Human Emotion Detection

Computer Vision & Deep Learning Project

Name: Muhammad Zubair Ahmed Khan

Matricola no.: VR509196

Date: 24-June-2024

Table of Contents

1. Introduction 3

2. Motivation.. 4-5

3. Objectives 5-6

4. Methodology 6-8

- 5.1 Data Loading and Preprocessing 6-7
- 5.2 Data Augmentation 7
- 5.3 Model Definition 7-8
- 5.4 Learning Rate Scheduler 8

5. Experiments and Results 9-11

- 6.1 Model Training 9
- 6.2 Model Evaluation 9
- 6.3 Training and Validation Loss 10
- 6.4 Detailed Evaluation 10
- 6.5 Qualitative Results 10-11

6. Conclusion 11-12

7. References 12

1. Introduction:

The Human Expressions Recognition Project aims to work on the deep learning ways to classify the facial expressions from different labels in orders of anger, disgust, fear, happiness, sadness, surprise, and neutrality. The project proceeds forward using Convolutional Neural Networks (CNNs) to determine the facial images and their displayed emotion. Facial expression recognition is in important in multiple fields for enhancing the human- computer relations, with the operations ranging from emotion analysis to perfecting the experience in interactive systems, security, and surveillance.

The project starts with the way loading of the data and preprocessing it afterwards, data augmentation, model building, training the model, and assessing its performance is processed. The dataset is comprised of images of faces with their different feelings, is preprocessed to regularize the input of the neural network. Data Augmentation approach has been adopted, such as rotations, zooms, shifts, and flips are used to increase the diversity of its training data, therefore it improves the robustness and generalization of the model.

The CNN model is used in multiple convolutional layers, batch normalization, pooling layers, and dropout layers for extracting and learning the important features from the images while preventing it to be overfitted. A learning rate scheduler is deployed to flexibly changes learning rate during the training, which ensures the optimal convergence to the model.

When we will start the training process, the data which is augmented is injected into the CNN model and adjusting the model parameters accordingly for reducing the loss function. The model's performance will be evaluated from a test set provided in our data set separately, for which we will use the metrics such as accuracy, precision, recall, and F1-score, including the confusion matrix and classification report of the model.

Subsequently, qualitative analysis is conducted by visualizing the model's prediction against the true label on the arbitrary test images. This helps us in understanding the strength of the model and further enhancement in our model. The project concludes with a summarization of results.

2. Motivation:

The motivation to start this project on Human Expressions Recognition Project has embedded from the significant role the facial expressions play in human behavior on the communication and interaction. It recognizes these expressions to accurately enhance the effectiveness and experience of a person in various applications.

- **Enhancing Human-Computer Interaction**

  Facial expression recognition model can improve the human-computer interaction (HCI) by enabling the computers to understand and respond to the human emotions accordingly. This is a crucial for the development of the responses the virtual assistants provide, interactive gaming experiences, and personalized user interfaces, thus it enhances the user satisfaction and engagement.

- **Applications in Mental Health and Well-being**

  Detection of emotion in the systems can monitor and analyze our facial expressions to detect to heads up our emotional vulnerabilities or mental health issues, facilitating timely interventions and support for our feelings. These systems can also be used in therapeutic to monitor the patient progress and adjust the treatments accordingly.

- **Security and Surveillance**

  For security and surveillance, Human expression can also be helpful in identifying suspicious behavior to indicate a potential threat. Integrating emotion detection into surveillance systems can enhance the public safety by providing security personnels with their additional insights into the emotional states of individuals.

- **Customer Experience and Marketing**

  Facial expression recognition can also be used in businesses to gain further insights into customer reactions and preferences of a particular product. Analysis of customer expressions can improve the product placement and its marketing strategies. In digital marketing, emotion recognition can also personalize be advertised and content can be generated according to the customers emotion status.

- **Advancements in Artificial Intelligence**

  By developing emotion detection, it can be used in AI to further enhance its complexities in the field of machine learning, image recognition and posture behavior. By integrating with the AI it can be used in large gathering to recognize suspicious behavior of an individual or group of individuals.

3. Objectives

In this project our objective is to develop a robust, effective and accurate Convolutional Neural Network (CNN) model which can classify the human expressions against the predefined emotional categories. To achieve this goal, our objective can divide into different approaches. Following are the different techniques which is applied in the project:

- **Data Acquisition and Preparation:**

  Objective: For gathering the dataset of facial images which are categorized by its different emotions for each section and preprocessed them to ensure they have suitable model for training.

  Outcome: For training the model a well standardized dataset of images is prepared for the evaluation.

- **Data Augmentation:**

  Objective: To improve the model's ability to normalize the unexplored data we will apply Data Augmentation to enhance its diversity.

  Outcome: The augmented dataset will be prepared that include variations in rotations, shifts, and flips, to enhance the model's robustness.

- **Model Architecture Design:**

  Objective: We will create an enhanced CNN architecture by using multiple convolutional layers, batch normalization, pooling layers, and dropout layers which would be effective in learn and generalize the features of the face.

Outcome: Using these approaches, we will get an enhanced model for detecting human emotions.

- **Model Training:**

  Objective: For Training the CNN model we must use the prepared and augmented dataset, by applying the learning rate schedule for optimizing our training process.

  Outcome: A well-trained CNN model will be prepared which would be capable of accurately differentiating the facial expressions of humans.

- **Model Evaluation:**

  Objective: The evaluation of the trained model would be calculated on a separate test dataset by using the metrics of accuracy, precision, recall, and F1-score. Subsequently a graph of confusion matrix and its classification report will be generated.

  Outcome: The detailed output of the metrics will be generated for the user to provide an insight into the model's strengths and weaknesses.

- **Qualitative Analysis:**

  Objective: A visualized quality analysis of the model will be conducted on the model's prediction by selecting 25 random test images.

  Outcome: By observing the visual images, it will help us in studying the model's accuracy and identify any consistent errors or biases occurring during the training process.

4. Methodology

We will study the methodology of the Human Emotion Detection which includes several essential steps from loading the data and preprocessing its, using data augmentation, defining model, training, and its evaluation. Every step is very critical to develop a robust and accurate Convolutional Neural Network (CNN) for facial expression recognition.

4.1 Data Loading and Preprocessing

We will first prepare the dataset to ensure that it is acting along for training the model.

- **Data Collection**: We will gather the dataset of facial images which are categorized with its emotion expressions. We have used the available dataset from Kaggle. These datasets are divided into seven classes which includes anger, disgust, fear, happiness, sadness, surprise, and neutrality.
- **Image Resizing**: We will resize all the images into a similar dimension of 48x48 pixels. It would standardize our input size for the model, which is a necessity for processing the model efficiently and robustly.
- **Normalization**: We have included normalization in our model to scale our pixel values on the images between the range of 0 to 1. This step will be helpful in speeding the convergence of the model during training.
- **Dataset Splitting**: We have split our dataset into training and validation data from the pre-determined folder of training data folder. A testing dataset is also given separately from the Kaggle website.

4.2 Data Augmentation

Data augmentation is used to enhance and improve its ability to generalize the model by increasing diversification of the training dataset. Different techniques are applied on the training dataset to introduce variability. Following are the techniques which have been applied on the training dataset.

Rotation: Images have been randomly rotated with a range of -10° to +10° to make the model become invariant from the orientation of the faces.

Zoom: Randomly zooming in and out with a range of 10% has been applied on the model to make it robust in different scales of the faces.

Shifts: Images are randomly shifting horizontally and vertically within the range 10% to simulate the real-world scenarios where faces might not be perfectly placed.

Flips: Images are random exploiting horizontal flips to make the model invariant to its direction while facing the faces.

These augmentations help our model to become more invariant to common variations in the input data, while making it more robust and better for generalizing the new, unseen data.

4.3 Model

In this section we will a design a model for creating an effective Convolutional Neural Network (CNN) architecture, which will be optimized to classify the seven classes of facial expressions.

- **Convolutional Layers**: The model is started by implementing the multiple convolutional layers which is used to filters for the detection of various features containing the following features of edges, textures, and shapes of the input images. These layers are very essential for learning the different hierarchies of features in the model.
- **Activation Functions**: The Rectified Linear Unit (ReLU) has been used as the activation function in our model. It makes our model to learn complex patterns due to its nonlinearity approach in the image.
- **Batch Normalization**: We will improve the stability and efficiency by reducing the covariate shift on the dataset by using batch normalization as well as its activations.
- **Pooling Layers**: Max-pooling layers has been used in our model to down sample the featured maps, by reducing the different dimensions but retaining the essential features of it. This step will help us in making the model steady from small disagreements in the training image.
- **Dropout Layers**: To reduce the overfitting in the training process beside using the validation dataset to compute it, we have used a dropout layer. For computing this layer, we have randomly set up the input values to zero.
- **Fully Connected Layers**: After the convolutional and pooling layers process are completed, from fully connected layers the featured maps are flattened and passed through it. These layers are combined in the features to be learned by the convolutional layers to predict the result.
- **Output Layer**: The probability distribution is used by applying SoftMax in CNN over the emotion classes. The class with the higher probability is chosen as its predicted emotion.

4.4 Learning Rate Scheduler

We will adjust the learning rate during the training process on the model, we have used learning rate scheduler. For implementation of this function, we have applied ReduceLROnPlateau. This function automatically reduces the learning rate when the validation data in the model stops decreasing or improving. By applying this function our model is prevented from overshooting for a optimal output. It is a kind of a fine-tuning method while the training model is being processed to make it more efficient and robust.

5. Experiments and Results

5.1 Model Training

The training process for the Convolutional Neural Network (CNN) model has the most important phase in the project where the model learns to identification of different images and classify accordingly from expressions classes accurately. In the start, the augmented dataset is injected inside the model in batches. By processing the dataset in batches and injecting in the model makes it very efficient for the system to handle the large datasets, by reducing it into small chunks. On every epoch, the training dataset passes through it for the model parameters such as weights and biases are adjusted accordingly to reduce the loss function of the model. It measures the difference between the predicted and actual emotion labels.

A validation dataset is used in the model continuously on each epoch to monitor the model's performance and preventing it from being overfitted in the testing phase of the project. A function of Early Stopping is used in the model for a predetermined period of epoch, if the validation loss is not improving. This approach helps us in ensuring the model to not be overfitted and maintains a generalization for unseen data.

In this project, we have used a Adam optimizer which is very important because of its efficiency and adapting the learning rate. It ensures that our learning rate and other parameters are faster and stable during the time of the convergence.

5.2 Model Evaluation

After completing the training process on the model, we will evaluate the performance of the system by using the testing dataset. This evaluation on the model will provides us with an in-depth analysis on the accuracy of the model. We have used various performance metrics to evaluate on the trained model to evaluate the accuracy, precision, recall and F1-score for understanding the effectiveness and efficiency of the model trained from our input dataset.

As there are a lot of classes in the project so we have visualized our predictions against its true classes. The confusion matrix is very essential in identifying the strengths and weakness of each class as the classes are not proportionally divided on the dataset. Subsequently, a separate classification report which includes the precision, recall, and F1-score for each class is also generated. This helps us in identifying in potential weaknesses in a particular class.

5.3 Training and Validation Loss

It is very important in a model to calculate and have some knowledge about the losses occurred during the time of learning. We have visualized the losses in the training and validation dataset occurred over each epoch and the performance of the model over time. Ideally, both the dataset should be decreased over every epoch and converge to a low value. This will help us in learning the effectiveness and efficiency on the unused data. If the loss on training decreases but the loss in the validation dataset starts to increase, it will signal us for the model of being overfitted. This graph helps us in diagnosing these issues during the training the model next time.

5.4 Detailed Evaluation

A detail evaluation of the model has been provided in the project by analysing the classification report to gain insights into the performance of the trained model. In the evaluation a comprehensive detail of model's accuracy with the confusion matrix of true positive, true negative, false positive and false negative for each class in the dataset has been provided. This helps us in evaluating the model's behaviour with each class in the dataset.

In the classification report, it provides us the details of precision which indicates the true positive predictions from all positive predictions from the model. Consequently, recall measures the predictions of true positive from all true positive images. The F1-score measure the harmonic between precision and recall. These metrics for the evaluation helps us in identifying in determining the strengths and the areas where improvement is further required.

5.5 Qualitative Results

For visualizing the images on the dataset, we have a conducted a comprehensive analysis on the quality of the model performance. For conducting this process, we have chosen 25 random images to visualize it with the prediction from the model versus the true labels from the given random image. By displaying these images of human emotions will give an insight into the performance of our model for a real-world scenario.

Qualatative analysis identifies in understanding any consistent error or biases occurring in the models' predictions. As an example, our disgust emotion class is giving a systematic error because of the two reasons. Firstly, the dataset in this class is very less as compared to their classes and secondly the face expression of anger and neutrality is very similar to the face of disgust, which is mismatched. Furthermore, when these errors occur in the system, it informs us with the particular part on which parameter should be adjusted and fine tuning should be processed.

After completing the analysis, it is also observed that the models accuracy is increasing from 62% despite extensively training and evaluating the model. This moderate performance of the model could be attributed to several factors. Firstly, I have observed that the dataset is imbalanced where some of the classes have significantly more sample than the other classes. This could be reason for the model to behave biasedly for some classes. Secondly, the subtle difference between certain emotion classes with other such as fear, anger, neutrality and disgust could confuse the model for deciding the classes of the image. Consequently, using custom CNN model which is effective but not robust as the other pre trained models like VGG16, or ResNet, which are trained with very large and diverse dataset. By overscoming the shortcoming mentioned in this paragraph we can enhance our model beyond the accuracy of 62%..

After completing the comprehensive training, evaluation and analysis if the model with CNN, this Human Emotion Detection will help us to develop a effective and accurate system for the classification of different emotions of the humans.

6. Conclusion

The Human Emotions Project has been developed to provide a customized made model using Convolutional Neural Network (CNN) for the classification of facial emotions of the human into seven different categories, which includes anger, disgust, fear, happiness, sadness and neutrality. Through the pipeline of preprocessing, data augmentation, model designing, training and evaluation, we have built a model which will classify the different emotions using deep learning approaches.

The project starts with the collection of data and classification of the images into there particular classes. We then preprocess our dataset by a resizing the data into a constant dimension for every image and normalizing their pixel values, which would be efficient for processing in the CNN model. Some data augmentation techniques are also used in the data, by changing or adjusting the rotations, zooms, shifts, and flips, of the dataset. It will increase the diversity of the training data, which would eventually enhance the robustness and generalization of the model.

The model architecture was well designed by involving multiple convolutional layers to extract the features, batch normalization for improving the stability of training the model. We have down sample the features maps and used dropout layers to prevent it from over-fitting the model and fully connected layers to combine all the features for the classification on the testing dataset. The model has used Adam optimizer for training and a learning rate scheduler to adjust the learning rate.

Despite it has achieved a test accuracy of 62%, the performance of the model was comprehensively to identify the fields where more improvement is required to enhance the system. After studying the qualitative analysis of the model, we have observed that there is imbalance in the dataset, similarity between emotion classes and using of custom-made CNN instead of using pre trained model. After visually displaying the images, it is observed that the model has struggled to differentiate between some classes.

To further enhance the model, we can use several strategies to overcome the shortcomings in the model. We would start with equally balancing the dataset by collecting the data for dataset which are heavily distinct. In the data augmentation we can further compute with advanced techniques on the training data for diversity.

By concluding the report, the Human Emotions Detection project has demonstrated the working from the CNN model using deep learning techniques to compute recognition of Human emotions through the face expressions. While the model achieved moderate results and the observation gained from this project has cleared our path for future improvement in our model. This model is very essential in our life which can be used in multiple fields from medical, businesses to security jobs as well. This is a very important project in pivotal for the advancement in the study artificial intelligence.

7. References

- *FER-2013 Dataset*. (n.d.). https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer/data
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1409.1556
- ChatGPT. (2024). Assistance with syntax and algorithms for the code.
- Human Emotion Detection Project, GitHub Link: https://github.com/zuba1rkhan/EmotionDetection