Traffic Signs Recognition System

By,

Tanay Singh – PES2UG20CS364 , Section F

Supreeth Shivakumar – PES2UG20CS357 , Section F

Shriya Y S – PES2UG20CS333 , Section F

Objective/Aim

- The project aims at developing a deep neural network model to recognize road signs such as speed limit signs , pedestrian crossing signs , hazard signs etc. This is achieved by using Convolutional Neural Networks to make the machine learn different types of road signs, and then providing the machine with Test data (unseen data) in order to check how well it can recognize different signs.

Technical Details

- This Project is developed using the TensorFlow Library of Python developed by Google.

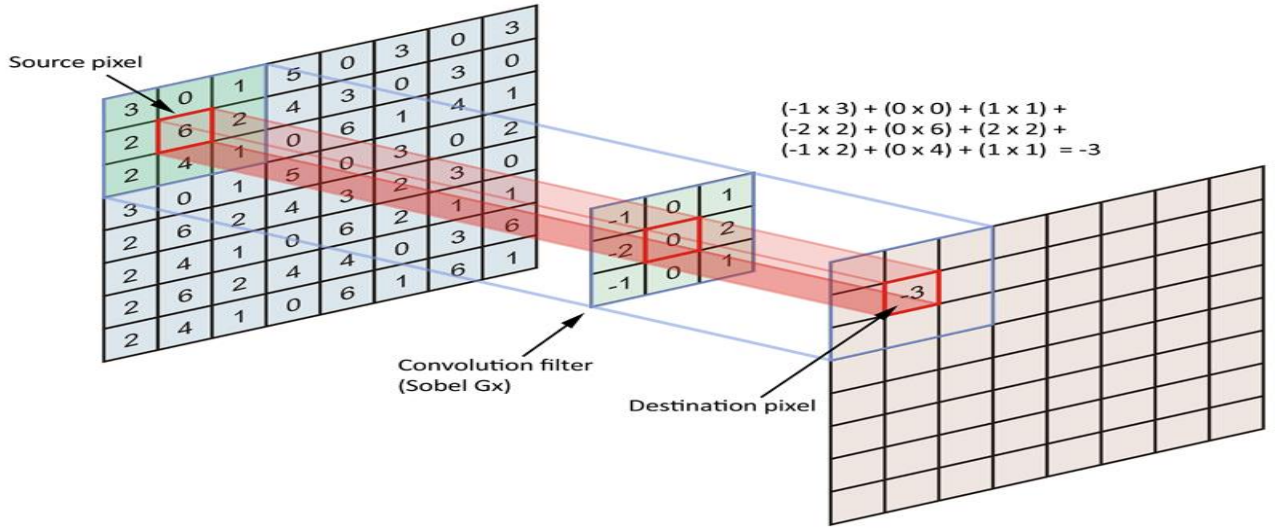
- The Project makes use of Convolutional Neural Networks, in order to train and recognize different Traffic Road Signs.

- Libraries used :

i) NumPy
ii) Pandas
iii) Matplotlib
iv) TensorFlow
v) Pillow (Python Library to handle Images)
vi) Scikit Learn
vii) Keras

What is Convolution?

- Convolution is the process of adding each element of the image to its local neighbors, weighted by the kernel. This is related to a form of mathematical convolution. The matrix operation being performed—convolution—is not traditional matrix multiplication, despite being similarly denoted by *.For example, if we have two three-by-three matrices, the first a kernel, and the second an image piece, convolution is the process of flipping both the rows and columns of the kernel and multiplying locally similar entries and summing.



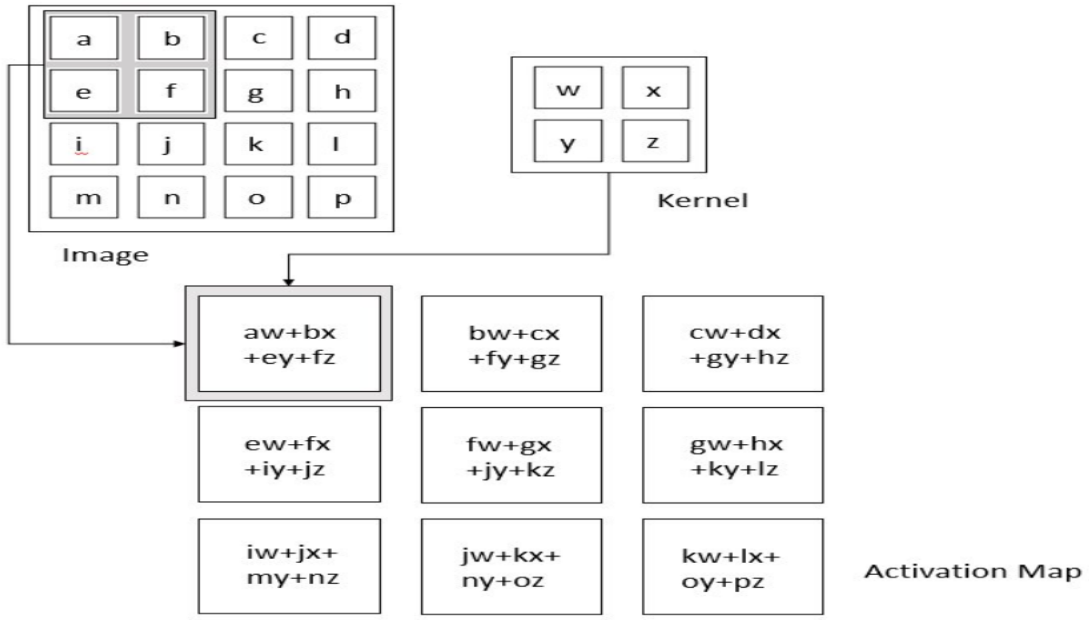
What are Convolutional Neural Networks?

- In deep learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of artificial neural network (**ANN**), most commonly applied to analyze visual imagery. CNNs are also known as **Shift Invariant** or **Space Invariant Artificial Neural Networks** (**SIANN**), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps. Counter-intuitively, most convolutional neural networks are only equivariant, as opposed to invariant, to translation. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain–computer interfaces, and financial time series.
- They mainly consist of three layers:

1. Convolution Layer
2. Pooling Layer
3. Fully Connected Layer

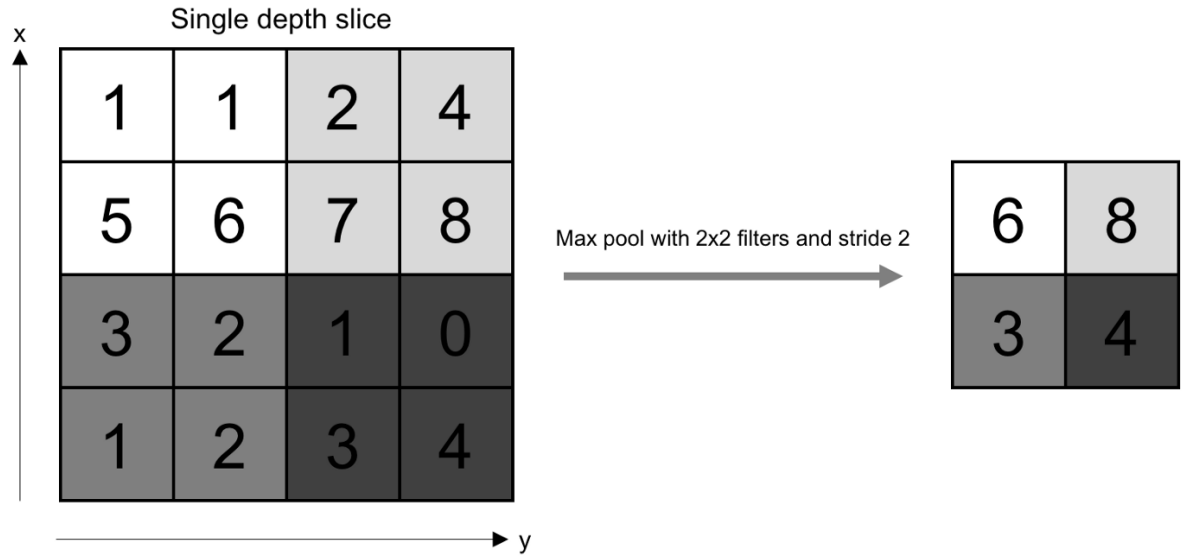
Convolution Layer

- In the Convolutional Layer, the layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a **kernel**, and the other matrix is the restricted portion of the **receptive field** (a small chunk of the entire image). The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels.



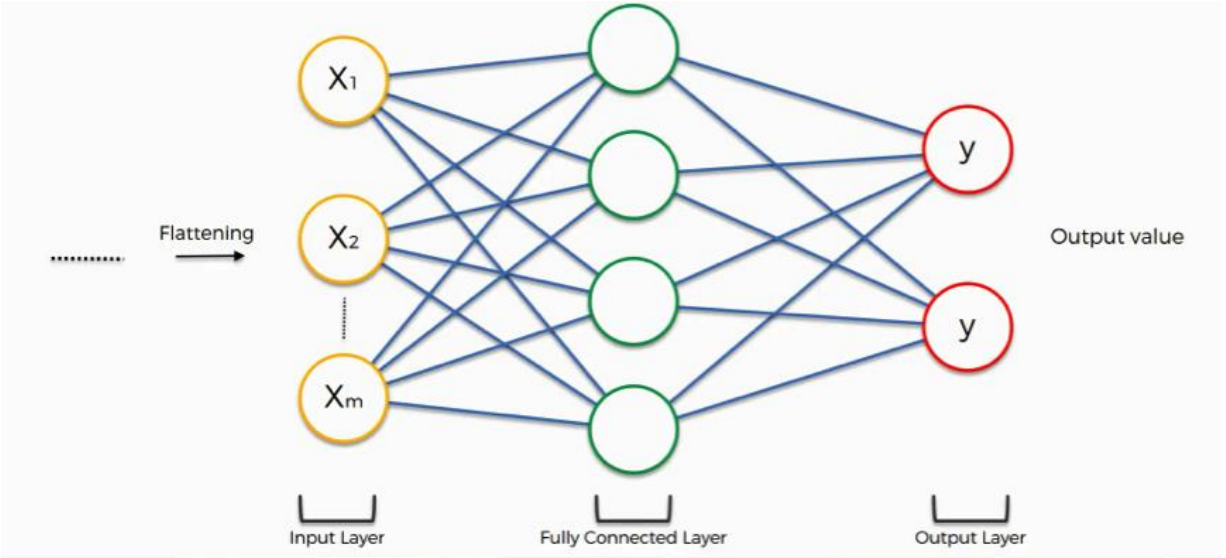
Pooling Layer

- The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.



Fully Connected Layer

- Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular Neural Network. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.
- The FC layer helps to map the representation between the input and the output.



Model Training

- Dataset used for this project is the GTSRB German Traffic Signs Dataset consisting of 39.2k training images and 12.6k testing images.

- *The architecture of our model consists of :*

1. Two Conv2D layers with a kernel size of 5x5

2. A MaxPool2D layer with a pool size of 2x2

3. A Dropout Layer with a dropout rate of 0.25

4. Two Conv2D layers with a kernel size of 3x3

5. A MaxPool2D layer with a pool size of 2x2

6. A Dropout Layer with a dropout rate of 0.25

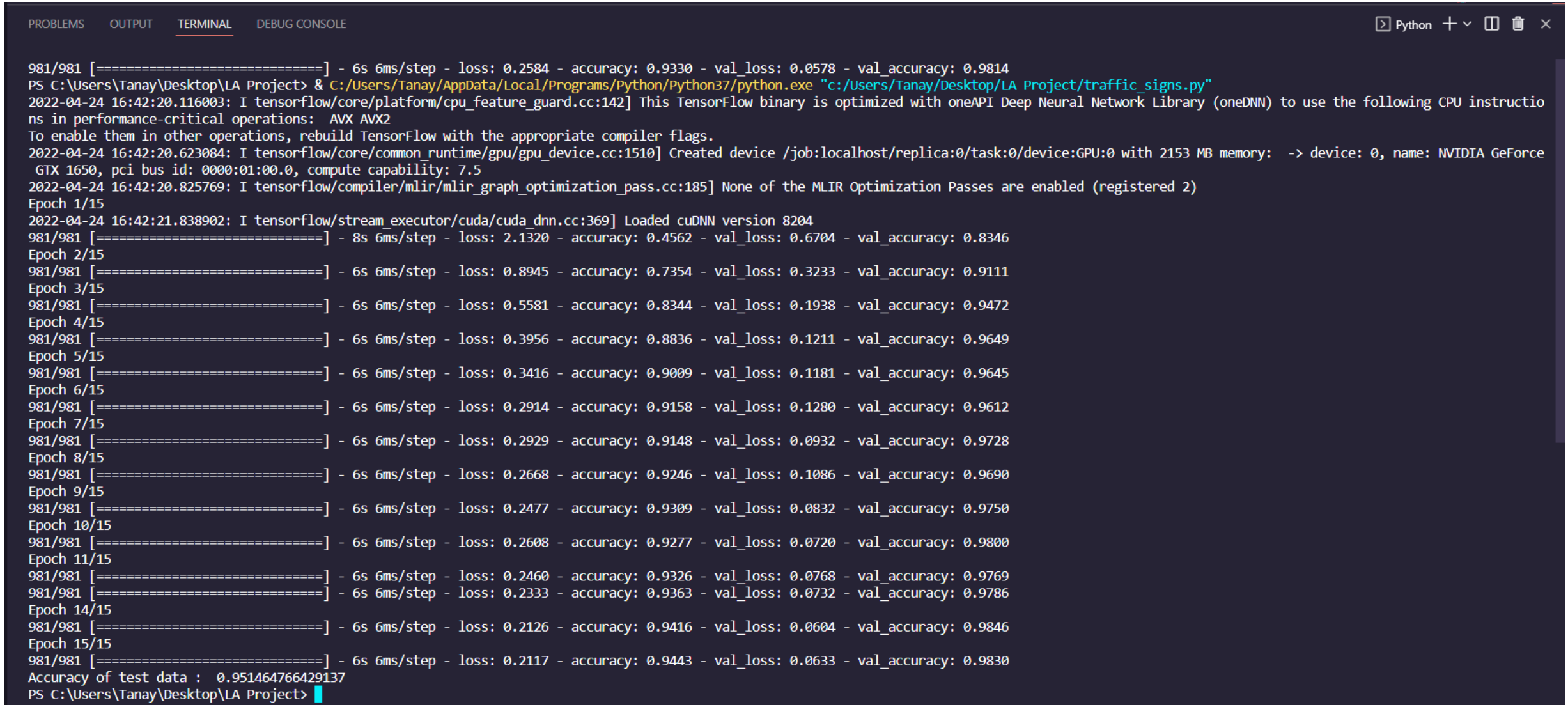
7. A Flatten layer

8. A Dense layer with 256 neurons and activation function of relu

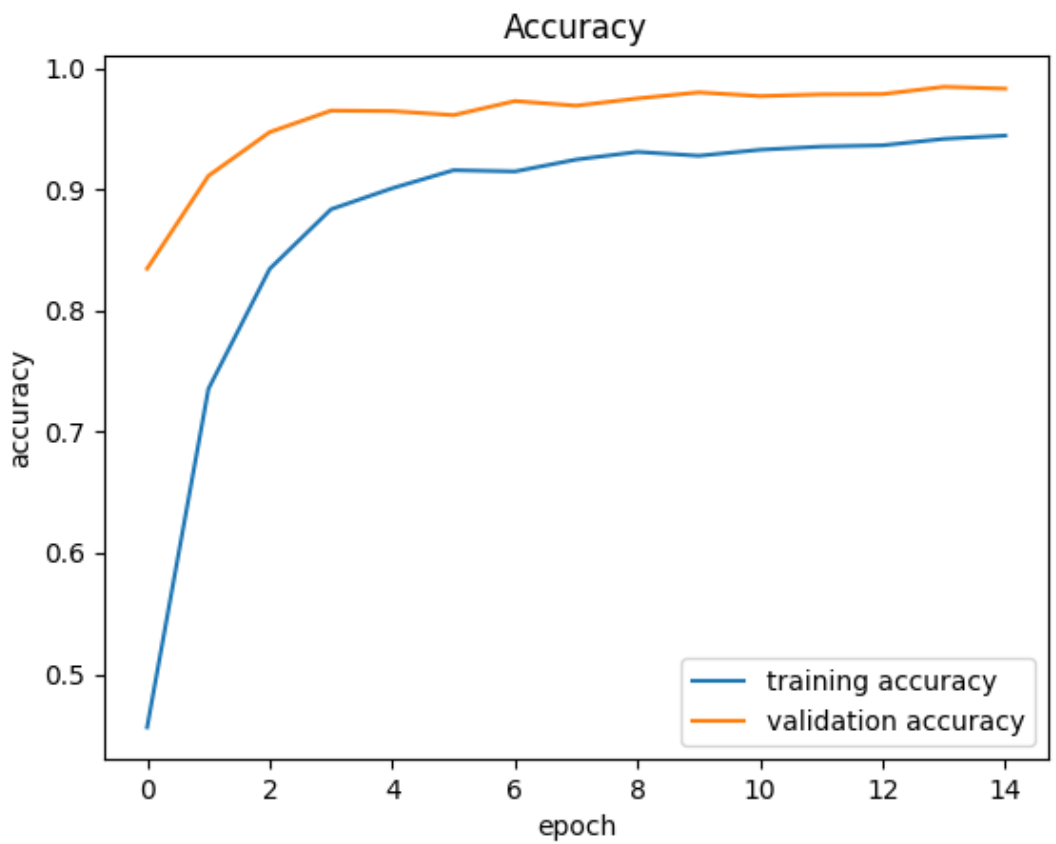
9. A Dropout layer with a dropout rate of 0.5

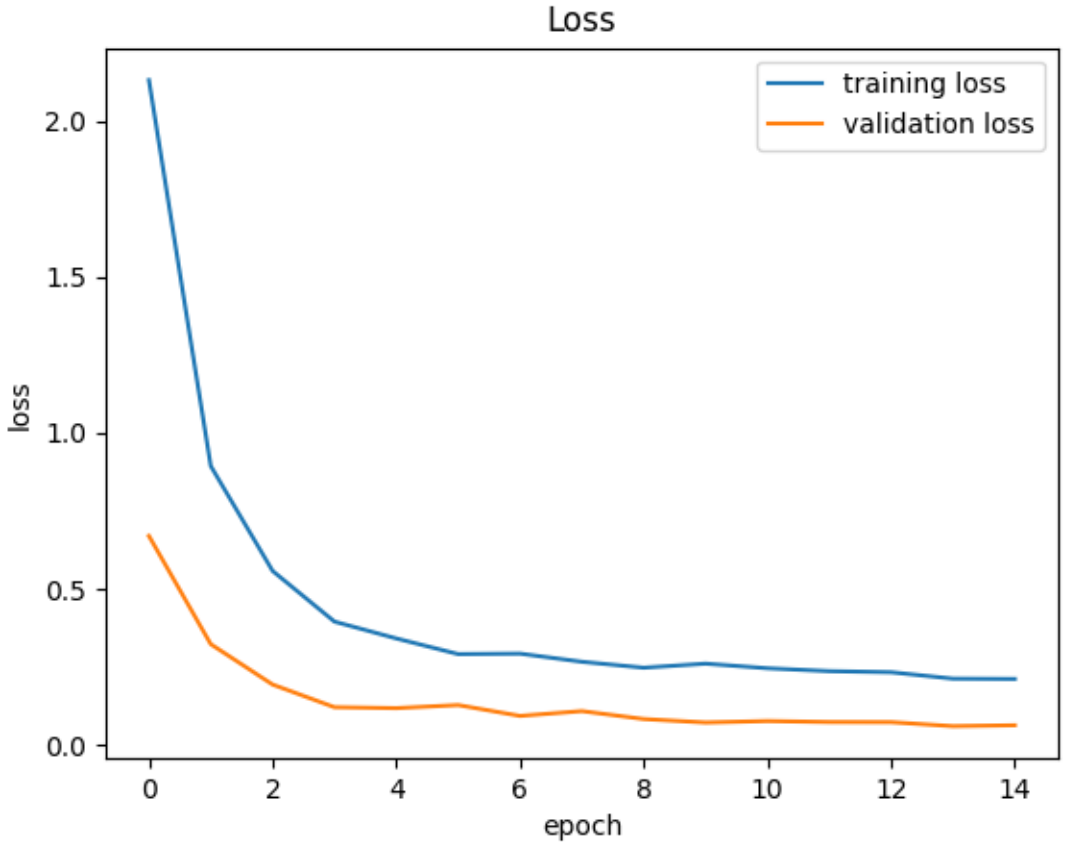
10. A Dense layer with 43 neurons and activation function of softmax

Model Training Screenshot



Model Training Graphs





Model Working

- GUI screenshots:

Refrences

- https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign
- https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939#:~:text=A%20Convolutional%20Neural%20Network%2C%20also,binary%20representation%20of%20visual%20data.
- Stanford University's Course — CS231n: Convolutional Neural Network for Visual Recognition by Prof. Fei-Fei Li, Justin Johnson, Serena Yeung
- https://cs231n.github.io/convolutional-networks/


THANK
YOU!