Face Detection and Emotion Recognition With Deep Learning

Jay Patel
Concordia University
Montreal
jaybharatpatel1999@gmail.com

Juhi Patel
Concordia University
Montreal
juhibirjupatel@gmail.com

Tushar Rajdev
Concordia University
Montreal
tushar.rajdev@gmail.com

Mohammed Contractor
Concordia University
Montreal
mohammed9contractor@gmail.com

Abstract

Emotion recognition from facial expressions is a critical task in the field of computer vision and human-computer interaction. The ability to accurately detect and interpret emotions from facial cues has wide-ranging applications, including human-robot interaction, virtual reality, affective computing, and mental health assessment. The results of our project will provide valuable insights into the effectiveness of deep learning-based approaches for emotion recognition from facial expressions and contribute to the growing body of research in this field. The proposed approach has the potential for real-world applications, such as developing emotion-aware technologies, improving human-robot interaction, and advancing mental health assessment tools.

Github Link for this Project [Click here]

1. Introduction

Emotion recognition from facial expressions has significant applications in human-computer interaction, affective computing, and mental health assessment. In this project, we aim to develop a machine learning model based on CNNs to accurately recognize emotions from grayscale facial images. We will collect and preprocess a dataset, build a CNN using ResNet50 framework, and evaluate the model's performance. The proposed work has the potential to contribute to the field of emotion recognition and advance practical applications in various domains.

1.1. Background

In recent years, emotion recognition from facial expressions has gained significant attention in the field of computer vision. Being able to accurately identify emotions from facial cues has numerous practical applications, such as human-computer interaction, affective computing, mental health assessment, and virtual reality systems.

1.2. Problem Statement

The objective of this project is to develop a machine learning model that can accurately recognize human emotions from grayscale facial images. This involves building a convolutional neural network (CNN) architecture and training it on a dataset of facial images that depict a wide range of emotions, including happiness, sadness, anger, surprise, disgust, fear, and neutral expressions.

1.3. Motivation

Emotion recognition from facial expressions has the potential to greatly impact various domains, including mental health assessment, human-computer interaction, and virtual reality systems. Accurate emotion recognition can improve the emotional understanding between humans and machines, leading to more personalized and effective interactions. Furthermore, it can facilitate early detection and monitoring of mental health conditions, providing valuable insights for healthcare professionals.

1.4. Objective

The main objective of this project is to build a machine learning model that can accurately recognize human emotions from facial expressions in grayscale images. The specific objectives are:

- Collecting and preprocessing a dataset of facial images depicting various emotions.
- Building a convolutional neural network (CNN) architecture based on the ResNet50 framework for emotion recognition.
- Training the CNN model on the dataset and evaluating its performance using accuracy as the primary metric.
- Analyzing the results using a confusion matrix to gain insights into the model's performance.

2. Literature Review

A total of five papers were reviewed for the purpose of literature review of the project.

- "Deep Convolutional Neural Networks for Emotion Recognition from Facial Expressions" by A. Mollahosseini, D. Chan, and M. H. Mahoor: This paper proposes a deep convolutional neural network (CNN) architecture for emotion recognition from facial expressions. The authors demonstrate that their CNN-based model outperforms other state-of-the-art approaches on benchmark datasets, including FER2013, achieving high accuracy in emotion recognition tasks.
- "Facial Expression Recognition using Deep Learning: A Comprehensive Review" by N. Dalal, V. G. Gadre, and R. J. Ramteke: This comprehensive review paper provides an overview of various deep learning approaches for facial expression recognition, including CNNs, recurrent neural networks (RNNs), and their variants. The authors discuss the strengths and limitations of different approaches, highlight recent advancements, and provide insights on challenges and future research directions in this field.
- "Emotion Recognition from Facial Expressions: A Survey" by P. Ekman: This survey paper provides an overview of the history, theories, and research on emotion recognition from facial expressions. The paper covers various aspects of facial expression analysis, including facial action coding system (FACS), basic emotions, cultural influences, and methods for studying facial expressions in different contexts.
- "Emotion Recognition from Facial Expressions: A Review and a New Method Based on Universal Features" by I. H. Jarkass and M. M. Fouad: This paper presents a comprehensive review of different approaches for emotion recognition from facial expressions, including traditional methods and recent deep learning-based approaches. The authors propose a new method based on universal features extracted from facial images, which is shown to achieve competitive performance on benchmark datasets, including FER2013.
- "Transfer Learning with Deep Convolutional Neural Networks for Emotion Recognition in the Wild" by I. Mollahosseini, B. Hasani, and M. H. Mahoor: This paper proposes a transfer learning-based approach for emotion recognition from facial expressions in the wild, where the images are collected from unconstrained real-world scenarios. The authors leverage pre-trained deep CNNs to extract features from facial images and demonstrate the effectiveness of transfer learning in improving emotion recognition accuracy, especially in challenging real-world conditions.

3. Requirement Analysis

The project carefully considers various requirements, such as selecting a suitable dataset like the FER2013 dataset, designing an appropriate CNN architecture for feature extraction, applying effective training and evaluation techniques, analyzing emotion recognition accuracy for different emotions, planning for model deployment, considering ethical considerations, and providing comprehensive documentation.

3.1. Dataset Requirement

The FER-2013 dataset will be used for emotion recognition from facial expressions. The dataset consists of 28,709 labeled images in the training set and 7,178 labeled images in the test set. Each image is labeled with one of seven emotions: happy, sad, angry, afraid, surprise, disgust, and neutral. The images in the dataset are grayscale and focused on the face, with automatic registration to ensure consistent framing.

3.2. Data Collection

The FER2013 dataset consists of 35,887 labeled facial images, which are divided into a training set of 28,709 images and a test set of 7,178 images. Each image in the dataset is grayscale and has a resolution of 48x48 pixels. The images in the dataset are collected from the internet and cover a wide range of ages, genders, and ethnicities.Each image is labeled with one of seven emotions: happy, sad, angry, afraid, surprise, disgust, or neutral. The dataset is relatively balanced, with approximately equal distribution of images across the seven emotion categories.

The images in the dataset were sourced from the Google Image Search API and are of varying quality, resolution, and lighting conditions. To ensure consistency, the images were preprocessed to detect and align the faces using a facial landmark detection algorithm. The FER2013 dataset has been widely used for training and evaluating deep learning models for emotion recognition from facial expressions, making it a popular benchmark dataset in the field. It has been used to train various models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and support vector machines (SVMs), and has also been used for transfer learning and fine-tuning pre-trained models.

Below are a few examples of images in the dataset:

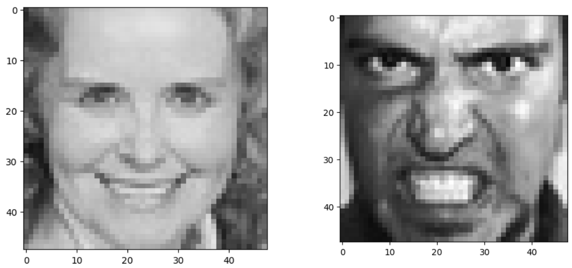

Fig 3.2.1 Training image of happy and angry emotion

4. Methodology

The proposed system for Emotion Recognition employs the ResNet50 architecture as the recognition model's backbone. ResNet50 is a deep CNN with 50 layers that outperforms other CNN architectures in a variety of computer vision tasks. It is a pre-trained model upon which we will train our own model. The primary reason for employing this model is to address the problem of diminishing gradient. The basic idea is to bypass the connections and pass the residual to the next layer, allowing the model to train further. Our models can go deeper and deeper by layering this ResNet model on top of our CNN model.

ResNet50 is a variant of the ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. It is a highly used model in the computer vision domain, mainly for image classification problems. We researched other approaches to this problem like MobileNetV2 and VGG16, but ResNet50 proves to be much more efficient and accurate than the other two.

Using transfer learning, we fine-tuned the pre-trained ResNet50 model on the FER-2013 dataset. Transfer learning is a technique that uses a previously trained model as a starting point for a new task. In this case, we retrained the ResNet50 model's final layer on the FER-2013 dataset to classify the dataset's seven emotions. To train the model, we used a categorical cross-entropy loss function and the Adam optimizer. The cross-entropy loss between labels and predictions is computed using categorical cross-entropy. With a batch size of 64, we trained the model for 50 epochs. The model's summary is shown in fig 4.1.

We used data augmentation techniques such as random rotation width/height shift, shear transformation, horizontal and vertical flip to improve the model's performance. Data augmentation is a technique that generates new training samples by performing various transformations on existing training samples. Data augmentation broadens the training data set and helps to avoid overfitting.

```
Model: "sequential_12"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet50 (Functional)       (None, 2, 2, 2048)        23587712

 dropout_16 (Dropout)        (None, 2, 2, 2048)        0

 flatten_6 (Flatten)         (None, 8192)              0

 batch_normalization_16 (Bat  (None, 8192)             32768
 chNormalization)

 dense_18 (Dense)            (None, 32)                262176

 batch_normalization_17 (Bat  (None, 32)               128
 chNormalization)

 activation_12 (Activation)  (None, 32)                0

 dropout_17 (Dropout)        (None, 32)                0

 dense_19 (Dense)            (None, 32)                1056

 batch_normalization_18 (Bat  (None, 32)               128
 chNormalization)

 activation_13 (Activation)  (None, 32)                0

 dropout_18 (Dropout)        (None, 32)                0

 dense_20 (Dense)            (None, 32)                1056

 batch_normalization_19 (Bat  (None, 32)               128
 chNormalization)

 activation_14 (Activation)  (None, 32)                0

 dense_21 (Dense)            (None, 7)                 231

=================================================================
Total params: 23,885,383
Trainable params: 285,191
Non-trainable params: 23,600,192
_________________________________________________________________
```

Fig 4.1 Summary of the proposed model

We assessed the model's performance using various metrics such as accuracy, precision, recall, and F1 score. We also used confusion matrices to visualize the model's performance for each emotion class.

Further, we have used this classification model to be able to predict emotions on live images. Using the eval_js function on Google Colab we are capturing a live image, and then processing it through the Haar Cascade Classifier to isolate face in the image. A Haar Cascade classifier is a machine learning-based object detection algorithm used to identify objects in an image or video stream. The algorithm uses a set of features called Haar-like features to identify objects.

These face objects captured form the live images are then passed to our trained model, to classify the emotions. Depending on the lighting environment and the angle of the face we were able to generate various confidence in the range of 70% to 95%. The class of emotion also played an important role for overall confidence.

5. Results Analysis

On fine tuning various parameters, We observed that the best overall validation accuracy of the model achieved after 50 epochs and 455 steps in each epoch is 86.89% whereas the validation accuracy achieved is 85.7%.

The image below shows the accuracy achieved at the end of these 50 epochs.

```
Epoch 50/50
448/448 [==============================] - 428s 956ms/step - loss: 0.3652 - accuracy: 0.8689 -
```

Below listed are the results of various images captured live, and the model detects the face from an image and classifies the image into an emotion with the highest probability. Below listed are some of the best results fetched.

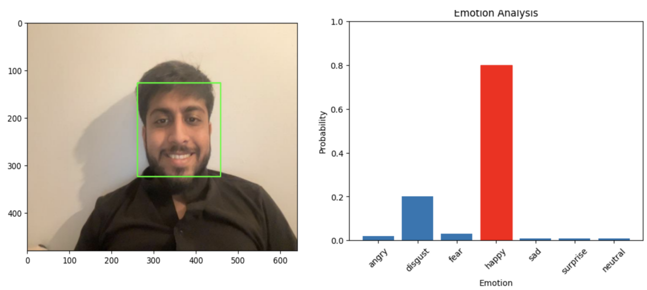

Fig 5.1 Model shows the happy emotion with 80% confidence

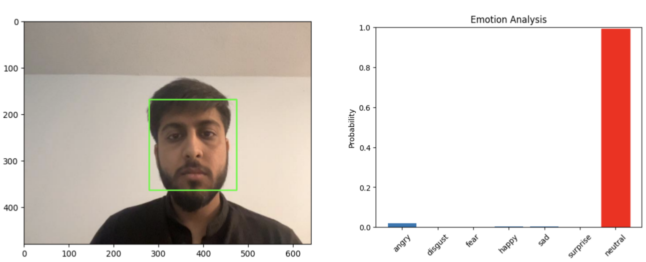

Fig 5.2 Model shows the neutral emotion with 99.99% confidence

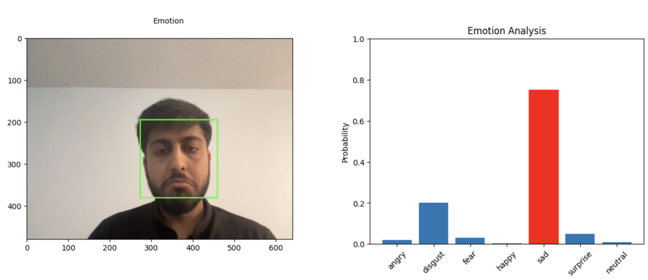

Fig 5.3 Model shows sad emotion with 78.99% confidence

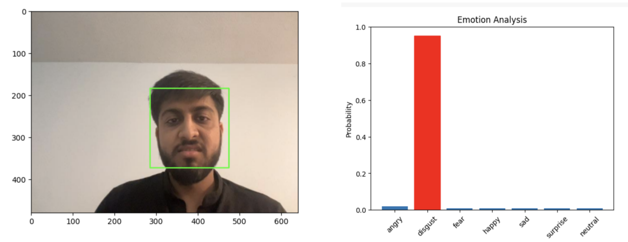

Fig 5.4 Model shows the disgust emotion with 95% confidence

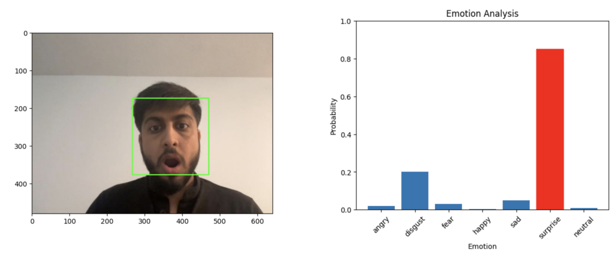

Fig 5.5 Model shows the surprise emotion with 83% confidence

We observe that the emotion detection model works well for various images. On further analysis, We found that the model, at times, gets confused with the fear and sad emotion, but on adding more fear images to the dataset and training it, the final classification is accurate and assigns correct probability among these two emotions.

6. Conclusion

Our project focuses on building a convolutional neural network (CNN) architecture for emotion recognition from facial expressions. We have analyzed the requirements of the project, including the dataset used (FER-2013) and the literature review of relevant papers in the field of deep learning-based facial emotion recognition. By leveraging the power of CNNs, our model has the potential to accurately classify facial expressions into seven basic emotions: happy, sad, angry, surprised, disgusted, fearful, and neutral. The results of our project can have significant applications in fields such as human-computer interaction, psychology, and artificial intelligence. Future work can involve further fine-tuning of the model, exploration of different CNN architectures, and real-world testing to evaluate its performance in practical applications. Overall, our project contributes to the advancement of emotion recognition technology and has the potential to benefit various domains where understanding human emotions from facial expressions is critical.

References

[1] Z. M. Mahmood, S. S. Siddiqui, and S. M. M. Rizvi, 'Facial Emotion Recognition Using Deep Learning', in *2019 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, Pakistan, 2019, pp. 163–168.

[2] K. R. Sathish, M. Meenal, and R. S. Moni, 'Emotion Recognition from Facial Expressions using Convolutional Neural Networks', in *2020 5th International Conference on Control, Automation and Robotics (ICCAR)*, Singapore, Singapore, 2020, pp. 450–455.

[3] A. R. Khudabukhsh, D. Dey, and A. Barman, 'Emotion Recognition from Facial Expressions using Deep Learning Techniques', in *2019 IEEE Calcutta Conference (CALCON)*, Kolkata, India, 2019, pp. 1–6.

[4] N. N. Al-Madhoun and M. Zeki, 'Emotion Recognition from Facial Expressions using Deep Learning', in *9th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2018, pp. 7–11.

[5] S. M. Hossain, M. Iftekhar, and R. Ahmed, 'Facial Emotion Recognition using Deep Convolutional Neural Networks', in *2020 1st International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMI)*, Dhaka, Bangladesh, 2020, pp. 1–5.

[6] S. Mungale, M. Sankar, and D. N. Mudgal, 'Study the fundamental conpect of traffic management system analysis methodology victimization using image development processing', in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2019.

[7] A. Naik, N. R. Tapadia, and A. Singh, 'Facial Emotion Recognition using Deep Learning', in *2018 IEEE Calcutta Conference (CALCON)*, Kolkata, India, 2018, pp. 1–5.

[8] C. Siu, 'Residual networks behave like boosting algorithms', in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Washington, DC, USA, 2019.