Facial Emotion Recognition

Aadi Waghray
Department of Computer Science
Emory University
Decatur, U.S.A.
aadi.waghray@emory.edu

Alex Forbes
Department of Computer Science
Emory University
Decatur, U.S.A.
alex.forbes@emory.edu

Abstract—
Index Terms—

I. INTRODUCTION

Emotion is an abstract concept, which doesn't lend itself well to quantification. This makes classifying emptions a classic task that is hard for computers but easy for humans. Following early developments in the field of psychology that backed the notion of a few cross-cultural emotions that are mostly visualized through facial features, computer scientists were compelled to attempt Facial Emotion Recognition (FER) [2] [4]. Early attempts relied heavily on Haar Cascades, a series of weakly learned filters to identify facial features, or geometric information stored in Action Unites (AUs.) After the rise of convolutional neural networks, which maintain spatial information while generating their own feature maps, most researchers switched to developing and improving convolutional neural networks (CNNs) [1]. A few approaches rose from this development: autoencoders and transfer learning CITE.

While these approaches attempted to solve fundamental issues with CNNs (EXPLAIN), they also improved performance on edge case: occluded images, side profiles, or over-contrasted images CITE. Earlier classical or pre-CNN techniques relied on determining a feature space, such as action units, and learning the connections between features to reconstruct features in the case of occlusion [3]. CNNs are able to learn their own feature maps based on input which allows them to develop a better representation. Both of the aforementioned approaches improve this ability of CNNs. Autoencoders prevent the CNN from falling into local minima by

II. MODELS

A. Convolutional Neural Networks

Convolutional neural networks, CNNs, have a fairly simple structure convolutions followed by sub-sampling and a non-linear activation function. As the kernel sweeps over the data in a convolutional layer, it scales and shifts the scalar output by some weight. These weights are what give the CNN its ability to learn spatial structure from data. CLOSE THIS THOUGHT. The following sub-sampling layer reduces dimensionality to make the cost of computation more manageable. This can come at the loss of information, which must be balanced. Finally, the non-linear activation function serves the same function as it would in any other neural network: allow the model to learn some irregular boundary in feature space.

B. Autoencoders

Autoencoders attempt to resolve the local minima problem. Rather than initializing the weights of the neural network to random values, Autoencoders start with an unsupervised task: encoding and decoding the input using a symmetric network. In the case of a single Convolutional Autoencoders, the network attempts to reconstruct the input image. Through this process, the network is supposed to learn significant commonalities in order to minimize information lost through encoding. This makes decoding easier, but more importantly, it allows the weights on the encoder to be initialized to values more suitable for classification training with similar data. SOMETHING ABOUT NOT OVER TRAINING HERE.

C. Transfer Learning

III. EXPERIMENTAL SETUP

A. Dataset

Given the nature of the project, we were unable to obtain datasets that are standard in the literature such as JAFFE, CK+, or AFEW CITE. But, Kaggle contained many subsets or similar datasets: LIST.

While we had passable general training data, we weren't able to identify a dataset created explicitly with occluded or messy images such as AFEW. This meant we had to either develop our own by manually selecting images in each class or artificially manipulate the given images to rotate the facial orientation, skew saturation/ contrast, or introduce occlusions. To make the best of the data we had, we combined both of these approaches.

B. Convolutional Neural Network

THIS AND SECTIONS BELOW ARE DEDICATED TO 1.) APPROACHES IN THE LITERATURE 2.) DETAILS OF HYPERPARAM 3.) DIAGRAMS

C. Sparse Stacked Autoencoder

D. Transfer Learning

IV. Results

V. Conclusion

References

[1] Felipe Zago Canal, Tobias Rossi Müller, Jhennifer Cristine Matias, Gustavo Gino Scotton, Antonio Reis de Sa Junior, Eliane Pozzebon, and Antonio Carlos Sobieranski. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582:593–617, January 2022.

[2] Albert Mehrabian and Susan R. Ferris. Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 31(3):248–252, 1967.

[3] Yoshihiro Miyakoshi and Shohei Kato. Facial emotion detection considering partial occlusion of face using Bayesian network. In *2011 IEEE Symposium on Computers & Informatics*, pages 96–101, March 2011.

[4] Disa A. Sauter, Frank Eisner, Paul Ekman, and Sophie K. Scott. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6):2408–2412, February 2010.