Detecting Dangerous Scenarios From Body Language and Emotion Using a 2D Convolutional Neural Network

WINONA STATE UNIVERSITY

Alexander Kruschek
Computer Science Department
Advisors: Iyengar Sudharsan and Ming Ma

Abstract

This work proposes using a 2D CNN to classify images including people based on the emotion that is being presented in the image. The emotions that will be primarily focused on in this study will be pain, anger, fear, and suffering.

- May result in a CNN model that could be installed into security cameras.
- The time that could be gained from detecting these situations early may save people from injuries or possibly even save lives.
- Attempts to improve upon existing emotional detection research by focusing more on the person's body and environment.
- Data comes from EMOTIC database of about 24,000 images

Methods

1. The images used from the EMOTIC database were separated into two groups of images, one that contained the images showing pain, fear, anger, or suffering and one that contained every other emotion (26 total). Groups were labeled PFAS and other respectively.
2. An equal number of images were selected from each group and were then used to create the datasets used to train and test the CNN model shown in Figure 1.
3. The model was trained with and without data augmentation, with the best results coming from training without augmentation.

Model Architecture

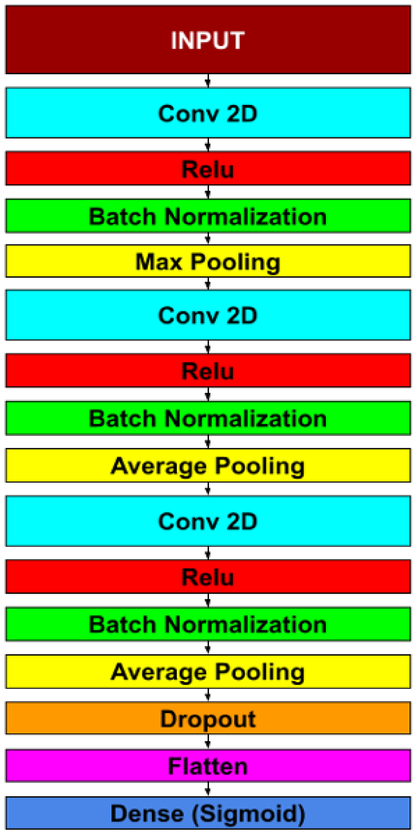

Figure 1. Architecture of the trained model

Dataset Examples

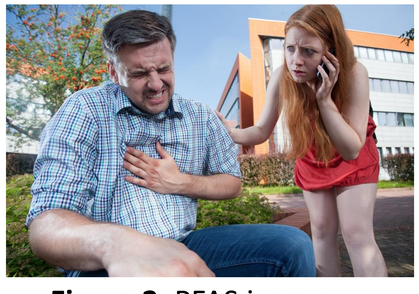

Figure 2. PFAS image

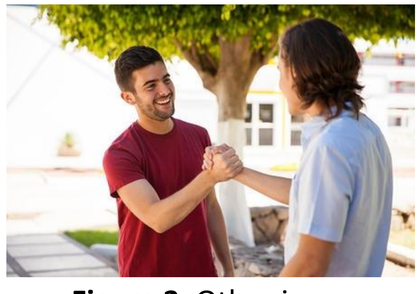

Figure 3. Other image

Above are examples of images from the PFAS group (left) and the other group (right).

Results

Model	Accuracy	Precision	Recall	AUC
Proposed	.5914	.5880	.6171	.6287

Table 1. Model Metrics

*Precision = Quality (TP/(TP+FP))
*Recall = Quantity (TP/(TP+FN))
*AUC (Area Under Curve) = Ability to distinguish different classes

The model was trained using 1282 images over 60 epochs, resulting in an accuracy rate of 59%.
This result is only slightly better than randomly guessing between the groups, which is much too low to be considered for any real-world applications.

Conclusion

The metrics of the model are lower than would be desired, most likely due to limited data. Current model predicts only marginally better than random guesses, but could be improved by:

- Collecting more PFAS images to train model with
- Using videos to detect violent actions more accurately than with images alone

References

1. Kosti, Ronak, Jose M. Alvarez, Adria Recasens, and Agata Lapedriza. "EMOTIC: Emotions in Context dataset." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 61-69. 2017.
2. Akhand MAH, Roy S, Siddique N, Kamal MAS, Shimamura T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics*. 2021; 10(9):1036. https://doi.org/10.3390/electronics10091036