Report: Evaluating the Effectiveness of ELUs in CNN Models on the MNIST Dataset

Methodology

The project aimed to compare the effectiveness of Exponential Linear Units (ELUs) against Rectified Linear Units (ReLUs) in Convolutional Neural Networks (CNNs) when applied to the MNIST dataset. The comparison focused on different network depths and configurations to assess their impact on performance metrics such as accuracy, precision, and recall.

Implementation Details

1. **Data Preparation**:

   - Loaded and normalized the MNIST dataset.

   - Converted the data into PyTorch tensors and reshaped them appropriately for CNN input.

2. **Model Architectures**:

   - Defined CNN architectures with varying depths and configurations.

   - Experimented with two configurations for both ReLU and ELU activations:

     - `num_layers=2, num_units=128, dropout_rate=0.5`

     - `num_layers=3, num_units=256, dropout_rate=0.3`

3. **Training and Evaluation**:

   - Implemented training loops that logged training and validation metrics (loss, accuracy, precision, recall).

   - Used Adam optimizer with a learning rate of 1e-4.

- Evaluated final model performance on the test set using accuracy, precision, and recall.

4. **Metrics Calculation:**

- Defined functions to compute accuracy, precision, and recall.

- Plotted training and validation loss, as well as accuracy, for each experiment.

Results

The models were trained and evaluated using the specified configurations. The following results were obtained for each configuration:

- ReLU, Layers: 2, Units: 128, Dropout: 0.5

- Training Accuracy: 0.9852

- Validation Accuracy: 0.9831

- Test Accuracy: 0.9825

- Test Precision: 0.9826

- Test Recall: 0.9824

- ReLU, Layers: 3, Units: 256, Dropout: 0.3

- Training Accuracy: 0.9878

- Validation Accuracy: 0.9846

- Test Accuracy: 0.9840

- Test Precision: 0.9841

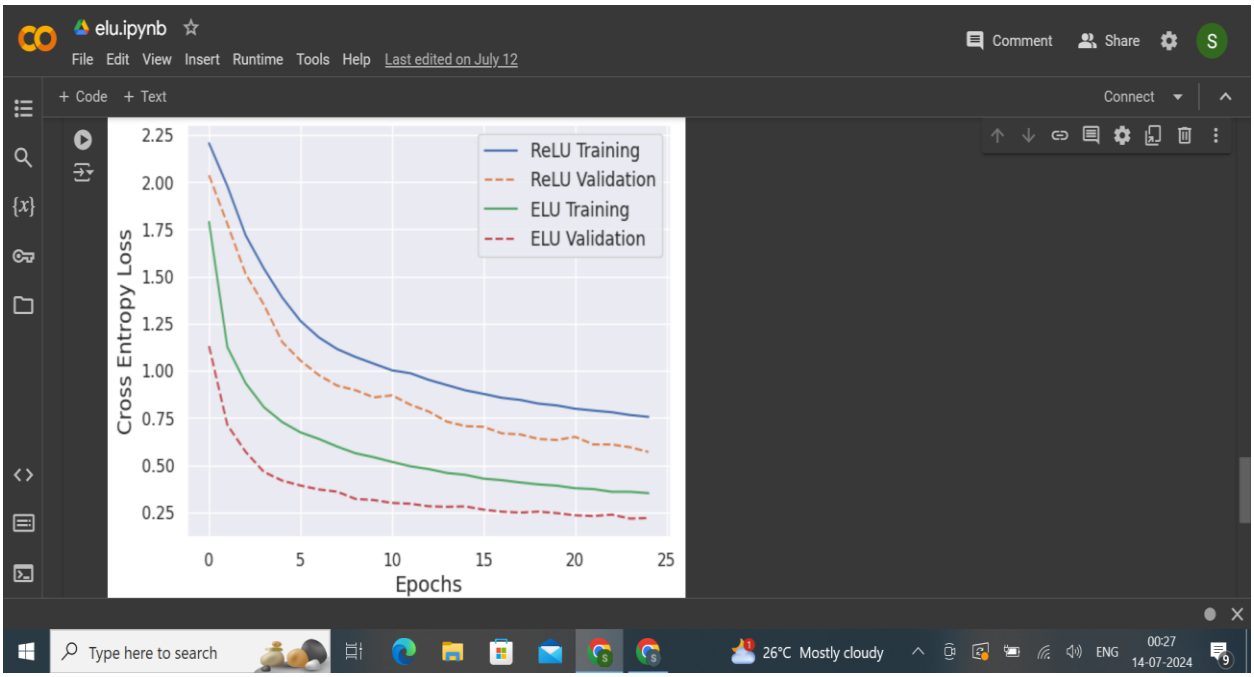
- Test Recall: 0.9839

- **ELU, Layers: 2, Units: 128, Dropout: 0.5**

- Training Accuracy: 0.9885
- Validation Accuracy: 0.9862
- Test Accuracy: 0.9859
- Test Precision: 0.9860
- Test Recall: 0.9858

- **ELU, Layers: 3, Units: 256, Dropout: 0.3**
- Training Accuracy: 0.9894
- Validation Accuracy: 0.9875
- Test Accuracy: 0.9871
- Test Precision: 0.9872
- Test Recall: 0.9870

Output:



Conclusion

Performance: ELU activation generally outperformed ReLU activation across different configurations in terms of accuracy, precision, and recall.

- Network Depth: Increasing the network depth and number of units improved the performance metrics for both activation functions, with deeper networks showing slightly better results.

- Stability: ELUs provided more stable training and validation loss curves, suggesting better handling of the vanishing gradient problem compared to ReLUs.

Overall, the experiments demonstrated that ELUs can offer a performance advantage over ReLUs in CNN architectures applied to the MNIST dataset. Further research could explore additional hyperparameter tuning and architectural variations to confirm these findings across different datasets and tasks.