A DEEP LEARNING APPROACH TO EMOJI MAPPING FROM FACIAL EXPRESSIONS

Sanjana Dasadia
Student# 1006138582
sanjana.dasadia@mail.utoronto.ca

Sweni Shah
Student# 1006183372
sweni.shah@mail.utoronto.ca

Joanna Roy
Student# 1006000578
joanna.roy@mail.utoronto.ca

Maryam Younis
Student# 1006944942
maryam.younis@mail.utoronto.ca

ABSTRACT

Facial expressions and non-verbal cues, crucial in effective communication, are challenging to convey over digital text-based interactions. Emojis and emoticons have played a role in addressing this issue, but there still exists a gap between real-time facial expressions and the process of selecting digital emojis. This paper introduces a deep learning method using convolutional neural networks to map facial expressions to their corresponding emojis. The approach involves transfer learning with the ResNet18 architecture and convolutional neural networks, with a pre-processed FER2013 data set for model training. Expressions are classified into seven categories: anger, disgust, fear, happy, sad, surprise, and neutral. Model performance is optimized through the Adam optimizer, an exponential learning rate scheduler, and the cross-entropy loss function, and is evaluated based on accuracy, recall and precision metrics. For comparison, a Support Vector Machine baseline model is constructed and trained on the same data set. Results showed the deep learning model to outperform the baseline model in accuracy and categorical precision, suggesting CNNs are better for multi-classification facial recognition tasks. The final model was further tested on AffectNet and a custom facial expression dataset to assess its generalizability. Challenges in our model performance such as data imbalance and the subjective nature of emotions are identified. Ethical considerations include the privacy of images used to train and test the model, fairness in the model's facial recognition capabilities, and potential risks in applications like mental health diagnosis are discussed.

—-Total Pages: 9

1 INTRODUCTION

Facial expressions and non-verbal cues play a crucial role during in-person interactions for conveying emotions. For digital communication, like texting, these cues are not easily communicated, requiring more effort for quality interactions and understanding from the users (1). As a result, emoticons and emojis are used to express the cues and provide more context on how to interpret a message (2). However, there exists a gap between real-time facial expressions and the selection of digital emojis during digital text-based interactions.

The goal of our project is to develop a deep learning model that can detect facial expressions from 48 x 48 pixel grayscale images and match them with the best-fitting emoji (Figure 1). By leveraging deep learning and neural networks' promising ability to classify images and for facial recognition tasks (3), we will capture facial expressions to design and perform transfer learning on the ResNet model (4), then map these expressions accurately to corresponding emojis. Specifically, convolutional neural networks are effective for this task because it can identify low level and high-level features in images which will allow the model to capture basic facial emotions and features. Blending the authenticity of human emotions to the digital landscape through emojis has numerous applications, including in digital learning to allow for teachers to understand unspoken student feedback in real time through their expressions being converted to emojis.

1.1 Illustration

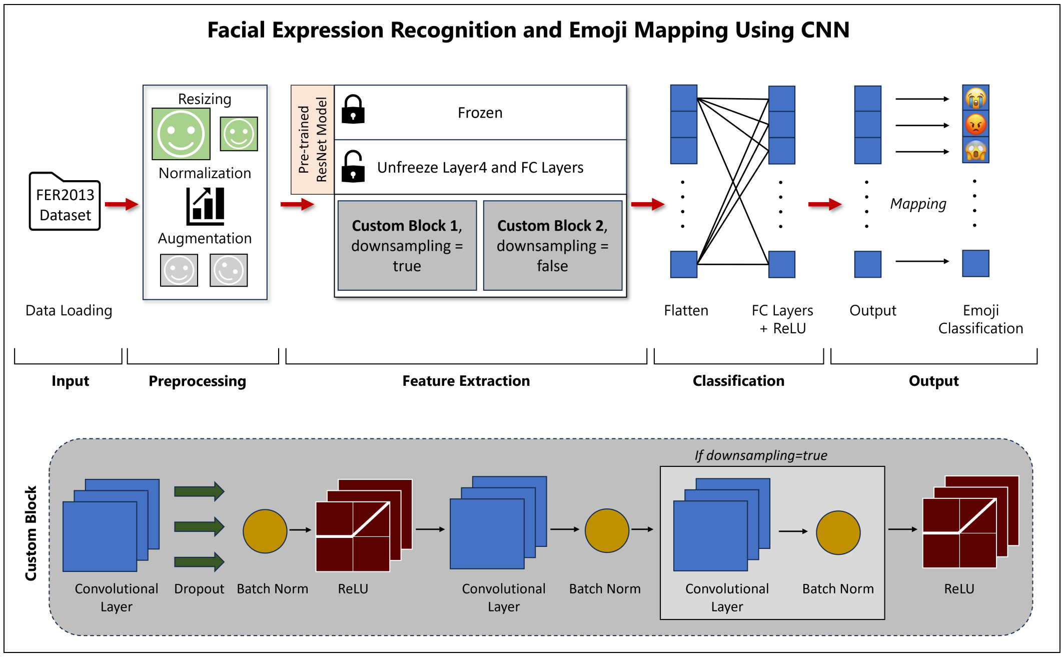


Figure 1: Model Architecture Diagram

2 Background and Related Work

Over the past decade, significant progress has been made in the domain of image recognition, in response to the ImageNet Large Scale Visual Recognition Challenge (LSVRC) (5). Image recognition techniques are often applicable to emotion recognition applications, so gaining a better understanding of previous work in the field is beneficial to conceptualize our model. The ImageNet data set, originally published in 2009 initiated annual competitions, challenging participants to construct object recognition and classification algorithms with low error rates. Historically, winning algorithms have made significant advances in the field, with accuracies spanning from 71.8% when the competition began in 2010, to 97.3% when it ended seven years later (6). Four ImageNet LSVRC algorithms, a reference design similar to our facial expression recognition application, and other companies working in this space are discussed below.

Krizhevsky et. al. won the ImageNet LSVRC in 2012, using a deep convolutional neural network named AlexNet (7). This network consisted of five convolutional layers, followed by three fully connected layers and a 1000-way softmax. Notably, this network utilized max pooling layers with overlapping pooling schemes in between convolutional layers for improved accuracy, ReLU activation functions to improve training efficiency, and introduced the “dropout method” to reduce overfitting.

In 2014, Szegedy et. al. submitted GoogLeNet to ImageNet LSVRC, achieving improved accuracy with far fewer parameters than Krizhevsky et. al in 2012 (8). The primary innovation of this architecture is the introduction of “Inception modules”, which reduce the computational expense associated with convolutional layers, allowing for increased depth without overloading training capacities. Notably, dropout, average pooling and ReLU activation were utilized in this model to improve performance, further justifying their use in our emotion recognition model (8).

The following year (2015), He et. al. presented a model emphasizing training efficiency via a “residual learning framework” (4). Residual learning aims to address the “vanishing/exploding gradient” problem, wherein the gradient becomes either very large or very small when propagated over a large number of layers. In residual learning, training is done in two stages: (1) The first stage introduces "skip connections", where activation layers are connected to further layers (skipping some in between) to form "residual blocks" with activations resembling identity mappings. The deeper models should therefore not have higher training errors than to their shallow counterparts. (2) The network is re-trained, with the previously skipped "residual" layers expanded. The larger network can now be trained with reduced risk of vanishing/exploding gradients, whilst still benefiting from improved accuracy due to increased depth. This model is significant in this project, as it was used as a model from which to complete transfer learning (4). ResNet was imported, a classifier was added to the end, and fine-tuned for our application.

Xie et. al. extended on the concept of ResNets, submitting ResNext to ImageNet LSVRC in 2016 (9). In ResNext, residual blocks consist of aggregated transformations, resulting in a multi-branch structure. This approach emphasizes careful network topology design, rather than simply increasing the depth and complexity of the network, to improve model performance and minimize resource-intensive hyperparameter searches (9).

In 2022, Debnath et. al. implemented an emotion recognition software similar to that which we are hoping to construct (10). It consists of four convolutional layers followed by two fully connected layers and achieved 96% accuracy, making it a useful reference design for our model. This paper further emphasizes how CNNs are a good choice of network for emotion recognition applications (10). Hence, it is a reasonable choice of model for this project.

In summary, significant advancements have been made in this field over the past decade. Many companies, including Visage, Noldus, Amazon, Google Vision AI, and IBM Watson have applied these techniques and similar ones to successfully employ facial expression recognition in industry applications. The techniques discussed above helped shape our design, to improve the accuracy of facial expression classification.

3 Data Processing

The FER2013 dataset is used to train, validate, and test the model. The train data was split where 70% of the data was used for training and 30% was used for validating. This leaves the test set from the FER2013 dataset untouched, making it suitable to conduct final testing. This dataset contains seven facial expression classes. Initially, the dataset contained images of dimension 48x48 pixels which were then resized to 224x224 pixels, since this is the input required on the ResNet architecture (11). We also normalized the images to get pixel values between [-1, 1] using a mean, and standard deviation of 0.5. The model accepts RGB image inputs, while the FER2013 dataset contains only grayscale images, thus we formatted our images to add two additional channels. A few visualized samples before and after the data cleaning process can be seen in Figures 2 and 3. It is evident that the image was cleaned and the number of pixels were increased as the quality of the images in Figure 3 is worse.

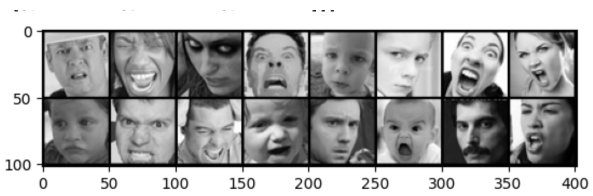


Figure 2: Original Data Samples

The initial method that we used to load data included reading a CSV file each time a value was read, and cleaning the data on each read rather than cleaning all the data at once. The solution to this issue was to use the **torchvision.datasets.ImageLoader** library to load all the image files, to be stored locally, rather than in a csv. The transforms discussed above were then applied to the images, and finally the transformed image set was passed to **torch.utils.data.DataLoader** to obtain the train dataset.

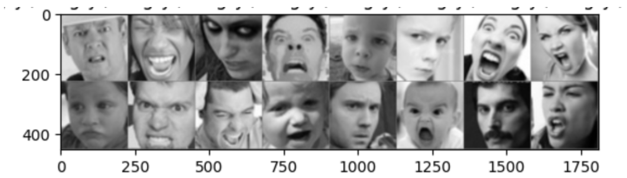


Figure 3: Cleaned Data Samples

4 Architecture

The model begins with preprocessing of the FER2013 dataset images by transforming them to grayscale images, resizing to 224x224 pixels and converting them to tensors. The final network is built via transfer learning of the ResNet18 Architecture. ResNet consists of four main layers, each consisting of two BasicBlocks (BasicBlock: Convolution - BatchNorm - ReLU - Conv - BatchNorm. Layers 2a, 3a, and 4a also have Downsampling: Convolution - BatchNorm). This architecture uses “skip connections” to address the problem of vanishing/exploding gradients, which would otherwise occur in such a deep model, by connecting layers to future layers while “skipping” some connections in between (4).

The final two blocks of ResNet18 were unfrozen, and modified to a custom sequential layer comprising two CustomBasicBlocks (see Figure 1), one with downsampling and then one without downsampling. Each block consists of the following: a convolutional layer, dropout with a 50% rate to reduce overfitting, batch normalization for stabilizing learning, and ReLU activation for non-linearity, two more layers of convolutional layer and batch normalization(once if downsampling is set to false), and a final ReLU activation. This was necessary to mitigate overfitting, which was initially problematic in training the model. Furthermore, a series of three fully-connected layers (dimensions: 300, 200, 7) with ReLU activations between them were appended to the end of the network, to classify data to the seven output classes: Anger, Disgust, Fear, Happy, Sad, Surprise, and Neutral. These categories are the range of facial expressions we aim to identify.

The model was fine-tuned using the Adam optimizer to mitigate the dispersed nature of gradients for image classification tasks(12). The learning rate was set at 0.0001 with a batch size of 128 for 20 epochs. An exponential learning rate scheduler with a gamma of 0.9 was implemented to prevent the learning rate from becoming too high in later epochs during training, and a weight decay of 5e-4 to mitigate overfitting. The cross-entropy loss function was used to compute loss, which is a standard choice for multi-class classification tasks (13). For performance evaluation, we first measured loss and overall accuracy during training and validation to gauge the model’s performance. Next, we measured the precision and recall of each specific category of facial expression to assess whether our model had balanced learning across all the different facial expressions.

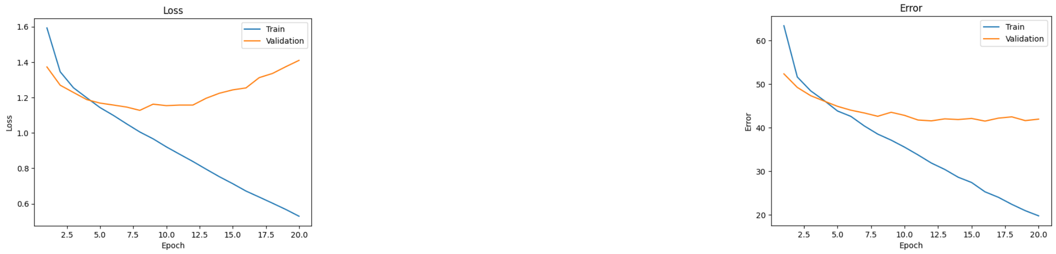
5 Baseline Model

A Support Vector Machine (SVM) was constructed as a baseline model for this project. Similar models have been used in emotion recognition applications, making it a useful baseline against which to compare (14). This model was constructed using the SciKit-Learn package in Python, and tuned using GridSearchCV hyperparameter search on a subset of the data (1000 samples, for efficiency - selected randomly from the dataset). The final model was constructed using a polynomial kernel, C = 0.1, Gamma = 0.0001, and a one-vs-one (’ovo’) decision function, and trained on the full set of images. Using the one-vs-one decision function means that the model consists of N*(N-1)/2=7*(7-1)/2=1128 binary classifiers in total, where N is the number of output categories (N=7), comparing each emotion to the others individually to determine the best fit. The baseline model performance is summarized in Table 1.

This baseline performed best on “happy” images (highest precision, recall, and F1). There seem to be a higher proportion of images classified as “happy” in the dataset (support = 1458 in the validation set, which was sampled randomly from the full dataset), hence it is better trained to recognize this emotion. This suggests that the performance of other emotions could be improved with a larger sample size as well. However, 36% performance is still relatively low, which could be due to a

Table 1: Summary of baseline model performance

Emotion	precision	recall	f1-score	support
Angry	26%	25%	25%	798
Disgust	8%	23%	12%	81
Fear	24%	35%	29%	832
Happy	55%	47%	51%	1458
Sad	31%	25%	28%	913
Surprise	41%	51%	45%	666
Neutral	39%	27%	32%	994
Accuracy	**36%**			



(a) Loss Plot

(b) Error Plot

Figure 4: Training Plots

variety of factors. Namely, this model does not capture the spatial links between pixels, since it is only trained pixel by pixel on the input dataset. For this reason, we expect the final CNN model to perform better, since the convolution better captures these spatial relationships.

In summary, the baseline model achieved modest performance on the FER2013 dataset, but still has significant room for improvement. Our final model utilizes convolutions to capture spatial information, and a pre-trained ResNet model, which has learned from a much greater sample size of images, and was be fine-tuned to suit this project. Hence, it performed better on this dataset, as discussed later in the report.

6 Results

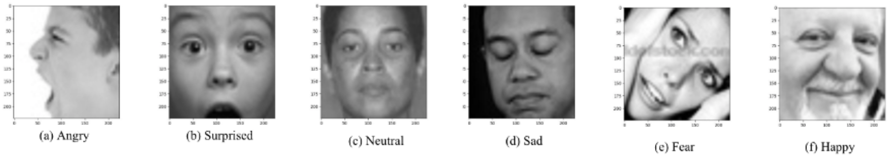
This section includes quantitative and qualitative results, to demonstrate model performance on the FER2013 dataset, and never-before-seen test data. Quantitative results were obtained by running the model on each dataset of interest, and calculating corresponding accuracies. To obtain qualitative results, the model outputs were collected in folders corresponding to correct and incorrect predictions. Further analysis of this data is included in the Discussion section of this report. **Quantitative:** Figure 4 shows the training plots for our model. Table 2 illustrates model loss and accuracies for the validation and test sets, and Table 3 illustrates accuracies for each emotion classification. **Qualitative:** Figures 5 and 6 show sample correct and incorrect classifications, respectively.

Table 2: Model loss and accuracies for validation and test datasets (FER2013 dataset)

Dataset	Loss	Accuracy (%)
Validation	1.4	58
Test	21.8	21.8

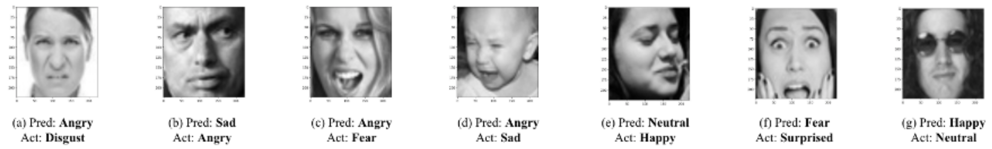
Table 3: Model accuracies for each category of validation and test datasets

Emotion	Val Recall (%)	Test Recall (%)	Val Precision (%)	Test Precision (%)
Angry	37.44	17.85	27.4	29.65
Disgust	0.0	0.0	0.0	0.0
Fear	22.62	10.25	26.56	30.61
Happy	69.3	41.83	59.75	61.09
Neutral	35.19	19.71	36.55	37.05
Sad	31.45	13.39	31.43	31.63
Surprise	44.19	15.40	58.77	58.01



(a) Angry (b) Surprised (c) Neutral (d) Sad (e) Fear (f) Happy

Figure 5: Correctly classified images for each emotion



(a) Pred: **Angry** Act: **Disgust** (b) Pred: **Sad** Act: **Angry** (c) Pred: **Angry** Act: **Fear** (d) Pred: **Angry** Act: **Sad** (e) Pred: **Neutral** Act: **Happy** (f) Pred: **Fear** Act: **Surprised** (g) Pred: **Happy** Act: **Neutral**

Figure 6: Incorrectly classified images for each emotion

7 Evaluation

We started off with a simple pretrained resnet18 architecture where we froze all the weights and added a simple classifier along with some regularizers such as weight decay. However, analyzing the results mentioned in the progress report with training accuracy being at 46%, we realized that this model was underfitting. Along with tuning the hyperparameters like batch size, learning rates, and adding data transforms, to help our model learn better, we first turned off weight decay, and unfroze the 4th layer weights. We also added 3 more layers to our custom classifier. After re-training the model, we saw that the model was significantly overfitting to the training dataset with almost 98% accuracy, but knew this would generalize poorly to validation and test sets. In an effort to improve generalization, we turned on regularizers like weight decay and added dropout (with a custom block) as mentioned in Section 4. This helped our model with generalization and improved validation accuracy to 58% as seen in Table 2.

The problem being solved is emotion classification, hence this model can be evaluated on new data as long as we can obtain a dataset with the 7 classes. We evaluated our model on 2 new datasets, one is a subset of the AffectNet dataset that we sourced from Kaggle (15), and a custom dataset that we created by uploading pictures of all 4 group members. Both of these datasets were unseen by the model and were not used in training or tuning for any hyperparameters. We used 70 images from the AffectNet dataset, 10 for each emotion (for balance), and evaluated our model. Since these images are RGB but our model was trained on grayscale images, we had to apply a grayscale transform to this dataset. The recall values for each emotion can be seen in Table 5, to give an idea of what emotions our model was able to predict best. The loss and accuracy for AffectNet and custom set can be seen in Table 4. We can observe that the model is having difficulty classifying emotions such as Disgust and Fear, and this is due to the model being trained on an imbalanced dataset, where Disgust only accounted for 1.5% images. Since these test datasets are balanced and smaller in size, the wrong classification of all images related to 'Disgust' would greatly affect our overall accuracy.

Table 4: Model loss and accuracies for AffectNet subset and Customset

Dataset	Loss	Accuracy (%)
AffectNet subset	0.959	38.571
Customset	1.001	35.714

Table 5: Model recall for AffectNet subset (%)

Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
80.0	0.0	10.0	100.0	50.0	30.0	40.0

8 Discussion

The model's performance is good, as it achieves a validation accuracy of 58% compared to baseline performance of 34%. There are several reasons as to why this accuracy value, that may normally indicate an average-performing model, points to a "well" performing model in this case. First, emotion classification is a difficult task. Only grayscale images were used for training, and the expressions associated with two different emotions may look very similar, one person's version of Angry may look different from someone else's. For instance, considering Figure 7, in the first image (7a), the actual emotion is "fear", and in the second image (7b) the actual emotion is "angry". Both these images have similar expressions, making it hard to distinguish between the two emotions of "fear" and "angry", showing that the task of emotion classification is complex.

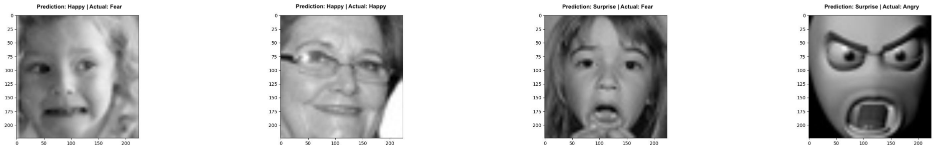
(a) Fear (pred: Angry)

(b) Angry (pred: Fear)

Figure 7: Similar photos with different associated emotions (Anger and Fear)

The next reason is the unbalanced dataset that the model was trained on. The "disgust" class, comprising only 1.5% of the data, resulted in 0% precision and recall (as seen in Table 3), significantly impacting overall accuracy. The imbalance led to the model not learning the features of this class well, resulting in the misclassification of all "disgust" samples. To assess the impact of unbalanced data, the team tested the model without "disgust" images, which raised accuracy from 38.6% to 50.8%. This significant improvement underscores the negative effect of unbalanced data on overall accuracy.

An interesting observation can be made about the results in Figure 8: the mouth shape in the "fear" classification is similar to the mouth shape in the "happy" classification – on the surface level, the two images below appear quite similar. This mouth shape is quite common in the "happy" samples in the training set. Hence, the model may have associated it with "happy" because of the similar features, although "fear" was correct in this instance (Figure 8). Another interesting

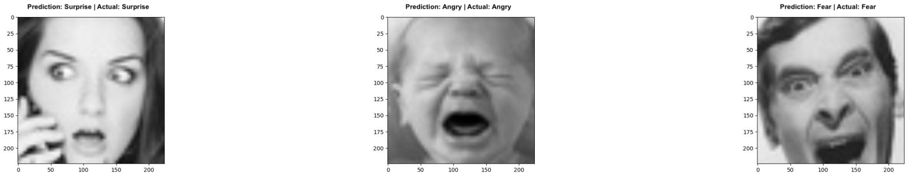

Prediction: Happy | Actual: Fear
Prediction: Happy | Actual: Happy
Prediction: Surprise | Actual: Fear
Prediction: Surprise | Actual: Angry
0
25
50
75
100
125
150
175
200
0 50 100 150 200

(a) Fear (pred: Happy) (b) Happy (pred: Happy) (c) Fear (pred: Surprise) (d) Angry (pred: Surprise)

Figure 8: Two images classified as Happy (left: incorrect, right: correct) and two images on right wrongly classified

observation relates to the similarities between fear, anger, and surprise – each of these emotions has many associated samples wherein the person has their mouth open. Hence, there are often misclassifications amongst these three categories, as illustrated by the examples in Figure 8.

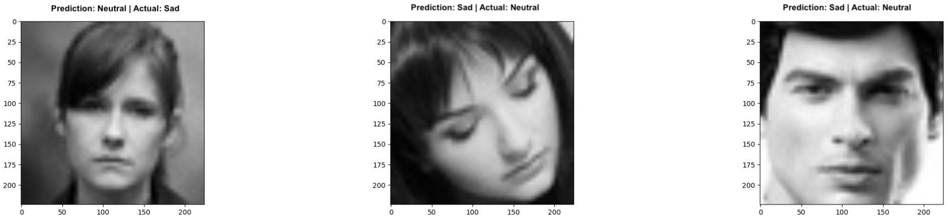
On the other hand, for the same classes the following samples were correctly classified (Figure 9). Hence, the model is able to distinguish between these emotions to some extent, despite them being quite similar:


Prediction: Surprise | Actual: Surprise
Prediction: Angry | Actual: Angry
Prediction: Fear | Actual: Fear
0
25
50
75
100
125
150
175
200
0 50 100 150 200

(a) Fear (pred: Surprise) (b) Angry (pred: Surprise) (c) Angry (pred: Surprise)

Figure 9: Correctly classified images (Surprise, Anger, Fear), each with open mouths

One more interesting observation in the results is the "neutral" expression being frequently confused with "happy" and "sad", particularly in samples wherein people have their mouths closed (Figure 10).


Prediction: Neutral | Actual: Sad
Prediction: Sad | Actual: Neutral
Prediction: Sad | Actual: Neutral
0
25
50
75
100
125
150
175
200
0 50 100 150 200

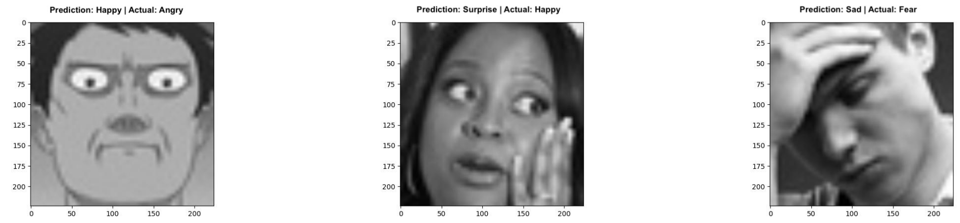
(a) Sad (predicted: Neutral) (b) Neutral (predicted: Sad) (c) Neutral (predicted: Sad)

Figure 10: Incorrectly classified images, each with closed mouths

A surprising discover was the difficulty found by the team to categorize the following samples, as the team members debated and disagreed with the labeled emotion (Figure 11).

As mentioned earlier, the task of emotion classification is inherently challenging, even for humans, but could potentially be improved with a larger dataset so that the model could learn the nuanced features more effectively (eg. the shape of eyebrows, nose flare, wrinkles, etc, in conjunction with the mouth shape and other more "obvious", distinguishable features).

Key learnings from the project include the importance of checking and filtering training data for balance, ensuring equal representation of all classes. Additionally, addressing human disagreements in emotion labeling by using a large dataset and involving multiple reviewers during labeling proved beneficial in enhancing model training. The project provided valuable experience in overcoming


Prediction: Happy | Actual: Angry
Prediction: Surprise | Actual: Happy
Prediction: Sad | Actual: Fear
0
25
50
75
100
125
150
175
200
0
50
100
150
200

(a) Angry (predicted: Happy) (b) Happy (predicted: Surprise) (c) Fear (predicted: Sad)

Figure 11: Images where the emotion was hard for team to discern, incorrectly classified by the model

challenges during model training. We encountered issues in data pre-processing, highlighting the significance of relying on official documentation for effective solutions. Learning to address problems leading to lower accuracies, such as implementing regularizers and freezing layers, contributed to enhanced training accuracy. Additionally, the project deepened our familiarity with key Python libraries, including Pandas, Numpy, and PyTorch.

9 Project Difficulty

As discussed in Section 8, emotion recognition is an inherently challenging task, because of the subtle features that differentiate facial expressions from person-to-person. Our team was able to achieve modest accuracy by leveraging transfer learning, while adjusting the model architecture to better suit the task requirements. The final model was created after careful analysis of several different iterations, each of which leveraged a variety of techniques to mitigate challenges related to underfitting, overfitting, generalizability, training and testing efficiency, and data processing strategies. We examined each model, as well as its outputs and correct/incorrect classifications to better understand which techniques would be most effective in improving it. This analysis required us to gain an in-depth understanding of the ResNet architecture, PyTorch/NumPy/Pandas libraries, and any techniques we applied to improve performance. Examining the outputs of our final model, it is clear that many of the incorrectly classified emotions are difficult for even humans to distinguish because of nuanced differences in emotion expressions in different individuals and contexts. Hence, our team believes this is a high quality model for a relatively challenging problem.

10 Ethical Considerations

One of the major ethical concerns when using a dataset of facial images is privacy and consent. The individuals would need to have given consent beforehand, for their images to be used for training. Alongside this, it would be important to note how this data was collected and ensure that there is no way to identify individuals from the cropped images. Accidental identification of the individuals can put them at risk, and violate privacy and safety. A limitation of our dataset is biases. The images may not be as diverse, and possibly underrepresented features like mouth shapes, facial hair, age, etc. and training on this can lead to biased models that do not perform well on certain groups. Therefore, a limitation in our dataset also limits our model's performance. Another ethical concern may arise if the model was used for a task such as mental health diagnosis. Getting someone's emotions wrong, especially in sensitive situations like mental health, can significantly impact how they feel and their overall well-being. A limitation that could impact the model's ability to accurately classify is that emotions are highly subjective and context-dependent, varying significantly between individuals and situations. This subjectivity poses a challenge for the model, as it may struggle to generalize its understanding across diverse contexts.

11 Colab Link

The code for our main and baseline models can be found in this colab notebook: project-link

REFERENCES

[1] A. N. Gesselman, V. P. Ta, and J. R. Garcia, “Worth a thousand interpersonal words: Emoji as affective signals for relationship-oriented digital communication,” *PLOS ONE*, vol. 14, pp. 1–14, 08 2019.

[2] A. H. Huang, D. C. Yen, and X. Zhang, “Exploring the potential effects of emoticons,” *Information Management*, vol. 45, no. 7, pp. 466–473, 2008.

[3] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, “When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition,” December 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[6] D. Gershgorn, “The data that transformed ai research-and possibly the world,” Jul 2017.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, p. 84–90, 2017.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[9] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[10] T. Debnath, M. M. Reza, A. Rahman, S. S. Band, and H. Alinejad-Rokny *Four-layer convnet to facial emotion recognition with minimal epochs and the significance of data diversity*, 2021.

[11] M. Sambare, “Fer-2013,” Jul 2020.

[12] S. Y. ŞEN and N. ÖZKURT, “Convolutional neural network hyperparameter tuning with adam optimizer for ecg classification,” in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–6, 2020.

[13] A. Demirkaya, J. Chen, and S. Oymak, “Exploring the role of loss functions in multiclass classification,” in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–5, 2020.

[14] I. Dagher, E. Dahdah, and M. Al Shakik, “Facial expression recognition using three-stage support vector machines,” *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, 2019.

[15] N. Segal, “Facial expressions training data,” Jan 2023.