AN INTRODUCTION TO

MACHINE LEARNING

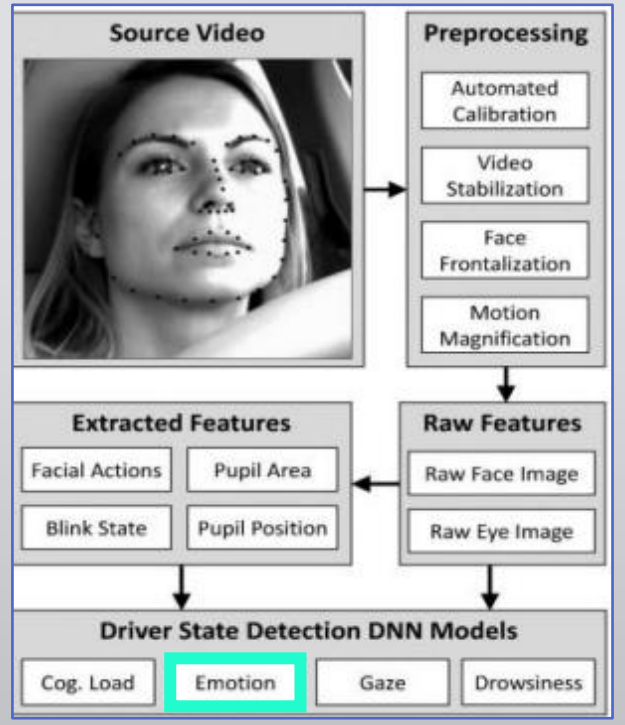
Facial Emotion Recognition with Convolutional Neural Networks: A Study

AUTHOR: LEONEL RAMIREZ III

Motivation

- **Humans often get distracted easily, which leaves the opportunity for automation to step in:**
  - The four D's: Drunk (~31%), Drugged (~23%), Distracted (~10%), Drowsy driving (~3%).
  - Automation results in safer driving but the thing is that we do not understand it, since we lack the data such as forward and backward roadway data, data on the driver. A LOT of data is needed.
- **Driver State Detection Challenge:**
  - Real-world data is "messy", must deal with vibration, lighting variation, body, head, eye movement.

DRIVE STATE DETECTION



SOURCE: HTTPS://DEEPLEARNING.MIT.EDU/

Possible Solutions

- **How is the car supposed to build trust in you when it cannot even perceive you?**
  - Camera facing the driver in every vehicle:
    - The safety benefits are huge.
    - Imagine the data it would have on you; it would know your reactions to just about everything
      - Must have a forward camera to capture outside world and internal to harbor your reactions to that environment.
  - Semi-Automated:
    - Ask a human for help when the machine is not confident.

SOURCE: HTTPS://DEEPLEARNING.MIT.EDU/

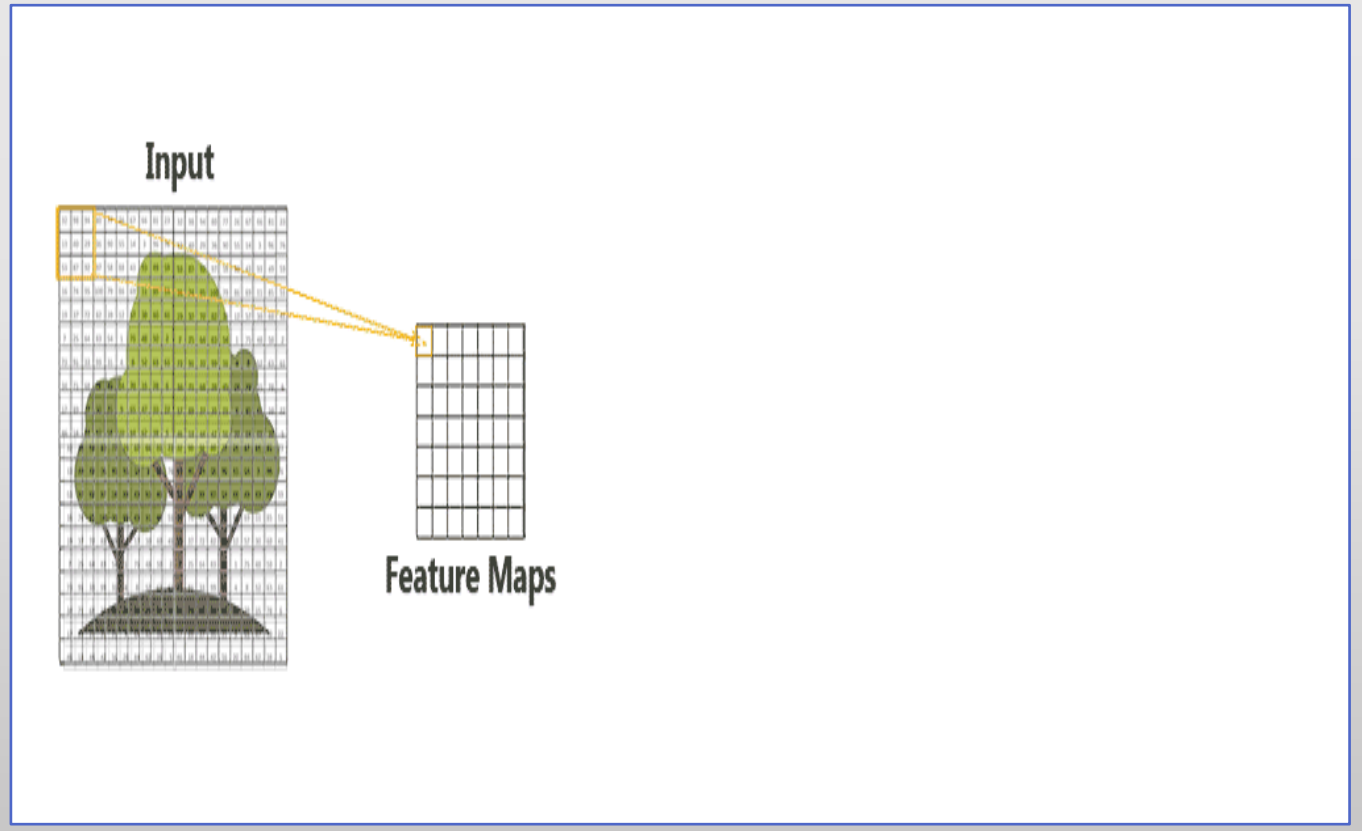
How does a Convolutional Neural Network work?

1. Input layer: This is where the raw image data is fed into the network.

2. Filters: Filters are used to extract features from the image data. They are small matrixes that are convolved with the input image to produce feature maps.

3. Max Pooling: Reduces spatial dimensions of the and retains only the most important features from the feature maps.

4. Fully Connected Layer: This layer takes the output from the previous layer and applies weights to it, producing a predicted class score for the image.



SOURCE: HTTPS://ADATIS.CO.UK/INTRODUCTION-TO-ARTIFICIAL-NEURAL-NETWORKS-PART-FOUR-CONVOLUTIONAL-NEURAL-NETWORKS/

Overview of Research Paper

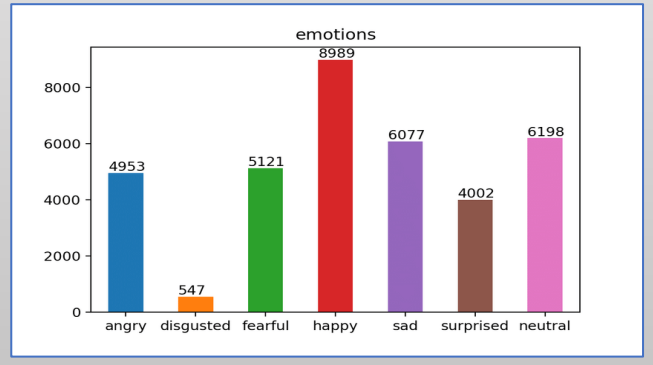
- "**Convolutional Neural Network Hyperparameters Optimization for Facial Emotion Recognition**" by A. Vulpe-Grigoraşi and O. Grigore.
- The authors aim to maximize the performance of a Convolutional Neural Network (CNN) that recognizes emotions from facial expressions by optimizing its hyperparameters to improve the accuracy of the model.
- The paper describes the data and methodology used for this task, including the dataset and the techniques used to optimize the hyperparameters of the CNN. The results of the study show the impact of hyperparameter optimization on the accuracy of facial emotion recognition and provide insights into the most effective hyperparameter values for this task.

SOURCE: A. VULPE-GRIGORAŞI AND O. GRIGORE, "CONVOLUTIONAL NEURAL NETWORK HYPERPARAMETERS OPTIMIZATION FOR FACIAL EMOTION RECOGNITION,"

Data and Methodology

IMAGES BY CATEGORY FOR FER2013

DISTRIBUTION OF IMAGES



- **The FER2013 database contains a total of 35,887 images of facial emotions which grayscale format, measuring 48x48 pixels.**
  - These images are divided into three categories: 28,709 for training, 3,589 for testing, and 3,589 for validation.
  - All pictures in the database are labeled so that each picture falls into one of the seven main categories of facial emotions: anger, disgust, fear, happiness, sad, surprised and neutral.
  - The data was developed by Ian Goodfellow et al. for a Kaggle competition on facial emotion recognition in 2013 (FER2013).

SOURCE: P A RIYANTOKO ET AL 2021, "FACIAL EMOTION DETECTION USING HAAR-CASCADE CLASSIFIER AND CONVOLUTIONAL NEURAL NETWORKS,"

TORCH.SUMMARY() OF PROPOSED MODEL

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 48, 48, 1)]	0
conv2d (Conv2D)	(None, 48, 48, 256)	2560
conv2d_1 (Conv2D)	(None, 48, 48, 512)	1180160
batch_normalization (BatchNormalization)	(None, 48, 48, 512)	2048
max_pooling2d (MaxPooling2D)	(None, 24, 24, 512)	0
dropout (Dropout)	(None, 24, 24, 512)	0
conv2d_2 (Conv2D)	(None, 24, 24, 384)	1769856
batch_normalization_1 (BatchNormalization)	(None, 24, 24, 384)	1536
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 384)	0
dropout_1 (Dropout)	(None, 12, 12, 384)	0
conv2d_3 (Conv2D)	(None, 12, 12, 192)	663744
batch_normalization_2 (BatchNormalization)	(None, 12, 12, 192)	768
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 192)	0
dropout_2 (Dropout)	(None, 6, 6, 192)	0
conv2d_4 (Conv2D)	(None, 6, 6, 384)	663936
batch_normalization_3 (BatchNormalization)	(None, 6, 6, 384)	1536
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 384)	0
dropout_3 (Dropout)	(None, 3, 3, 384)	0
flatten (Flatten)	(None, 3456)	0
dense (Dense)	(None, 256)	884992
batch_normalization_4 (BatchNormalization)	(None, 256)	1024
dropout_4 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 7)	1799

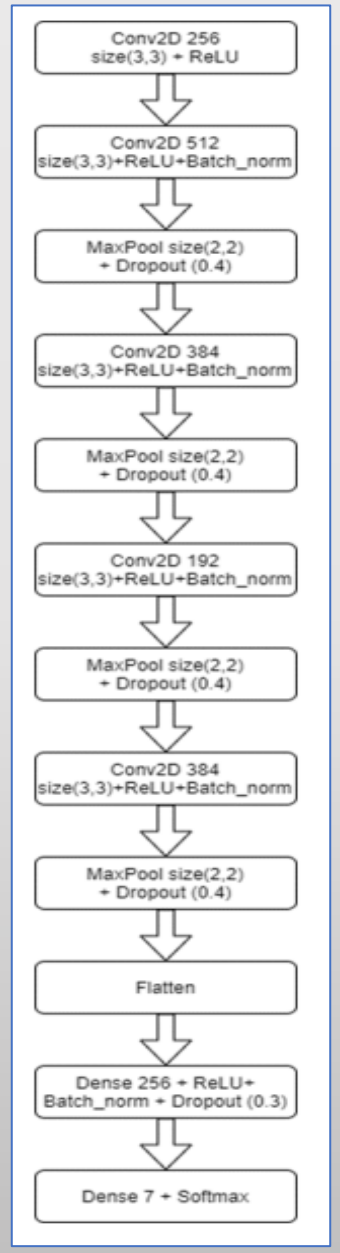
Total params: 5,173,959
Trainable params: 5,170,503
Non-trainable params: 3,456

Proposal

- **Hyperparameter Optimization using Random Search.**
- **Parameters bounded as followed:**
  1. Number of kernels in the first convolutional layer minimum 32 and maximum 256 with a step of 32.
  2. Maximum number of convolutional layers beside the first layer: 4
  3. Number of kernels in the convolutional layers: minimum 64 and maximum 512 with a step of 64.
  4. Dropout in convolutional layers: minimum 0.1, maximum 0.4, with a step of 0.1.
  5. Dropout in the fully-connected layer: minimum 0.1, maximum 0.4, with a step of 0.1.

SOURCE: A. VULPE-GRIGORAŞI AND O. GRIGORE, "CONVOLUTIONAL NEURAL NETWORK HYPERPARAMETERS OPTIMIZATION FOR FACIAL EMOTION RECOGNITION,"

PROPOSED MODEL ARCHITECTURE



Results and Findings

- **Set Hyperparameters:**
  - Learning rate: 0.001 | Optimization algorithm: Adam | Batch size: 128
  - Activation function: Rectified Linear Unit (ReLU)
  - Size of kernels in the convolutional layers: 3x3
  - For classification 2 fully-connected layers
- **According to the text, the following hyperparameters were set for the solution bounded by the proposed search-space:**
  1. Number of kernels in the first convolutional layer: 256
  2. Number of hidden Convolutional layers: 4
  3. Number of kernels in the convolutional layers: 192, 384, 512
  4. Dropout rate: 0.4 in convolutional layers and 0.3 in fully connected layer

SOURCE: A. VULPE-GRIGORAŞI AND O. GRIGORE, "CONVOLUTIONAL NEURAL NETWORK HYPERPARAMETERS OPTIMIZATION FOR FACIAL EMOTION RECOGNITION,"

FIGURE 1: PROPOSED CNN MODELS

Model	Accuracy%
VGG [2]	72.7
Proposed model	72.16
Fine-tuned VGG in fusion CNN [8]	72.11
Best individual model in ensemble [10]	71.86
Inception [2]	71.6
Tang [1], [5]	71.16
CNN only model [6]	70.8
VGG-16 [9]	68.2
Best individual model in ensemble [7]	68.18

FIGURE 2: TRAIN AND VALIDATION ACCURACY

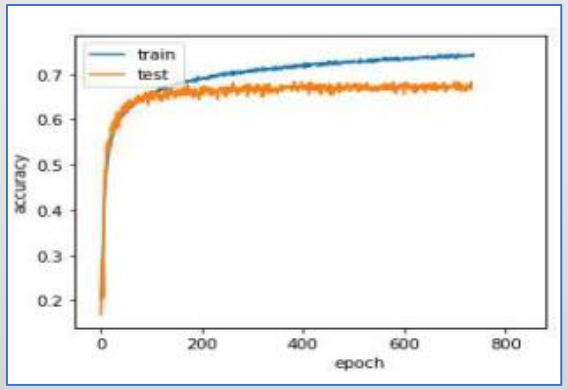
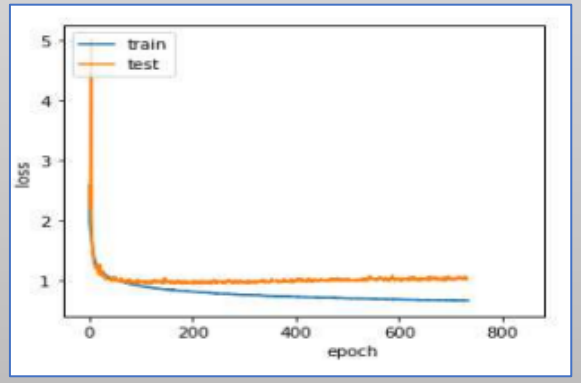


FIGURE 3: TRAIN AND VALIDATION LOSS



Conclusion

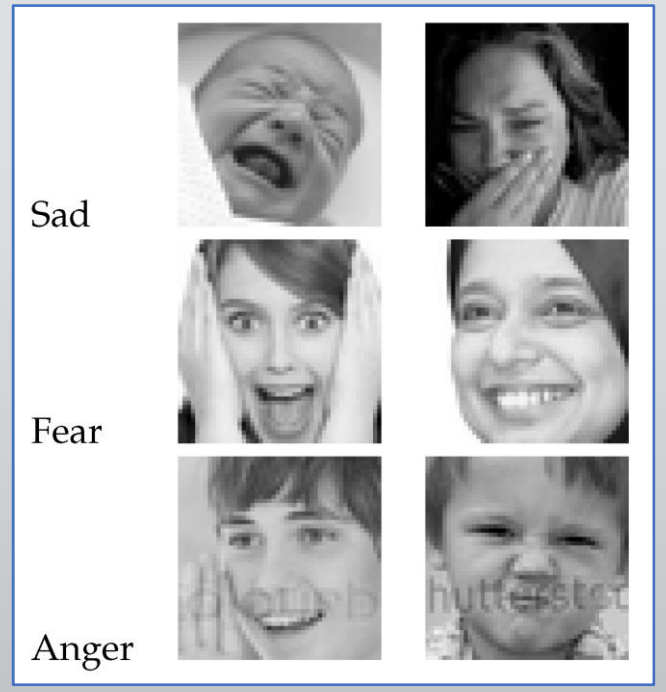
- The proposed model for facial emotion recognition has a total of 5,173,959 parameters. In comparison, VGG, which has a total number of parameters ranging from 133 to 144 million parameters, achieved 72.7% accuracy.
- The proposed model achieved an accuracy of 69.96% with a loss of 1.08 on the validation data and 72.16% accuracy on the training data from the FER2013 database.

SOURCE: A. VULPE-GRIGORAŞI AND O. GRIGORE, "CONVOLUTIONAL NEURAL NETWORK HYPERPARAMETERS OPTIMIZATION FOR FACIAL EMOTION RECOGNITION,"

Challenges and Possible Improvements

- **Key Takeaway:** CNNs are great at finding patterns and using them to classify images.
- Future work includes further fine-tuning and optimizing the model to improve its accuracy and efficiency.
- Another avenue for improvement is to explore alternative architectures and hyperparameter tuning techniques
- This is a complex task due to the variability of emotions and facial expressions
    - Imagine trying to develop a model that generalizes to everyone's face!

INCORRECTLY CLASSIFIED IMAGES



SOURCE: LUNA-JIMÉNEZ, C.; CRISTÓBAL-MARTÍN, J.; KLEINLEIN, R.; GIL-MARTÍN, M.; MOYA, J.M.; FERNÁNDEZ-MARTÍNEZ, F. GUIDED SPATIAL TRANSFORMERS FOR FACIAL EXPRESSION RECOGNITION. APPL. SCI. 2021, 11, 7217. HTTPS://DOI.ORG/10.3390/APP11167217

References

1. https://deeplearning.mit.edu/
2. https://e2eml.school/blog.html#193
3. https://adatis.co.uk/introduction-to-artificial-neural-networks-part-four-convolutional-neural-networks/
4. A. Vulpe-Grigoraşi and O. Grigore, "Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition," 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 2021, pp. 1-5, doi: 10.1109/ATEE52255.2021.9425073.
5. Riyantoko, P. A., Sugiarto, & Hindrayani, K. M. (2021). Facial emotion detection using haar-cascade classifier and convolutional neural networks. Journal of Physics: Conference Series, 1844(1) doi:https://doi.org/10.1088/1742-6596/1844/1/012004
6. Luna-Jiménez, C.; Cristóbal-Martín, J.; Kleinlein, R.; Gil-Martín, M.; Moya, J.M.; Fernández-Martínez, F. Guided Spatial Transformers for Facial Expression Recognition. Appl. Sci. 2021, 11, 7217. https://doi.org/10.3390/app11167217

Thank you for your time!

Any Questions?