International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 3366–3375

Real Face Emotion Identification Using Stacked Convolutional Neural Networks

1.**V.Srinivasagopal** , , PG Scholar, Department of ECE,BVC Engineering College(Autonomous), Odalarevu , Email id:v.srinivasagopal@gmail.com
2. **Dr. K.B.S.D.Sarma**, Professor, Department of ECE, BVC Engineering College (Autonomous), Odalarevu Email id:bsdsarma.kompella007@gmail.com
3. **Dr.Pravin Akula** ,Professor , Department of ECE, BVC Engineering College (Autonomous),Odalarevu Email id :akula.pravin@gmail.com

Abstract
The use of machines to perform different tasks is constantly increasing in society. Providing machines with perception can lead them to perform a great variety of tasks; even very complex ones such as elderly care. Machine perception requires that machines understand about their environment and interlocutor’s intention. Recognizing facial emotions might help in this regard. During the development of this work, deep learning techniques have been used over images displaying the following facial emotions: happiness, sadness, anger, surprise, disgust, and fear. In this research, a pure convolutional neural network approach outperformed other statistical methods' results achieved by other authors that include feature engineering. Utilizing convolutional networks involves feature learning; which sounds very promising for this task where defining features is not trivial. Moreover, the network was evaluated using two different corpora: one was employed during network's training and it was also helpful for parametertuning and for network's architecture definition. This corpus consisted of facial acted emotions. The network providing best classification accuracy results was tested against the second dataset. x While the results achieved were not state-of-the-art; the evidence gathered points out deep learning might be suitable to classify facial emotion expressions. Thus, deep learning has the potential to improve human- machine interaction because its ability to learn features will allow machines to develop perception. And by having perception, machines will potentially provide smoother responses, drastically improving the useexperience.
Key Words: Fake face identification, CNN, DNN, Emotion recognition, Python, Open cv, Harcascading face detection

1. Introduction

The use of machines in society has increased widely in the last decades. Nowadays, machines are used in many different industries. As their exposure with humans increase, the interaction also has to become smoother and more natural. In order to achieve this, machines have to be provided with a capability that let them understand the surrounding environment. Specially, the intentions of a human being. When machines are referred, this term comprises to computers and robots. A distinction between both is that robots involve interaction abilities into a more advanced extent since their design involves some degree of autonomy. When machines are able to appreciate their surroundings, some sort of machine perception has been developed [95]. Humans use their senses to gain insights about their environment. Therefore, machine perception aims to mimic human senses in order to interact with their environment [65][68].Nowadays, machines have several ways to capture their environment state trough cameras and sensors. Hence, using this information with suitable algorithms allow to generate machine perception. In the last years, the use of Deep Learning algorithms has been proven to be very successful in this regard [1][31][35]. For instance, Jeremy Howard showed on his Brussels2014

ISSN: 2233-7857 IJFGCN

TEDx's talk [43] how computers trained using deep learning techniques were able to achieve some amazing tasks. These tasks include the ability to learn Chinese language, to recognize objects in images and to help on medical diagnosis. Affective computing claims that emotion detection is necessary for machines to better serve their purpose [80].For example, the use of robots in areas such as elderly care or as porters in hospitals demands a deep understanding of the environment. Facial emotions deliver information about the subject's inner state [74]. If a machine is able to obtain a sequence of facial images, then the use of deep learning techniques

Implementation Framework

Nowadays, many frameworks have been developed for deep learning. Some of the most popular ones include librariessuch as: Caffe, Theano, and TensorFlow. Also, implementing a frameworkfrom scratch using a programming language was never considered.Itwouldhavebeenoutofscope since it requires a big amount of effort, and the duration of such a project usually takes years. The use of Python as the front-end API on all these frameworks shows that itis thepreferredlanguageformachinelearning. Usually, Python is combined with a programming language that provides support for low level operations such as: C or C++, to act on the back end.

Fig 1.1 Examples of face recognition

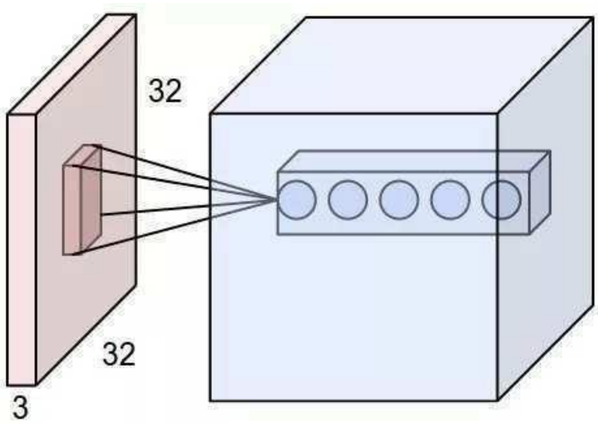

Fig1.2 Local receptive field of size 5x5x3 for a typical CIFAR-10 image, 32x32x3

As it was previously stated, after the convolution operation a plane composed of the results of applying the same filter through the entire input is generated. This plane is named feature map. Each feature map is the result of a convolutional operation with a kernel.

Kernels are initialized with different weights in order to perceive different features. Thus, the feature found is kept through the whole feature map, and its position is irrelevant for the network.

2.LiteratureSurvey

ANNrevivalDuringthe1970sand1980s, the shift on AI went exclusively to symbolic processing. However, a set of factors helped ANN to be on the spot again on the mid 1980s. These factors belong to different areas. One factor was that symbolic processing showed slowprogress and it was near- rowed to small simulations. Another factor was computer accessibility and hard- ware improvement compared to the one on previous decades.Nowadays, this factor is still relevant since experiments demand a lot of computational power. Most of current simulations would not be feasible in those times. Finally, connectionist re- searchers started to show some interesting results.

Deep Learning

The latest reincarnation of ANN is known as Deep Learning (DL). According to Yann Le Cun, this term designates"...any learning method that can train a system with more than 2 or 3 non-linear hidden layers."[32]. DL has achieved success on fields such as computer vision, natural language processing, and automatic speech recognition. One of the main strengths of using DL techniques is that there is no need for feature engineering. The algorithms are able to learn features by themselves over basicrepresentations.Forinstance,onimage recognition, an ANN can be feed withpixel representations of images. Then, the algorithm will determine if certain pixel combination represents any particular feature that is repeated through the image. As the data is processed through the layers, the features will go from very abstract forms to meaningful representationof objects. DL started to become popular after somebetterthanstate-of-the-artresultswere achieved on several fields. For instance, the first paper containing information about a major industrial application was one related to automatic speech recognition[31].Inthis paper from 2012, ANN outperformed Gaussian mixture models in several benchmarking tests. This paper is collaboration between four research groups: University of Toronto, Microsoft Research, Google Research and IBM Research

3.BlockDiagram

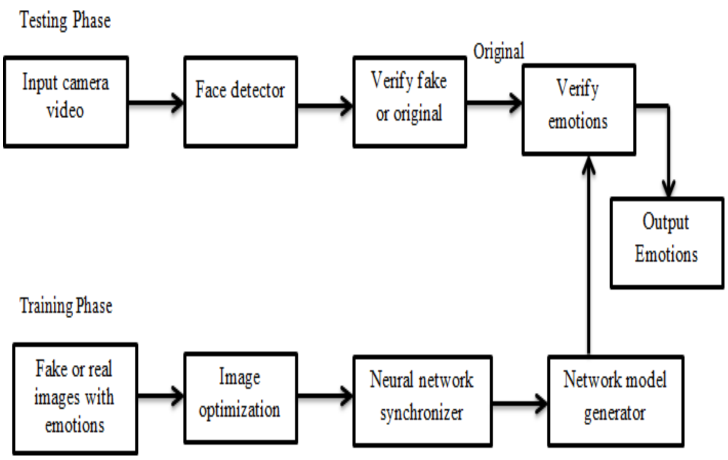

Fig 3.1 Block diagram

International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 3366–3375

4. Methodology

This chapter is divided into two main sections; each corresponding to the two work phases mentioned on eachsection, itisdescribedhowimageandvideoarepre- processed, the reasons behind choosing a particular topology and values for several parameters,andhowthenetwork'saccuracy is evaluated. This chapter describes how techniquesandconceptsdescribedsofaron the report interact during the experimental part.

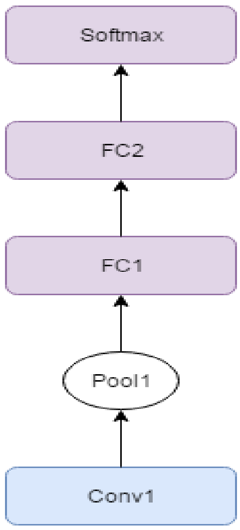

Fig 4.1: Network topology diagram

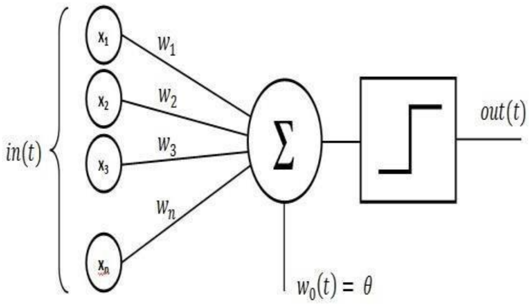

Fig 4.2 Control neural network Mux

5. Existing System

Convolutional Neural Networks

The study of the visual cortex is closely related to the development of the convolutional neural networks. Back in 1968, Hubel and Wiesel presented a study focused on the receptive fields of the monkey's visual cortex. This study was relevant because of the striate cortex (primary visual cortex) architecture description and the way that cells are arranged on it. Moreover, it also presented two different types of cells: simple, and complex. The simple ones are focused on edge-like shapes; while the complex ones cover a broader spectrum of objects and they are locally invariant. Therefore, the different sets of cell arrangements in the cortex are able to map the entire visual field. by exploiting the

ISSN: 2233-7857 IJFGCN

correlation of objects and shapes in local visual areas. One of the first implementations inspired by Hubel and Wiesel ideas was one called Neocognitron. Neocognitron [is a neural network model developed by Kunihiko Fukushima in 1980. The first layer of the modeliscomposedbyunits

6. PythonSoftware

Python is an interpreted, high-level, general-purpose programming language. Created by Guidovan Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notableuse of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects Python is dynamically typed and garbage-collected. It supports multiple programming paradigms including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released in 2008, was a major revision of thelanguage that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3. The Python 2 language was officially discontinued in 2020 (first planned for 2015), and "Python 2.7.18 is the last Python 2.7 release and therefore the last Python 2 release." No more security patches or other improvements will be released for it. With Python 2'send-of-life, only Python 3.5.x and later are supported. Python interpreters are available for many operating systems. A global community of programmers develops and maintains Python, an open sourcereferenceimplementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and Python development

7. Proposed System

Artificial Neural Networks

Supervised learning has a set of toolsfocusedonsolvingproblemswithinits domain. One of those tools is called Artificial Neural Networks (ANN). An ANNisasetoffunctionsthatperformlabel prediction. If the ANN is analyzed as a black box; the input would consist of labeled examples, and the output would be a vector containing a set of predictions. Usually, these predictions are expressed asa probability distribution for all labels [7]. Other definitions of ANN emphasize on other aspects such as its processing properties: “A massively parallel distributed processor made up of simple processing units that has a natural propensityfor storing experiential knowledge and making it available foruse.” [37]. However, an ANN might not necessarily be massively. Small implementations are made just for the sake of trying new ideas. Engel Brecht provided a with a different intention, more focused on the topology: “It is a layered network of artificial neurons. An artificial neural net- work may consist of an input layer, hidden layers, and an output layer. An artificial neuron is a model of a biological neuron.”

ISSN: 2233-7857 IJFGCN

8. FlowChart

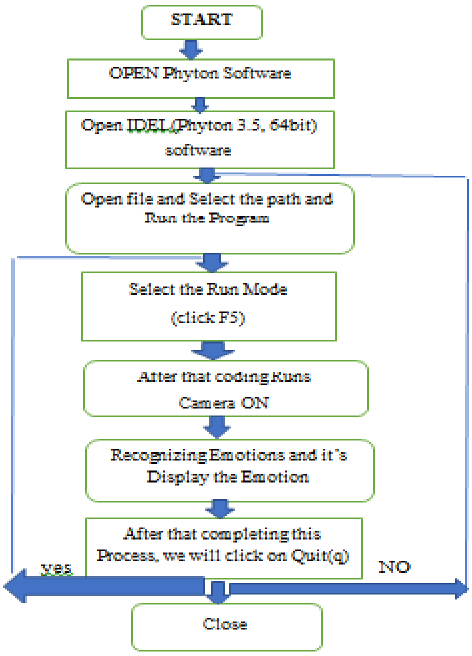

Fig 8.1: Flow chart

9. FutureScope

Further improvement on the network's accuracyandgeneralizationcanbeachieved through the following practices. The first one is to use the whole dataset during the optimization. Using batch optimization is more suitable for larger datasets. Another techniqueistoevaluateemotionsonebyone. This can lead to detect which emotions are more difficult to classify. Finally, using a larger dataset for training seems beneficial. Nowadays, optimizers work on a stochastic basis because of large datasets (millions of examples). However, this was true for our project. Given a limited dataset, trying on the whole data set could have lead to a better feature learning. Also, the use of some optimizers reported on this research would have had a different behavior. This behavior can be displayed on the loss curve having a smoother shape or by avoiding an early convergence. Second, due to time constraints, it was not possible to evaluate each emotion. On this way, it would have been possible to detect which emotions are easier to classify, as well as, which ones are more difficult. Moreover, pre-training on each emotion could lead to a better feature learning. After that, the network could have received this learning (transfer learning). This could have improved on reducing the training time; as well as, minimizing

10. Results

In this chapter, experimental results are displayed. A set of baselines are presented to better understand the performance of the experiments compared to other publications.

Baselines

In this section, a set of baselines are introduced.Thefirstofthemexhibitstheac- curacy of a random guess proposal. The following baselines were extracted fromthe EmotionRecognitionintheWildChallenge and Workshop (EmotiW) [23]. More- over, a researchpresented in Related Work section is introduced as a state-of-the- art baseline. All the numbers accounted on this section refers to the corresponding model pre- diction accuracy. Random guess A random guess computed over the 7 labels provided by CK+

ISSN: 2233-7857 IJFGCN

gives a prediction performance of 14.27%. This result is equivalent to calculate anuniform probability for each label. As it was mentioned before, the contempt emotion was removed from the dataset. In that case, thebaselinefor6labelsonlyis16.67%.The most common label in CK+ is surprisewith 83sequences,ifwewouldassignalwaysthis label, thepredictionperformancewouldbeof25.38%. The same approach for the 3most common labels (surprise, joy and disgust) would have resulted in 64.5% of prediction accuracy.

Program

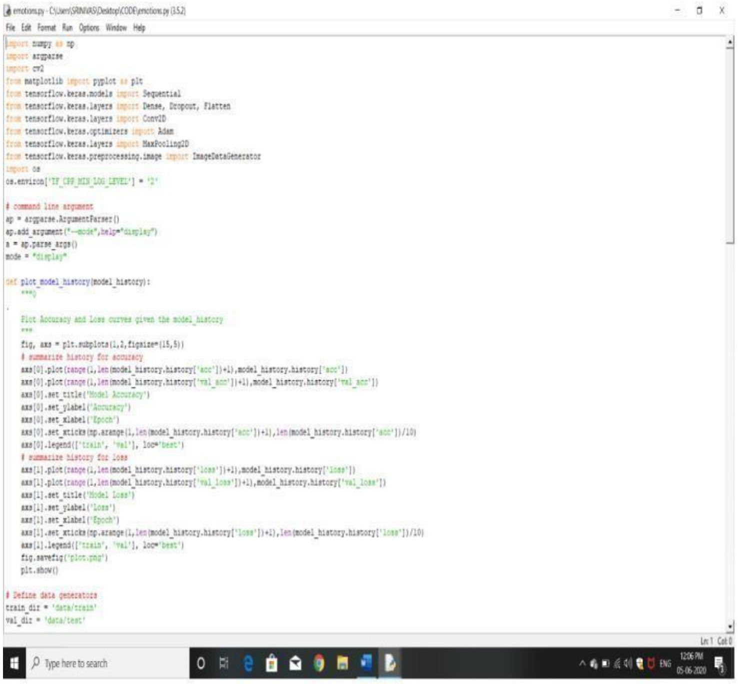

Fig10.1 Existing code

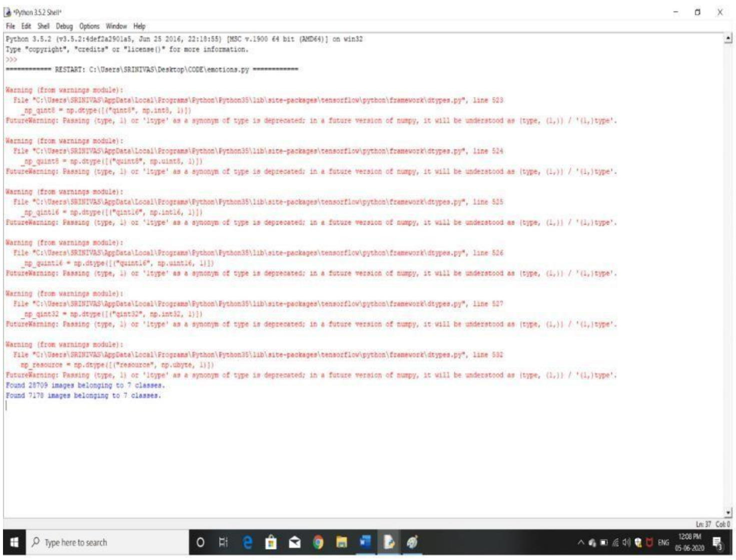

Fig10.2 Extension code

Experimental Results

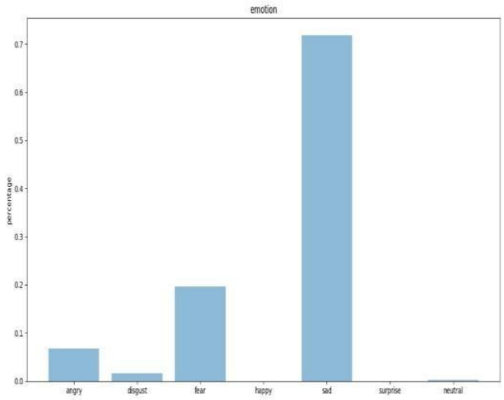

Fig 10.3 Analysis if the expressions

Fig 10.4 Happy expression

Fig 10.5 Surprised expressions

Fig 10.6 Angry expressions

Fig 10.7 Fear full expressions

References

[1] TensorFlow. Convolutional Neural Networks. https://www.tensorflow.org/versions/r0.8/tutorials/deep_cnn/index.html. Accessed:2016-05-15.

[2] TimeDettmers.Deep learning in a nutshell: History and training.https://devblogs.nvidia.com/parallelforall/ deep-learning-nutshell-history- training, 2015. [Online; Accessed 24 May2015].

[3] Andrej Karpathy. Stanford university: Cs231n convolutional neural networks for visual recognition.
http://cs231n.github.io/neural-networks-1,2016. Online; Accessed 07 June 2015. [4]AmazonWeb Services, Inc. Ec2 instance types. http://aws.amazon.com/ec2/instance-types/, 2016. [Online; Accessed27
May 2016].

[5] Andrej Karpathy. Cs231n convolutional neural networks for visual recogni- tion. http://cs231n.github.io/convolutional- networks/, 2015. [Online; Accessed 24 May 2016].

[6] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static fa- cial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conferenceon,pages2107–2112.IEEE,2016.

[7] Matthew Turk. Perceptive media: machine perception and human computer interaction. CHINESE JOURNAL OF COMPUTERS- CHINESE EDITION-,23(12):1235–1244, 2016.

[8] Vandal, T., McDuff, D., & El Kaliouby, R. Event detection: Ultra large- scale clustering of facial expressions. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International ConferenceandWorkshopsonVol.1,pages 1–8,2016.

[9] Jun Wang and Lijun Yin. Static topographic modeling for facial expres- sion recognition and analysis. Computer Vision and Image Understanding, 108(1):19–34,2017.

[10] Paul J Werbos. Applications of advances in nonlinear sensitivity analysis. InSystemmodelingandoptimization,pages 762–770. Springer,1982.2018

[11] Paul J Werbos. Backpropagation throughtime: whatitdoesandhowtodoit. ProceedingsoftheIEEE,78(10):1550–1560, 1990,2018

[12] Ren Wu, ShengenYan, Yi Shan, Qingqing Dang, and Gang Sun. Deep image: Scaling up imagerecognition. arXiv preprintarXiv:1501.02876,22:388,2018.

[13] YaserYacooband Larry Davis. Computing spatio-temporal representations of human faces. In Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference On, pages 70–75. IEEE,2018.

[14] OmryYadan, Keith Adams, Yaniv Taigman, and MarcAurelioRanzato. Multi- gpu training of

ISSN: 2233-7857 IJFGCN

convnets. arXivpreprint arXiv:1312.5853, 9,2018.

[15] Matthew D Zeiler. Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701,2019.

[16] Matthew D Zeiler and Rob Fergus. Visualizingandunderstandingconvolutional networks. In Computer vision–ECCV 2014, pages 818–833. Springer,2019.

[17] Matthew D Zeiler, Marc'AurelioRanzato, Rajat Monga, Min Mao,Kun Yang, Quoc Viet Le, Patrick Nguyen, Alan Senior, Vincent Vanhoucke, Jeffrey Dean, et al. On rectified linear units for speech processing. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pages 3517– 3521. IEEE, 2019.

Author details

Mr.VADAPALLI SRINIVASAGOPAL.
My Roll N018221D7001.Iam Doing M.TECH in
the Department of Electronics and Communication Engineering,
BVC Engineering College (AUTONOMOUS),Odalarevu.
v.srinivasgopal@gmail.com.

Dr.K.B.S.D.SARMA working as a Professor in the Department of Electronics and Communication Engineering, BVC Engineering College (Autonomous), Odalarevu. He is published 30 paper publications in various national and International journals and conferences. His research interest includes Signal Processing and VLSI System designs and its Applications.

Dr.K.B.S.D.SARMA
Professor,
BVC Engineering College (AUTONOMOUS), Odalarevu
bsdsarma.kompella007@gmail.com

Dr. Pravin Akula obtained his B. E Degree in Electronics and Communications Engineering from SRKR Engineering College and M. Tech degree in Instrumentation and Control systems from JNTUK, Kakinada, Ph.D. at PRIST University, Tanjavuru He is presently working as Professor in the Department of Electronics and Communication Engineering, B V C Engineering College (Autonomous), Odalarevu. He has 20 years of teaching experience, Registered for IPR Patent, published a Textbook on STBC, published more than 27 papers in international and national journals, attended more than 6 National and international conferences. His current research interests include VLSI Low Power. He is the Life Member of MISTE.

Dr. Pravin Akula
Professor,
BVC Engineering College (Autonomous), Odalarevu
akula.pravin@gmail.com