Facial Expression Recognition with Deep Learning

TEAM MEMBERS

Boyina Ganesh – ganesh.20bcr7108@vitap.ac.in

Gumperla Vaishnavi – vaishnavi.20bcd7055@vitap.ac.in

Shaik Mohammad Waseem Akram – akram.20bcd7141@vitap.ac.in

Mudit Dhoundiyal –mudit.dhoundiyal2020@vitstudent.ac.in

OUTLINE

- INTRODUCTION
- OBJECTIVE
- RELATED WORK
- DATASET
- DEEP LEARNING MODEL
- RESULTS
- CONCLUSION

INTRODUCTION

- One of the easiest ways people communicate is through facial expressions. In this project, we will be implementing a Deep learning model for facial expression recognition.
- Facial Emotion Recognition (FER) is a technology used for analyzing sentiments from different sources, such as pictures and videos.
- FER analysis comprises two steps:

  a) face detection.

  b) facial expression detection.
- Emotion detection is based on the analysis of facial landmark positions

  Example: end of nose, eyebrows.

- Depending on the algorithm, facial expressions can be classified as basic emotions such as anger, disgust, fear, joy, sadness, and surprise or compound emotions such as happily sad.
- Facial expression recognition has been widely used for lie detection and **human-machine interaction**.

OBJECTIVE

- To develop an application that understands and improves the performance of the emotion recognition model and also applies them to real-world situations using the VGG16 Model.

PREVIOUS WORK

- The experimental results based on CK + dataset show that the improved VGG16 network has strong supervised learning ability. It can extract features well for different expression types, and its overall recognition accuracy is close to 90%.

DATASET

- Facial Expression Recognition is a well-studied field with numerous available datasets. In this project, we will be using FER2013 as our main dataset.
- FER2013 is a well-studied dataset and has been used in research papers. It is one of the more challenging datasets with accuracy only at 65±5%.
- The dataset's 35,887 contained images are normalized to 48x48 pixels in grayscale. FER2013 is not a balanced dataset, as it contains images of 7 facial expressions, with distributions of Angry (4,953), Disgust (547), Fear (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral (6,198).

IMAGES FROM FER2013 DATASET

CONVOLUTIONAL NEURAL NETWORKS

- CNN is used primarily in computer vision and image classification applications that can detect features and patterns within an image, enabling tasks, like object detection or recognition.
- CNN is a regularized version of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks makes them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include: penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.)
- CNN takes a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme.

METHODS

Data Preparation:

- It is the process of preparing raw data so that it is suitable for further processing and analysis.
- Key steps include collecting, cleaning, and labeling raw data into a form suitable for machine learning (ML) algorithms and then exploring and visualizing the data.
- One of the primary purposes of data preparation is to ensure that the raw data being readied for processing and analysis is accurate and consistent.
- Data is commonly created with missing values, inaccuracies, or other errors, and separate data sets often have different formats that need to be reconciled when they're combined.
- Correcting data errors, validating data quality, and consolidating data sets are big parts of data preparation projects.

METHODS

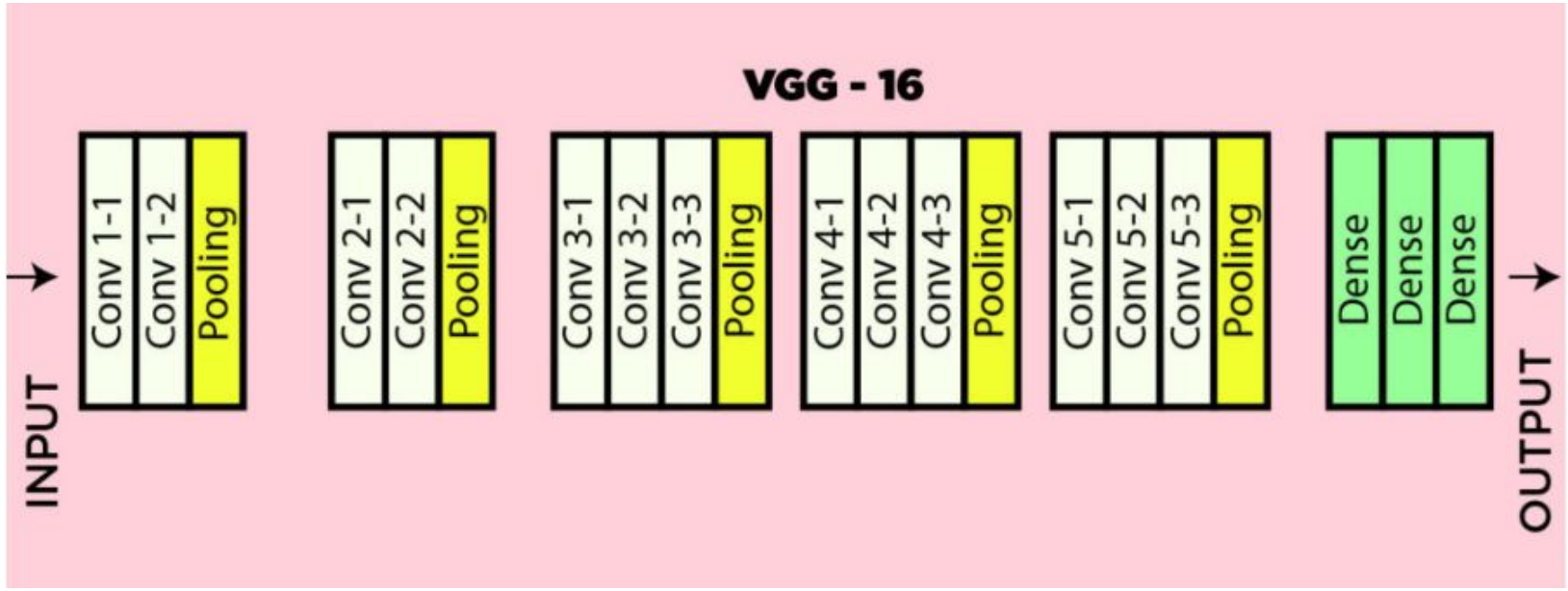
Data Augmentation:

- Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points.
- Data augmentation increases the number of examples in the training set while also introducing more variety in what the model sees and learns from. Both these aspects make it more difficult for the model to simply memorize mappings while also encouraging the model to learn general patterns.
- While it is possible to collect more real-world data, this is much more expensive and time-consuming than using data augmentation techniques. So while it's always better to grow the real-world dataset, data augmentation can be a good substitute when resources are constrained.

VGG16 Model

- The VGG model stands for the **Visual Geometry Group.**
- VGG-16 is a convolutional neural network that is 16 layers deep.
- VGG16 is an object detection and classification algorithm which can classify images of different categories with high accuracy.
- It is one of the popular algorithms for image classification and is easy to use with transfer learning.
- VGG16 has a large number of parameters on its first fully connected layer, which makes the amount of calculation huge.

VGG16 MODEL ARCHITECTURE



CODE

- We have executed the code in the **google colab** platform.
- Here is the link to access the code:

  https://colab.research.google.com/drive/1SlmXNLzHCe5mJfZ6mEYbyco3jgxIbRp

RESULTS

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 48, 48, 64)        640

conv2d_1 (Conv2D)            (None, 48, 48, 64)        36928

batch_normalization (BatchN  (None, 48, 48, 64)        256
ormalization)

activation (Activation)      (None, 48, 48, 64)        0

max_pooling2d (MaxPooling2D  (None, 24, 24, 64)        0
)

dropout (Dropout)            (None, 24, 24, 64)        0

conv2d_2 (Conv2D)            (None, 24, 24, 128)       73856

conv2d_3 (Conv2D)            (None, 24, 24, 128)       147584

batch_normalization_1 (Batc  (None, 24, 24, 128)       512
hNormalization)

activation_1 (Activation)    (None, 24, 24, 128)       0

max_pooling2d_1 (MaxPooling  (None, 12, 12, 128)       0
2D)

dropout_1 (Dropout)          (None, 12, 12, 128)       0

conv2d_4 (Conv2D)            (None, 12, 12, 256)       ...
```

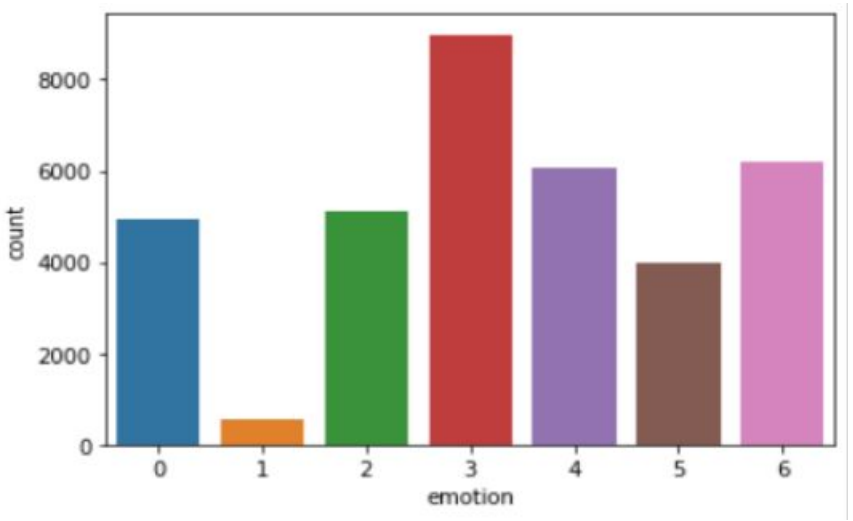
```
449/449 - 1064s - loss: 1.3119 - accuracy: 0.5003 - val_loss: 1.4064 - val_accuracy: 0.4508 - 1064s/epoch - 2s/step
```

```
113/113 [==============================] - 32s 283ms/step
VGG16 Model Accuracy on test set: 0.4377
```

RESULTS

- The given table shows the accuracy of the VGG-16 Model achieved on the FER2013 private test dataset.

S.No	Models	Accuracy
1	Basic CNN	54
2	VGG-16	60



Interpretability:

- To better understand our network's behavior, we employed various methods including Grad-CAM, XRAI, vanilla gradients, and occlusion maps.
- Running occlusion maps on correctly classified images in our web app model, we observed the network had learned to focus on the nose and mouth to make predictions for disgust, the mouth for happiness, and the eyes and nose for a surprise.
- For neutral images, we found that the network focused on all parts of the face except for the nose, which made sense given that small changes in non-nose regions tend to correspond to emotion changes.

CONCLUSION

- We explored a model including shallow CNNs and pre-trained networks based on VGG16. To reduce FER2013's inherent class imbalance, we employed class weights, data augmentation, and auxiliary datasets.
- By using seven models we achieved 60% accuracy, which is the highest to our knowledge. We also found through network interpretability that our models learned to focus on relevant facial features for emotion detection.
- The improved VGG16 network model is effective in training and testing expression recognition. With the expansion of the depth and structure of the network, the application of certain networks in some fields has shown better results.

THANK YOU