CSCE 4604 - Advanced Machine Learning

Facial Expression Recognition
Project Proposal

Dr. Moustafa Youssef
Malak Gaballa - 900201683
Masa Tantawy - 900201312

Table of Contents

Introduction and Problem Statement

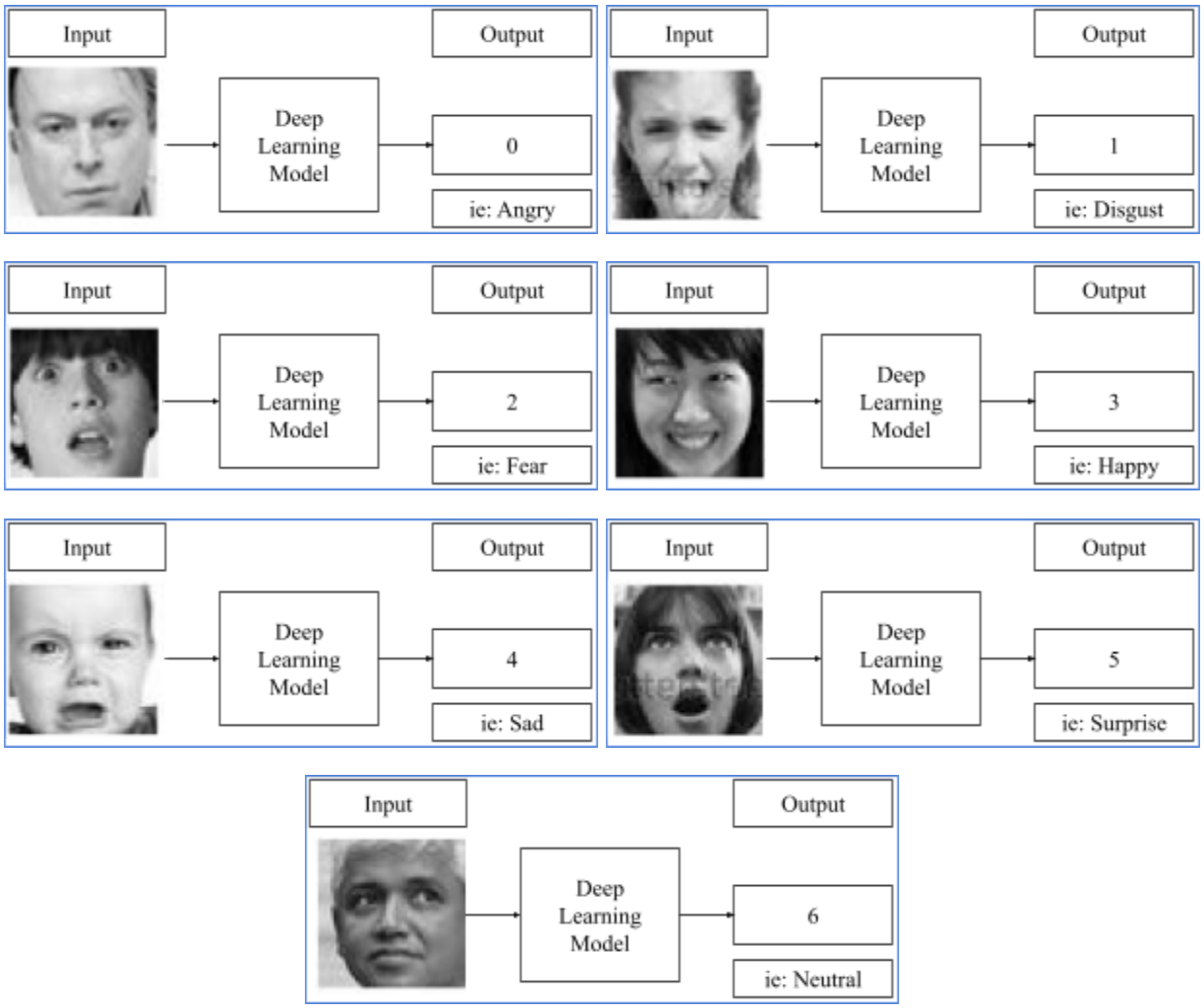
Human communication is a very complex process that involves multiple elements. Among these elements, facial expressions are considered to play a critical role. This is because just like verbal expression, facial expression can be used to determine the emotional state of humans. Particularly, "Face changes during a communication are the first signs that transmit the emotional state, which is why most researchers are very interested by this modality"(source). Facial expressions recognition (FER) has a diverse range of applications including in the fields of robotics, surveillance, or human-computer interaction systems (source). Applications also extend to biometrics, detection of mental illness, understanding of human behaviour, and psychological profiling (source).

Accordingly, it is crucial to leverage advances in deep learning to be able to categorise facial expressions from existing images of human faces. However, this problem is challenging as facial expressions vary tremendously across individuals; each human expresses emotions differently. Multiple factors such as human age, gender, or race can be obstacles or even other factors such as background, sunglasses, or scarves can hinder the process of accurate recognition of facial expressions (source).

The proposed problem statement is using deep learning to identify facial expressions from images. Given images of human faces showing different expressions, the model should be able to categorise each image into one of seven categories, each representing a facial expression. These are: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral); the images below from the FER2013 dataset each represent one of the categories respectively.

Figure 1: Images from each emotion class in the FER2013 dataset.

The input to the model is an image, expressed as a vector of pixels for a 48x48 pixel grayscale image, and the model's output should be a number from 0 to 6 which indicates the facial expression illustrated in the image. For example:



The primary goal of the model is to be able to accurately classify each image to a facial expression via outputting the correct label. The model results should also be generalizable, such that it is able to correctly identify the facial expression given the image of any individual regardless of their age, gender, or ethnicity.

Evaluation Metrics

The only metric found that is used to evaluate the performance of different models on FER is accuracy. Accuracy is used to determine how likely the model is to correctly classify an instance (Source). This is done, as shown below, by calculating the ratio of correctly classified instances from all the dataset, which is a number on a scale of 0 to 1 or as a percentage.

$$Accuracy = \frac{number\ of\ correct\ classifications}{total\ number\ of\ classifications}$$

In order to calculate the number of correct classifications, the confusion matrix shown below is used. This matrix is applied on binary classification, such that there are 2 possible labels to each data point - positive or negative, however this can be generalised to multinomial classification as each class can be broken down into a binary classification. In the case of FER, if the model's output is equal to the image's actual label, this instance is correctly classified. If the model's output is not the same as the image's actual label, this instance is incorrectly classified.

	Actual Value		
Predicted Values		**Positive (1)**	**Negative (0)**
	Positive (1)	True Positive (TP)	False Positive (FP)
	Negative (0)	False Negative (FN)	True Negative (TN)

- **TP:** The number of data points with actual labels as positive and predicted as positive.
- **TN:** The number of data points with actual labels as negative and predicted as negative.
- **FP:** The number of data points with actual labels as negative and predicted as positive; this is known as type 1 error.
- **FN:** The number of data points with actual labels as positive and predicted as negative; this is known as type 2 error.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

However, one problem with this metric is that it may be misleading if the data is imbalanced as it only provides an overall evaluation of the model. Particularly, it does not take into account the proportion of misclassification for each individual class (source). Hence, this problem is avoided by calculating the *weighted accuracy* such that the size of each class is taken into account.

Current State-of-the-Art Results

As aforementioned, accuracy is the only and most common evaluation metric in the problem of FER. Hence, the accuracy for different deep learning models that have worked on this problem will be stated. It is important to note that FER has been gaining popularity thus many models have been implemented, among which some were able to achieve state-of-the-art (SOTA) results in literature. Each model described below claims to have achieved SOTA accuracy in FER, which was the case when the corresponding paper was released, yet at the moment, the first model below has achieved the best performance, making it the SOTA model.

Ensemble ResMaskingNet with 6 other CNNs

- Research Paper: Facial Expression Recognition using Residual Masking Network (Not Accessible)
- Repository: Github link
- Frameworks: PyTorch

The *Ensemble ResMaskingNet with 6 other CNNs* was able to achieve an accuracy of 76.82% in 2021, a state-of-the-art performance**.** To achieve SOTA accuracy, the public dataset *FER2013* was used in addition to extra training done using private *VEMO* dataset.

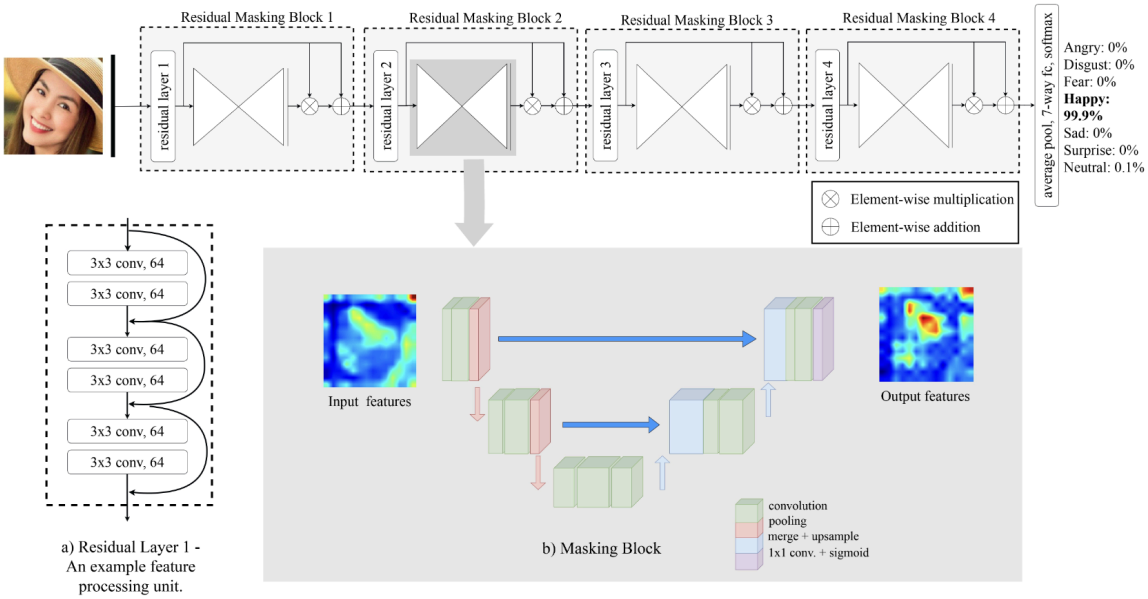
This model is mainly an ensemble of the Residual Masking Network, which is explained below, in addition to 6 other convolution Neural Networks (source).

Residual Masking Network

Although this model does not have the best performance, it is incorporated of the previous model which has the best performance (the research paper, respiratory and frameworks are the same). Thus, it is included in details although the. The *Residual Masking Network*, introduced in the same paper mentioned above, was able to achieve an accuracy of 74.14% in 2021 using the public dataset *FER2013* and extra training done using private *VEMO* dataset. This model was also benchmarked on another dataset named *ImageNet* and achieved a Top-1 Accuracy of 74.16% and a Top-5 Accuracy of 91.91%

This model follows a deep architecture with the attention mechanism; a novel Masking Idea was proposed to boost the performance of convolution neural networks (CNN) in FER. “It uses a segmentation network to refine feature maps, enabling the network to focus on relevant

information to make correct decisions" (source). A combination of Deep Residual Network and Unet-like architecture produce the Residual Masking Network, whose architecture is shown below.



EmoNeXt

- Research Paper: EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition (Not Accessible)
- Repository: Github link
- Frameworks: PyTorch

The model *EmoNeXt* was able to achieve an accuracy of 76.12% in 2023, claiming to achieve superiority of our model over SOTA deep learning models on the FER2013 dataset for emotion classification accuracy. To achieve this accuracy, the public dataset *FER2013* was used without any extra training.

This novel deep learning framework for FER is an adapted ConvNeXt architecture network. Spatial Transformer Network (STN) was integrated to focus on feature-rich regions of the face and Squeeze-and-Excitation blocks to capture channel-wise dependencies; this is in addition to the introduction of a self-attention regularization term, encouraging the model to generate compact feature vectors (source). More details about this model's architecture cannot be obtained as the research paper is not accessible.

Datasets

The realm of datasets for facial expression recognition is experiencing a significant boom, reflecting the growing interest and advancements in the field of computer vision. What's particularly intriguing is the diversity in the formats of these datasets, with some comprising raw images, others presented in CSV format with each image represented as a vector of pixels, and some offering both. Amidst surveying the available datasets, five datasets have emerged as prevalent fixtures in research papers: FER2013, CK+, JAFFE, AffectNet, and ExpW. Each of these datasets offers unique insights and challenges, making them invaluable resources for researchers aiming to push the boundaries of facial expression recognition technology.

Facial Expression Recognition 2013 (FER2013)

FER2013, short for Facial Expression Recognition 2013, is a widely used dataset containing 35,887 facial grayscale images restricted to the size 48x48 pixels. The main labels of the images are divided into 7 categories or facial expressions : Angry (4,953), Disgust (547), Fear (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral (6,198). Thus, the dataset is highly imbalanced since the expressions are not equally distributed. This dataset is available in both formats, raw images and csv, with storage size of 63 MBs. Originally, the dataset was split into a training set containing 28,709 images and a testing set containing 7,178 images. A sample of the available images are provided below and in the following link.

Extended Cohn-Kanade (CK+)

The extended Cohn-Kanade (CK+) facial expression database is a public dataset for action unit and emotion recognition. It includes both posed and non-posed (spontaneous) expressions. The CK+ comprises a total of 593 video sequences across 123 subjects, ranging from 18 to 50 years of age with a variety of genders and heritage. Each sequence of images contains 10 to 60 frames of a subject transitioning from neutral to the target emotion and each frame is roughly 640x480 with grayscale and/or color values. From the 593 frames or sequences, 327 are labelled with one of seven expression classes: anger (135), contempt(54), disgust(177),

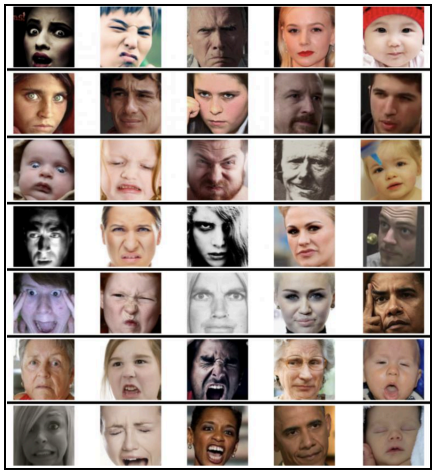
fear(75), happiness(207), sadness(84), and surprise(249). The dataset storage size is 4 MB with 981 images in total. A sample of the the available images are provided below and in the following link.

JAFFE

The Japanese Female Facial Expression (JAFFE) is a relatively small dataset containing 213 images of 10 Japanese female models. Similar to FER2013, the images are labeled with 7 facial expressions. Each subject was asked to do 7 facial expressions (6 basic facial expressions and neutral) and the images were annotated with average semantic ratings on each facial expression by 60 annotators. The dataset wasn't available anywhere except this link. When requested to gain access to the data, it was immediately declined since it cannot be provided for undergraduate projects. Therefore, we weren't able to get a hold of the dataset storage size,examples size or splits sizes that were defined in the papers. However, a sample of the images is provided.

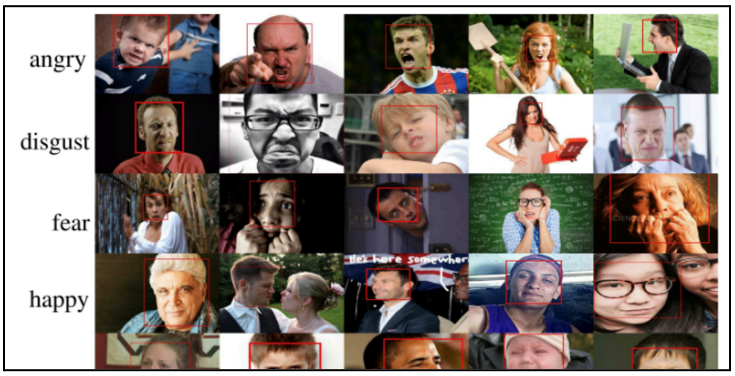
AffectNet

AffectNet is a large facial expression dataset that contains more than one million images manually labeled for the presence of eight (neutral, happy, angry, sad, fear, surprise, disgust, contempt) facial expressions along with the intensity of valence and arousal.This is the largest available dataset; however, only a sample of the data is available online while the real data could only be requested by lab managers or professors from this link. The available sample only covers 5,10, and 15 year olds which took up a dataset storage size of 5 GBs. This dataset can be found here. It contains 1822 angry faces, 1833 contempt, 1740 disgust, 1839 fear, 1862 happy, 1880 neutral , 1821 sad and 1851 surprised faces which makes the dataset slightly balanced.



Expression in-the-Wild (ExpW)

The Expression in-the-Wild (ExpW) dataset is for facial expression recognition and contains 91,793 faces manually labeled with expressions. Each of the face images is annotated as one of the seven basic expression categories: “angry”, “disgust”, “fear”, “happy”, “sad”, “surprise”, or “neutral”. The dataset storage size is approximately 8 GBs without any training/validation/testing splits. Raw images are available and a csv folder is available listing all the pictures with its corresponding expression category. An example is shown below and the link for downloading the data is provided.



Selected Dataset

Considering all available datasets, FER2013 emerges as the preferred choice for our research endeavors. One of its standout features is its complete availability on open-source platforms without any access restrictions. This accessibility makes it incredibly convenient for researchers like us to work with. Moreover, FER2013 is extensively utilized in numerous research papers, featuring multiple models, which facilitates benchmarking and comparison for proposing improvements. With a collection of 35,887 facial grayscale images, each restricted to the size of 48x48 pixels, FER2013 offers a comprehensive representation of facial expressions. These expressions are categorized into seven main labels: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. However, it's important to note that the dataset suffers from imbalanced distribution among these categories. Despite this, FER2013 is available in both raw image and CSV formats, with a manageable storage size of 63 MBs. Originally split into a training set of 28,709 images and a testing set of 7,178 images, FER2013 provides a solid foundation for our research endeavors in facial expression recognition. Observations from csv file can be found above.

# emotion	A Usage	A pixels
		153 158 171...
0	Training	151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 178 185 185 189 187 186 193 194 185 183 ...
2	Training	231 212 156 164 174 138 161 173 182 200 106 38 39 74 138 161 164 179 190 201 210 216 220 224 222 218...
4	Training	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59 65 12 20 63 99 98 98 111 75 62 41 ...

Models and Solutions

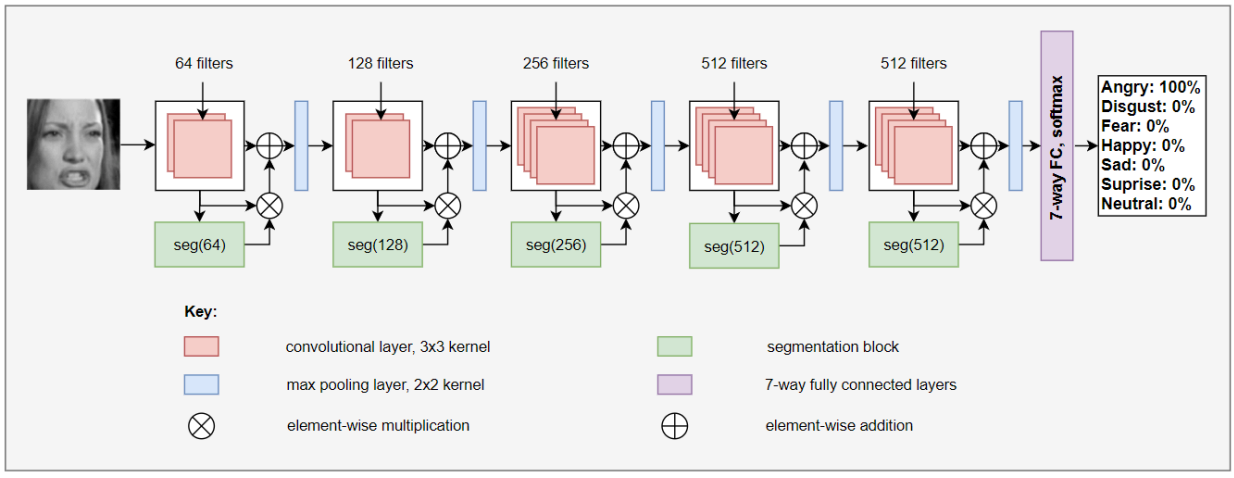
After explaining the available datasets for the FER problem and the best models in the Current State-of-the-Art Results, other implemented models and solutions will be discussed. Next, our baseline model will be chosen depending on the comparative analysis at the end of this survey. To avoid redundancy, the SOTA models will not be discussed in this section, but are included in the comparative analysis.

Segmentation VGG-19

- Research Paper: A novel facial emotion recognition model using segmentation VGG-19 architecture (Not Accessible)
- Repository: Github link
- Frameworks: PyTorch

The model *Segmentation VGG-19* was able to achieve an accuracy of 75.97% in 2023, claiming to achieve SOTA single network accuracy compared with other well-known FER models on the FER2013 dataset. To achieve this accuracy, the public dataset *FER2013* was used plus extra training on the extended Cohn-Kanade (CK+).

This model proposed a novel CNN architecture, shown below, by interfacing U-Net segmentation layers in-between Visual Geometry Group (VGG) layers to allow the network to emphasize more critical features from the feature map, which also controls the flow of redundant information through the VGG layers.



Ensemble of 7 models

- Research Paper: Facial Expression Recognition with Deep Learning
- Repository: Github link
- Frameworks: TensorFlow

In 2020, an ensemble model created by students was able to achieve a classification accuracy of 75.8% on the FER2013 test set. To achieve this performance, the *FER2013* dataset was used as the main dataset, extra training was done using *CK+* and *JAFFE* as auxiliary datasets in addition to a privately created dataset.

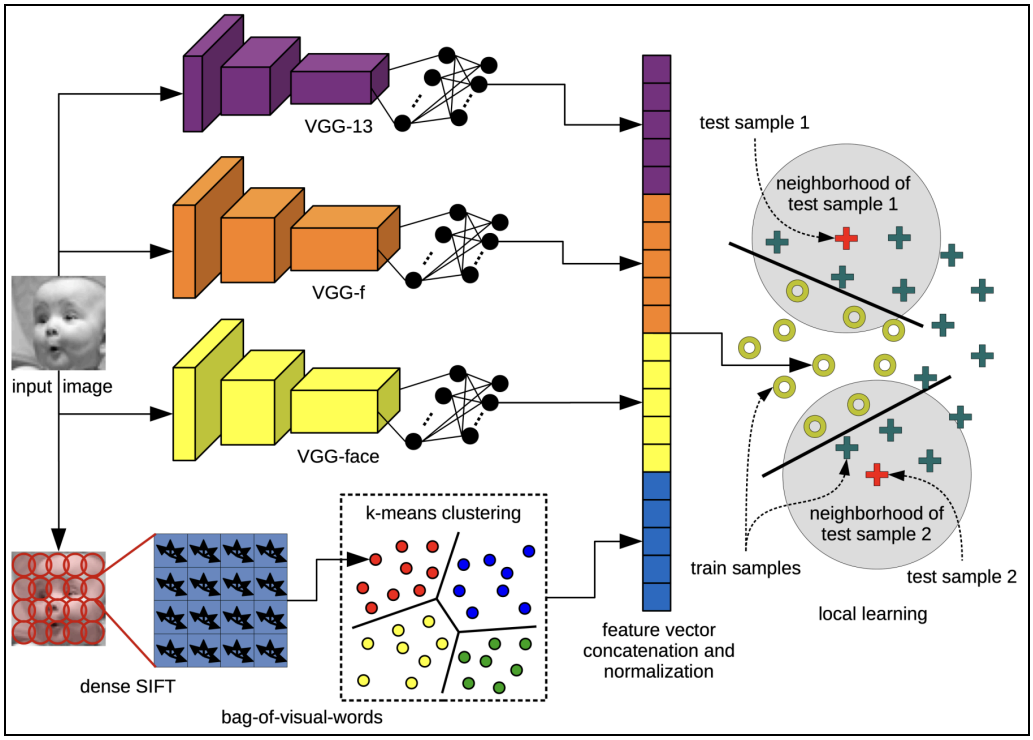
This model is an ensemble of seven different models including shallow CNNs and pre-trained networks based on SeNet50, ResNet50, and VGG16. Also, this performance was achieved via employing class weights due to the dataset class imbalance, data augmentation, and auxiliary datasets.

Local Learning Deep + BOVW

- Research Paper: Local Learning with Deep and Handcrafted Features for Facial Expression Recognition
- Repository: Not Available
- Frameworks: Not Available

The model *Local Learning Deep+BOVW* was able to achieve an accuracy of 75.42% in 2018. This accuracy was achieved on the *FER2013* dataset; accuracies achieved on other datasets were 87.76% on the *FER+*, 59.58% on *AffectNet 8-way classification* and 63.31% on *AffectNet 7-way classification*. Hence, it is claimed that this model surpasses SOTA methods by more than 1% on all data sets. Also, extra training was used.

This model combines automatic features learned by three CNN models with handcrafted features computed by the bag-of-visual-words (BOVW). A local learning framework is used for classification, consisting of 3 steps. First, a k-nearest neighbors model is applied in order to select the nearest training samples for an input test image. Second, a one-versus-all Support Vector Machines (SVM) classifier is trained on the selected training samples. Finally, the SVM classifier is used to predict the class label only for the test image it was trained for. The model architecture is shown below.



VGG, Res-Net, and Inception

- Research Paper: Facial Expression Recognition using Convolutional Neural Networks: State of the Art
- Repository: Github link
- Frameworks: Not Available

In a comparative study in 2016, multiple image-based FER models that are CNNs were compared. Trained on the *FER2013* dataset without using extra training data, *VGG* achieved a classification accuracy of 72.7%, *Res-Net* achieved a classification accuracy of 72.4%, and *Inception* achieved a classification accuracy of 71.6%.

Through overcoming one of bottlenecks of CNNs - the comparatively basic architectures of the CNNs utilized in FER, this paper introduced an ensemble of modern deep CNNs that was able to obtain a FER2013 test accuracy of 75.2%.

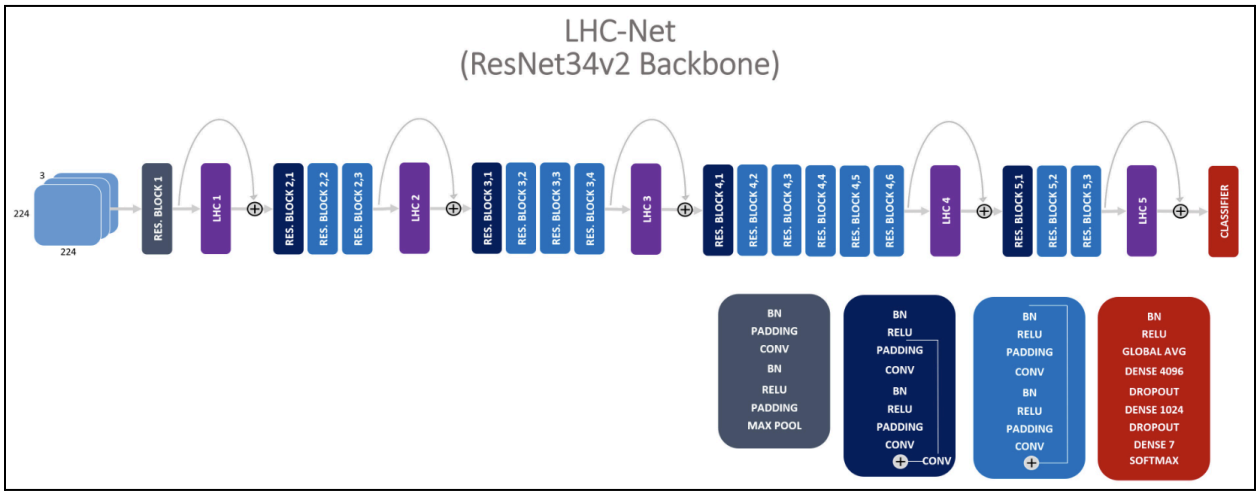
LHC-Net

- Research Paper: Local Multi-Head Channel Self-Attention for Facial Expression Recognition
- Repository: Github link

- Frameworks: TensorFlow

The model *LHC-Net* was able to achieve an accuracy of 74.42% in 2021. To achieve this accuracy, the public dataset *FER2013* was used plus extra training. This model claims to achieve a new state of the art in this famous dataset with a significantly lower complexity and impact on the "host" architecture in terms of computational cost when compared with the previous SOTA.

LHC-Net is a novel self-attention module that can be easily integrated in virtually every convolutional neural network and that is specifically designed for computer vision. The LHC: Local (multi) Head Channel (self-attention) model relies on 2 main ideas: first, leveraging the self-attention paradigm through the channel-wise application instead of the more explored spatial attention; second, a local approach has the potential to better overcome the limitations of convolution than global attention. The architecture of LHC-Net, shown below, likely consists of convolutional layers for feature extraction, followed by the proposed local multi-head channel self-attention mechanism. This is likely followed by additional layers for further feature processing and classification, possibly including fully connected layers and softmax classification.

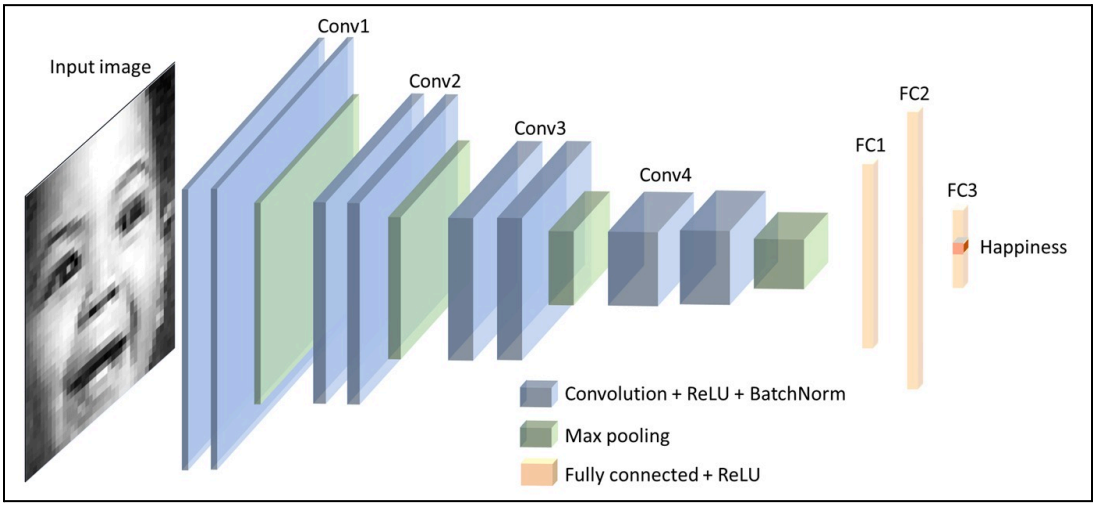


VGGNet

- Research Paper: [Facial Emotion Recognition: State of the Art Performance on FER2013]
- Repository: [Github link]
- Frameworks: PyTorch

The model *VGGNet* achieved an accuracy of 73.28% in 2021 on the *FER2013* dataset without using extra training data based on a single-network classification. This model adopts an

optimised VGGNet architecture, with fine-tuned hyperparameters. VGGNet,short for Visual Geometry Group Network, is a classical convolutional neural network architecture used in large-scale image processing and pattern recognition.The network consists of 4 convolutional stages and 3 fully connected layers. Each of the convolutional stages contains two convolutional blocks and a max-pooling layer. The convolution block consists of a convolutional layer, a ReLU activation, and a batch normalization layer. Batch normalization is used to speed up the learning process, reduce the internal covariance shift, and prevent gradient vanishing or explosion. The first two fully connected layers are followed by a ReLU activation. The third fully connected layer is for classification. The convolutional stages are responsible for feature extraction, dimension reduction, and non-linearity. The fully connected layers are trained to classify the inputs as described by extracted features.



Methods		Testing Accuracy
Trained VGGNet		73.06 %
Regular split	CosineWR	72.64 %
	Cosine	73.11 %
Combine training and validation	CosineWR	73.14 %
	Cosine	**73.28 %**

CNN Hyperparameter Optimisation

Through optimising the hyperparameters of CNN in 2021, this model's classification accuracy is 72.16%, trained on the *FER2013* dataset without extra training. The optimum hyperparameter values were found through the Random Search algorithm applied on a search space defined by discrete values of hyperparameters (source).

- Research Paper: Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition (Not Accessible)
- Repository:Github link
- Frameworks: TensorFlow

Ad-Corre

- Research Paper: Ad-Corre: Adaptive Correlation-Based Loss for Facial Expression Recognition in the Wild
- Repository: Github link
- Frameworks: TensorFlow

This model achieved a classification accuracy of 72.03% in 2022 through training on the datasets *AffectNet, RAF-DB*, and *FER-2013* and no auxiliary datasets.

In this model, the Adaptive Correlation (Ad-Corre) Loss is proposed, such that embedded feature vectors with high correlation for within-class samples and less correlation for between-class samples are generated. The model's backbone is Xception network and it has 3 elements: Feature Discriminator, Mean Discriminator, and Embedding Discriminator.

The Feature Discriminator component guides the network to create the embedded feature vectors to be highly correlated if they belong to a similar class, and less correlated if they belong to different classes. The Mean Discriminator component leads the network to make the mean embedded feature vectors of different classes to be less similar to each other, and the Embedding Discriminator component penalizes the network to generate the embedded feature vectors, which are dissimilar. The embedding feature space that contains k feature vectors, and the model's loss function was a proposed combination named Ad-Corre Loss jointly with the cross-entropy loss.

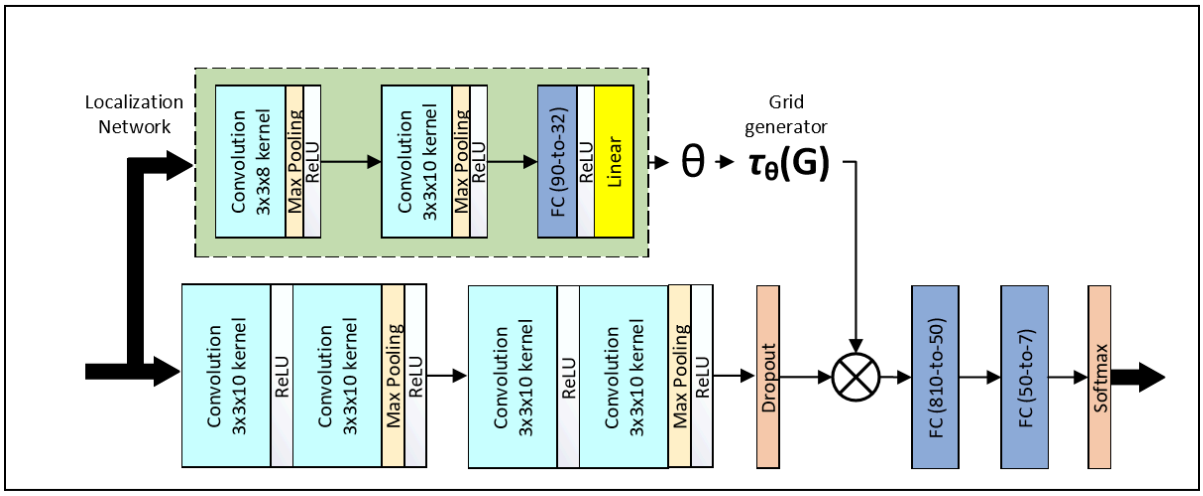
DeepEmotion

- Research Paper: Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network
- Repository: Github link

- Frameworks: PyTorch

The model *DeepEmotion* achieved a classification accuracy of 70.02% in 2019 through training on the datasets *FER-2013, CK+, FERG,* and *JAFFE.3* and no auxiliary datasets.

This model is based on attentional convolutional network, which is able to focus on important parts of the face utilizing the fact that different emotions seem to be sensitive to different parts of the face. The model's architecture starts with a feature extraction part consisting of four convolutional layers, each two followed by max-pooling layer and rectified linear unit (ReLU) activation function. They are then followed by a dropout layer and two fully-connected layers. The spatial transformer (the localization network) consists of two convolution layers (each followed by max-pooling and ReLU), and two fully-connected layers. After regressing the transformation parameters, the input is transformed to the sampling grid T(θ) producing the warped data. The spatial transformer module essentially tries to focus on the most relevant part of the image, by estimating a sample over the attended region. One can use different transformations to warp the input to the output, here we used an affine transformation which is commonly used for many applications.



Comparative Analysis

Model	Dataset	Accuracy(%)
Ensemble ResMaskingNet with 6 other CNNs	FER2013 and VEMO	76.82
Residual Masking Network	FER2013 and VEMO	74.14
EmoNeXt	FER2013	76.12
Segmentation VGG-19	FER2013 and CK+	75.97
Ensemble of 7 Models	FER2013, CK+ , JAFFE	75.8
Local Learning Deep + BOVW	FER2013	75.42
	FER+	87.76
	AffectNet 8-way classification	59.58
	AffectNet 7-way classification.	63.31
VGG,Res-Net, and Inception	FER2013	75.2
LHC-Net	FER2013	74.42
VGGNet	FER2013	73.28
CNN Hyperparameter Optimization	FER2013	72.16
Ad-Corre	FER2013, AffectNet , RAF-DB	72.03
DeepEmotion	FER2013, CK+, FERG, JAFFE	70.02

Selected Model (Baseline) and Evaluation Metric

After going over all the proposed models and their repositories, **the VGGNet model**, short for Visual Geometry Group Network, will be our chosen baseline model upon which we will build to elevate its performance. This is due to multiple reasons; other than the code successfully running, the model's selection was also based on its understandable architecture. This is explained in details in the *models and solutions* section. It consists of stacked convolutional layers and max-pooling layers, offering simplicity and interpretability. The network consists of 4 convolutional stages and 3 fully connected layers. Each of the convolutional stages contains two convolutional blocks and a max-pooling layer. The convolution block consists of a convolutional layer, a ReLU activation, and a batch normalization layer. Batch normalization is used to speed up the learning process, reduce the internal covariance shift, and prevent gradient vanishing or explosion. The first two fully connected layers are followed by a ReLU activation. The third fully connected layer is for classification. The convolutional stages are responsible for feature extraction, dimension reduction, and non-linearity. The fully connected layers are trained to classify the inputs as described by extracted features.

The model is available and pre-trained in the PyTorch framework which provides quick access to SOTA implementations and faster model development. The source code for the model is found here while the model weights or model zoo will be found here. With the source code and parameters becoming accessible, we are optimistic that the model's performance will improve after implementing our proposed updates. Furthermore, the model is trained on the *FER2013* dataset, which is perfect for our use due to its accessibility, size, and other factors explained earlier.

To evaluate the model effectiveness, we opt for the **weighted accuracy metric**, which accounts for class imbalance by computing the average accuracy weighted by the number of instances in each class. This metric promises a more comprehensive assessment of the model's performance, particularly in scenarios with imbalanced class distributions. To point out comparison results, confusion matrices will display the classification results and identify areas for the model's improvement.

Proposed Updates

As illustrated by the comparative analysis of different models discussed above, it is conspicuous that our chosen model, *VGGNet*, does not currently hold SOTA performance although it is still among the top performing ones. Accordingly, we aim to enhance its performance to supersede other models for FER in order to achieve better classification in an optimised manner.

First, we aim to enhance the model''s performance via different data augmentation techniques. These include adding auxiliary datasets to train the model, the dataset by incorporating additional datasets or auxiliary data sources to provide the model with a more diverse training set. This is highly feasible since more than 1 of the datasets relevant to FER explained earlier are publically available for use. Another approach is to manipulate the images themselves, such as mirroring/reflecting them, adding background noise, or other appropriate approaches.

Furthermore, to address the data imbalance problem, we intend to explore various techniques to handle this issue which mainly fall under undersampling, oversampling, or oversampling followed by undersampling, such as Random Undersampling, Tomeklinks, Random Oversampling, SMOTE, Smote + Tomek, and Smote + ENN to rebalance the dataset and enhance model generalisation.

Moreover, we plan to hypertune the model hyperparameters through trying different combinations. This could be done through either a random search to find the optimal ones, or, more efficiently, employing grid search to systematically hyper-tune model hyperparameters, improving the model's accuracy. For example, the number of epochs could be also increased from 300 to 500 in an attempt to make the model better performing. Similarly, regularisation of different forms can be used if the model appears to be overfitting.

Also, we propose creating an ensemble of models to leverage their collective strengths and improve overall performance. This could result in an overall better classification accuracy and higher generalisability. In this case, model interpretability will play a critical role; although our baseline model holds high interpretability, adding other models could decrease this understandability. Accordingly, it is important to work on the model's interpretability analysis using the available techniques.

Finally, if time permits, we aim to test the model on real-life data. In Particular, we can create a local application or website such that users can upload pictures of their faces and await the model's classification. This will ensure that the model has high generalizability since the Egyptian/Arab race is not common in the public dataset.

Graduation Project and Data Science Projects

Our graduation project is titled "*Data-Level Imbalance Handling Techniques: A Comparative Study on Credit Card Fraud Detection*". The aim of our paper is to compare multiple data imbalance handling techniques and machine learning models that yield an efficient credit card fraud detection system that can identify whether each transaction is legitimate or fraudulent. To handle data imbalance before model training, oversampling, undersampling, and oversampling followed by undersampling are used. Particularly, these are the techniques used:

1. Undersampling: Random Undersampling (RUS) and Tomeklinks
2. Oversampling: Random Oversampling (ROS) and Smote
3. Oversampling followed by Undersampling: Smote + Tomek (SmoteTomek) and Smote + ENN (Smoteen)

After the data is pre-processed, three supervised machine learning models are used: Random Forest, Extreme Gradient Boosting (XGBoost), and LightGBM. These models are evaluated based on a set of explained performance measures.

As for other data science projects, we have a diverse portfolio of a range of projects that we have worked on during our undergraduate study as data science students. These projects, including the description and code, can be found here; the most recent and relevant to machine learning among our projects are *MetroPT-3 Predictive Maintenance* and *Credit Card Fraud Detection System*. Both of us have worked together on all these projects.

Member Contribution

We had a clear and fair division of work, such that both members met together to work on the proposal. We searched for topics together and equally worked on finding the datasets as well as reference papers. Masa was responsible for the state-of-the-art and model descriptions while Malak was responsible for surveying the dataset and implemented models on the proposed problem, and making sure that the baseline model code runs successfully. We chose the baseline model as well as other parts of the report together. Hence, it is safe to say that all parts of the proposal were a combined effort since we only divided a small part of the requirements but worked on everything else together.

Resources

All the external resources used in this proposal are indicated within the paper. The links are hyperlinked to the word *source* when used respectively. No additional sources were used other than the ones included in this paper.