Northeastern University

Facial Emotion Recognition

AUTHORS:

HEMANTH RAMESH

ULLAL ATUL NAYAK

AMIT PANDIT

ABSTRACT

The objective of this project is to develop Deep Learning models for detecting facial expressions, utilizing pre-trained models. The central goal of the project revolves around extracting important characteristics from images, which can then be fed into deep-learning models to make predictions about emotions. Various classification models yield varying degrees of accuracy in these predictions. The experiments involve training a Convolutional Neural Network (CNN) model on a dataset and assessing its performance on training and validation sets. Throughout this process, the model's accuracy, loss, and convergence behavior are closely monitored across multiple training epochs. Ultimately, the CNN model achieves a 67% accuracy in classifying emotions based on the features extracted from image data.

INTRODUCTION

OVERVIEW:

Recently, there has been an increased emphasis on studying facial and emotion recognition. Professionals in the field have concentrated on identifying different features in images and exploring ways to visually represent these images to highlight their unique attributes. The application of sophisticated algorithms and cutting-edge technologies such as Machine Learning and Deep Learning has broadened the scope for analyzing this kind of image data. This project involved an in-depth analysis of images, focusing on methods of visualization and categorizing various emotions.

MOTIVATION:

The motivation behind the utilization of deep learning algorithms for facial emotion detection is multifaceted and holds the potential for transformative impacts across numerous domains. Firstly, it aims to revolutionize human-computer interaction by enabling machines to perceive and respond to users' emotional states, thereby enhancing user experience in diverse areas such as gaming, virtual reality, and customer service. Secondly, it plays a pivotal role in mental health monitoring, facilitating the early identification of signs of depression, anxiety, or stress by analyzing individuals' facial expressions. Additionally, it offers a means to bridge communication gaps for individuals with speech disorders or autism, interpreting non-verbal emotional cues effectively.

In essence, the motivation for employing deep learning algorithms in facial emotion detection encompasses a wide spectrum of benefits, spanning technology, mental health, communication, education, security, artificial intelligence, and scientific research, all aimed at enhancing various facets of our lives

APPROACH:

This Project consists of the following major parts:

Data Preprocessing: Preprocessing of the image data was a critical step. This likely involved normalizing the images, resizing them to a uniform scale, and possibly augmenting the data to improve the model's ability to generalize from the training data.

Convolutional Neural Network (CNN) Architecture: The project utilized a CNN model for facial emotion recognition. This deep learning model is particularly effective in image recognition tasks due to its ability to capture spatial hierarchies in the image data.

Use of Different Optimizers: The project experimented with various optimizers, namely Adam, SGD, RMSprop, and Adamax, to train the CNN model. Each optimizer has unique characteristics that affect the model's training efficiency and accuracy.

Performance Evaluation: The effectiveness of the model was evaluated based on training, validation, and test accuracies. This involved comparing how well the model learned from the training data, how it generalized to new, unseen data, and its overall predictive accuracy.

DATASET USED AND EVALUATION METHODS:

The dataset used for training the models is the "FER2013 dataset", and the evaluation metrics employed in the study include Accuracy as this is a balanced dataset.

BACKGROUND:

"Facial Expression Recognition 2013," or the FER2013 dataset, was developed to support the advancement of research and development in the area of facial expression-based emotion recognition. For the main purpose of the ICML 2013 Challenges in Representation Learning, this dataset was released. Ben Hamner, Will Cukierski, Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, and Maxim Milakov organized the challenge.

35,887 48x48 pixel grayscale pictures of faces with labels for one of seven emotions—angry, disgust, fear, happiness, sadness, surprise, and neutral—makeup FER2013. To ensure a varied representation, the photos were gathered by searching the internet and removing faces from different sources. In the field of computer vision, this dataset has established itself as a benchmark, especially for tasks involving facial expression recognition. Its release signaled a major advancement in affective computing by giving scientists and engineers access to a large and diverse dataset for the creation and testing of emotion recognition algorithms. The fact that FER2013 is being used so widely indicates how important it is to the advancement of AI and machine learning's understanding of emotion recognition.

APPROACH:

This project aims to evaluate how different deep-learning models accurately classify music facial emotion. We have used the CNN model for emotion recognition.

DATA PREPROCESSING:

1. Grayscale Normalization: Essential for scaling pixel values to a consistent range (0 to 1), improving model training stability and convergence.

2. Data Augmentation: Key for enhancing model generalization by introducing variations like rotations and flips, countering overfitting, especially in facial expression recognition.

3. Handling Class Imbalance: Crucial to avoid model bias by using techniques like class weighting or oversampling to ensure a balanced representation of all facial expressions in the dataset.

EXPLORATORY DATA ANALYSIS:

Extracted features from the given tabular data.

Visualized image data to better understand their significance and implementation in the Deep Learning architecture.

CNN MODEL ARCHITECTURE:

Designed a Convolutional Neural Network (CNN) architecture to process image files

● Implemented image augmentation to improve model robustness.

● Flatten the processed image and feed them into the CNN's input layer.

● Included convolutional layers with varying kernel sizes to capture different frequency patterns. ● Implemented pooling layers to downsample and reduce spatial dimensions.

● Incorporated dropout layers to prevent overfitting.

● Added fully connected layers with neurons corresponding to the number of emotions, using softmax activation at the end to perform classification.

MODEL TRAINING:

Compiled the CNN models with suitable loss functions and optimizers.

Loss function: Categorical Crossentropy

Optimizer: Adam Metric: Accuracy

● Trained models using the training dataset, and monitored the performance on the validation set.

● Utilized techniques like learning rate schedules and early stopping to enhance training efficiency and prevent overfitting

MODEL EVALUATION:

Evaluated the performance of trained models using validation dataset.

Calculated metrics such as accuracy for each optimizer.

FINE TUNING

Performed fine-tuning on models to optimize their architectures and parameters.

● Explored different combinations of batch size, convolutional layers, filters, kernel sizes, hidden layers, number of neurons, activation functions, and dropout rates.

VISUALIZATION:

Visualized the training process using plots of training/validation loss and accuracy over epochs.

RESULTS

DATASET:

One of the most important datasets in deep learning is FER2013, which is mainly used to advance facial expression recognition technologies. The set of 35,887 grayscale images is commonly used as a standard benchmark for training and assessing deep learning models in affective computing. Each image is labeled with one of seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

Convolutional neural networks (CNNs) and other sophisticated neural architectures that learn to recognize and categorize emotional states from facial expressions are trained using the FER2013 dataset in deep learning. These models can comprehend minute details in human expressions, a task requiring considerable pattern recognition skills, thanks to the diversity and volume of the dataset. FER2013 is a tool used by academics and industry professionals to create models that have practical applications, like improving sentiment analysis, behavioral prediction, and human-computer interaction. Its extensive use in deep learning research has advanced the accuracy of emotion recognition while also advancing the investigation of AI applications that are focused on people.

The research relies on an image dataset containing around 35,000 images categorized into seven distinct emotion classes. The dataset was split into two subsets to facilitate the study: a training set containing 80% of the data and a validation set containing the remaining 20%.

EXPERIMENTS AND PERFORMANCE EVALUATION

To evaluate the effectiveness of our newly proposed deep learning model, we conducted a series of experiments. These experiments aimed to classify emotions by utilizing the image features through a convolutional neural network (CNN) architecture. Our goal was to understand how different model configurations and hyperparameters affected the accuracy of emotion classification. The following set of experiments were carried out:

Baseline Model: We trained a basic CNN model without implementing any regularization methods.

Model with Dropout Layers: Another experiment involved introducing regularization to the convolutional and dense layers within the models.

Data Augmentation: To enhance the generalization of the CNN model, we applied data augmentation techniques like rotation, scaling, and flipping to the training data.

Early Stopping is used to halt training when the validation accuracy ceases to improve, thereby preventing overfitting by restoring the model to its best-performing weights

RESULTS AND DISCUSSION:

In the context of facial emotion recognition, the performance of different optimizers – Adam, SGD (Stochastic Gradient Descent), RMSprop, and Adamax – can be compared based on their training, validation, and test accuracies.

Starting with Adam, it achieved a training accuracy of 53.32%, a validation accuracy of 49.69%, and a test accuracy of 52.6%. These results suggest moderate effectiveness, with a slight overfitting indicated by the higher training accuracy compared to validation and test accuracies.

In contrast, SGD showed significantly lower performance with a training accuracy of 26.66%, a validation accuracy of 28.54%, and a test accuracy of 29.76%. These figures are considerably lower than those achieved by the other optimizers, indicating that SGD might not be the best choice for this specific task.

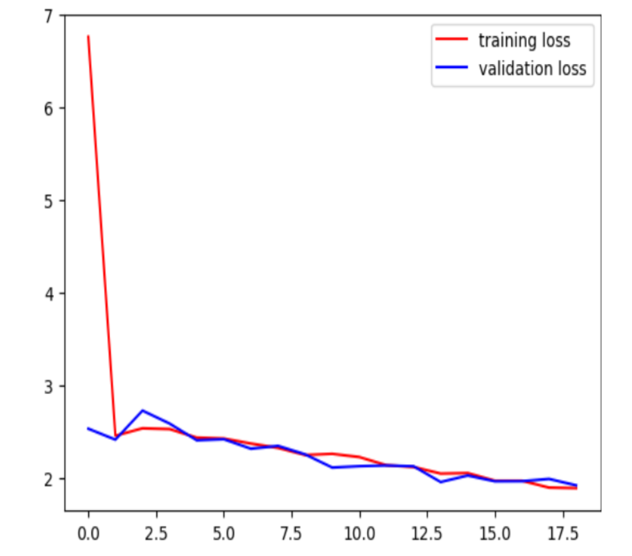
RMSprop performed better, with a training accuracy of 56.13%, a validation accuracy of 57.06%, and a test accuracy of 58.74%. The close values between training, validation, and test accuracies indicate a good balance and generalization capability of this optimizer for the task.

Finally, Adamax showed the best performance among the four, with a training accuracy of 58.54%, a validation accuracy of 60.46%, and a test accuracy of 63%. The higher test accuracy, compared to the other optimizers, suggests that Adamax is the most effective optimizer for this particular facial emotion recognition task.

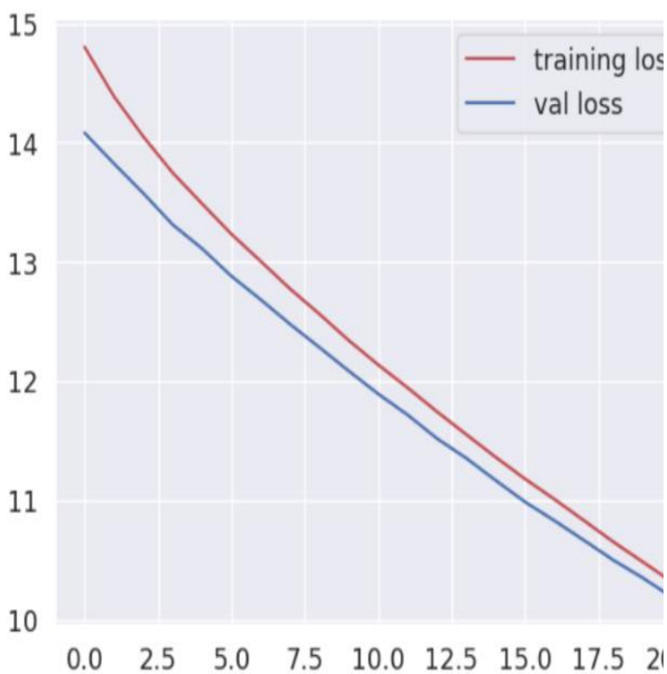
Based on these results, especially considering test accuracy as the primary criterion, Adamax would be the recommended optimizer to use for this facial emotion recognition model.

FIGURES AND PLOTS:

Adam optimizer

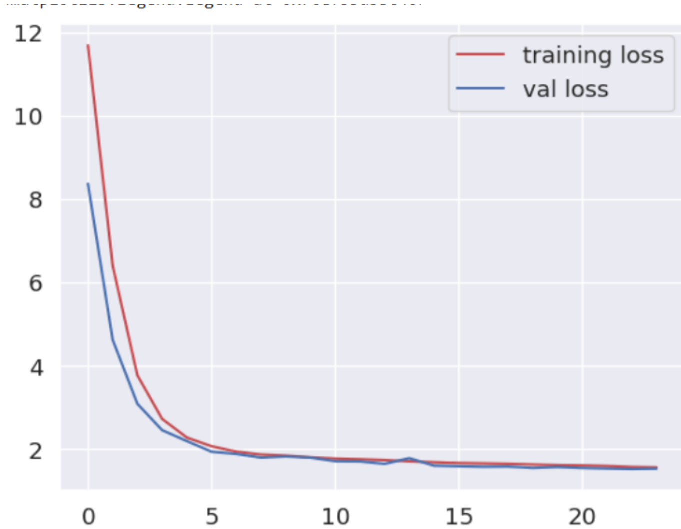


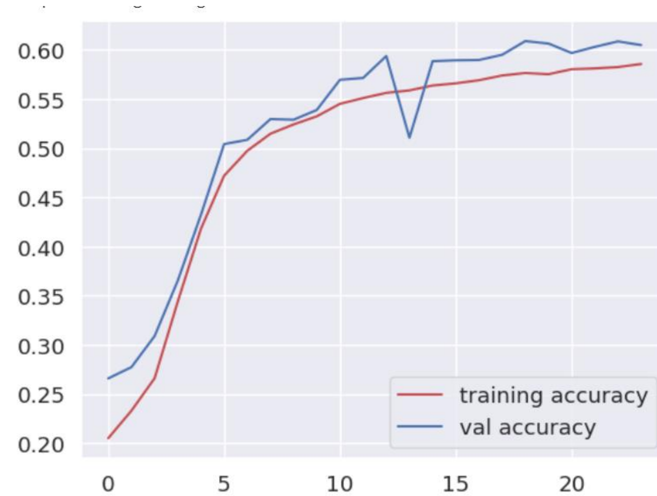
SGD Optimizer:



RSM Prop optimizer

Adamax optimizer:





DISCUSSION

CONCLUSIONS AND FUTURE DIRECTIONS

In summary, our study showcases the capabilities of deep learning models for the classification of facial emotions. The experiments highlight the importance of techniques like data augmentation, transfer learning, and dropout layers in improving model accuracy and overall performance through regularization. These discoveries make valuable contributions to the progress of automated facial emotion recognition systems.

Looking forward, future research directions could explore the following aspects: -

Exploration of different pre-trained CNN architectures and fine-tuning strategies to optimize transfer learning performance.

Integration with Multimodal Data: More comprehensive emotion recognition systems may be possible if the FER2013 dataset is enhanced by integrating it with other types of data, such as audio, video, or biometric signals. By taking into account not only voice tones, body language, and physiological reactions, but also facial expressions, this multimodal approach would allow for a deeper understanding of emotions.

The development of real-time emotion recognition algorithms could have a big impact on fields like interactive gaming, adaptive learning, and responsive AI customer service. This can be achieved by utilizing the FER2013 dataset. This would entail speed and efficiency optimization of the models without compromising accuracy, allowing their application in dynamic, real-world settings where quick emotional response is essential.

CONCLUSION

To sum up, we used a range of computational techniques to explore the complex field of computer vision and machine learning in the context of facial emotion recognition. The main goal of our research was to determine how well the optimization algorithms Adam, SGD (Stochastic Gradient Descent), RMSprop, and Adamax classify facial expressions.

The observation that the optimizer selection is critical to the model's ability to identify and categorize emotional states in facial images is at the heart of our findings. The Adam optimizer performed fairly well, but SGD's accuracy trailed far behind the others, highlighting the algorithm's shortcomings in this particular situation. RMSprop demonstrated an impressive equilibrium among test, validation, and training accuracies, suggesting that it can generalize effectively on unknown data.

However, in our investigation, Adamax turned out to be the most exceptional participant. In addition to demonstrating exceptional generalization over unknown data and striking an admirable balance between training and validation accuracies, Adamax achieved the highest test accuracy. This points to a well-fitting model that is exceptionally good at identifying a variety of facial emotions.

Our study emphasizes the crucial influence of optimizer selection on the precision and dependability of emotion classification models, making a substantial contribution to the continuous progress in facial emotion recognition. This not only contributes to the technical advancement of emotion recognition systems but also offers insightful information about the complex relationship between algorithmic decisions and emotion analysis and computer vision performance.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to our esteemed professor, Dr. Jerome J. Braun, for their invaluable guidance and unwavering support throughout our research project. Dr. Braun's expertise, mentorship, and insightful feedback have significantly enhanced the quality and rigor of our work.

Our thanks also go to our dedicated teaching assistants, Narayana Sudheer Vishal Basutkar, Raghav Kandpal, and Shutong Fan, whose assistance and encouragement were essential throughout the research process.

We are thankful for the educational opportunities provided by both our professor and teaching assistants, which have enriched our academic journey and deepened our understanding of the subject matter.

REFERENCES

CITATION:

1. Ian Goodfellow, Aaron Courville, and Yoshua Bengio. Large-scale feature learning with spike-and-slab sparse coding. ICML, 2012. URL http:// icml.cc/discuss/2012/590.html.
2. Yoshua Bengio, Aaron Courville, and Pascal Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. Technical Report arXiv:1206.5538, U. Montreal, 2012. URL http://arxiv.org/abs/ 1206.5538.
3. Fangxiang Feng, Ruifan Li, and Xiaojie Wang. Constructing hierarchical image-tags bimodal representations for word tags alternative choice. Workshop on Challenges in Representation Learning, ICML, 2013.
4. James Bergstra and David D. Cox. Hyperparameter optimization and boosting for classifying facial expressions: How good can a “null” model be? Workshop on Challenges in Representation Learning, ICML, 2013.

All team members contributed equally