DEEP LEARNING PROJECT
SECTION A

Sumaiyah Zahid

Muhammad Waleed Gul 20K-0259

Muhammad Fahad 20K-0441

Traffic Sign Classification

1. Introduction:

Traffic sign recognition is a critical component of advanced driver-assistance systems (ADAS). Accurate identification of traffic signs allows vehicles to understand road regulations, improving safety and navigation. This project investigates the effectiveness of both CNNs and ViTs for this task.

2. Data Acquisition and Preprocessing:

The project utilizes a standard traffic sign dataset, such as the GTSRB dataset, containing labeled images of various signs. The script loads the training labels and defines paths to training and testing images. Images are pre-processed with techniques like resizing and converting to NumPy arrays to prepare them for model training.

3. Model Architecture:

Convolutional Neural Networks (CNNs):

CNNs are a well-established architecture for image classification tasks. They utilize convolutional layers to extract features from images by learning spatial relationships between pixels. This project implements a CNN model with:

Three convolutional layers with ReLU activation: These layers progressively extract lower-level to higher-level features from the images.
Max pooling layers: Downsample the feature maps, reducing spatial dimensions while preserving important information.
Dropout layers: Introduce randomness to prevent overfitting by randomly dropping out neurons during training.
Dense layers: Fully-connected layers that process the flattened feature maps from the convolutional layers.
Final output layer with softmax activation: Classifies the input image into one of the traffic sign categories.

Vision Transformers (ViTs):

ViTs offer an alternative approach to image classification, processing images by dividing them into patches. These patches are then:

Transformed into embeddings using a linear projection layer.
Fed into a transformer encoder: This component performs self-attention, allowing the model to attend to any part of the image and capture long-range dependencies between features.

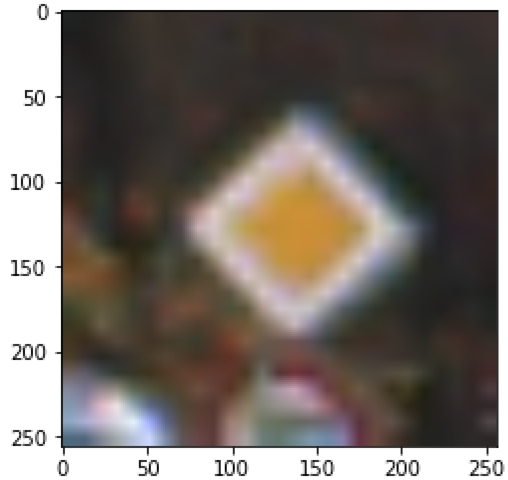
Positional encoding is incorporated to inject information about the relative positions of patches within the image.
A final layer performs classification based on the combined features extracted from all patches

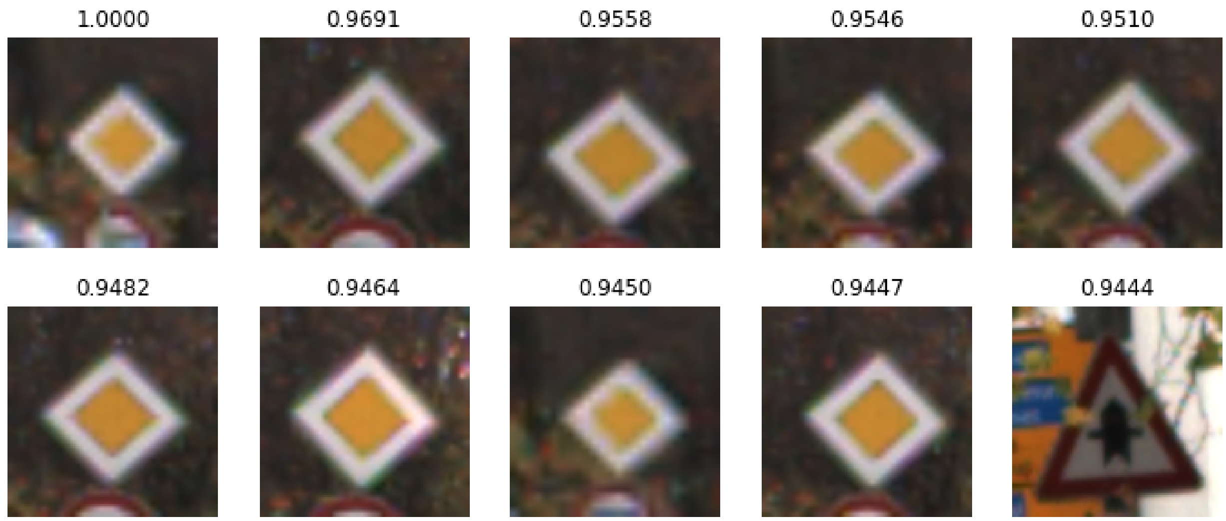
4. Training and Evaluation:

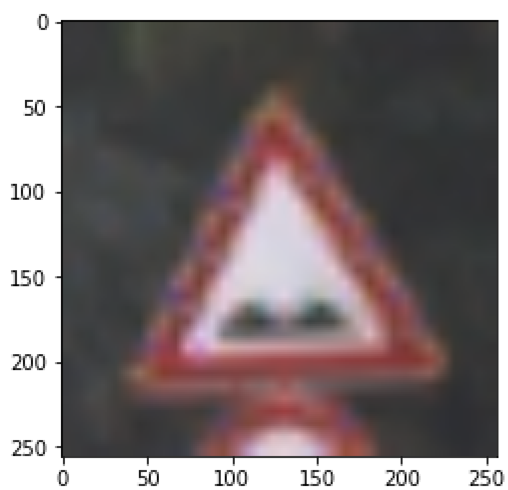
The model is trained using the Adam optimizer and categorical cross-entropy loss function. Early stopping is employed to prevent overfitting. The script evaluates the trained model on the test set and prints a classification report that summarizes performance metrics like precision, recall, F1-score, and support for each traffic sign class.

5. Results

The achieved accuracy of 97% indicates the model's effectiveness in classifying traffic signs. However, further exploration can be done using techniques like data augmentation and hyperparameter tuning to potentially improve performance.


0
50
100
150
200
250
0 50 100 150 200 250


1.0000
0.9691
0.9558
0.9546
0.9510
0.9482
0.9464
0.9450
0.9447
0.9444


0
50
100
150
200
250
0 50 100 150 200 250


1.0000
0.9993
0.9987
0.9986
0.9986
0.9985
0.9976
0.9970
0.9970
0.9967