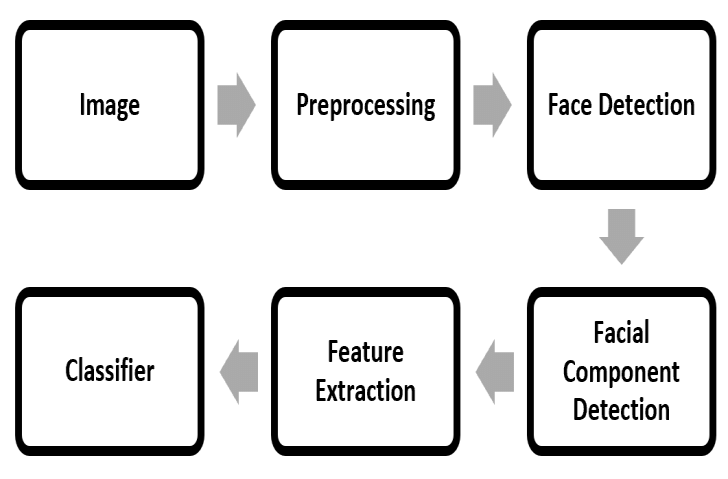
**Abstract**

Automatic facial emotion recognition (FER) has become an increasingly important field in computer vision due to its potential applications in human-computer interaction and social signal processing. Deep learning approaches have achieved remarkable success in FER, surpassing traditional methods. This paper reviews the state-of-the-art deep learning techniques for FER. We discuss the typical pipeline of a FER system, including face detection, preprocessing, feature extraction, and classification. We explore deep learning architectures employed for FER, such as convolutional neural networks (CNNs). The trained model is used in web development. Finally, we discuss the challenges and future directions of deep learning-based FER.

Introduction

Automatic emotion recognition is a large and important research area that addresses two different subjects, which are psychological human emotion recognition and artificial intelligence (AI). The emotional state of humans can obtain from verbal and non-verbal information captured by the various sensors, for example from facial changes [1], tone of voice [2] and physiological signals [3]. In 1967, Mehrabian [4] showed that 55% of emotional information were visual, 38% vocal and 7% verbal. Face changes during a communication are the first signs that transmit the emotional state, which is why most researchers are very interested by this modality.

Facial Emotion Recognition (FER) is a flourishing study topic in which many breakthroughs are being made in industries, such as automatic translation systems and machine-to-human contact. The classical FER consists of two main steps: feature extraction and emotion recognition. In addition, image pre-processing, including face detection, cropping, and resizing. Face detection crops the facial region after removing the backdrop and non-face areas. Finally, the retrieved characteristics are used to classify emotions, which is commonly done with the help of neural networks (NN) and other machine learning approaches. The challenge of facial emotion recognition is to automatically recognize facial emotion states with high accuracy. Therefore, it is challenging to find the similarity of the same emotional state between different people since they may express the same emotional state in various ways. As an example, the expression may vary in different situations such as the individual’s mood, skin colour, age, and the environment surrounding. Generally, FER is separated into following stages as shown in Figure 1:



In the first, stage, which is a pre-processing stage, an image of a face is detected in second stage and facial components of the face will be detected from the region in the third stage.

In the second stage, an informative feature will be extracted from different parts of the face. In the last stage, a classifier needs to be trained before been use a classifier needs to be trained before being used to generate labels for the Emotions using the training data. Deep learning is a part of machine learning approaches that can be adapted to emotion recognition and facial expression analysis. However, deep learning depends on data size which may affect its performance.

**Literature Review**

Facial emotion recognition (FER) has emerged as a crucial area of computer vision, fueled by its potential to revolutionize human-computer interaction (HCI) and social signal processing. Traditional methods relied on manually crafted feature extraction, limiting performance and requiring significant domain expertise. However, deep learning approaches have propelled FER forward, achieving remarkable success by enabling automatic feature learning from vast datasets. This review delves into the state-of-the-art deep learning techniques employed for FER.

**FER System Pipeline:**

A typical FER system follows a multi-stage process:

1. **Face Detection:** The initial stage identifies and locates human faces within an image or video frame.
2. **Preprocessing:** Images undergo normalization for size consistency, noise reduction, and facial alignment to ensure uniformity.
3. **Feature Extraction:** Deep learning models act as powerful tools, automatically extracting features that capture the emotional information embedded within facial images.
4. **Classification:** The extracted features are then classified into distinct emotional categories, ranging from basic emotions (happiness, sadness, anger) to more nuanced emotional states.

**Deep Learning Architectures:**

Convolutional Neural Networks (CNNs) reign supreme as the dominant deep learning architecture for FER tasks. CNNs excel at extracting spatial features from images, making them ideal for capturing the subtle variations in facial expressions that convey emotions. Variants of CNNs, such as VGGNet, ResNet, and Inception, have demonstrated impressive performance in FER applications.

**Beyond Research: Web Development Integration**

While the core of FER lies in research and model development, the ultimate goal is its practical application. Integration of the trained model into web development opens doors for real-world scenarios. Imagine an interactive platform that adapts to a user's emotional state through facial recognition, personalizing the user experience in real-time.

By addressing these challenges and actively exploring these promising research avenues, deep learning has the potential to significantly enhance FER accuracy and pave the way for its seamless integration into real-world applications, fundamentally transforming human-computer interaction and social signal processing.

Background Information

1. Emotion Recognition

Facial Recognition is a branch of computer science that deals with methods and strategies for detecting emotions in facial expressions. It is expected that expressions can be the next communication medium with computers the majority of this field's research focuses on recognizing human emotions from movies or auditory data. The majority of the research has focused on recognizing and matching faces, but no convolutional neural networks have been utilized to infuse emotions into photos. Emotion Recognition is the study of identifying emotions, as well as the strategies and procedures utilized to do so. Emotions can be detected through facial expressions, verbal signals, and other indicators. Machine learning, neural networks, artificial intelligence, and emotional intelligence are just a few of the methods that have been used to infer emotions. Emotion Recognition is gaining traction in the study, which is critical to solving a variety of challenges.

1. Facial Emotion

Recognition Facial Emotion Recognition is a research area that tries to identify the emotion from the human facial expression. The surveys state that developments in emotion recognition make complex systems simpler. Emotion Recognition is a difficult process since emotions can differ depending on the environment, appearance, culture, and facial reaction, resulting in unclear data.

1. Deep Learning

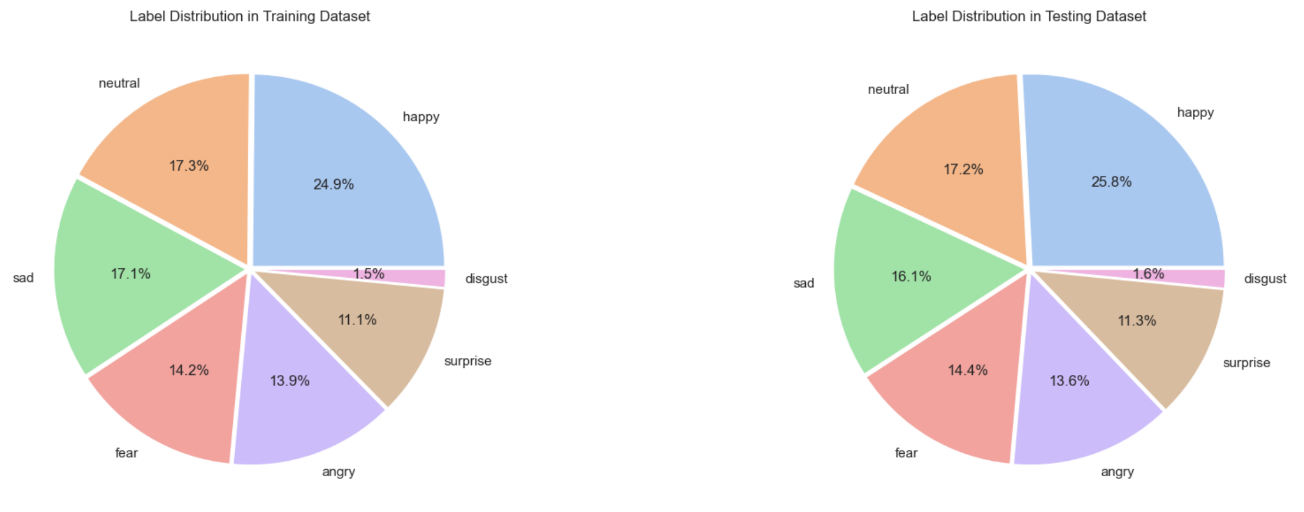
Deep Learning is a machine learning technique that models the data that are designed to do a particular task. Deep learning in neural networks has wide applications in the area of image recognition, classification, decision making, pattern recognition, etc.

PROPOSED METHODOLOGY

The proposed technique, the emotion database used for the study, and the iinception model are all explained in this section. This project uses a Haar classifier for human detection. The Haar classifier is trained by Haar-like small features and also the Haar-like may be a commonly used texture descriptor, and its main features are linear, edge, center, and diagonal. The Haar-like feature can reflect the grey level change of image, so it's very effective to explain the face because many features of external body parts have obvious contrast change characteristics.

Data Description

The dataset used is Facial Expression Recognition(FER) dataset. It contains two files: train and test. The train contains total of 28821 images and the test files have 7066 images. The images are labelled into following classes: anger,sad,happy,neutral,disgust,fear,surprise. The distribution of various classes are as follows:



The images in the dataset looks like:



**Preprocessing**

Splitting data: The train dataset is further divided into two parts: train containing 23056 images; and validate containing 5765.

Data Augmentation: All the three dataset went through data augmentation. The images are standardized,flipped, rotated, and shifted to artificially expand the dataset and improve model generalisation.

Deep Learning Model

Building

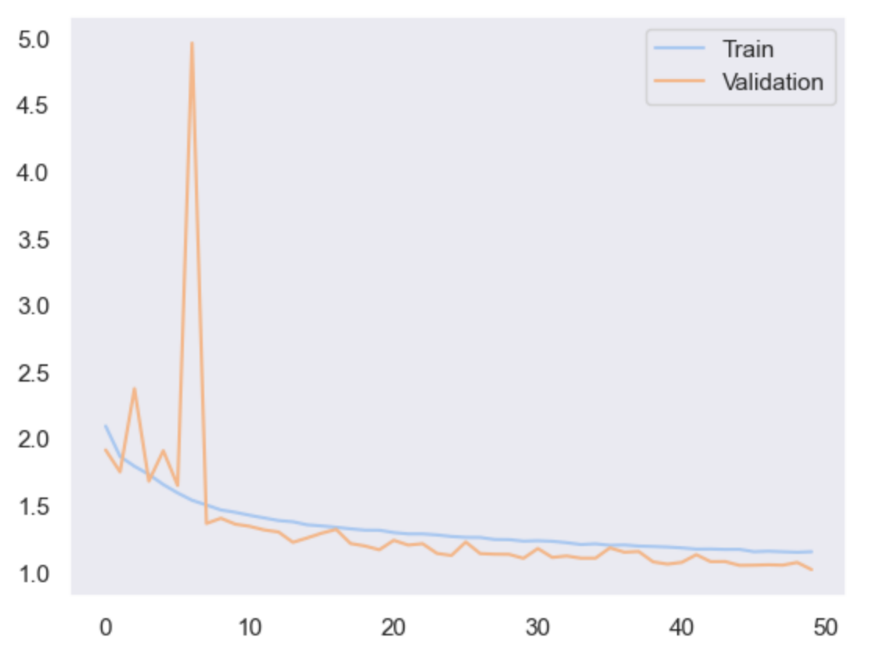
* **1st Layer**: Convolutional layer with 64 filters of size (5, 5), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
* **2nd Layer**: Convolutional layer with 128 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
* **3rd Layer**: Convolutional layer with 512 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
* **4th Layer**: Convolutional layer with 512 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
* **Flatten Layer**: Flatten the output from the convolutional layers to be fed into the fully connected layers.
* **Fully Connected Layer 1**: Dense layer with 256 units and ReLU activation, followed by batch normalization and dropout for regularization.
* **Fully Connected Layer 2**: Dense layer with 512 units and ReLU activation, followed by batch normalization and dropout for regularization.
* **Output Layer**: Dense layer with softmax activation for multi-class classification, with the number of units equal to the number of classes in the dataset.
* **Compilation**: The model is compiled with the Adam optimizer, categorical cross-entropy loss function, and accuracy metric.

Training

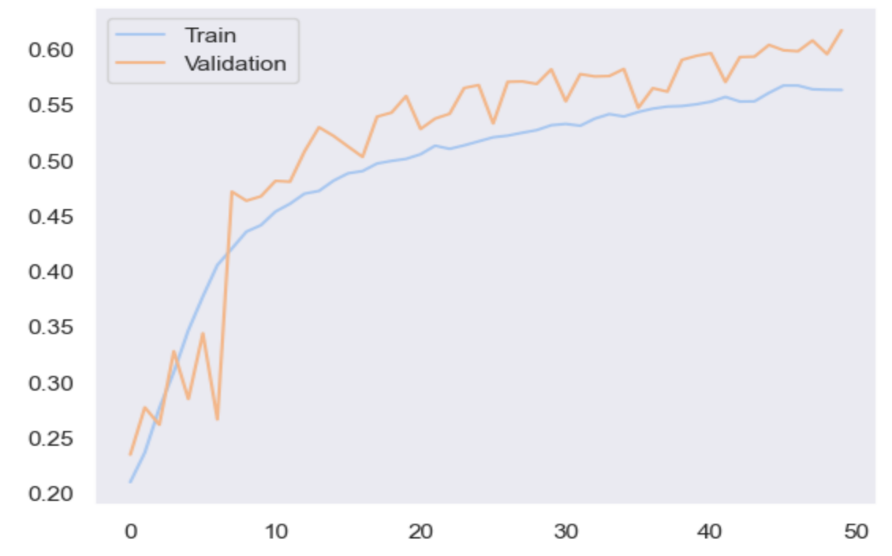
Model was trained on 50 epochs with 23056 train set and 5765 validate set.

Model Evaluation

1. Loss: Train loss and validate loss decreases with time. The graph shows the model is not overfitted.

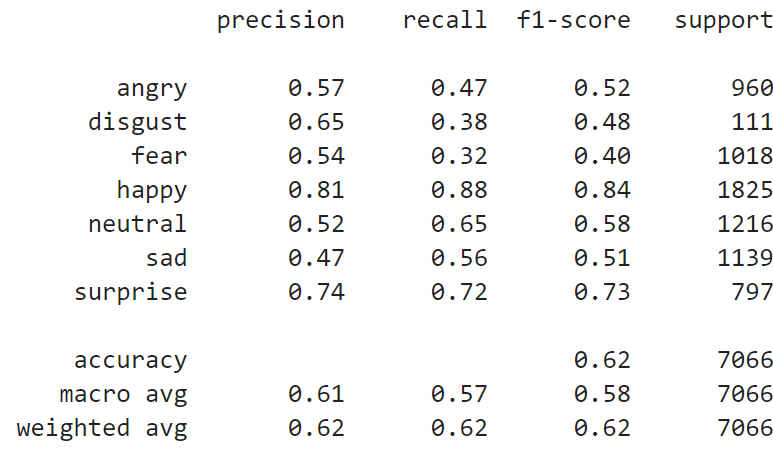


1. Accuracy: Train as well as validate accuracy increases with epochs. Train set accuracy at last epoch is 56.28% and validate set accuracy is 61.67%

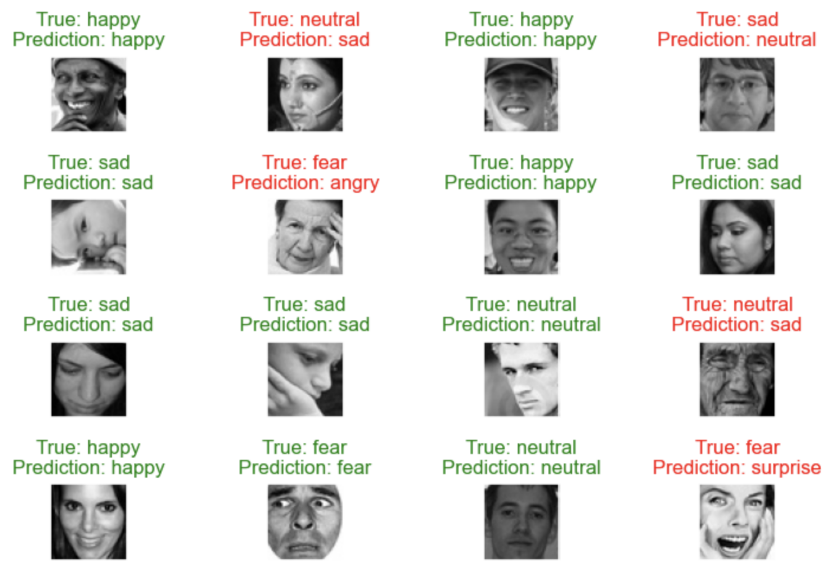


Prediction

Test set have a loss of 1.01 and accuracy of 62.5%. The classification report is as shown:



The prediction on test set is as shown:



Model Deployement

Face Detection

Face detection could be a pre-processing phase to acknowledge the facial expressions of humans. a picture is segmented into two parts which have faces and other non-face regions [5]. There are numerous methods used for face detection.

The first step in face detection was to load the image and convert it into gray scale. After that, a haar cascade classifier runs through this gray image and detect the face. The face is then standardize. The standardize image is used for prediction.