Facial Expression Recognition using Recurrent Neural Network

1st Anvesh Radharapu
Data Sceience Major
Dept. of Information Science
University of North Texas
AnveshReadharapu@my.unt.edu

2nd Sairam Mandarapu
Data Sceience Major
Dept. of Information Science
University of North Texas
sairammandarapu@my.unt.edu

3rd Reddy Narendranath Reddy
Data Sceience Major
Dept. of Information Science
University of North Texas
narednranathreddyreddy@my.unt.edu

4th Jahnav Jayanth Reddy Kukkala
Data Sceience Major
Dept. of Information Science
University of North Texas
JahnavJayanthReddyKukkala@my.unt.edu

5th Sai lohith konakanchi
Data Sceience Major
Dept. of Information Science
University of North Texas
sailohithkonakanchi@my.unt.edu

Abstract—**The goal of emotion recognition technology is to determine a person's emotions from their facial expressions. Facial expression recognition research and public interest have a long history. The Three major steps are face detection, facial feature extraction, expression classification are applied on CK+48 (Extended Cohn-Kanade dataset). For Face detection and extract high-level information from the image we use Candide-3 face model with a learned objective function, and Recurrent Neural Network in combination with convolution Network, support vector machines (SVM) are utilized in recognize and classifying facial expressions. The benefit of utilizing a recurrent network is that the classification process can taken into account the temporal relationships contained in the image sequences. The method would be great for real-time recognition because everything is automatic and recurrent networks may be utilized to create predictions online. Different businesses already utilize emotion recognition extensively to determine how customers feel about their goods, brands, marketing initiatives, employees, or on-site interactions, this technology offers options that go beyond market research and digital advertising. Our Experimental results are promising and reached an accuracy of 85% which is 5% more compared to traditional Neural Nets.**

Index Terms—**CK+, SVM, Candide-3, Reccurent Neural Network, Facial expression recognition**

I. Introduction

A. Emotion and Facial Expression

Facial expressions are a form of nonverbal communication that help individuals understand the basic emotions of sadness, happiness, fear, disgust, anger, and tiredness, etc. Facial expressions are often misunderstood. Facial expressions of an emotion are the patterned facial muscle movements that reflects the associated emotional states experienced internally. The process of learning about facial expressions allows one to develop a detailed understanding of how an emotion is expressed on the face. It involves different ways of comprehending elements like age, gender, culture, and fundamental emotions, as well as the development of a role in how an emotion is expressed through the face. Or to put it another way, certain facial expressions can be used to understand and assess emotions.

In contrast, For the basic understanding as well as better communication in between the people the facial expressions play a crucial role. The expressions such as Sadness, happiness, fear, disgust, anger, and tiredness are meant to be the basic emotions. Moreover, the concept of complicated emotions centres on the feelings that are present at the time. Complex feelings including happiness, embarrassment, envy, surprise, and grief. Complicated expressions are important in a routine, but basic expressions are the ones mostly used by the people to express facial expressions.

Despite the emphasis on facial expressions, it seems that Basic Emotion Theory is one of the most popular theories for comprehending facial expression and emotions (BET). The advantage of comprehending the theory is not just knowing what the six basic emotions are, but also comprehending the rationale behind why these basic emotions were given more attention than others. The study of basic emotions examines the connection between facial expressions and emotions as perceived by various cultural groups. So, Facial expressions can convey feelings and reveal a person's intentions in a social setting. Emotions and expressions are vital to people's ability to interact socially.

B. Facial Expression Recognition

Using a virtual image or video body from a video source, a computer program called a facial expression recognition framework may automatically identify or verify someone. The evaluation of certain facial capabilities using a face database and a photo is one method used to do this. Face discovery, the process of identifying faces in a scene, facial capacity extraction from the detected faces, and facial expression recognition are all parts of the facial expression recognition process that may be carried out by humans or computers.

In order to automatically understand and recognize human emotions like happiness, sadness, anger, fear, surprise, disgust, and so on, an automatic facial expression recognition system is being developed. This research area is currently very active in the fields of signal processing, pattern recognition, and artificial

C. Recurrent Neural Networks

Recurrent neural networks are artificial neural networks that use sequential data or time series data (RNN). Several well-known apps, like Siri, voice search, and Google Translate, use these deep learning techniques. They are frequently used for ordinal or temporal problems in speech recognition, picture captioning, and natural language processing (nlp). Recurrent neural networks (RNNs), like feedforward and convolutional neural networks (CNNs), use training data to learn.

Because of their ”memory,” which enables them to use information from earlier inputs to influence the current input and output, they stand out from other systems. Unlike normal deep neural networks, which assume that inputs and outputs are independent of one another, recurrent neural networks' outputs are dependent on the previous components in the sequence. Even though they would be helpful in determining the output of a specific sequence, unidirectional recurrent neural networks are unable to incorporate future events into their projections.

Recurrent neural networks have a connectionist design in which one or more of the network layers are connected to one another. These self connections enable the network to create an internal model of previous events, enabling it to utilize temporal context in a variety of ways. Additionally, compared to static techniques, the internal representation is typically more resilient to changes and distortions in the input sequence (such as the same expression performed at various rates).

Using a recurrent network allows the classification process to take into account the temporal relationships present in the image sequences, which is advantageous. Due to the fact that everything is automated

The network may be immediately fed the sequence of extracted Candide features because it is designed to handle temporal patterns. The ability of a recurrent neural network classifier, known as Long Short-Term Memory, to leverage the dynamic temporal behavior of the sequence for classification was presumptively used.

D. Related Work

Facial Expression Recognition(FER) has a substantial amount of research that has been carried out over the last two decades. Most of the related research on FER is usually divided into three modules, namely feature extraction, model training, feature detection. Feature Extraction module usually comes under data preprocessing phase, where numerous feature extraction techniques has been used in multiple researches to get a comparative performance of the various efficient extraction techniques. Previous works have used a range of these feature extraction techniques but the most used ones among them are Facial Acting Coding System(FACS), Principal Component Analysis(PCA), Linear Discriminant Analysis(LDA), Local Binary Pattern(LBP), Local Gradient Code(LGC), and so on. As feature extraction plays a vital role in FER, it is recommended to use the appropriate technique for the process and if in doubt, a comparative analysis based on the performance of these techniques is always helpful in making the right decision.

Coming to model fitting module, this phase is generally responsible for training a selected model with the extracted features from the images. Though numerous models are available to process this extracted features, neural network approach is the most used way in majority of relevant research for their efficient mechanism and the Convolutional Neural Networks(CNN) as well as the Recurrent Neural Networks(RNN) are the most used models for processing and training the features that are extracted in the previous module. Finally, the feature detection and classification module usually involves classification models to classify and predict the facial expression. There are only three classification models that are being used extensively for the detection, classification and prediction of the featured facial expressions in the previous research, which are Support Vector Machine(SVM), k-Nearest Neighbor(KNN), and Long Short-Term Memory(LSTM).

Facial Expression Recognition Using Frequency Neural Network [16] Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity [3] Facial emotion recognition using convolutional neural networks (FERC) [10] Emotion Recognition by Facial Features using Recurrent Neural Networks [11] Facial Expression Recognition with Recurrent Neural Networks [5] Facial expression recognition using SVM classifier [17] Facial Expression Recognition: A Survey [9] A Compact Deep Learning Model for Robust Facial Expression Recognition [13] Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network. [8] Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure [15] Facial Expression Recognition from a Single Face Image Based on Deep Learning and Broad Learning [2] Facial Image based Emotion Recognition System using Neural Network [4] A Novel Based 3D Facial Expression Detection Using Recurrent Neural Network [7] Spatial–Temporal Recurrent Neural Network for Emotion Recognition [19] facial expression recognition via deep learning [20]

II. Methodology

A. Dataset Description

The Extended Cohn-Kanade (CK+) dataset consists of 593 video clips from 123 distinct people who range in age from 18 to 50 and are of various genders and ethnic backgrounds. Each video depicts a change in face expression from neutral to a specified peak expression. It was shot at 30 frames per second (FPS) at either 640x490 or 640x480 resolution. One of the seven expression classes—anger, contempt, disgust, fear, pleasure, sorrow, and surprise—is assigned to 327 of these movies. The majority of facial expression classification algorithms employ the CK+ database, which is widely recognized as the most frequently used laboratory-controlled facial expression classification database currently available.

AU	*Name*	N	AU	*Name*	N	AU	*Name*	N
1	*Inner Brow Raiser*	173	13	*Cheek Puller*	2	25	*Lips Part*	287
2	*Outer Brow Raiser*	116	14	*Dimpler*	29	26	*Jaw Drop*	48
4	*Brow Lowerer*	191	15	*Lip Corner Depressor*	89	27	*Mouth Stretch*	81
5	*Upper Lip Raiser*	102	16	*Lower Lip Depressor*	24	28	*Lip Suck*	1
6	*Cheek Raiser*	122	17	*Chin Raiser*	196	29	*Jaw Thrust*	1
7	*Lid Tightener*	119	18	*Lip Puckerer*	9	31	*Jaw Clencher*	3
9	*Nose Wrinkler*	74	20	*Lip Stretcher*	77	34	*Cheek Puff*	1
10	*Upper Lip Raiser*	21	21	*Neck Tightener*	3	38	*Nostril Dilator*	29
11	*Nasolabial Deepener*	33	23	*Lip Tightener*	59	39	*Nostril Compressor*	16
12	*Lip Corner Puller*	111	24	*Lip Pressor*	57	43	*Eyes Closed*	9

Table 1. *Frequency of the AUs coded by manual FACS coders on the CK+ database for the peak frames.*

Emotion	Criteria
Angry	AU23 and AU24 must be present in the AU combination
Disgust	Either AU9 or AU10 must be present
Fear	AU combination of AU1+2+4 must be present, unless AU5 is of intensity E then AU4 can be absent
Happy	AU12 must be present
Sadness	Either AU1+4+15 or 11 must be present. An exception is AU6+15
Surprise	Either AU1+2 or 5 must be present and the intensity of AU5 must not be stronger than B
Contempt	AU14 must be present (either unilateral or bilateral)

Table 2. *Emotion description in terms of facial action units.*

The given data is divided into training and test datasets using the K-fold approach. The input dataset is divided several times in this procedure. The model will be assessed using a different test set in every cross-validation step because every time a different split is generated (in the case of cross-validation, it is called the validation set or development set). We decrease the possibility of over-fitting thanks to this method of producing the validation sets. We must divide the original dataset into a training set and a test set before applying K-fold. The K-fold split uses the training set as its input to create additional training and validation sets. We want to utilize the test set to predict the model's performance in a real-life scenario, so we won't be using it for K-fold validation. Stratification, a technique for choosing examples that keeps the amount of target classes in the supplied data constant.

B. Implemented Approach

Model parameters that best represent the face in a given image must be estimated in order to retrieve high-level information. This task is resolved through model fitting, which frequently involves minimizing an objective function that assesses how well a model parameterisation matches a particular image. The goal function, such as the pixel difference between the generated surface of the model and the underlying image content, is frequently created manually. Contrarily, we suggest that the objective function be learned from annotated According to Candide-3 face model [1], these parameters are calculated using learnt objective functions. Human face data is extracted using model-based picture interpretation algorithms. Models force the user to know information about the item of interest and distill high-dimensional visual data into a few expressive model parameters. The approach used in our model is the deformable 3D wire frame Candide-3 face model [1]. The Candide parameter vector,

$$p = (rx, ry, rz, s, tx, ty, \sigma, \alpha)^T$$

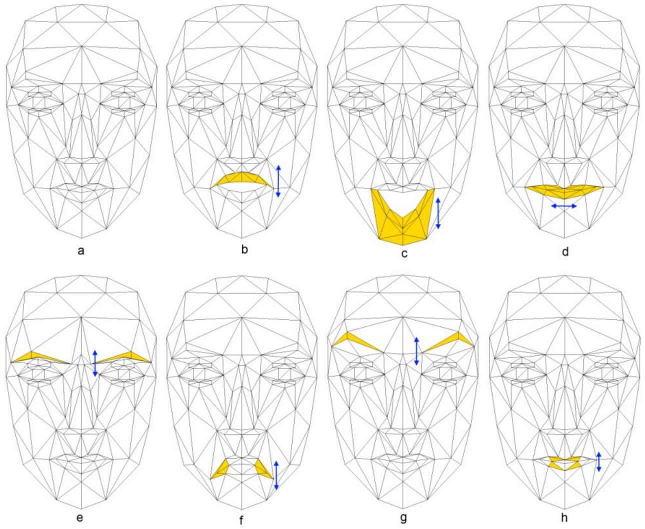
explains the linear transformation

$$(rx, ry, rz, s, tx, ty)$$

and the deformation

$$(\sigma, \alpha)$$

. Through a set of 116 anatomical markers, the 79 deformation parameters reveal the shape of face features like the lips, the eyes, or the brows.



Facial expression recognition is typically broken down into three smaller tasks, face detection, feature extraction, and expression classification [12]

C. Objective Function

In order to extract high-level information, model parameters have to be estimated that best describe the face within a given image. Model fitting solves this task and is often addressed by minimising an objective function that evaluates how well a model parameterisation fits a given image. The objective function is often designed manually, such as the pixel error between the model's rendered surface and the underlying image content. In contrast, we propose to learn the objective function from annotated These parameters are estimated with help of learned objective functions as described in [18]

D. Face Model Fitting

Model parameters that best represent the face in a given image must be calculated in order to retrieve high-level information. This job is resolved through model fitting, which frequently involves minimizing an objective function that assesses how well a model parameterisation matches a particular picture. The goal function, such as the pixel difference between the generated surface of the model and the underlying picture content, is frequently created manually. Contrarily, we suggest that the goal function be learned from annotated

How well a parameterized model p fits to an image I is determined by the objective function f(I, p), which produces a similar value. The model parameters p that optimize the objective function are sought for by the fitting method. For a current summary and classification, this study will refrain from going into further detail on them. The key element of the model fitting process, the objective function, is frequently created manually utilizing the designer's subject expertise and intuition. After that, its suitability is evaluated subjectively by looking at the output on sample photographs and model parameter

E. Feature Extraction

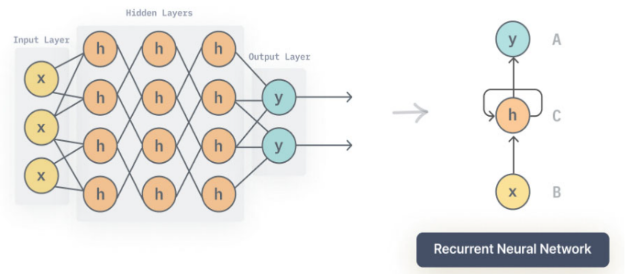
Action Units (AUs) identify the face muscles used to move certain facial areas. The visible face's structure and the locations of the 116 landmarks are described by the deformation parameters (σ,α). The examples in Figure below show how the value of (σ,α) and the facial expression are related. So, in our opinion, (σ,α) contributes high-level knowledge to the interpretation process. We fit the model to each image in the sequence and extract the model parameters to serve as training data for the classification phase. Keep in mind that the extracted characteristics only apply to a single image, however they might be acquired for a collection of independent photos as well. However, the classifier will take use of the fact that the characteristics were taken from pictures that were connected semantically.

F. Classification

We want to classify the face expression based on the sequence of retrieved features. This can be accomplished directly by producing a single class for the entire sequence, or indirectly by first combining the individual frames of the photo sequence from the database to obtain the overall expression. Because the categorization is only provided at the end of the sequence, real-time recognition is likewise impossible.

As previously stated, face expressions are intrinsically temporal. The majority of classifiers, however, such as support vector machines, decision trees, and feedforward neural networks, are developed for static patterns. This means that the series of input features must be preprocessed into a single, fixed-length vector before they can be used as expression classifiers. This strategy, in addition to demanding great effort on the part of the experimenters, frequently discards significant temporal relationships.

Recurrent neural networks is a type of network where one or more network layers are connected to itself as shown below



The network's self connections enable it to construct an internal representation of past occurrences, allowing it to make flexible use of temporal context. Furthermore, the internal representation is more resistant to shifts and distortions in the input sequence (for example, the same expression enacted at various speeds) than static approaches. Because the network is built to detect temporal patterns, the sequence of retrieved Candide features can be given directly to it. The use of long short-term memory (LSTM) [6] cells is one improvement to the fundamental recurrent architecture. LSTM cells, as seen in Figure below, use linear units protected by multiplicative gates to retain information over long periods of time. This broadens the scope of temporal context available to the internet.

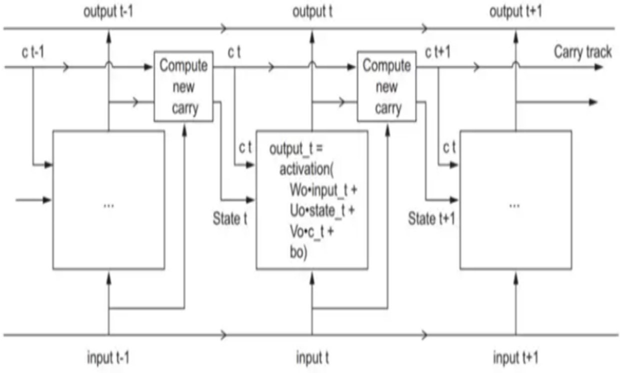
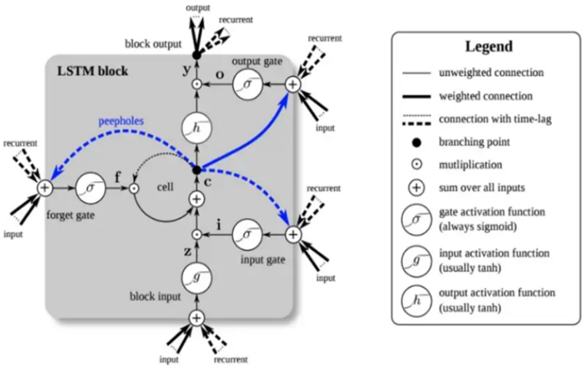


Figure 6.15 Anatomy of an LSTM

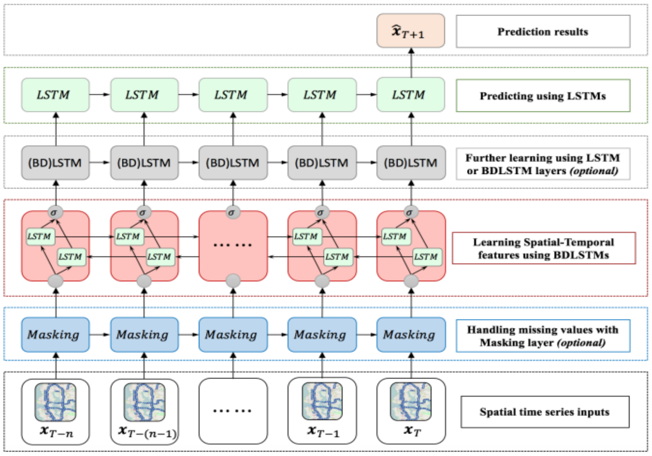
LSTM (Long Short Term Memory) Cell (Source)

The application of bidirectional recurrent networks is to provide both future and past context Bidirectional networks, as shown in Figure below, scan the same sequence forwards and backwards with two independent hidden layers, both of which are coupled to the same output layer. As a result, the output classifications can be determined in light of all surrounding context in the input sequence.



A Bi-Directional LSTM Cell (Source)

A bidirectional LSTM network with 128 layers in both the forward and backward hidden layers was utilized completely, as a unidirectional LSTM network in the hidden layer. All hidden layers were completely connected to each other, as well as to the input and output layers of the network.



III. Experimental Results

The Extended Cohn-Kanade face expression recognition database was used to test our system. The challenge was to categorize each of the video sequences generated by merging three instances of the visuals at each certain time for an expression into one of the five types of standard expressions: happiness, anger, sadness, fear, and surprise.

A. Data

The Cohn-Kanade Facial Expression Database is publicly available and referenced by several research groups [14] It features short image sequences of roughly 100 people of both genders ranging in age from 18 to 30 years. The image sequences begin with the neutral face and progress to the peak expression. All photographs are taken from the front. The database also includes AU activation details as well as the portrayed face expression.

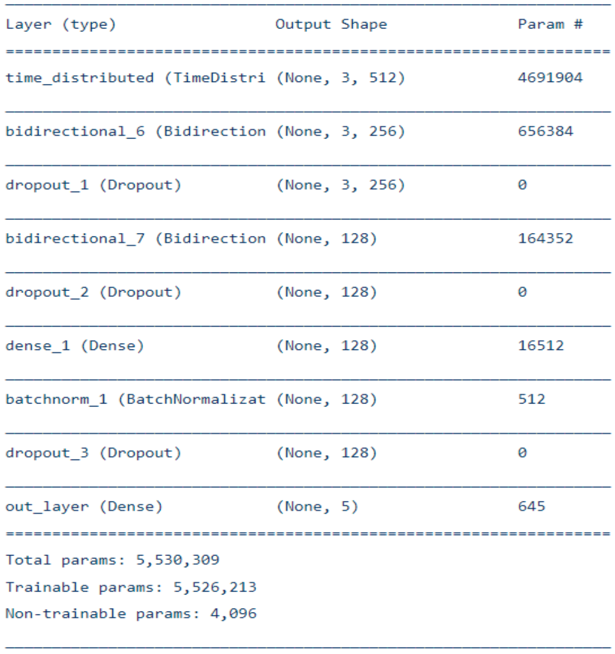
B. Classifier Parameters

A bidirectional LSTM network with 128 layers and also 64 layer Bidirectional LSTM layer in both the forward and backward hidden layers was utilized, as was a unidirectional LSTM with in the hidden layer. Each network contained 35 input units and six output units, one for each target class. All hidden layers were completely connected to each other, as well as to the input and output layers.

The output layer had a softmax activation function and was trained for classification using the cross-entropy objective function. The data collection contains the last three frames of each such video clip. As a result three frames associated with each video clip must be combined into a single short three-frame video clip (although this clip is very short, in milliseconds, because a single second contains 30 frames in a 30fps video). We will then feed these three-frame films into our model.

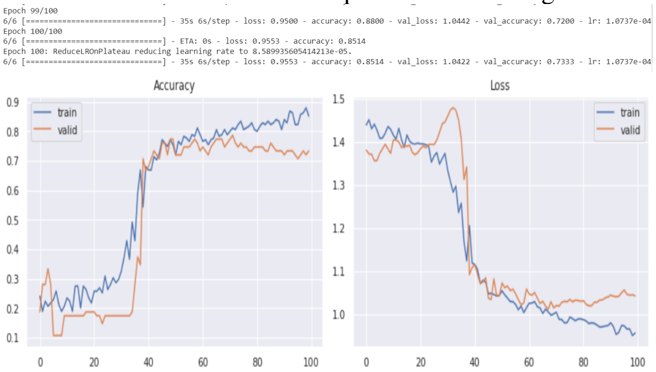
Dropouts are employed at regular intervals for generalization and ELU is selected as the activation function since it not only avoided the dying relu problem, but it also outperformed LeakyRelu. he_normal is chosen as the kernel initializer because it is compatible with ELU.

For improved results, batch normalization is also performed. For a time-distributed CNN model we stacked few Bidirectional LSTMs over a unidirectional LSTM and finally a few dense layers. To the dence layer we added a SVM classifier for more accuracy with the help of kernel_regularizer=tf.keras.regularizers.l2(0.01) and squared_hinge in the loss factor of model compiler. The Compiled model parameters as shown below



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
time_distributed (TimeDistri (None, 3, 512)            4691904
_________________________________________________________________
bidirectional_6 (Bidirection (None, 3, 256)            656384
_________________________________________________________________
dropout_1 (Dropout)          (None, 3, 256)            0
_________________________________________________________________
bidirectional_7 (Bidirection (None, 128)               164352
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               16512
_________________________________________________________________
batchnorm_1 (BatchNormalizat (None, 128)               512
_________________________________________________________________
dropout_3 (Dropout)          (None, 128)               0
_________________________________________________________________
out_layer (Dense)            (None, 5)                 645
=================================================================
Total params: 5,530,309
Trainable params: 5,526,213
Non-trainable params: 4,096
_________________________________________________________________
```

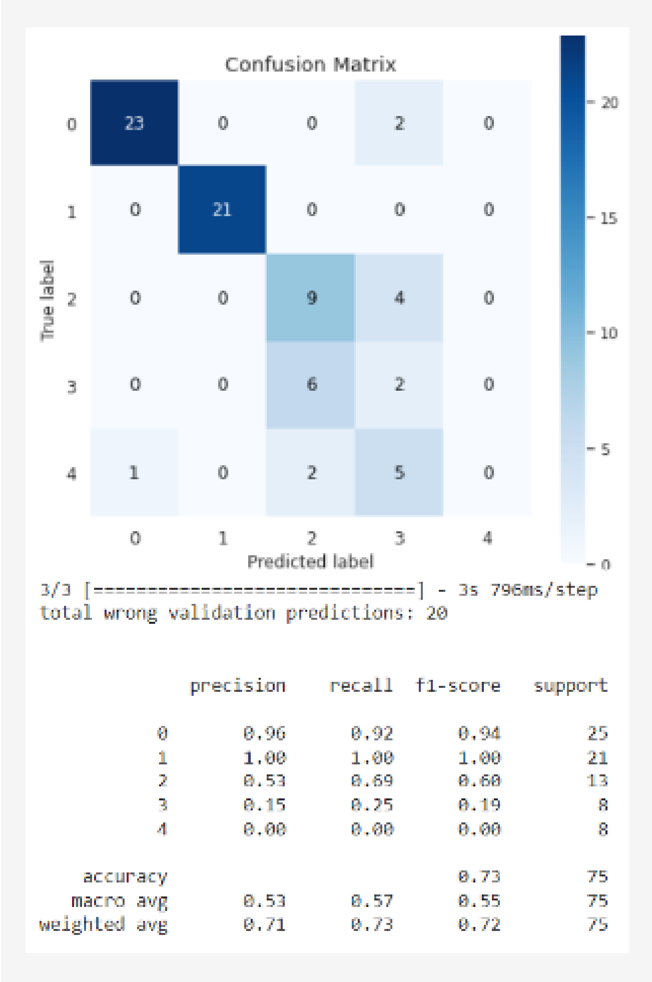
The epoch's history reveals that accuracy gradually improves and reaches +85% on both the training and validation sets. ReduceLROnPlateau is also termed anytime the accuracy reaches a plateau. The fluctuations in the epoch metrics are due to a lack of data for such a complex task and it is given below



```
Epoch 99/100
6/6 [==============================] - 35s 6s/step - loss: 0.9500 - accuracy: 0.8800 - val_loss: 1.0442 - val_accuracy: 0.7200 - lr: 1.0737e-04
Epoch 100/100
6/6 [==============================] - ETA: 0s - loss: 0.9553 - accuracy: 0.8514
Epoch 100: ReduceLROnPlateau reducing learning rate to 8.589935605414213e-05.
6/6 [==============================] - 35s 6s/step - loss: 0.9553 - accuracy: 0.8514 - val_loss: 1.0422 - val_accuracy: 0.7333 - lr: 1.0737e-04
```

C. Results

The confusion-matrix is a popular evaluator for multi-class classification. It provides us with a solid overview of the model's performance across all classes.



Confusion Matrix

True label \ Predicted label	0	1	2	3	4
0	23	0	0	2	0
1	0	21	0	0	0
2	0	0	9	4	0
3	0	0	6	2	0
4	1	0	2	5	0

```
3/3 [==============================] - 3s 796ms/step
total wrong validation predictions: 20


              precision    recall  f1-score   support

           0       0.96      0.92      0.94        25
           1       1.00      1.00      1.00        21
           2       0.53      0.69      0.60        13
           3       0.15      0.25      0.19         8
           4       0.00      0.00      0.00         8

    accuracy                           0.73        75
   macro avg       0.53      0.57      0.55        75
weighted avg       0.71      0.73      0.72        75
```

It's worth noting that the classifier learned the training set quickly in both situations, with nearly all sequences properly classified. This suggests that the main challenge of the task is generalizing from training to test cases. With such a tiny number of training data, generalization is especially tough. We anticipate a significant improvement in performance if a larger dataset is employed.

D. Conclusion and Future Work

We conclude this work presenting an approach for facial expression estimation that combines a convolution-Recurrent Neural Net model which predicts 5 Facial Expressions with SVM as a classifier included.

Future study will present classifier training data taken from multiple publicly available databases in order to reflect a greater range of face expressions. Furthermore, our approach will be tailored for real-time applicability. This approach will be used to produce a working demonstrator. The goal of any model is to test and apply it in the actual world, not merely train and validate it. I advanced my project significantly by experimenting with numerous models and emotion classes.

References

[1] Jörgen Ahlberg. Candide-3 - an updated parameterised face. 2001.

[2] Mei Bie, Huan Xu, Yan Gao, and Xiangjiu Che. Facial expression recognition from a single face image based on deep learning and broad learning. *Wireless Communications and Mobile Computing*, 2022, 2022.

[3] Tanoy Debnath, Md Mahfuz Reza, Anichur Rahman, Amin Beheshti, Shahab S Band, and Hamid Alinejad-Rokny. Four-layer convnet to facial emotion recognition with minimal epochs and the significance of data diversity. *Scientific Reports*, 12:2045–2322, 2022.

[4] Parnal Dudul, Shital Tayade, and Ajay K. Talele. Facial image based emotion recognition system using neural network. *International Journal of Advanced Research in Computer and Communication Engineering*, 6:52–66, 2017.

[5] Alex Graves, Christoph Mayer, Matthias Wimmer, Jürgen Schmidhuber, and Bernd Radig. Facial expression recognition with recurrent neural networks. In *Proceedings of the International Workshop on Cognition for Technical Systems*, 2008.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[7] KS Jaswanth and D Stalin David. A novel based 3d facial expression detection using recurrent neural network. In *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, pages 1–6. IEEE, 2020.

[8] Ji-Hae Kim, Byung-Gyu Kim, Partha Pratim Roy, and Da-Mi Jeong. Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. *IEEE access*, 7:41273–41285, 2019.

[9] Jyoti Kumari, Reghunadhan Rajesh, and KM Pooja. Facial expression recognition: A survey. *Procedia computer science*, 58:486–491, 2015.

[10] Ninad Mehendale. Facial emotion recognition using convolutional neural networks (ferc). *SN Applied Sciences*, 2(3):1–8, 2020.

[11] Amr Mostafa, Mahmoud I. Khalil, and Hazem Abbas. Emotion recognition by facial features using recurrent neural networks. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 417–422, 2018.

[12] Maja Pantic and Leon J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1424–1445, 2000.

[13] R Reji and P Sojan Lal. A compact deep learning model for robust facial expression recognition. *International Journal of Engineering and Advanced Technology*, 8(6):2956–2960, 2019.

[14] Roland Schweiger, Pierre Bayerl, and Heiko Neumann. Neural architecture for temporal emotion classification. In *Tutorial and Research Workshop on Affective Dialogue Systems*, pages 49–52. Springer, 2004.

[15] Jie Shao and Yongsheng Qian. Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, 355:82–92, 2019.

[16] Yan Tang, Xingming Zhang, Xiping Hu, Siqi Wang, and Haoxiang Wang. Facial expression recognition using frequency neural network. *IEEE Transactions on Image Processing*, 30:444–457, 2021.

[17] PC Vasanth and KR Nataraj. Facial expression recognition using svm classifier. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 3(1):16–20, 2015.

[18] Matthias Wimmer, Freek Stulp, Sylvia Pietzsch, and Bernd Radig. Learning local objective functions for robust face model fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1357–1370, 2008.

[19] Tong Zhang, Wenming Zheng, Zhen Cui, Yuan Zong, and Yang Li. Spatial–temporal recurrent neural network for emotion recognition. *IEEE transactions on cybernetics*, 49(3):839–847, 2018.

[20] Xiaoming Zhao, Xugan Shi, and Shiqing Zhang. Facial expression recognition via deep learning. *IETE technical review*, 32(5):347–355, 2015.