FER-2013 Facial Expression Classification: Report

Georgios Athanasiou

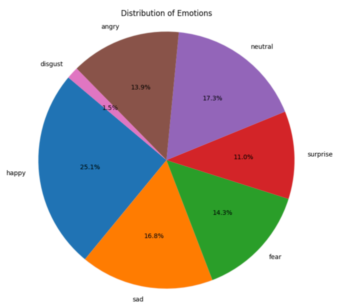
1. Introduction

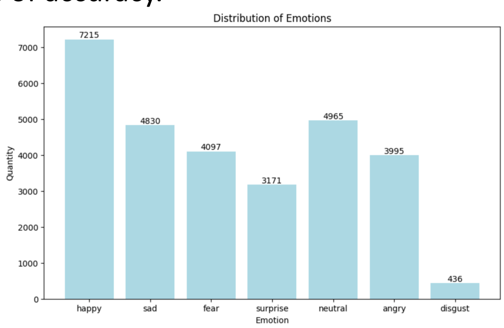
The FER-2013 dataset, a collection of 35,887 grayscale images showcasing seven facial expressions, was introduced in the Kaggle competition "Challenges in Representation Learning: Facial Expression Recognition Challenge." It serves as a pivotal benchmark in the domain of automated emotion recognition. This report delves into our efforts to harness this dataset, detailing our approach to design an innovative classifier, the challenges faced, and the subsequent results achieved.

2. Data preparation

Upon obtaining the FER-2013 dataset, an initial exploration was embarked upon to gain insights into its structure and intricacies. Key observations during this exploratory phase included:

- Class distribution
  An insight of the dataset revealed a skewed distribution among the facial expression classes. Particularly, the 'disgust' class was notably underrepresented compared to other expressions, indicating potential challenges in training a model to recognize this emotion with a high degree of accuracy.





- Ambiguous cases
  During the exploration, certain images were encountered that presented ambiguities in their expression labels. These ambiguities could arise from subtle facial expressions that blur the boundaries between different emotion categories, potentially complicating the classification task.

Given the aforementioned challenges, specific data preparation strategies were employed:

- Oversampling
  To address the class imbalance issue, an oversampling technique was applied.
- Exclusion of 'disgust' class
  Post-oversampling, a strategic decision was made to eliminate the 'disgust' class from training. Given its limited representation and the inherent complexities associated with distinguishing it, it was deemed more prudent to focus on achieving higher accuracies with the remaining, more distinct emotion categories

3. Strategy

Given the nature of the FER-2013 dataset and its inherent complexity, deep learning emerged as the preferred approach to tackle the facial expression classification task. Convolutional Neural Networks (CNNs), with their proven prowess in handling image data, were employed.

- Initial model
  An elementary CNN architecture was the first contender in this task. Upon training, it demonstrated a validation accuracy of approximately 60%. While promising for a starting model, there was clear room for improvement.
- Benchmark research
  To set a performance context, an analysis of the Kaggle competition's leaderboard was conducted. The top 10 models exhibited a mean accuracy of 65%. Further literature review[1] unearthed a notable paper claiming a best-in-class accuracy of 67.2%. Their success was attributed to an ensemble approach, combining the strengths of vgg16 and ResNet50 architectures.
- Exploratory phase
  Drawing inspiration from the benchmarks and keen on pushing the envelope further, several architectures were experimented with vgg13 and vgg16, given their reputation for robustness in image classification tasks.
- Complex architectures
  Beyond the conventional, intricate architectures were tested, harboring the potential to capture nuanced patterns within the dataset.
- Framework evaluation
  This journey of model experimentation was embarked upon using two of the most popular deep learning frameworks - TensorFlow and PyTorch. While both frameworks yielded competitive results, TensorFlow consistently edged out with superior performance. This differential could be attributed to several factors, one of which might be the initialization of weights in TensorFlow, although this remains a point for further investigation. The decision inclined towards TensorFlow, but there are implementation also in PyTorch that are not considered path of the main solution package.

The code is available in the following GitHub page:
https://github.com/ga83wuw/facial_expression_recognition

4. Novelty

To optimize model performance, a hierarchical, two-tiered approach was conceptualized:

- Binary emotion classification
  Objective: Categorize facial expressions into broader emotional spectrums
  'Negative' ('Angry', 'Sad', 'Fear') and 'Positive' ('Happy', 'Surprise', 'Neutral').
  Outcome: A validation accuracy of 85% was
- Granular emotion classification
  For Positive Expressions: A dedicated model trained solely on 'Positive' expressions registered a validation accuracy of 90%.
  For Negative Expressions: When trained exclusively on the 'Negative' expressions, the model attained a validation accuracy of 70%.

Under this paradigm, an image first navigates through the broader classification network, being tagged as either 'Positive' or 'Negative'. Subsequent to this, it's routed to the respective granular network, which refines this categorization into one of the specific facial expressions. This strategy, rooted in simplifying the models and their respective tasks, proved efficacious in enhancing overall accuracy.

5. Results

Upon assembling and deploying the hierarchical training strategy in a pipeline, the outcomes offered intriguing insights:

- Positive expressions
  The performance on the test set mirrored that of the validation set, suggesting that the model generalized well for positive facial expressions without great improvements
- Negative expressions
  For the negative emotions, the pipeline showcased a modest enhancement in accuracy on the test set compared to the literature model, validating the potential of the two-tiered approach for this category.
- Neutral expression
  Contrarily, the neutral expression exhibited a decline in performance on the test set.

6. Discussion

Concluding, the approaches undertaken were preliminary due to limited time, leaving ample avenues for enhancement.

- Transfer learning: While the VGG16 experiment incorporated aspects of transfer learning, the final model remained untouched by its potential benefits.
- Exploration of architectures and loss functions: The architectural landscape of deep learning is vast, with many architectures and loss functions yet to be explored for this specific problem. Diving deeper into these could uncover configurations that further optimize accuracy.
- Generative Adversarial Networks (GANs): A GAN model was conceptualized for data augmentation, envisioning the creation of more training data. However, its performance fell short of expectations and was sidelined. Refining this GAN be crucial in future endeavors.
- Sophisticated approaches: Facial expressions are an amalgamation of subtle cues, be it the arch of the eyebrows, the curvature of the mouth, or the glint in the eyes. An intriguing avenue of exploration could involve dissecting these constituent elements. By building a detection system that recognizes these subtleties, one could superimpose this understanding onto the broader image classification task, potentially enhancing accuracy and robustness.

In summary, while the strides made in this venture are commendable, the realm of facial expression recognition remains expansive, with numerous sophisticated strategies waiting to be harnessed.

[1] https://www.semanticscholar.org/paper/Recognizing-Facial-Expressions-Using-Deep-Learning-Savoiu-Wong/59b50e396e657627cc503a0747c5b84bd17c5468