First Project Submission

Deep Learning

Code: 2598038

Group: 01

Date: 29/ 09/ 2024

By

Melissa Galeano Ruiz, CC: 1000416463

Professor

Raúl Ramos Pollán

UNIVERSIDAD
DE ANTIOQUIA
1 8 0 3

Faculty of Engineering

Mechanical Engineering Department

University of Antioquia

Contents

1 Application context

In mechanical engineering, the integration of advanced technologies such as artificial intelligence and computer vision is transforming the way complex problems are addressed. This project aims to develop a real-time facial and emotion recognition model, intended for application in process automation and improved safety within industrial environments.

Leveraging convolutional neural networks or CNNs, this model will detect the emotional states of operators and personnel in industrial plants, optimizing human-machine interaction and allowing automated systems to adjust their responses based on the emotions identified. This innovation can potentially improve workplace safety and optimize mechanical processes by adapting operational conditions to the emotional state of personnel, contributing to a safer and more efficient work environment.

By recognizing emotions such as happiness, sadness, anger, fear, disgust, neutrality, and surprise, the model can provide valuable insights into the workforce's emotional well-being. The detection of surprise, in particular, may indicate that something is not going well, allowing for timely intervention. This information facilitates more informed decision-making regarding task assignments, workload distribution, and the implementation of support and safety measures, ultimately fostering a more responsive and empathetic workplace culture. By addressing the emotional dynamics within industrial settings, the project aims to enhance productivity, satisfaction, and employee safety.

2 Machine Learning objective

The objective of the Machine Learning component of this project is to accurately predict the emotional state of individuals based on facial images presented to the model. This involves training a CNN to recognize and classify different emotions.

This project is a classification task, a common category of supervised learning in machine learning, where the model is trained to map input data (facial images) to specific output labels (emotions). The model aims to achieve high accuracy in real-time emotion detection by leveraging advanced feature extraction techniques and robust training methodologies. This will enhance human-machine interaction in industrial settings, enabling systems to respond appropriately to the emotional states of personnel and improving overall workplace dynamics.

3 Dataset

The dataset used in this project is ***FER-2013***, originally sourced from a Kaggle competition. It is divided into training and test sets. It consists of 48x48 pixel grayscale images of faces, with the training set containing 28,709 examples and the test set comprising 3,589 examples, for a total of 32,298 images stored in *PNG* format. The dataset takes up to 140 MB of storage in total. Additionally, 10% of the training set will be used for validation.

The dataset is organized into folders based on seven emotion categories: *Happy, Sad, Neutral, Angry, Disgust, Fear,* and *Surprise*. While the goal is to work with all seven emotions, if training time becomes excessive due to the high number of classes, the scope will be reduced to focus on four key emotions. This adjustment will help manage computational costs while ensuring the model remains relevant to the mechanical engineering context. Samples from the dataset are presented in Figure 1.

Figure 1: Sample of the training dataset.

The following table illustrates the data distribution of the entire training set within the dataset.

Table 1: DATA DISTRIBUTION OF TRAINING DATASET

Emotion Class	Number of samples
Happy	7,215
Sad	4,830
Angry	3,995
Neutral	4,965
Surprised	3,171
Fearful	4,097
Disgusted	436
Total	**28,709**

4 Performance metrics

The main evaluation metric for this model is *Accuracy*, which is used to assess classification models and is based on the number of true positives or the percentage of correct predictions.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{1}$$

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{2}$$

Where:

- TP : True Positives

- TN : True Negatives
- FP : False Positives
- FN : False Negatives

In addition to accuracy, other metrics such as *Precision, recall,* and the *F1-score* will be used to assess the model's ability to differentiate between different emotions:

- **Precision:** proportion of correct predictions made by the model.
- **Recall (Sensitivity):** proportion of true positives identified out of all actual positives.
- **F1-Score:** the harmonic mean of precision and recall, providing a balance between the two.

On the other hand, in the context of mechanical engineering and industrial environments, safety incident reduction is a primary business metric. By integrating real-time emotion recognition into process automation, the system aims to reduce the occurrence of unsafe conditions, particularly those linked to operator stress, confusion, or surprise. Another relevant business metric is operational efficiency, measured through reduced downtime or enhanced decision-making during emotionally charged situations. Improved safety and operational responses based on emotional recognition directly contribute to fewer errors, higher productivity, and cost savings.

5 Previous results

Numerous researchers have worked with this dataset, achieving varying levels of performance. The CNN model by Ketan Sarvakar and R. Senkamalavalli reached 55.61%. However, FERNet and fine-tuned VGGNet surpassed it with accuracies of 69.57% and 73.28%, respectively. These models utilized advanced techniques like data augmentation, transfer learning from ImageNet, and fine-tuning to enhance their performance. It is important to note that several authors agree that the *FER2013* dataset is one of the most challenging for expression or emotion recognition.

References

[1] *FER-2013 | Kaggle.* (n.d.). Retrieved Sept. 16, 2024, from `https://www.kaggle.com/datasets/msambare/fer2013`

[2] K. Sarvakar *et al.*, "Facial emotion recognition using convolutional neural networks," *Materials Today: Proceedings*, vol. 80, pp. 3560–3564, Jan. 2023, doi: `https://doi.org/m79z`.

[3] J. D. Bodapati *et al.*, "FERNet: A Deep CNN Architecture for Facial Expression Recognition in the Wild," *J. Inst. Eng. India Ser. B*, vol. 103, no. 2, pp. 439–448, Apr. 2022, doi: `https://doi.org/m792`.

[4] Y. Khaireddin *et al.*, "Facial Emotion Recognition: State of the Art Performance on FER2013," May 08, 2021, arXiv:2105.03588, doi: `https://doi.org/m793`.