# CNN FACIAL EXPRESSION DETECTION MODEL

**Hecan Zhou**
Student# 1006069326
zhouheca@mail.utoronto.ca

**James Kwok**
Student# 1005772601
james.kwok@mail.utoronto.ca

**Alan Lau**
Student# 1006055260
alanw.lau@mail.utoronto.ca

**Angelica Wittenberg**
Student# 1005757922
a.wittenberg@mail.utoronto.ca

## 1 INTRODUCTION

Facial expression detection has many different applications such as marketing and psychology, human-computer interaction. Leveraging deep learning, our project aims to develop a facial expression detection model capable of accurately identifying and categorizing emotions in static facial images and dynamic image sequences. For marketing purposes, facial expression detection can help personalize content delivery based on users' emotional responses. An example would be to continue showing content that has regularly made a user "happy". For psychology purposes, facial expression detection can help with mental health diagnosis such as detecting signs of depression, anxiety, trauma. Deep learning has demonstrated extraordinary success in image classification. In essence, this is an image classification problem by using facial features in order to classify an emotion. Due to this, deep learning is well suited for a facial expression detection project.

## 2 ILLUSTRATION

Illustration of our facial expression model

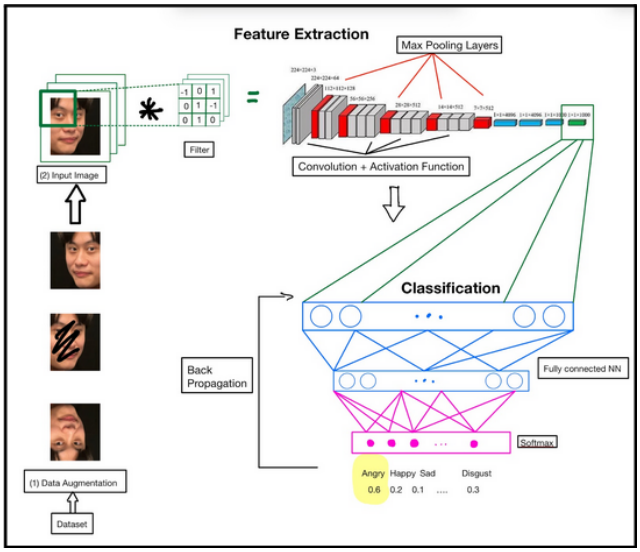


Figure 1: Illustration of architectureal concept

## 3 BACKGROUND/RELATED WORK

This GitHub page discusses using Convolutional Neural Networks (CNNs) and transfer learning for facial recognition, *Facial recognition via transfer learning* **?** available at `https://github.com/jeffprosise/Deep-Learning`. focusing on classifying faces by specific metrics. Our project could adapt this framework to detect emotions like smiles, anger, and sadness in facial expressions. We can use a similar dataset to the project above, applying a new labeling scheme to annotate emotions for training our CNN. This involves adjusting the neural network to capture the emotional expressions and using a different labeling system to train and monitor losses. In essence, for our project, we could repurpose this facial recognition model and code for emotion detection, using a shared dataset and transfer learning to develop a model that classifies facial expressions into specific emotions. We can take a pre-trained model that uses architecture such as ResNet and apply adjustment to make it fit for our usage. There also needs to be a pre-processing step that our team could develop and look into. Preprocessing will prepare all data input for better recognition, base on paper from deep learning survey, a preprocessing step to align and highlight facial structure and area allows for the data input to have better quality and accuracy, leading to trained model that has better prediction and less losses.

Another project we have taken a look into that exists is *facenet* **?** available at `https://github.com/davidsandberg/facenet`. This project again is a facial recognition model that uses ResNet to train a facial detection and recognition model. From this project, we could modify the label to have it fit for emotions instead of specific character and person. The repo also mentions a use of a pre-processing alignment to pre-align and modify input data for better accuracy and something we can look into in our project to incorporate for better reliability and less losses.

The DISFA dataset provided in the paper *DISFA Denver Intensity of Spontaneous Facial Action* **?** available at `https://paperswithcode.com/dataset/disfa`. comprises 27 videos, totaling 130,788 frames of facial expressions, and focuses on spontaneous facial expressions to more accurately capture the natural nuances of these expressions. It serves as a dataset for models to train on. The author of this paper annotated the dataset based on the Facial Action Coding System (FACS), marking action units (AUs) for different levels of intensity. For our purposes, we can utilize these intensity annotations to develop new models that recognize not only the presence of an emotion but also its intensity. This approach enhances the classification problem by allowing the model to capture more intricate details of human emotions.

The RAF-DB mentioned in the paper *Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild* **?** available at `https://paperswithcode.com/dataset/raf-db`. consists of 29,672 facial images tagged with basic or compound expressions, collected under real-world conditions. It encompasses a wide range of variability in age, gender, ethnicity, head poses, lighting conditions, and occlusions. While the DISFA dataset provides depth, the RAF-DB offers breadth. It is instrumental in training models to identify facial features in photos that represent real-world scenarios. By utilizing this dataset, we can explore how occlusions, lighting variations, and head poses affect the recognition process. This, in turn, increases the reliability of the model in real-world applications.

The following paper *Deep Facial Expression Recognition: A Survey* **?** available at `https://ieeexplore.ieee.org/abstract/document/9039580`. goes through many different models used in Facial Expression Recognition (FER). Some of which are described below. Network Ensemble methods in (FER) involve the integration of spatial and temporal information from various networks and combining their outputs to improve performance. One example is a multi-channel network that extracted spatial and temporal information from emotion-expressing faces and changes between emotional and neutral faces. Deep Spatial temporal Networks are specifically designed to capture both spatial and temporal dependencies in consecutive frames, making them well-suited for FER tasks on dynamic sequences. These networks typically combine architectures such as Recurrent Neural Networks (RNNs) and Convolutional 3D (C3D) models. RNNs capture temporal dynamics of the learned features. C3D extends traditional CNNs by adding a 3D filter over very short video clips.

## 4 DATA PROCESSING

For our project, the team will use the CK+ data set from Kaggle *CK+dataset* **?** available at `https://www.kaggle.com/datasets/shuvoalok/ck-dataset`. Which is sourced from the Extended Cohn-Kanade Database. It is a comprehensive collection of facial expression sequences from 123 subjects that contains posed expressions of seven basic emotions: anger, contempt, disgust, fear, happiness, sadness and surprise. Each image is annotated with the corresponding emotion label, making it ideal for training and evaluating facial expressions through our model. The dataset is publicly available so there is no need for additional data collection efforts. Some data cleaning will need to be done such as resizing images to a consistent resolution, converting to grayscale and normalizing pixel values to the [0,1] range. In the preprocessing phase, the data will be checked for any corrupt or unuseable images. Since the CK+ dataset is relatively small, the team will consider augmenting the data to increase diversity and robustness by using rotation, flipping or adding noise to the dataset. The data will need to be formatted into directories based on the different emotions, as well as ensure that each image is properly labeled with the correct emotion category for supervised learning. The dataset will be split into training, validation and testing sets, either 80-10-10 or 70-15-15.

## 5 ARCHITECTURE

We plan on using a convolutional neural network with max pooling to reduce the spatial dimensions of the feature map. Padding should be used to preserve the edges of the images where distinct facial features may be located. Hyperparameters will also be tuned such as learning rate, batch size, and the number of epochs. For the activation function, ReLU will be used.

## 6 BASELINE MODEL

In the development of a Convolutional Neural Network (CNN) for face detection and facial recognition of facial expressions, our team plans to leverage established baseline models that utilize CNN architectures, such as AlexNet and ResNet. These will serve as a solid foundation for comparison, enabling us to test and benchmark our results effectively. Additionally, we will consider utilizing models pre-trained on architectures like VGG, time permitting. Employing pre-trained models based on ResNet or VGG as the backbone architecture allows us to use datasets such as the FERDataset or VGGFace2. This approach facilitates the evaluation of the impact of hyperparameter modifications, as well as the effects of increasing or decreasing the number of layers in our architecture, to fine-tune and adjust our model for enhanced accuracy.There are pre-existing models available in open-source GitHub repositories that have been trained with ResNet. For instance, the FaceNet project hosted at *facenet* **?** available at `https://github.com/davidsandberg/facenet`. offers a pre-trained model using ResNet, boasting high accuracy and providing valuable data for comparison. Furthermore, the project available at *pytorch-facial-recognition* **?** available at `https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch`. presents models trained with VGG19 and ResNet18, offering an additional benchmark for our results

## 7 ETHICAL CONSIDERATION

The primary ethical concern for the topic of this project involves privacy and consent. Our model requires the input of photos or videos, which contain facial data that is inherently personal and sensitive. The collection and analysis of individuals' images should be conducted with consent. We must ensure that individuals are fully informed about the data being collected, its intended use, and with whom it will be shared. This is particularly important in the current era, given the potential misuse of deepfake technology and other tools that can be exploited for impersonating individuals.

Another significant ethical issue is the potential for inherent bias within the detection models. These biases often come from imbalanced training datasets that fail to adequately represent minorities within the human population across dimensions such as race, gender, age, and culture. Quoting Joy Buolamwini's findings in "The Coded Gaze," it is evident that artificial intelligence systems, especially those involving facial analysis technologies, can reflect the biases of their creators and the

Table 1: Project Plan and Breakdown

| Person | Task | Deadline | Progress |
|---|---|---|---|
| | Proposal - Introduction | Sunday 11th February | Completed |
| | Proposal - Illustration | Sunday 11th February | Completed |
| Team | Proposal - Background & Related Work | Sunday 11th February | Completed |
| Angelica | Proposal - Data Processing | Sunday 11th February | Completed |
| | Proposal - Architecture | Sunday 11th February | Completed |
| | Proposal - Baseline Model | Sunday 11th February | Completed |
| | Proposal - Ethical Considerations | Sunday 11th February | Completed |
| | Proposal - Project Plan | Sunday 11th February | Completed |
| | Proposal - Risk Register | Sunday 11th February | Completed |
| Hecan | GitHub Setup | Monday 5th February | Completed |
| Team | Proposal - Structure, Grammar & Mechanics | Sunday 12th February | Completed |
| Hecan | Proposal - Format to Latex | Monday 12th February | Completed |
| | Progress Report - Project Description | | Not Started |
| | Progress Report - Summary of Responsibilities and Contributions | | Not Started |
| | Progress Report - Data Processing | | Not Started |
| | Progress Report - Baseline Model | | Not Started |
| | Progress Report - Primary Model | | Not Started |
| | Data Splitting & Processing | | Not Started |
| | Model Creation | | Not Started |
| | Demonstration | | Not Started |
| | Results | | Not Started |
| | Takeaways | | Not Started |
| | Report Write-Up | | Not Started |
| | Presentation Slides | | Not Started |

datasets they are trained on. Many models have been shown to exhibit lower accuracy for women of color, highlighting a troubling discrepancy. This issue underscores the importance of fairness and equity in the field and points to the need for increased transparency and accountability within the machine learning community. To mitigate these issues, it is essential to utilize a diverse and representative dataset during the training phase. Ensuring that the dataset includes sufficient representation from various demographic groups and ethnicities is vital to achieving a fair training process. This approach guarantees that the resulting model performs consistently across all individuals without inherent biases.

## 8 PROJECT PLAN

Table 1 above is a table summarizing each team member's tasks and deadlines.

To ensure the completion of the project, the team has implemented a divide and conquer approach whereby each member has assigned tasks with internal deadlines. Each member is expected to complete their tasks by the internal deadline. The team will have weekly touch points on Sundays at 6PM to review and discuss the project progress, challenges and next steps. Communication will primarily be done through the team's APS360 Discord server for quick updates and discussions. For our weekly touchpoints and any important decisions or complex problems, the team will connect over audio calls via Discord. To maintain code versions and avoid overwrites, the team will use Git for version control. Each team member will work on separate branches, and before merging to the main branch, a code review will be conducted to ensure code quality and prevent conflicts.

## 9 RISKS

Starting with logistic risks, if a teammate drops the course, the rest of the teammates will divide the tasks of the dropped teammate equally. Time constraints also pose another risk as the team may not

be able to meet deadlines as a result of various reasons such as model failures or a model taking too long to train. To mitigate this risk, the team will have internal deadlines that are reasonably placed ahead of the real deadlines for some extra time to deal with these sorts of setbacks. A model that takes too long to train may be encountered and even with an internal deadline, the model will have to keep training past the deadline. In this scenario, the team will closely monitor checkpoints of the model and verify whether the model is progressing according to expectations. If this is not the case, hyperparameters will have to be tuned and model complexity will have to be decreased to speed up the training. The availability of team members may be impacted due to unforeseen circumstances such as illness. While it is not possible to predict these circumstances, the onus is on the unavailable team member to offload some of their tasks onto the other team members. In this way, the team will still be able to meet the internal deadlines that are planned. Another risk could be the quality of the dataset. A poorly chosen data set could lead to biased or inaccurate models. To mitigate this risk, appropriate research has been done on the chosen dataset and was chosen on the criteria that it has been widely used.

## 10 LINKED TO GITHUB

https://github.com/hecan1234/Aps360Project

## 11 REFERENCES

## 12 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

### AUTHOR CONTRIBUTIONS

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

### ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.