulu-CASIA and NCUFE do not provide specified training and testing sets, we employ 5-fold cross-validation protocol in these four datasets. The proposed
Conclusions and future work
This paper presents a novel convolutional neural network with attention mechanism for facial expression recognition. The method fuses LBP features and convolution features, and then is combined with attention mechanism to improve the performance of the network. In order to prevent overfitting and ensure the generalization ability of the network, we apply data augmentation in the datasets we used in the experiments. In addition, we collected a new dataset called Nanchang University Facial
CRediT authorship contribution statement
Jing Li: Conceptualization, Methodology, Writing - original draft. **Kan Jin:** . **Dalin Zhou:** . **Naoyuki Kubota:** Writing - review & editing. **Zhaojie Ju:** Conceptualization, Methodology, Writing - original draft.
Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Acknowledgements
This work is supported by National Natural Science Foundation of China under Grant 61963027, 61703198, and 51575412, Natural Science Foundation for Distinguished Young Scholars of Jiangxi Province under Grant 2018ACB21014.
Jing Li received the B.E. degree in Electronic Information Engineering from Nanchang University, China, in 2005, and obtained her PhD degree in Electronic and Electrical Engineering from the University of Sheffield, UK, in 2011. From 2011 to 2012, she was a Research Associate with the Department of Computer Science at the University of Sheffield. Currently, she is an Associate Professor with the School of Information Engineering at Nanchang University. Her research interests include visual

References (44)
 * Sun _et al._
An efficient unconstrained facial expression recognition algorithm based on stack binarized auto-encoders and binarized neural networks
Neurocomputing
(2017)
 * Z. Yu _et al._

Spatio-temporal convolutional features with nested LSTM for facial expression recognition[J]
Neurocomputing
(2018)
 * T. Ojala _et al._
A comparative study of texture measures with classification based on feature distributions
Pattern Recogn.
(1996)
 * W. Sun _et al._
A visual attention based ROI detection method for facial expression recognition
Neurocomputing
(2018)
 * G. Zhao _et al._
Facial expression recognition from near-infrared videos
Image Vision Comput.
(2011)
 * J. Shao _et al._
Three convolutional neural network models for facial expression recognition in the wild
Neurocomputing
(2019)
 * Z. Zhang _et al._
From facial expression recognition to interpersonal relation prediction
Int. J. Comput. Vis.
(2018)
 * Y.-H. Lai, S.-H. Lai, Emotion-preserving representation learning via generative adversarial network for multi-view...
 * A. Mollahosseini, D. Chan, M.H. Mahoor, Going deeper in facial expression recognition using deep neural networks,...
 * C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, IEEE Conference on Computer Vision and Pattern...
* K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, (2015)...
* C. Szegedy, S. Ioffe, V. Vanhoucke, et al., Inception-v4, Inception-ResNet and the Impact of Residual Connections on...
* B. Hasani, M.H. Mahoor, Spatio-temporal facial expression recognition using convolutional neural networks and...
* T. Ojala, M. Pietikinen, T. Menp, Multiresolution grayscale and rotation invariant texture classification with local...
* P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): a complete...
* M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, in: Proceedings Third IEEE...
* G. Huang, Z. Liu, K. Q. Weinberger, L. van der Maaten, Densely connected convolutional networks. arXiv preprint...
* T. Zhao, X. Wu, Pyramid feature attention network for saliency detection, Proceedings of the IEEE Conference on...
* R.R. Varior, B. Shuai, J. Tighe, et al., Scale-Aware Attention Network for Crowd Counting. arXiv preprint...
* T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2011)...
* W. Wang, Q. Sun, T. Chen, et al., A Fine-Grained Facial Expression Database for End-to-End Multi-Pose Facial Expression...
* C.M. Kuo, S.H. Lai, M. Sarkis, A compact deep learning model for robust facial expression recognition,

Proceedings of...
View more references
Cited by (263)
* ### Emotion recognition and artificial intelligence: A systematic review (2014?2023) and research recommendations
2024, Information Fusion
Show abstract
Emotion recognition is the ability to precisely infer human emotions from numerous sources and modalities using questionnaires, physical signals, and physiological signals. Recently, emotion recognition has gained attention because of its diverse application areas, like affective computing, healthcare, human?robot interactions, and market research. This paper provides a comprehensive and systematic review of emotion recognition techniques of the current decade. The paper includes emotion recognition using physical and physiological signals. Physical signals involve speech and facial expression, while physiological signals include electroencephalogram, electrocardiogram, galvanic skin response, and eye tracking. The paper provides an introduction to various emotion models, stimuli used for emotion elicitation, and the background of existing automated emotion recognition systems. This paper covers comprehensive searching and scanning of well-known datasets followed by design criteria for review. After a thorough analysis and discussion, we selected 142 journal articles using PRISMA guidelines. The review provides a detailed analysis of existing studies and available datasets of emotion recognition. Our review analysis also presented potential challenges in the existing literature and directions for future research.
* ### A hybrid approach for forecasting ship motion using CNN?GRU?AM and GCWOA
2022, Applied Soft Computing
Show abstract
The motion of a ship, which has six degrees of freedom, is a complex nonlinear dynamic process with variable periodicity and chaotic characteristics. With the development of smart ships, modern high-precision equipment needs the help from high accuracy of ship motion (SHM) forecasting. Existing models will not easily be able to satisfy future accuracy requirements. Therefore, to improve the accuracy of SHM forecasts, by firstly determining the sequence features of SHM time series, a convolutional neural network (CNN) was used herein to extract automatically spatial feature vectors. Considering the variable-period characteristics of SHM time series, a gated recurrent unit (GRU) was used to learn the inherent time characteristics and to extract temporal feature vectors. The attention mechanism (AM) was developed to control the effect of feature vectors on the output to solve the problem of the contribution of feature vectors. Integrating the above methods, an SHM hybrid forecasting model, the SHM CNN?GRU?AM (SHM-C&G&A) model, was established. Secondly, in view of the difficulty of selecting the hyperparameters of a hybrid model, on account of the defects of the whale optimization algorithm (WOA), a normal cloud local search (NCLS) algorithm was developed. Exploiting the advantages of the normal cloud search (NCS) and the genetic algorithm (GA), a genetic random global search (GRGS) algorithm was developed. Then, a hybrid genetic cloud whale optimization algorithm (GCWOA) was developed and used to optimize the hyperparameters of the SHM-C&G&A model. Accordingly, a hybrid forecasting approach that integrates SHM-C&G&A and GCWOA was proposed; it is referred to as GCWOA-SHM-C&G&A. Finally, ship heave and pitch time series data are used to analyze an example to test the forecasting effectiveness of SHM-C&G&A and the optimization performance of GCWOA. The experimental results reveal that the proposed SHM-C&G&A model is more robust that the other models that are

considered in this paper, and exhibits better nonlinear characteristics. The proposed GCWOA yields a better combination of hyperparameters than contrast algorithms in the forecasting process.

* ### Distract Your Attention: Multi-Head Cross Attention Network for Facial Expression Recognition
2023, Biomimetics
* ### Deep Learning-Based IoT System for Remote Monitoring and Early Detection of Health Issues in Real-Time
2023, Sensors
* ### Facial Expression Recognition With Visual Transformers and Attentional Selective Fusion
2023, IEEE Transactions on Affective Computing
* ### Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation
2020, Complexity

View all citing articles on Scopus

Jing Li received the B.E. degree in Electronic Information Engineering from Nanchang University, China, in 2005, and obtained her PhD degree in Electronic and Electrical Engineering from the University of Sheffield, UK, in 2011. From 2011 to 2012, she was a Research Associate with the Department of Computer Science at the University of Sheffield. Currently, she is an Associate Professor with the School of Information Engineering at Nanchang University. Her research interests include visual tracking, behavior recognition, and crowd scene analysis. She has authored or coauthored in various journals, such as IEEE Transactions on Industrial Informatics, IEEE Transactions on Image Processing, Information Sciences (Elsevier), etc.

Kan Jin received the B.E. degree in Mechatronic Engineering from Hubei Polytechnic University, China, in 2017. He is currently pursuing the M.S. degree in the School of Information Engineering at Nanchang University, China. His research interests include facial expression recognition and crowd counting.

Dalin Zhou received the B.S. degree in automation from the University of Science and Technology of China, Hefei, China, in 2012, the Ph.D. degree in computing from the University of Portsmouth, UK, in 2018 and is currently a Lecturer in Computer Science with the School of Computing, the University of Portsmouth, UK. His current research interests include signal processing, machine learning, computational intelligence, biomedical robotics and multimodal sensor fusion. His current research contributes to the monitoring and rehabilitation of limb motor function improving the daily life activity and working capability for both the disadvantaged group of limb-impaired patients and the aging community.

Naoyuki Kubota, Faculty of Systems Design, Tokyo Metropolitan University, Tokyo, Japan Naoyuki Kubota received the B.Sc. degree from Osaka Kyoiku University, Kashiwara, Japan, in 1992, the M.Eng. degree from Hokkaido University, Hokkaido, Japan, in 1994, and the D.E. degree from Nagoya University, Nagoya, Japan, in 1997. He joined the Osaka Institute of Technology, Osaka, Japan, in 1997. He joined the Department of Human and Artificial Intelligence Systems, University of Fukui, Fukui, Japan, as an Associate Professor in 2000. He joined the Department of Mechanical Engineering, Tokyo Metropolitan University, Tokyo, Japan, in 2004. He was an Associate Professor from 2005 to 2012, and has been a Professor since 2012 at the Department of System Design, Tokyo Metropolitan University.

Zhaojie Ju received the B.S. in automatic control and the M.S. in intelligent robotics both from Huazhong University of Science and Technology, China, and the Ph.D. degree in intelligent robotics at the University of Portsmouth, UK. He held a research appointment at the University College London, London, U.K.,

before he started his independent academic position at the University of Portsmouth, U.K., in 2012. His research interests include machine intelligence, pattern recognition, and their applications on human motion analysis, multi-fingered robotic hand control, human?robot interaction and collaboration, and robot skill learning. He has authored or co-authored over 180 publications in journals, book chapters, and conference proceedings and received four best paper awards and one Best AE Award in ICRA2018.
View full text