Facial Emotion Recognition System using FER-2013 with Deep Learning Approaches

by: Mohamed Adel **supervisor**: Professor Amr S. Ghoneim

The BRITISH UNIVERSITY IN EGYPT

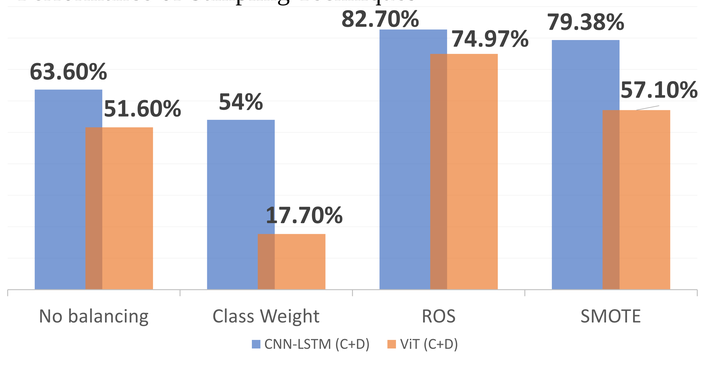
Abstract

Facial Emotion Recognition *(FER)* is a challenge by the ICML in 2013. since then, researchers challenge is to create a best fitting model for FER. In this study, we examine the related work of the FER dataset, focusing on researches done on single-network architecture. The highest reported accuracy on this dataset is 73.2%. This was reported by Khaireddin et. al. in 2021 [2]. In this study, we show that following a data-centric approach easily leads to increase models performance. **This project does not only identify problems that reside within FER dataset that past researchers still need to consider but also reports the highest single network accuracy for the FER-2013 dataset.**

Methodology

Our approach is a data-centric approach showing the performance change on a set of experiments after cleaning the data and training a single-network model architecture. Then, apply fine-tuning to the model to reach highest accuracy. Our methodology include investigating three different sampling techniques **Class Weighting, Random Oversampling, and Synthetic Minority Oversampling Technique.**

Performance of Sampling Techniques



Dataset

FER2013: 35,887 grayscale, normalized images labelled as 7 classes “angry”, “disgust”, “fear”, “happy”, “sad”, “surprise”, and “neutral”. Collected by Google's API engine using 182 keywords like “blissful”, and “meaningful”[1].

- Our challenge includes only FER-2013 dataset with no auxiliary data.

Data-Centric Approach

Outliers

- Using a simple box-and-whisker plot that shows the average pixel value, we have managed to extract 28 misleading outlier image that only added bias and noise to the model.

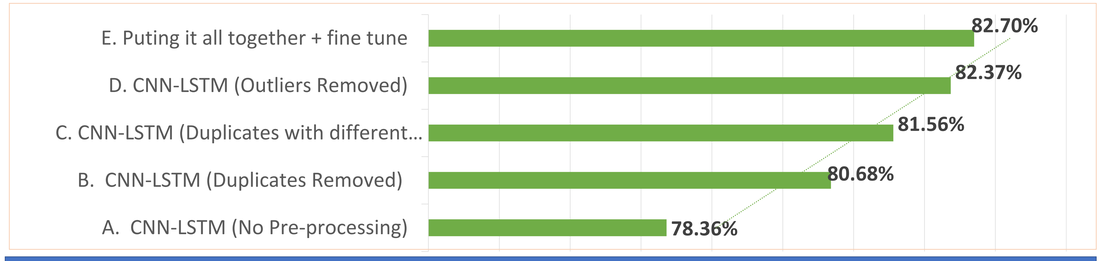
- Dropping 28 outliers led to 4.01 improvement in accuracy.

Duplicates with different emotions

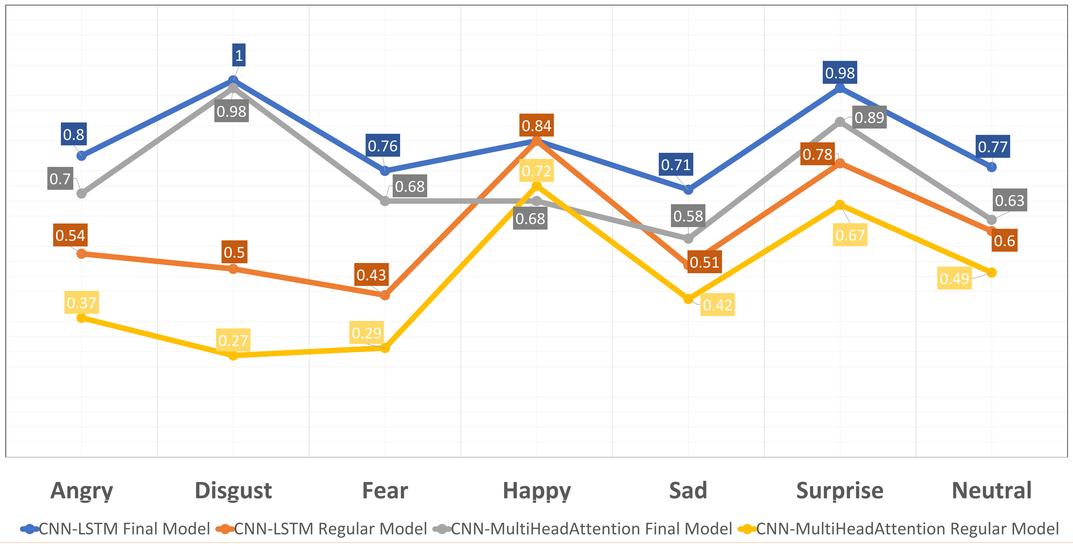
- Investigating 1853 duplicates led us to 160 duplicated image with non-matching emotion.

- Dropping 160 duplicate w/ non-match emotion led to 3.2 improvement in accuracy.

Improvements of Data-centric approach

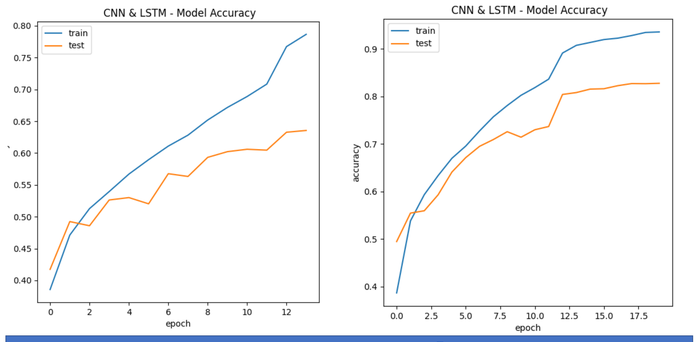


Performance in F1-Score For CNN-LSTM and CNN-MultiHeadAttention



Conclusions

In this work, we have not only achieved the highest attained accuracy of 82.7 % using a single network with no additional data but also, we have managed to introduce new issues that reside within the FER dataset that past authors still need to consider.



Future Work

1. Fine-tune and improve the Vision Transformer model for better performance on the FER dataset.
2. Explore the possibility of creating a cleaned FER dataset: Merge cleaned FER with auxiliary data to represent real-world scenarios of facial emotions & publish the cleaned dataset available for public use.
3. Fine-tune ViT with ROS variation, to test the model’s capacity for improvement.
4. Investigate ensemble models for future studies, Choosing models that oppose each other in terms of emotion classifications (i.e., Binary-Tree classification on multi-class problem).

References

[1] I. J. Goodfellow, et al.,“Challenges in representation learning: A report on three machine learning contests,” arXiv.org, 01-Jul-2013. [Online]. Available: https://arxiv.org/abs/1307.0414. [Accessed: 04-Apr-2023].

[2] Y. Khaireddin and Z. Chen, “Facial emotion recognition: State of the art performance on FER2013,” arXiv.org, 08-May-2021. [Online]. Available: https://arxiv.org/abs/2105.03588. [Accessed: 04-Apr-2023].