IDS 705 Project Proposal

Feeling It: Emotion Classification with Machine Learning

Using Convolutional Neural Networks to Detect and Analyze Human Emotions from Facial Images

Team Members:
Andrew Kroening
Chloe (Ke) Liu
Wafiakmal Miftah
Jenny (Yiran) Shen

Team Number: 7

Introduction

A significant amount of historical psychological research has explored how humans interpret the emotional states of one another by examining facial expressions. Of particular interest, research has sought to map emotional states to the activation of specific facial muscle groups giving rise to popular phrases such as, “it takes fewer muscles to smile than frown, so be happy!” In recent years, expanding computational power and advanced methods have also brought computers and computer vision to this space. Numerous computational methods have attempted to recreate or expand upon this knowledge by examining pictures or videos of faces using various techniques.

The COVID-19 pandemic introduced a new wrinkle to identifying emotions from observation. As much of the world adopted facemasks as a standard practice, questions naturally emerged about the possible impact of obstructing a portion of a face on human interaction and emotional inference. While this challenge has already existed in certain cultures where the wearing of items that obstruct a part of a person’s face might be tradition, the COVID-19 pandemic brought this question to the masses. To date, various studies have explored this topic, and we will seek to build upon this knowledge as a starting point for our experiments.

Goal/Objective

This project aims to gain insights into how specific machine learning models might infer human emotions from facial representation, compare those results to traditional psychological research, and then challenge that framework by obstructing portions of the faces to estimate those impacts on emotional inference.

Background

To date, considerable efforts have sought to explore the interpretation of facial emotions using modeling techniques. In a cursory review of those efforts, we found that most were worthy attempts to explore this space. However, some of the research we uncovered was outdated and would benefit from more advanced models and computing power developed in the years since publication, such as O’Toole and Abdi’s work[1] from 2001. We did find some promising contributions in recent years from Dajose[2] and Dores[3], and an interesting piece from Heaven[4] that we think has potential application to our efforts.

We seek to incorporate and leverage this literature by providing a baseline that will serve as a measure to compare certain aspects of our model’s performance. One area with keen interest is the regions of a

[1]O'Toole, A.J., and H. Abdi. “Face Recognition Models.” *International Encyclopedia of the Social & Behavioral Sciences*, 2001, pp. 5223-5226. *Science Direct*, https://www.sciencedirect.com/science/article/pii/B0080430767006902.
[2] Dajose, Lorinda. “Facial Expressions: How Brains Process Emotion.” *Caltech*, 24 April 2017, https://www.caltech.edu/about/news/facial-expressions-how-brains-process-emotion-54800. Accessed 9 March 2023.
[3] Dores, Artemisa, et al. “Recognizing Emotions through Facial Expressions: A Largescale Experimental Study.” *National Library of Medicine*, vol. 20, 2020, p. 7420. *National Library of Medicine*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599941/.
[4] Heaven, Douglas. “Why faces don’t always tell the truth about feelings.” *Nature*, 2020, https://www.nature.com/articles/d41586-020-00507-5. Accessed 08 March 2023.

face that are significant for predicting each emotional state. We already know that certain emotions trigger specific muscle groups, and we would like to incorporate a comparison of our model's insights with traditional psychological research. Put simply, we want to compare what our model sees versus what a human sees.

With that knowledge, we can then expand upon the problem space by obstructing portions of the pictures that our model sees to ascertain an effect on the overall performance. We do this to address a question that extends from the first. Once we know what the model is seeing and basing decisions on, does it remain equally effective when we challenge it by being unable to observe some of those critical areas? We see this as a key to understanding the impact of covering portions of a face on emotional inference and being able to make an educated guess about how this may also impact a human's ability to do the same.

Data

The facial expression dataset originated from Kaggle[5]. This dataset consisted of 7 facial expressions from various gender, race, and age. The list of facial expressions and the number of images for each train and validation data are stated in table 1.

Table 1. Dataset properties

Facial Expression	Train Images	Validation Images
Angry	3,993	960
Disgust	436	111
Fear	4,103	1,018
Happy	7,164	1,825
Neutral	4,982	1,216
Sad	4,938	1,139
Surprise	3,205	797

The images also have been resized to 48x48 pixel to focus on the facial features of a person, greyscaled, and labeled accordingly. Our group chose this dataset because it supports our goal of learning the features most important in facial expression recognition. This robust dataset will also allow us to experiment with data augmentation.

[5] Oheix, Jonathan. "Face expression recognition dataset." *Kaggle*, 2019, https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset. Accessed 9 March 2023.

Above is an excerpt of the dataset, with five samples of each of the seven emotions. As seen from the sample images, there is considerable variation within each class that our model(s) will have to contend with. For instance, we can observe hands-on or covering parts of faces, young people and older adults, angled faces with respect to the photo, and even some watermarks. While the source of the data we derived from Kaggle is unclear, the presence of many easily identified personalities suggests these are images collected from various internet sites and manually cataloged.

Methods

Convolutional Neural Network

For the task of facial expression recognition, a Convolutional Neural Network (CNN) was selected as the preferred method. CNNs are especially effective at identifying and extracting features from images, which makes them an ideal fit for tasks like image classification. According to Chang et al.[6], the architecture of a CNN, which includes multiple convolution and pooling layers, can effectively extract complex features from the entire face and specific local regions in the facial expression recognition task. This could therefore result in better classification performance for facial expression images.

[6] Chang, T., Wen, G., Hu, Y., & Ma, J. (2018). Facial expression recognition based on complexity perception classification algorithm. *arXiv preprint arXiv:1803.00185*.

Facial Image Preprocessing

Pre-processing input images before feeding them into a Convolutional Neural Network (CNN) can significantly improve the accuracy of image classification models[7]. In this facial expression recognition task, where images are generally low-resolution, pre-processing can be particularly helpful in enhancing the visibility of important facial features, such as the eyes, mouth, and eyebrows, which are crucial for distinguishing between different expressions. Techniques such as histogram equalization and image sharpening can be used to enhance the quality and visibility of these features[8], and data augmentation methods can be applied to artificially increase the size of the training dataset and expose the model to a broader range of input variations.[9] By pre-processing input images using these techniques, the CNN can more accurately identify and extract the relevant features from the image, improving facial expression recognition accuracy.

CNN Architectures For Feature Extractions

Kim et al.[10] suggested that using various network architectures, input normalization techniques, and random weight initialization methods during deep model training can improve the generalization performance of models. Therefore, different CNN architectures, such as ResNet and VGGNet, could be evaluated to help select the best architecture for our facial expression recognition task.

Additionally, the Attention Convolutional Network proposed by Minaee and Abdolrashidi can be applied to detect different facial expressions[11], as they emphasized that emotions are sensitive to other parts of the face. Hence, identifying salient regions of facial images using attention mechanisms can help us achieve more accurate results since some images in the dataset only show partial faces. By exploring different network architectures and incorporating attention mechanisms, facial expression detection models can provide more robust and accurate results.

Model Interpretability Improvement

Rather than directly incorporating attention-based interpretability into CNNs, which only highlight the input parts that the model focuses on, deep models' interpretability can be further enhanced by utilizing the Prototypical Part Network (ProtoPNet) introduced by Chen et al[12]. The ProtoPNet extends the focus beyond the input parts by identifying and highlighting similar prototypical cases related to those parts, thereby enhancing interpretability. In addition, we can utilize a hierarchical system of prototypes to

---

[7] Ranganathan, G. "A study to find facts behind preprocessing on deep learning algorithms." *Journal of Innovative Image Processing (JIIP)*, vol. 3, no. 01, pp. 66-74.
[8] Shin, M., et al. "Baseline CNN structure analysis for facial expression recognition." *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 724-729.
[9] Jeon, J., et al. "A Real-time Facial Expression Recognizer using Deep Neural Network." *International Conference on Ubiquitous Information Management and Communication*, 2016, pp. 1-4.
[10] Kim, B. K., et al. "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition." *Journal on Multimodal User Interfaces*, vol. 10, 2016, pp. 173-189.
[11] Minaee, S., et al. "Deep-emotion: Facial expression recognition using attentional convolutional network." *Sensors*, vol. 21, no. 9, 2021, p. 3046.
[12] Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., & Su, J. K. (2019). This looks like that: deep learning for interpretable image recognition. *Advances in neural information processing systems*, *32*.

provide multiple explanations for image classification at each level of taxonomy, further improving interpretability[13].

Experiments

In many real-life scenarios, facial recognition technology must face various challenges. For example, individuals may not always show their entire faces, wearing glasses or sunglasses, masks in hospitals, hats for fashion, or hijabs due to religious practices. Additionally, cameras may capture only a portion of an individual's face or may be affected by changes in lighting. To address these challenges, we propose conducting experiments on data augmentation techniques that can improve facial detection and emotion classification performance.

First, we will train our baseline model without any data augmentation techniques. Then we will conduct two experiments to test the effectiveness of different approaches:

Experiment 1 (Facial Obstruction)

In our first experiment, we will use masking techniques to cover different parts of the face:

- Top 1/3 mask: corresponds to the covering of the forehead.
- Middle ¼ mask: corresponds to the covering of the eyes.
- Bottom half mask: corresponds to the covering of the mouth and lower face.

Inspired by Minaee and Abdolrashidi's work[14], we will occlude a square region of size NxN in the top-left corner of an image and make a prediction using the trained model. The occluded area is considered significant if the model makes a wrong prediction. This procedure is repeated for different sliding windows of NxN with a stride of s to generate a saliency map of essential regions for detecting emotions in various images.

In this case, we can also use our masking techniques to create saliency maps of vital areas in detecting emotions in different images. We will compare the model's accuracy trained using masked data with the baseline model and (ideally) use Zeiler and Fergus's[15] method to visualize features in a fully trained model and compare them with masking results.

[13] Hase, Peter, Chaofan Chen, Oscar Li, and Cynthia Rudin. "Interpretable image recognition with hierarchical prototypes." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 7, pp. 32-40. 2019.
[14] Minaee, Shervin, and Amirali Abdolrashidi. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Networks." 2019, https://arxiv.org/pdf/1902.01019v1.pdf
[15] Zeiler, Matthew, and Rob Fergus. "Visualizing and Understanding Convolutional Networks." 2013. *Visualizing and Understanding Convolutional Networks*, https://arxiv.org/pdf/1311.2901.pdf.

Experiment 2 (Data Augmentation)

Next, inspired by Pei, Zhao, et al.[16], we will apply image rotation, translation, zoom, and brightness to our original image data to augment the training data further. To ensure the real-world applicability of our model, we will train it on various human stock photos and test it on laptop webcam images. We will use standard evaluation metrics, including accuracy, precision, recall, and F1, to compare and evaluate the results of our experiments.

Roles

Andrew Kroening
- Task Manager
- Data Cleaning and Pre-processing
- Reports Editing and Formatting

Chloe (Ke) Liu
- Project Technical Lead
- Model Evaluation
- Model Experimentation
- Model Predictions

Wafiakmal Miftah
- Github Repository Manager
- Data Exploration and Visualization
- “Red Team” Code Review

Jenny (Yiran) Shen
- Literature Review
- Model Building and Tuning
- Model Experimentation

[16] Pei, Zhao, et al. “Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments.” *Electronics*, 2009. *Semantic Scholar*, https://www.semanticscholar.org/paper/Face-Recognition-via-Deep-Learning-Using-Data-Based-Pei-Xu/8449af7f2950e1beacdb5d759ca743815bb59748.

References

Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., & Su, J. K. (2019). This looks like that: deep learning for interpretable image recognition. *Advances in neural information processing systems*, *32*.

Dajose, Lorinda. “Facial Expressions: How Brains Process Emotion.” *Caltech*, 24 April 2017, https://www.caltech.edu/about/news/facial-expressions-how-brains-process-emotion-54800. Accessed 9 March 2023.

Dores, Artemisa, et al. “Recognizing Emotions through Facial Expressions: A Largescale Experimental Study.” *National Library of Medicine*, vol. 20, 2020, p. 7420. *National Library of Medicine*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599941/.

Goodfellow, Ian J., et al. “Challenges in representation learning: A Report on Three Machine Learning Contests.” *Neural Information Processing: 20th International Conference*, vol. 8228, 2013, pp. 117-124. *Springer*, https://link.springer.com/chapter/10.1007/978-3-642-42051-1_16.

Hase, Peter, Chaofan Chen, Oscar Li, and Cynthia Rudin. "Interpretable image recognition with hierarchical prototypes." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 7, pp. 32-40. 2019.

Heaven, Douglas. “Why faces don’t always tell the truth about feelings.” *Nature*, 2020, https://www.nature.com/articles/d41586-020-00507-5. Accessed 08 March 2023.

Jeon, J., et al. “A Real-time Facial Expression Recognizer using Deep Neural Network.” *International Conference on Ubiquitous Information Management and Communication*, 2016, pp. 1-4.

Kim, B. K., et al. “Hierarchical committee of deep convolutional neural networks for robust facial expression recognition.” *Journal on Multimodal User Interfaces*, vol. 10, 2016, pp. 173-189.

Minaee, S., et al. “Deep-emotion: Facial expression recognition using attentional convolutional network.” *Sensors*, vol. 21, no. 9, 2021, p. 3046.

Minaee, Shervin, and Amirali Abdolrashidi. “Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Networks.” 2019, https://arxiv.org/pdf/1902.01019v1.pdf.

Oheix, Jonathan. “Face expression recognition dataset.” *Kaggle*, 2019, https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset. Accessed 9 March 2023.

O'Toole, A.J., and H. Abdi. “Face Recognition Models.” *International Encyclopedia of the Social & Behavioral Sciences*, 2001, pp. 5223-5226. *Science Direct*, https://www.sciencedirect.com/science/article/pii/B0080430767006902.

Pei, Zhao, et al. “Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments.” *Electronics*, 2009. *Semantic Scholar*, https://www.semanticscholar.org/paper/Face-Recognition-via-Deep-Learning-Using-Data-Based-Pei-Xu/8449af7f2950e1beacdb5d759ca743815bb59748.

Ranganathan, G. “A study to find facts behind preprocessing on deep learning algorithms.” *Journal of Innovative Image Processing (JIIP)*, vol. 3, no. 01, pp. 66-74.

Shin, M., et al. “Baseline CNN structure analysis for facial expression recognition.” *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 724-729.

Vepuri, Ksheeraj Sai. “Improving Facial Emotion Recognition with Image processing and Deep Learning.” *SJSU ScholarWorks*, 2021. *San Jose State University*, https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=2029&context=etd_projects.

Zeiler, Matthew, and Rob Fergus. “Visualizing and Understanding Convolutional Networks.” 2013. *Visualizing and Understanding Convolutional Networks*, https://arxiv.org/pdf/1311.2901.pdf.