International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

FACE RECOGNITION THROUGH CONVOLUTIONAL NEURAL NETWORK

Dr.R.Prema
Asst.Prof (CSE)
SCSVMV University
Kanchipuram

R.Rahul
B.E(CSE)
SCSVMV University
Kanchipuram

S.S.K.Chaitanya
B.E(CSE)
SCSVMV University
Kanchipuram

ABSTRACT:

Face recognition involves identifying a face and checking whether the face already exists in the system or not. The proposed work is focused on capturing face, predicting emotion, finding out location and storing date and time when an image is captured. This involves usage of a deep learning technique by the name convolutional neural network. A CNN is built to process the image and predict an emotion from the face. A data file is created comprising the stated details for the further assessment. An emotion plays an important role in finding out how a person is acting when doing a particular thing. It sometimes conveys the intention of a person. Date and time makes us aware of the time when an image was previously captured. The location attribute is the location of the internet service provider of an area. Python programming language is utilized to implement the idea of constructing a face recognition system. This comes with a set of useful libraries to ease our task. It is observed that these face recognition systems take a lot of time in training and the proposed work focuses on reducing it to a certain extent and to recognize faces thus paving a way for an improved face recognition rate. Fer2013 dataset is used to train the CNN model and it involves training it with a face and an associated emotion. The shortcomings of the recognition systems which include difficulty in finding faces, spending a considerable amount of time in training and testing are considered to be improved. Some of the face recognition systems built using R-CNN and FRR-CNN exhibited recognition rates less than 80%. The proposed work intend to improve the face recognition rate a bit more than the existing ones.

Key words:- Face recognition, Convolutional neural networks , Python,Fer2013 dataset.

INTRODUCTION:

There is a huge demand for things that simplify people's lives. The modern man tries hard to build systems to achieve that objective. Face recognition mechanism is one of those things and with advancement in technology, people started using surveillance systems to keep track of what's happening in a particular area. They can be utilized to find out ones who commit robberies, crimes and other illegal things. There is no storage option to store a set of wanted people in the earlier systems. Face recognition system will do that and surveillance systems are embedded with face recognition software these days.

Face recognition mechanism includes many tasks – identifying a face in an image or a recording, storing the face for future retrieval and reuse, recognizing the face which is already existing in the system. People use them mainly for recognizing faces from those of listed ones. Some examples include providing access only to the authorized employees into their cooperate campuses, recording daily attendance of students in educational institutions and many other.

The proposed work is focused on building face recognition system with better functionality setting it apart from those of existing ones. It is not a basic one as it holds location and time. The movements of a particular person can be traced by implementing it in distributed monitoring systems. The emotion attribute is stored. This will yield better results and providea different perspective to think about and work on.

The work involve usage of convolutional neural network (CNN) algorithm as it proved to be better than its other variants R- CNN and FRR – CNN. The use of CNN will create a structure resembling the neural moments of a human brain. The connections that get established between the layers will complete the process of building a CNN. A proper care has to be taken regarding the data which our work wants to feed into the model and system. One can go for preprocessing or can use existing dataset.

The approach to detect a face differ from one to another depending on the techniques and algorithms involved. Some systems work based on frontal face, full face view and others consider eyes too. A classifier is created using cascades to do this task of finding out faces. A classifier based on its learning, classifies whether a data item belongs to a particular set or not. The classifier in the stated context tries to find out the face. To get better results, stress has laid down on considering best approaches and works. Accuracy should not be comprised upon.

The face recognition rate is seen as a metric to evaluate the work and the proposed model works quite well in complex background too. This is further accompanied by improvement in time taken during training and testing. Training phase is the most important step in building these systems and there shouldn't be any room

for doing wrong things as it will affect the whole thing. These things place the proposed model at a higher level to that of frequency redundancy reduced convolutional neural network (FRR – CNN) and faster regions with convolutional neural network (R – CNN).

LITERATURE SURVEY:

The authors of [1] proposed

The idea of using both image edge computing and convolutional neural network to avoid the complex process of explicit feature extraction. It starts with processing the data that is required in the due process. The three basic steps that are in involved in preprocessing are locating and cutting out a face from an image, normalizing that face image, equalizing the normalized one, extracting the edge of each layer of the image. They stated the use of AdaBoost algorithm and Haar classifier. After this preprocessing step, all images will be in gray scale.

To build a network that recognizes face expression, they opted convolutional neural network. The structure for recognition contains convolution layer, pooling layer, full connection layer and soft max layer. The tasks that are carried out in the above stated layers include expressing an image in a numerical matrix form without disturbing the actual meaning reducing the number of parameters so as to increase recognition rate, creating a connection between existing layers to enable interactions and finding out class to which an expression belong.

After its build, the model is trained and tested to assess its working and efficiency. The results turned out to be quite satisfactory interns of training time, test time and recognition rate. They went on to state that their mechanism will provide a recognition rate of 88.56% and it is 1.5 times faster in training compared to the existing ones.

The authors of [2] proposed

They proposed the idea of deep locality preserving convolutional neural network (DLP – CNN) method. They used the existing data base of images by the name real world affective face data base. The main objectives stressed out by this work include enhancing the label estimation's readability, to assess difference between emotions captured under controlled and uncontrolled conditions, proposing a novel deep learning framework as an alternative to existing CNN based expression recognition and considering real world entities to work on and get appropriate result.

The mechanism involved includes choosing appropriate dataset as the further steps involve finding out expression, there should be no ambiguity in doing so. The dataset is an alternative to captured images as the latter brings in complexity's retaining to color, back ground to name a few. The proposed work used LBP and haar feature extractions methods. They are used to extract face feature. For face expression classification, support vector machine, near neighbor based classifiers are used.

The RAF – DB contains emotions changing from simple to complex the creation of those databases involved data collection, database annotation, getting metadata, reliabilityestimation and partitioning subsets. There have been a couple of algorithms proposed and one is optimization algorithm. The experiment is carried out with using pre – trained model Alexnet, deep learning framework caffe and training is done on NVIDIA Tesla K40 GPU. The work provided the research community a new real facial expression data base.

The authors of [3] proposed

A mechanism considering that there have been few sources available for data gathering to train models. Their work gives better results even though the training face tests. For a considerable time with few inputs they also tried to remove expression irrelevant factors from the input images so that the expression recognition task will become much easier. The outcomes have suggested that there is an improvement of 1.72% and 1.11% in accuracy over it's to competitivealgorithms. In building the system synthesizing frontal face and sampling key regions are involved. The spatial features are thus extracted and assessed. One point to mention hear is that unsupervised feature learning algorithm is encoded. The normalization face involvesconverting images into grayscale and localizing various landmarks so as to correct context thus paving a way to achieve better results. As in many other works, this is to create its own dataset by preprocessing the required images. The outcome of the work has been compared to other feature learning algorithm namely RAW and LWP it is evident from their work that their proposal achieved more accuracy than those existing. The various expressions consider include anger, disgust, fear, happiness, sadness, surprise. In the above stated each expression, it performed better. By proper adjustment, this work will also be able to explore the pose- invariant recognition.

The authors of [4] proposed

A collective work which throws a light on automated face analysis. It will be quite overwhelming when one thinks of it at first. It involves implementing those structure across an area to monitor people. There

have been many proposal stating what to consider in building a face recognition system. One can consider face attributes that include age, gender, race and ethnicity. The works include facial landmark detection in videos, gaze, redirection, detection of moods of variation, head pose estimation and tracking. This ones will work on existing datasets and if you created their own datasets by utilizing available resources. The datasets vernissage dataset, gaze, redirection data set, synthetic images are consider all the stated works areprimarily focused on wild imaginary which means that the real world scenarios worked upon.

EXISTING SYSTEM:

The existing systems use various approaches which involve neural networks: R-CNN and FRR-CNN. The thing with these ones are they are not a go to solution for our purpose. They serve a part but not whole theme of our work. They are considered to work decent in ideal conditions but when put to work in complex backgrounds, they turn out to be quite moderate. By adopting these methodologies, incompleteness will arise. The face recognition rates of R-CNN and FRR-CNN based systems are 79. 34% and 70. 63% respectively.

The various problem posed by existing models include taking a lot of time to train and test the model, finding it difficult to find out faces when subjected to complex backgrounds, incompleteness will be created by the inclusion of a lot of methods in the process of designing a system. There have been instances where researchers propose a new methodology, coming age algorithms to produce better results, but their accuracy has to be assessed upon to know how well it is working.

DRAWBACKS:

- A considerable number of face recognition mechanisms spend a lot of time in training and testing the model.
- After all this tedious task of building a system, the face recognition rate is not quite satisfactory.
- An incompleteness will be created in the process as it involves usage of various techniques and methodologies.
- There is no go to solution for our stated objective, if it is there, the face recognition rate may not be impressing, if it satisfies that also then it may not come at free of cost.

PROPOSED METHOD:

Haar cascade is a classifier that can be trained to efficiently recognize intended objects. There are various haar xml files for detecting frontal face, eye, full body, smile and many other. For the project, the frontal face file is used. All these things are not available beforehand. They come under opencv-python/cv2 module. One can get that module by running the following command in the Windows command prompt admin: 'pip install opencv-python' (Internet required). The stated haar xml files will be present in the location: 'Python38 > Lib > site- packages > cv2 > data'. To use the installed module, one have to import it as 'import cv2'. The desired file can be used in the program by using 'Cascade Classifier' class. The training is done to cascade with frontal face file, so that it will detect faces in an image. Further, the task wenton to crop those faces which can later be used for detecting emotions.

To correct uneven distribution of light and shade, one can adjust the gray level of the image to enhance the contrast of the image. One can use 'cvtColor' and 'equalizeHist' methods to get done that thing. The resultant images can be used for training by labeling them with an expression. The same can be used for testing. This image preprocessing is apt for a couple of images but to get best efficiency our model should be trained with thousands of images. In that case, it will become a tedious task. That's why it is appropriate enough to use 'fer2013' dataset. Fer2013 is a facial expression recognition dataset which consist 35887 training and testing images. As previously said, to overcome the tedious task of formatting images one by one, one can choose to use this dataset. The images in this dataset are labeled with seven different emotions- angry, disgust, fear, happy, sad, surprise and neutral. Almost all the images will be centered and occupies the same amount of space in each image with 48x48 size. One can customize the provided size to one's requirement. The task of using/accessing fer2013 can be done in a few different ways. By default, fer2013 is a csv file. One can access it using pandasor as a normal file as 'pandas.read_csv('fer2013.csv')', 'with open ('fer2013.csv')'. When one tries to display the contents of the file, they get displayed in the form of three columns labeled as emotion, pixels and usage. The actual images are represented in the form of corresponding pixel values.

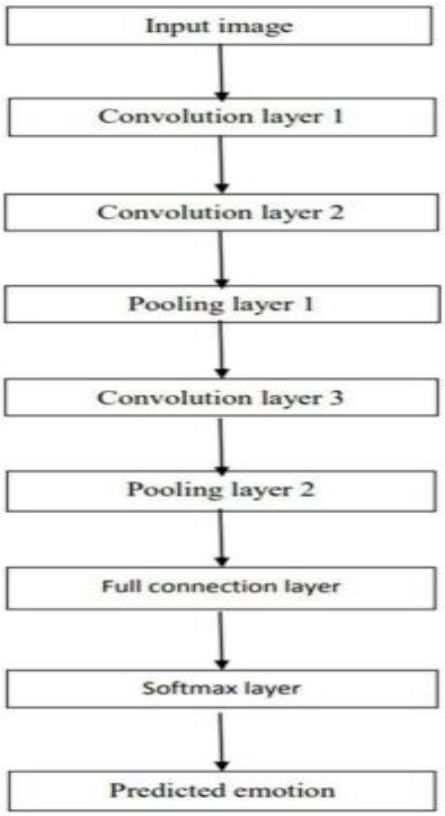
Building a CNN:

The first layer in the proposed network is the convolution layer. In this layer, an image is represented in a matrix of pixels. By the end, a particular field in that matrix will be brought down to a single cell in another matrix. This process is called convolution. It involves convolution of pixels from receptive field (input) to summed up pixels in filter (output). In normal terms, one can say that the values/ fields in the matrix are reduced without disturbing the original meaning. The number of values at the end depend on the size of

filter. The convolution layer is a part of convolutional neural networks (CNNs). The other layers in CNN are pooling layer, full connection layer and softmax layer.

To build a CNN, one has to make use of relevant libraries. Here, the libraries which will be using are tensorflow and keras. Keras works on top of tensor flow and provides various models and layers to help us build CNN. All the stated layers of CNN have predefined methodsin keras. Since the input for our network is two dimensional, our work used Conv2D layer of keras to process two dimensional pixel values. There are a total of three convolution layers in our network. They differ from one another in the usage of convolution kernels as they use 32,64 and 128 to generate the same number of feature maps respectively.

The activation function can be passed as a parameter to Conv2D layer or can be specified separately. Our work also specified the number of feature maps which one wants to get indirectly by stating convolution nuclei. There are a lot more parameters which one can pass to Conv2D but there are quite enough to carry out our intended purpose. Our work used MaxPooling2D by specifying pool size within it to carry out the task intended by pooling layer. By doing so, the number of outputs are reduced. This results in inputting a minimized output to the following convolution layer in the network. It has to be mentioned that at all times, a pooling layer is placed only after a convolution layer.



ADVANTAGES:

- Automatically detects the important features without any human supervision.
- Gives good accuracy.
- Computationally Efficient

IMPLEMENTATION:

The work has been implemented using Anaconda Jupyter and Python language. Anaconda Jupyter is an IDLE used to run python scripts. It can run online as well as offline. It comes with a set of useful packages installed. Python is a general purpose, interpreted language. Our work involved choosing Python because of its rich set of libraries which will ease our work a little bit in implementing the idea. Coming to developing code snippets, the whole work is carried out in Anaconda Jupyter Notebook. In the project timeline, many useful libraries are installed which were necessary to perform a particular task. There have been issues while doing that and with proper research and solving issues with the help from internet community, it became possible to complete the work.

The following text describes the various libraries which were used in our work

opencv-python/cv2:

It is used for image processing. It makes use of numpy library in carrying out certain operations. Our work used cv2 to enable video recording and to capture images accordingly. A set of appropriate classes and methods are used to process images.

numpy:

It is mainly used for matrix/array related tasks and operations. Using numpy, an image is represented in a numerical matrix without distributing the actual meaning. With proper methods, the image can be produced back from those numbers. Few matrices were even flattened (converting to a row matrix).

pandas:

It is used for data analysis and manipulation. It is used to access a CSV (comma separated values) file i.e., Fer2013 dataset. Its queries resemble those of a database management system's 18 query languages. It is a powerful library with a rich set of methods. It doesn't have much scope in the proposed work and is used only for accessing a data file.

datetime:

As the name suggests, it is used for tasks and operations involving date and time. There are various formats in which one can display and use those details. The use of datetime library made it possible to store date and time when an image is captured.

geolocation:

This library is used to find out the location of a device on which an image is captured. It results in getting location coordinates expressed in latitude and longitude. The thing with this library is that it will display the location of internet service provider by default.

glob:

It is used to traverse through files in a directory. One of the tasks involve going through each and every file/image at some point in the project execution for checking whether an image is present or not.

Tensorflow:

It is extensively used for dataflow and differentiable programming. As it is used for implementing neural networks, it has been considered to be embedded in the code implementation. It is involved in classifying, percepting, understanding, discovering, predicting and creating related tasks.

keras:

It is a neural network library and it is used to run on top of tensorflow. It enables faster and efficient creation and implementation of neural networks. The code in the appendix will list all the various classifiers used in building a classifiers used in building a convolutional neural network. The proposed work utilized various two dimensional matrix implementing layers namely conv2D and MaxPooling2D.

The data collected at the end of this process will be quite useful. There are no issues concerned in storing data. The objective of recognizing a force is achieved by giving enough scope to emotion, date, time and location. These are the other things which one has to worry a little about as they too are implemented and included in the dataset. This location field has to be worked on as it results in displaying a standard address across an area. Considering few useful methods, it is possible to find that one. With incorporation of other techniques, it may even be possible to live track a person. An enthusiast working on face recognition can use the model as it is or can customize the methodology so as to include new functionalities.

CONCLUSION:

The results produced by the model are quite satisfactory and it is able to detect emotions in a better way when subjected to complex backgrounds than its counterparts. Along with that, it is able to create and populate a dataset with the attributes face, emotion, date, time and location.It doesn't create any irregularities and imbalances while accessing and same is true that one can adopt this method so as to get quick and efficient results. The proposed work is a stand- alone task. It exhibits some meaning in the way as it is now. It can be embedded in complex systems for recognition related tasks. There are a few sources available which provide fully functional systems and it is difficult to find all things at one place.

The proposed work comprises a meaningful objective. A distributed monitoring system can be built using the proposed work. It can be installed in existing systems to get efficient results. On top of this work, some other functionalities have to be added to achieve that. Onecan adopt the proposed work as it is if they want to use a face recognition system or can addfew other things to build a customized system for one's

requirements. In coming days, with the entry of newer and better algorithms and techniques, the proposed work will find its competitor (or) the existing ones can be improved.

REFERENCES:

[1] H. Zhang, A. Jolfaei and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing", IEEE Access, 2019.

[2] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition", IEEE Trans. Image Process., vol. 28, no. 1, pp. 356–370, Jan. 2019.

[3] F. Zhang, Q. Mao, X. Shen, Y. Zhan, and M. Dong, "Spatially coherent feature learning for pose-invariant facial expression recognition", ACM Trans. Multimedia Comput., Commun., Appl., vol. 14, no. 1s, Apr. 2018, Art. no. 27.

[4] S. Escalera, X. Baró, I. Guyon, H. J. Escalante, G. Tzimiropoulos, M. Valstar, M. Pantic, J. Cohn, and T. Kanade, "Guest editorial: The computational face", IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 11, pp. 2541–2545, Nov. 2018.

[5] S. Li and J. Du, "Reliable crowdsourcing and deep locality – preserving learning for expression recognition in the wild", in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2584 – 2593. IEEE, 2017.

[6] E. Sariyanidi, H. Gunes and A. Cavallaro, "Learning bases of activity for facial expression recognition", IEEE Transactions on Image Processing, 2017.

[7] H. Ma and T. Celik, "FER-Net: Facial expression recognition using densely connected convolutional network", Electron. Lett., vol. 55, no. 4, pp. 184–186, Feb. 2019.

[8] Q. Mao, Q. Rao, Y. Yu and M. Dong, "Hierarchical Bayesian theme models for multipose facial expression recognition", IEEE Trans. Multimedia, vol. 19, no. 4, pp. 861 – 873, Apr. 2017.

[9] N. Kumar and D. Bhargava, "A scheme of features fusion for facial expression analysis: A facial action recognition", J. Statist. Manage. Syst., vol. 20, no. 4, pp. 693–701, 2017.

[10] S. Xie and H. Hu, "Facial expression recognition with FRR – CNN", Electron. Lett., vol. 53, no. 4, pp. 235 – 237, Feb. 2017.

[11] B. Mass e, S. Ba, and R. Horaud, "Tracking gaze and visual focus of attention of people involved in social interaction", IEEE Trans. Pattern Anal. Mach.Intell., 2018.

[12] J. He and X. Lu, "Multi – task mid – level feature learning for micro – expression recognition", Pattern Recognition, 66 : 44 – 52, 2017.

[13] T. Song, W.Zheng, C. Lu and Z. Cui, "MPED: A multi – modal physiological emotion database for discrete emotion recognition", IEEE Access, vol. 7, pp. 12177 – 12191, 2019.

[14] J. Booth, A. Roussos, G. Trigeorgis, E. Antonakos, S. Ploumpis, Y. Panagakis, and S. Zafeiriou, "3D reconstruction of 'in-the-wild' faces in images and videos", IEEE Trans. Pattern Anal. Mach. Intell., 2018. 28.

[15] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions", IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 1, pp. 96– 105, Feb. 2006.