CMPT419 Final Report
Adding Interaction to Games Using Expression Recognition

Sean Wallace
scw26@sfu.ca
301411475

Nagisa Nakamura
nagisan@sfu.ca
301416720

Aurora Yang
yya225@sfu.ca
301417395

ABSTRACT

Games are a great way for people to interact with and learn new things. But there are not a lot of games that allow you to interact with them by using your face or emotions. We wanted to change that by creating the game "Emotion Defender" where the main form of interaction is expression recognition. By using machine learning models we are able to predict expressions on the players face. This prediction then causes some effect on the game. Where different emotions correspond to different effects. This form of interaction creates a fun and novel way to play. Our game provides and example of how this form of interaction can create fun and exciting interaction.

ACM Reference Format:
Sean Wallace, Nagisa Nakamura, and Aurora Yang. 2026. CMPT419 Final Report Adding Interaction to Games Using Expression Recognition . In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Gaming has often been limited by traditional controls like keyboards and joysticks, which can make it hard for players to feel fully involved and express themselves. We think that there are a lot of different ways of interacting that we are missing out on. One of them being expression recognition. This form of interaction is currently being used as a way to control background systems and difficulty. As explored in many different papers, including ones by Akbar and Moniaga. Where they explore this idea how emotions can change different systems of a game. Finding improvements when adjusting systems to players using expression recognition [2, 5].

But our game, "Emotion Defender," aims to do something different by making facial expressions the main way players interact with the game. This method allows for a new form of gaming experience, instead of looking to enhance current interactive systems. We want to see if primary interaction through expression recognition is viable. To do this we look to build a responsive and accurate model, so that playing the game stays fun and enjoyable. And use simple emotions to it is natural and easy to interact with.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 APPROACH

To detect user's emotions fast and accurately, several models are developed for this game. Each using a dataset that consisted of FER-2013 dataset [4] and annotated pictures of friends making different emotional face that we collected. Each image we trained with, was converted to grayscale and resized to 48x48 pixels. Method 1 and Method 1.2 were neutral networks with multiple convolutional layers (Conv2D(32, (3,3))), pooling layer (MaxPooling2D(2,2)), normalization layer (BatchNormalization), and all-coupling dense layers (128 and 7) for output . ReLU and Softmax were used as activation functions, and Adam was chosen as the optimization algorithm. method 1 modeled all labels, while method 1.2 analyzed only the 5 emotions actually used in the game. In method 2, basemodel from method 1 was used and added a new layer on top of it to fine-tune the model to identify only certain emotions (anger, happiness, sadness, surprise, neutral). The additional layers included multiple dense(16,8,5) and BatchNormalization layers, and a Softmax layer was introduced to predict five categories. In Method 3, after dimensionality reduction to 150 with PCA we classify using an SVC model. The SVC model uses an RBF kernel and StandardScaler for data scaling. The dataset was split into training and test data, with the models trained primarily on the training data and the performance of the models evaluated on the test data. A ModelCheckpoint callback was used to store the best models in training. To visualize the results of each model, we utilized the matplotlib and seaborn libraries to display the confusion matrix. The best model was then used in the game to detect players emotions. To detect the players emotions the game accesses the player webcam using Open CV. Then as the game is running, the game takes the image of the player and runs it though a pre-built facial recognition library [3]. That will return the location of the face in the image. We then crop the image to only get the face, converting it to grayscale and scaling it to the correct size for the model. The model then takes the face and returns its expression prediction which is then passed to the game. This whole process runs many times per second allowing for responsive interaction. The game that we are linking the model too was developed in Python using Pygame and uses assets found on https://itch.io/game-assets[1, 6, 7]. The game is called "Emotion Defender" and it involves enemies spawning from the start of a path and if they get to the end of the path you lose. Each enemy is defeated by a corresponding emotion which is represented by their colour. The goal of the player is to defeat the enemies by changing facial expressions to match the enemies on screen, and survive as long as they can.

3 DATASET

The project uses a combined dataset consisting of FER-2013 and our own photos of various facial expressions. FER-2013 found on https:

//www.kaggle.com/datasets/yusufkorayhasdemir/fer2013csv contains thousands of samples labeled with 7 basic emotion categories of facial expressions, anger, disgust, fear, happiness, sadness, surprise, and neutral. Our own data includes a total of 16 images of group members and their friends, labeled anger, happiness, sadness, and surprise. The photos of group members and friends were labeled accordingly and ready to be trained with.

4 EXPERIMENTS AND RESULTS

The first method used a Conv2D network model. Trained and validated using our dataset, which consists of a combination of the FER-2013 dataset and our own facial expression images. Training was done in 20 epochs, with 10% of the dataset kept as a test set and the rest used for training. The validation split was set to 0.2 (20%). During the early stages of model training, the accuracy of the validation set gradually improved, with the highest value being about 55.73% at epoch 15. From the confusion matrix in Figure 1, we can see that happiness has the best accuracy at 71%. And sadness has the worst at 39%. Even though sadness had worse predictions than anger, we found it was difficult for the model to accurately predict our faces when we were angry, but it was much better at predicting them when we were sad. When using this model we found it work a lot better if we merged all of anger, fear and disgust into a prediction of just anger. Since anger was often being predicted as something else and we weren't using disgust or fear for the game. This turned out quite good, so we ended up using this model in our game, because we had not experimented with some of the later models in this section.

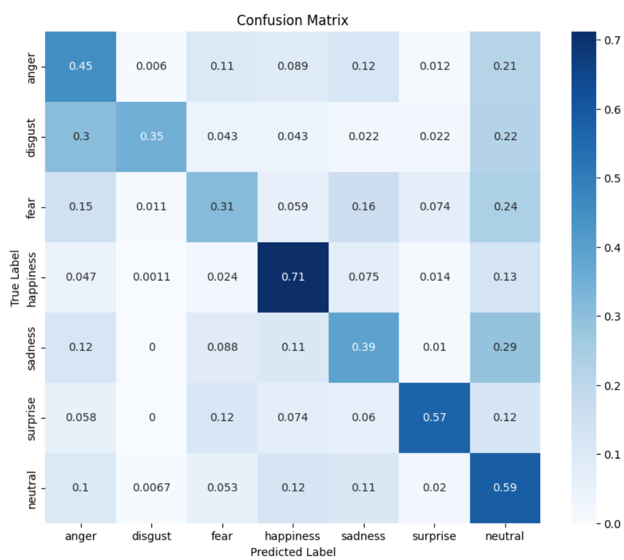

Figure 1: Method 1 Conv2D 7 emotion categories Confusion Matrix

In method 1.2, was trained in the same way as method 1 but trained with data from only the four emotion categories actually used in the game: anger, happiness, sadness, surprise with neutral. Compared to the first method, which used all emotion categories, the probability of correctly recognizing happiness, sadness, and surprise increased by around 10% each, but the probability of correctly recognizing neutrality decreased around 10% and anger didn't change much at all. We can see from Figure 2 that angry or sad faces would often be predict the other options. And so in actual use, the recognition capability for anger and sadness was low. We couldn't combine any outputs so it turned out to be annoying to try and get some emotions, so we didn't use it for the final version of the game.

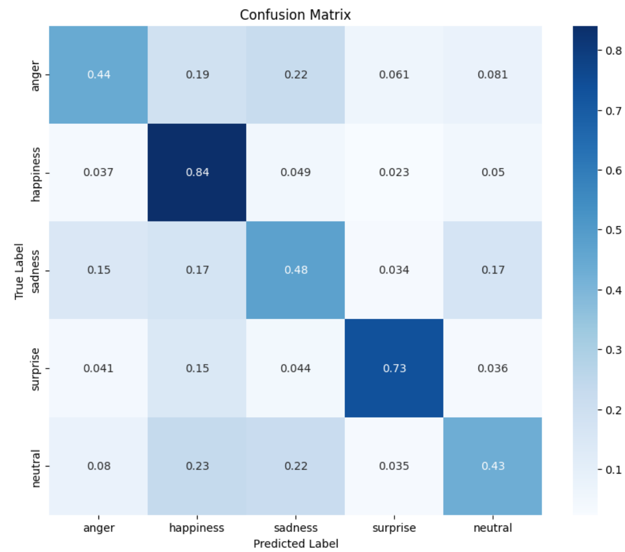

Figure 2: Method 1.2 Conv2d with 5 emotion categories Confusion Matrix

The second method was to fix the weights of the early layers of the basemodel trained in Method 1 and create a model with fine-tuning on the data of the five emotion categories used in the game. Training was run for 20 epochs, adding new layers on top of the basemodel, and accuracy on the validation data improved from an initial 80.62% to a final 83.74%. The final accuracy of the model on the test set was 83.30%. From Figure 3, we can see the probability of accurately recognizing each emotion category was also significantly improved, with the lowest accuracy being anger at 77%, and all others being 80% or better. The strength of this model is not only the high accuracy of each recognition, but also the very low probability of misidentification with other emotions. For example, the probability of mistakenly recognizing "anger" as "sadness" was 22% in the previous method 1.2, but it is much lower at 7.8% in this model.

In method 3, we used PCA and SVM to predict expressions. The principle component for PCA was set to 150, so we reduced the amount of data by about 95% speeding up training and prediction. The model used the entire dataset to predict all of the labeled emotions. The trained model implemented standardized scaling processes , PCA, and SVM. As we can see in Figure 5, the overall accuracy was 48%, which was significantly lower than the other models that we trained. We can see that the precision and recall score were similar for each category, so it wasn't biasing towards

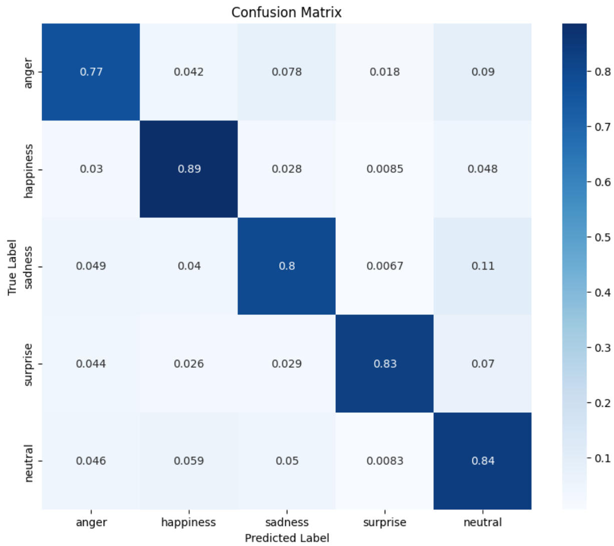

Figure 3: Method 2 Fine Tuning Confusion Matrix

predicting one expression. From Figure 4, we can see that even though this was a completely different machine learning technique it still found that happiness and surprise were the easiest to predict compared to other emotions.

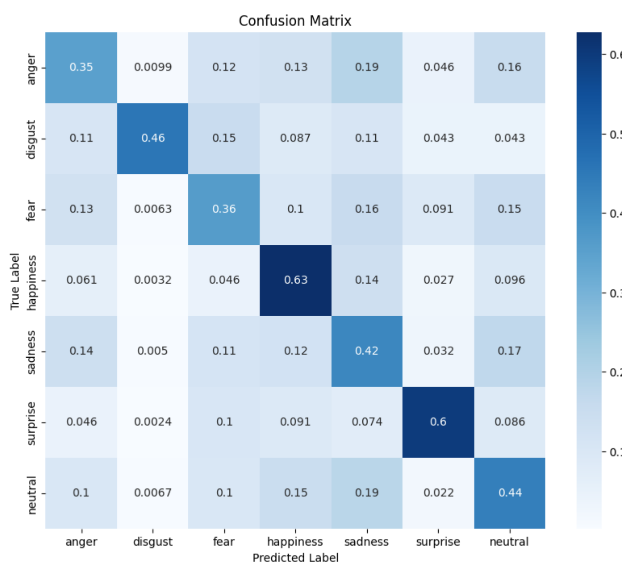

Figure 4: Method3 SVM with PCA Confusion Matrix

Even thought the second method, fine-tuning, predicted emotion categories the best, we did not have time to implement it into the game. And overall, when models that were created using all seven emotion categories were tested for real-time facial expression recognition, there was a tendency for low accuracy in recognizing

	precision	recall	f1-score	support
0	0.38	0.35	0.36	503
1	0.53	0.46	0.49	46
2	0.38	0.36	0.37	473
3	0.65	0.63	0.64	951
4	0.36	0.42	0.38	602
5	0.66	0.60	0.63	417
6	0.41	0.44	0.42	599
accuracy			0.48	3591
macro avg	0.48	0.46	0.47	3591
weighted avg	0.49	0.48	0.48	3591

Accuracy: 0.4803675856307435

Figure 5: Method3 Report

the expression of anger. Therefore, if the model analyzed an expression as disgust or fear, we treated it as anger, which allowed for a smoother gaming experience.

When creating the game we had to make sure that using your face wouldn't be a pain and that players would enjoy use it. When we used the models in the game, it turned out to be fun and engaging. The model would predict fast enough that we were able to update expression predictions many times per second. Even though the models weren't perfectly accurate, they were good enough since we are updating the predictions many times a second, the player has lots of time to play around and try to find a face that the model will predict correctly. The game was also quite forgiving since there was never any punishment for getting a face wrong. Only positive actions happened when the face was got right.

As you can see from Figure 6, the game included UI elements that allowed the player to see who facial expression was being predicted and what the current prediction was. Adding in these elements made the game easier to work with, as the player could check if they were being tracked or what type of faces would result in different emotions.

We thought that combining this game with a facial recognition model made for a fun game that was responsive and interactive. And making the information around detection clear to the player valuable to the experience.

5 DISCUSSION

In our project, we found some interesting results that were both expected and unexpected regarding facial recognition in gaming. Our system worked well in many situations by reacting to players' emotions, but it also showed some areas where it could be improved. One surprising finding was how differently the system performed with various emotions. It was very good at recognizing 'happiness' but had trouble with 'anger' and 'sadness'. This might be because 'anger' and 'sadness' have more subtle facial signs compared to the clear signs of 'happiness'. These errors point out that we need better training data that captures a wider range of emotional expressions.

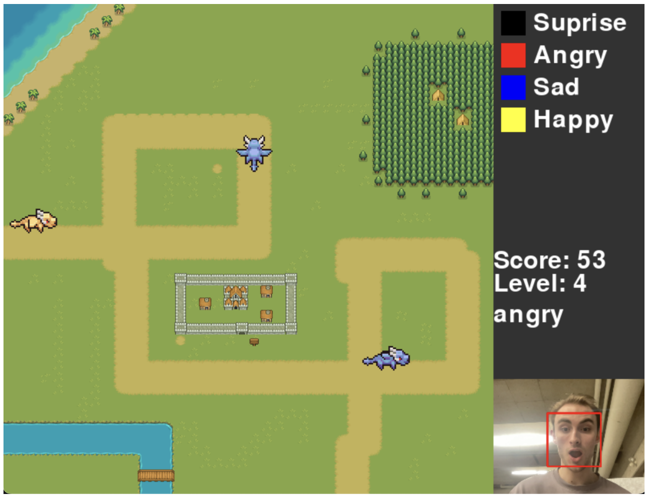

Figure 6: Image of the game

An unexpected aspect of our findings was that initially, we thought reducing the number of emotion categories from seven to four would simplify the system and increase accuracy. However, our results showed that the system was actually more accurate with the full seven emotions. This suggests that having a broader range of categories helps the system to better distinguish between different emotional states, even though it's more complex. Looking ahead, there are several ways we could make our recognition system better. Adding more diverse expressions to our dataset might help the system learn to tell similar emotions apart more accurately. Another idea is to combine facial expressions with other types of data, like voice tones or body language, for a fuller understanding of player emotions. This could make the system more reliable in understanding the subtle aspects of human emotions.

6 CONCLUSION

Our project, "Emotion Defender," demonstrated how expression recognition can add new ways of interacting by allowing players to control the game with their facial expressions. This technology proved to be engaging and introduced a novel way to interact with games. However, we encountered challenges with models accurately recognizing certain emotions, particularly anger and sadness. Our best results were achieved when we fine-tuned the system to improve recognition accuracy. When connected to the game it created a fun and responsive form of interaction between the player and the game. Overall, "Emotion Defender" shows that using expression recognition as a form of player interaction is a viable and fun way of playing games.

7 CITATIONS AND BIBLIOGRAPHIES

REFERENCES

[1] 2024. 8-bit / 16-bit Sound Effects (x25) Pack by JDWasabi. https://jdwasabi.itch.io/8-bit-16-bit-sound-effects-pack
[2] M. Taufik Akbar, M. Nasrul Ilmi, Imanuel V. Rumayar, Jurike Moniaga, Tin-Kai Chen, and Andry Chowanda. 2019. Enhancing Game Experience with Facial Expression Recognition as Dynamic Balancing. *Procedia Computer Science* 157 (Jan. 2019), 388–395. https://doi.org/10.1016/j.procs.2019.08.230
[3] Adam Geitgey. 2024. face-recognition: Recognize faces from Python or from the command line. https://github.com/ageitgey/face_recognition
[4] Manas Sambare. 2024. FER-2013. https://www.kaggle.com/datasets/msambare/fer2013
[5] Jurike V. Moniaga, Andry Chowanda, Agus Prima, Oscar, and M. Dimas Tri Rizqi. 2018. Facial Expression Recognition as Dynamic Game Balancing System. *Procedia Computer Science* 135 (Jan. 2018), 361–368. https://doi.org/10.1016/j.procs.2018.08.185
[6] octoshrimpy. 2024. 16x16 miniworld + by octoshrimpy. https://octoshrimpy.itch.io/miniworld-plus
[7] octoshrimpy. 2024. Free 16x16 Mini World Sprites by Shade, octoshrimpy. https://merchant-shade.itch.io/16x16-mini-world-sprites

A APPENDIX

A.1 Section 3.1: Motivation

For what purpose was the dataset created?
This dataset was created to help develop a game that uses facial expressions as inputs to control game interactions.
Who funded the creation of the dataset?
There was no external funding; this project was part of a university course.

A.2 Section 3.2: Composition

What do the instances that comprise the dataset represent?
The instances are images of different facial expressions.
How many instances are there in total?
The dataset combines around a few thousand instances from a public dataset and 16 custom images.
Does the dataset contain all possible instances or is it a sample of instances from a larger set??
Yes, it is a sample from a larger potential set of facial expressions.
What data does each instance consist of?
Each image is processed into a grayscale 48x48 pixel format.
Is there a label or target associated with each instance?
Yes, each image is labeled with an emotion from the following categories: anger, happiness, sadness, surprise, and neutral. These labels are used to train the emotion recognition model.
Is any information missing from individual instances?
No significant information is missing from the images. However, some images may have variations in lighting and background that were not controlled during collection.
Are relationships between individual instances made explicit?
No, the dataset does not contain explicit relational data between the instances.
Are there recommended data splits?
Yes, the dataset is typically split into 70% training, 15% validation, and 15% testing. This split helps in validating the generalization of the model on unseen data.
Are there any errors, sources of noise, or redundancies in the dataset?
Variations in lighting, slight occlusions, and different facial angles introduce noise. There are no known redundancies or errors as each image is unique.
Is the dataset self-contained, or does it link to or otherwise rely on external resources?
The dataset is self-contained and does not rely on external resources.

Does the dataset contain data that might be considered confidential?
No, there is no confidential data as all images are either from a public dataset or obtained with consent for use in this project.
Does the dataset contain data that, if viewed directly, might be offensive or otherwise cause anxiety?
No, the dataset only contains facial expressions and does not include any offensive or anxiety-inducing content.
Does the dataset identify any subpopulations?
No, the dataset does not specifically identify any subpopulations such as age, gender, etc.
Is it possible to identify individuals from the dataset?
No, it is not possible to identify individuals as the dataset focuses solely on facial expressions without linking to personally identifiable information.
Does the dataset contain sensitive data?
No, the dataset does not contain sensitive data related to race, ethnic origins, sexual orientations, religious beliefs, political opinions, or other sensitive information.

A.3 Section 3.3: Collection Process

How was the data associated with each instance acquired?
The data was acquired directly by capturing facial expressions using a digital camera and collecting from a public dataset, FER-2013.
What mechanisms or procedures were used to collect the data?
Digital cameras were used to capture facial expressions, and images were also downloaded from the FER-2013 dataset through its public API.
If the dataset is a sample from a larger set, what was the sampling strategy?
The sampling was opportunistic, capturing images from volunteers available at the time. FER-2013 provides a comprehensive pre-defined sample of facial expressions.
Who was involved in the data collection process?
Students and volunteers participated in the data collection process. They were not compensated as they volunteered for academic support.
Over what timeframe was the data collected?
Data collection occurred over several weeks during the academic semester.
Were any ethical review processes conducted?
Yes, the project was reviewed and approved by the university's ethics committee to ensure compliance with ethical standards for using human subjects.
Did you collect the data from the individuals directly, or obtain it via third parties or other sources?
Data was collected directly from individuals who volunteered and from the publicly available FER-2013 dataset.
Were the individuals notified about the data collection? Yes, all participants were fully informed about the purposes of the data collection and its usage.
Did the individuals consent to the collection and use of their data?
Yes, permission was obtained from all individuals before collecting their facial expressions.
If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?
Yes, participants were informed that they could withdraw their consent at any time without any consequences.
Has an analysis of the potential impact of the dataset and its use on data subjects been conducted?
An initial impact assessment was performed to ensure no harm to the participants, focusing on privacy and data security.

A.4 Section 3.4: Preprocessing / Cleaning / Labeling

Was any preprocessing/cleaning/labeling of the data done?
Yes, preprocessing was essential for our facial recognition model. All images were converted to grayscale to reduce complexity. Faces were detected and cropped to focus on the facial expressions. Each image was resized to a uniform size of 48x48 pixels, and normalization was applied to scale pixel values between 0 and 1. Labeling our own photos is done by ourselves.
Was the "raw" data saved in addition to the preprocessed / cleaned / labeled data?
Yes, the raw data was saved separately to ensure that we could revisit any stage of preprocessing for adjustments or to support future uses that might require raw images.
Is the software that was used to preprocess/clean/label the data available?
The preprocessing was conducted using Python with libraries such as OpenCV for image processing and NumPy for numerical operations. These are open-source tools and are freely available.

A.5 Section 3.5: Uses

Has the dataset been used for any tasks already?
Yes, the dataset has been used to train and test a facial recognition system designed to recognize different emotional expressions for our interactive game project.
Is there a repository that links to any or all papers or systems that use the dataset?
Currently, there is no public repository linking to papers or systems that use this dataset as it was developed specifically for academic coursework.
What (other) tasks could the dataset be used for?
Besides emotion recognition in gaming, this dataset could potentially be used in other applications requiring facial emotion analysis.
Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?
The dataset's focus on a limited set of emotions and the specific preprocessing techniques used might not generalize well to other facial recognition tasks or datasets with different emotional categories or environmental conditions.
Are there tasks for which the dataset should not be used?
The dataset should not be used for clinical or diagnostic purposes

as it was not designed with the necessary rigor for medical applications.

A.6 Contributions

Sean Wallace: game,document
Nagisa Nakamura: model,document
Aurora Yang: model,document