Final Year Project

Emotion Recognition By Inclusion Of Age And Gender Parameter By Using Deep Learning

**Literature Survey:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Title** | **Methodology** | **Conclusion** | **Future Scope** |
| **Facial emotion recognition using deep learning: review and insights** | They proposed deep CNN for FER across several available databases. After extracting the facial landmarks from the data, the images reduced to 48x 48 pixels.  The architecture used consist of two convolution-pooling layers, then add two inception styles modules, which contains convolutional layers size 1x1, 3x3 and 5x5  They propose a novel CNN for detecting AUs of the face. For the network, they use two convolution layers, each followed by a max pooling and ending with two fully connected layers that indicate the numbers of AUs activated.  For the detection of the essential parts of the face They used three CNN with same architecture each one detect a part of the face such as eyebrow, eye and mouth. Before introducing the images into CNN, they go through the crop stage and the detection of key-point facial. The iconic face obtained combined with the raw image was introduced into second type of CNN to detect facial expression. Researchers show that this method offers better accuracy than the use raw images or iconize face alone | This paper presented recent research on FER, allowed us to know the latest developments in this area. We have described different architectures of CNN and CNN-LSTM recently proposed by different researchers, and presented some different database containing spontaneous images collected from the real world and others formed in laboratories | FER are one of the most important ways of providing information about the emotional state, but they are always limited by learning only the six-basic emotion plus neutral. It conflicts with what is present in everyday life, which has emotions that are more complex. Researchers are now pushing their research to create and offer powerful multimodal deep learning architectures and databases |
| **Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks** | Our first deep model is trained on a large dataset of four million images for the task of face recognition. This model serves as the backbone to our facial attribute recognizers and is used to fine-tune networks for four tasks: real age estimation, apparent age estimation, gender recognition and emotion recognition we start by training a deep neural network for the task of face recognition using four million images of over 40, 000 identities This model serves as the backbone of our facial attribute recognition engine. We designed a highly optimized deep network architecture for accuracy and speed for each task. | this paper, we present an end to end system for age, gender and emotion recognition. We show that our novel deep architecture, along with our large, in-house collected data, can outperform competitive commercial and academic algorithms on several benchmarks |  |
| **Understanding and Comparing Deep Neural Networks for Age and Gender Classification** | One choice to be made for training and classification is regarding data preprocessing. The SVM-based system from improves upon by introducing a 3D face frontalization preprocessing step  In this paper, we compare models initialized with random weights to models starting with weights trained on other data sets, namely the ImageNet data set and the IMDB-WIKI data sets  The method operates iteratively from the model output to its inputs layer-by-layer in a backpropagation-style algorithm, computing relevance scores Ri for hidden units in the interim | In this paper we opened the black-box classifier using Layer-wise Relevance Propagation and investigated which facial features are actually used for age and gender prediction. We compared different image preprocessing, model initialization and architecture choices on the challenging Adience dataset and discussed how they affect performance |  |
| **Age and Gender Classification using Convolutional Neural Networks** | Age classification: Early methods for age estimation are based on calculating ratios between different measurements of facial features Once facial features (e.g. eyes, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and used for classifying the face into different age categories according to hand-crafted rules  Gaussian Mixture Models (GMM) [13] were used to represent the distribution of facial patches  Gender classification: used a neural network trained on a small set of near-frontal face images. In [37] the combined 3D structure of the head (obtained using a laser scanner) and image intensities were used for classifying gender. SVM classifiers were used by [35], applied directly to image intensities. Rather than using SVM, [2] used AdaBoost for the same purpose, here again, applied to image intensities. Finally, viewpoint-invariant age and gender classification was presented by [49]. | Two important conclusions can be made from our results. First, CNN can be used to provide improved age and gender classification results, even considering the much smaller size of contemporary unconstrained image sets labeled for age and gender. Second, the simplicity of our model implies that more elaborate systems using more training data may well be capable of substantially improving results beyond those reported here. |  |
| **Methods for emotions, mood, gender and age recognition** | This article offers an analytical review of the following software solutions in the field of recognition of emotions, mood, sex and age of a person.  Emotion recognition using Deep Convolutional Neural Network. A Compact Soft Stage wise Regression Network. Real-time Convolutional Neural Networks for Emotion and Gender Classification Age Recognition using CNNs. | As a result, we can conclude that such tasks as the recognition of emotions, mood, gender and age are very popular among researchers all over the world. Enthusiasts use different approaches to the implementation of intelligent systems that can solve such problems and achieve good results in accuracy of recognition, even with limited resources. The main means of the implementation of the tasks are convolutional neural networks of various architectures, trained in well-known in the network dataset images. |  |
| **Emotion recognition by inclusion of age and gender parameters with a novel hierarchical approach using deep learning** | The proposed method is to find the age, gender and emotion of a person standing in-front of the camera using the hierarchical approach. It first detects the face of the person in front of the camera using YOLOv2-tiny model (You Only Look Once i.e. used to detect the objects in real time) [5], then passes each frame as the input to the network architecture which consists of several deep neural networks which are trained on different type of publicly available datasets. This method is implemented in 2 modules, the first module is used for detection of emotion. The output of YOLOv2-tiny is used as input for emotion module. The input image is passed through different type of hidden neural networks which are used to map the features of the face and gives the output using softmax layer. The output of emotion module is stored in the HDF5 file (Hierarchical data format version 5) and is fed as input for the second module which is implemented for age and gender. The output from age gender module is the final integrated result of both the modules which shows the age, gender along with the emotion of the input data. | In this paper, the recognition of emotion, gender and age has been attempted in real time video stream. The proposed method of using the general architecture of CNN  along with the squeeze net and Xception architecture combined in a hierarchical approach has been successfully implemented. |  |
| Facial age estimation based on label-sensitive learning and age-oriented regression | Facial age estimation based on label-sensitive learning and age-oriented regression  Instead of applying the global regression techniques for age determination, we propose to utilize local regression mainly for two reasons: Local regression can generate comparably sophisticated mapping functions; hence is more likely to capture the complicated facial aging | In this paper, a new age estimation approach considering the intrinsic factors of human ages is proposed. After feature extraction, RCA is utilized to achieve a suitable space for neighbor searching. Then based on this adjusted space, LPP and MFA are trained to drastically reduce the feature dimensionality and learn the connections between features and age labels. To further consider the ordinal relationship of human ages as well as the imbalanced learning problem in RCA, LPP, and MFA  . |  |
| Automatic Age Estimation Based on Facial Aging Patterns | The aging function based methods regard age estimation as a conventional classification problem: the data are the face images, the target is their age labels.  A representative model for the aging patterns can be built up by the information theory approach of coding and decoding. One widely adopted technology is using PCA [10] to construct a subspace that captures the main variation in the data set. . Based on the characteristics of aging patterns, an EM-like algorithm is proposed here to learn a representative subspace.  During the age estimation process of AGES, the proper aging pattern for the test image is generated based on both the aging pattern subspace and the face image feature. The subspace defines the general trend of aging, and the face image feature represents the personalized factors. By placing the feature vector at different positions, candidate aging patterns specified to the test face are generated. Among these candidates, only one is consistent with the general aging trend, which can be detected via minimum reconstruction error by the aging pattern subspace. At the same time, the position of the test image in that aging pattern can be determined. | This paper proposes an automatic age estimation method named AGES, which improves our earlier work [5]. It is interesting to note that, at least under the experimental configuration in this paper, the performance of AGES is not only significantly better than that of the state-of-the-art algorithms, but also comparable to that of the human observers |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Emotion Detection and Characterization using Facial Features | The detection and recognition implementation proposed here is a supervised learning model that will use the one versus-all (OVA) approach to train and predict the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise).  The overall face extraction from the image is done first using a Viola-Jones cascade object face detector. The Viola Jones detection framework seeks to identify faces or features of a face (or other objects) by using simple features known as Haar-like features.  Once the faces are detected, they are extracted and resized to a predetermined dimensional standard  Next, the mean image for all training faces will be calculated. The entire training set is comprised of faces  The mean image is then subtracted from all images in the training set. Then using the mean-subtracted training set the scatter matrix S is formed  This allows us to proceed to the Fisher linear discriminant analysis (LDA) in a reduced dimensionality. For each emotion that we wish to train a predictor for, we will perform Fisher LDA, in which the goal is to optimize the objective function that minimizes within class variance and maximizes between class variance to gain clear class separation between the class of interest and the other classes. | The predictor is relatively successful at predicting test data from the same dataset used to train the classifiers. However, the predictor is consistently poor at detecting the expression associated with contempt. This is likely due to a combination of lacking training and test images that clearly exhibit contempt, poor pre-training labeling of data, and the intrinsic difficulty at identifying contempt. The classifier is also not successful at predicting emotions for test data that have expressions that do not clearly belong exclusively to one of the seven basic expressions, as it has not been trained for other expressions. | Future work should entail improving the robustness of the classifiers by adding more training images from different datasets, investigating more accurate detection methods that still maintain computational efficiency, and considering the classification of more nuanced and sophisticated expressions. |
| **Eye-Tracking Analysis for Emotion Recognition** | The purpose of the presented  research is to analyze whether it is possible to recognize  emotions by using eye-tracking signal.  ,tree classifiers were tested: SVM,  LDA, and k-NN. ,e leave-one-subject-out method was  used to assess the quality of the classification. Individual,  successive stages of the research were the following:  (i) acquisition of eye-tracking data while evoking  emotions  (ii) performing signal preprocessing  (iii) removal of the effect of luminance on pupil width  (iv) calculation of eye-tracking features related to eye  movements and pupil width  (v) classification of emotions using SVM, LDA, and k-  NN | ,e classification results confirm the ability of recognizing  emotions using eye movement features and pupil diameter. ,e  use of an eye-tracking requires, however, the elimination of  factors that may affect this classification. ,ese factors include  the effect of luminance on changes in the pupil diameter. It  should be ensured that the lighting conditions remain the same  throughout the experiment. ,e influence of luminance of the  presented materials should also be compensated. , is could be achieved by implementing appropriate methods, such as regression or principal component analysis  . |  |
| CNN BASED DETECTION OF EMOTION, AGE, GENDER | Normalizing the data between 0 and  We use 3 layers of convolution. for  each layer, we do Batch Normalization,  RELU activation function and use  Max Pooling. In fully connected layer we use RELU activation function and SOFTMAX function. By using Adam optimizer we calculate the loss function. To use the trained model later, save  the weights | Training the mannequin at greater epoch cost yields Better result. We have used resent structure as an alternative of VGG16 for age or inception v3 for  Gender classification. Resent helps in coping with the  Education error generated as the networks get deeper.  We have been in a position to obtain 95% accuracy  Rate. Replacing VGG-16 layers in Faster R-CNN  with Resent, we can study a relative enchantment of  28%. | 1.Further, this mannequin can be used to classify  Sufferers and their drugs based totally on age groups.  2.It can additionally be used for film advice gadget  in order to predict the frequency of a variety of age  organizations who come to watch films the most and  additionally to classify films into distinct genres  Primarily based on the viewers’ emotion. |
| Detection of Gender, Age and Emotion of a Human Image using Facial Features | we have the tendency to begin with downloading the image-set from a  dataset referred to as IMDB WIKI as a result of it being the Largest publicly available dataset with gender and age labels for training. Simultaneously, image-set is downloaded From a dataset referred to as FERC-2013. | This paper presented recent research on FER, allowed us to know the latest developments in this area. We have  described different architectures of CNN and CNN-LSTM recently proposed by different researchers, and presented  some different database containing spontaneous images collected from the real world and others formed in laboratories  (SeeTable.1), in order to have and achieve an accurate detection of human emotions. We also present a discussion that  shows the high rate obtained by researchers that is what highlight that machines today will be more capable of  Interpreting emotions, which implies that the interaction human machine becomes more and more natural. | FER are one of the most important ways of providing information about the emotional state, but they are always  Limited by learning only the six-basic emotion plus neutral. It conflicts with what is present in everyday life, which  Has emotions that are more complex. This will push researchers in the future work to build larger databases and create powerful deep learning architectures to recognize all basic and secondary emotions. Moreover, today emotion  Recognition has passed from unimodal analysis to complex system multimodal. Pantic et Rothkrantz show that  Multimodality is one of the condition for having an ideal detection of human emotion. Researchers are now pushing  their research to create and offer powerful multimodal deep learning architectures and databases, for example the  fusion of audio and visual studied by Zhang et al.and Ringeval et al. for audio-visual and physiological  Modalities. |
| SAF- BAGE: Salient Approach for Facial Soft-Biometric Classification - Age, Gender, and Facial Expression | The proposed approach uses the pre-trained weights of SALICON [35] and ImageNet [3] for saliency prediction and soft-biometric classification respectively. The fine-tuning is accomplished on a workstation with Intel Xeon core processor and accelerated by NVIDIA Titan 12 GB GPU. All experiments run in Tensor flow 1.6 | Direction for training a neural network, that using multiplied reweighted salient image, the performance could be further improved for facial soft-biometric classification. Also from the Figure 2, it can be seen that the salient regions - Eyes, Nose, and Mouth, these three are the dominant attributes which help the model to classify facial soft-biometric. Further, using Class Activation Maps (CAM) |  |
| Face Recognition with Age, Gender and Emotion Estimations | For training and testing images, we use the Haarcascade classifier in OpenCV to get the position of the face in the face detector. The earliest Haar features were proposed by Papageorgiou C. et al.Then, Paul Viola and Michal Jones proposed a method for quickly calculating Haar features using the integral image method. Later, Rainer Lien hart and Jochen Maydt extended the Haar signature library with diagonal features. The Haarcascade classifier in OpenCV is based on the extended feature library. | Through the face information, the system we proposed  can correctly detect faces in real-time, and estimate the  Age, gender, and emotion. Applied in retail industries, it  can help merchants build their own databases and Ana-  lye the shopping preference so that to adjust the pur-  Chase list and operation mode timely. Moreover, through emotion feedback, retailers can catch user experience to  Improve service quality. | At the same time, there are several considerations for  Future work. First of all, the prediction accuracy needs  To be improved, specifically the emotion estimation. Secondly, we want to build a user interface for easy operation  And data management. Finally, enter member information to improve personalized service. |