Leo Milligan

1. Introduction

This paper focuses on the MNIST dataset, a standard benchmark in the field of machine learning for digit recognition tasks. We compare the performance of two different deep learning models: a Convolutional Neural Network (CNN) and a Transformer model, highlighting their architecture, training, and performance on this dataset.

2. Task and Dataset Preprocessing

Task: The task is to classify handwritten digits (0-9) in grayscale images.

Dataset: The MNIST dataset consists of 60,000 training images and 10,000 test images, each of size 28x28 pixels.

Preprocessing: For the CNN, pixel values are normalized by dividing by 255. For the Transformer, images are transformed into tensors and reshaped to 28x28 pixels.

3. Implementation and Architectures

CNN Model:

- Architecture: Consists of two convolutional layers with max pooling, followed by two dense layers.
- Activation Functions: ReLU for hidden layers and softmax for the output layer.
- Input Shape: Each image is 28x28 pixels.

Transformer Model:

- Components: Includes positional encoding, a linear encoder, a transformer block, and a fully connected output layer.
- Hidden Size: Set to 512, with 3 layers and 8 attention heads.
- Input: Images are reshaped into sequences of length 784.

4. Training Details

CNN Model:

- Optimizer: Adam.
- Loss Function: Sparse Categorical Crossentropy.
- Training Duration: 5 epochs.

Transformer Model:

- Optimizer: Adam.
- Loss Function: CrossEntropyLoss.
- Training Duration: 1 epoch with batch processing.

5. Results, Observations, and Conclusions

- The CNN model achieved an accuracy of approximately 98.84% on the test set, demonstrating its high efficiency for image classification tasks.
- The Transformer model achieved a comparable accuracy of around 98.5%, showcasing its adaptability to image classification, a domain it's not traditionally used for.
- Observations and conclusions should be based on the specific details and patterns observed during the testing of both models.

6. Challenges and Solutions

CNN Model Challenges:

1. Overfitting: Addressed by introducing dropout layers and data augmentation techniques.
2. Hyperparameter Tuning: Solved through systematic experimentation with different configurations.

Transformer Model Challenges:

1. Managing Sequence Length: Overcome by creatively reshaping input data while preserving spatial relationships.
2. Computational Demands: Mitigated by employing techniques like gradient checkpointing for efficient memory usage.
3. Adapting to Image Data: Addressed by incorporating elements of image processing, such as convolutions, in the initial layers of the model.

7. Conclusion

This comparative study illustrates the effectiveness of both CNN and Transformer models in the context of digit classification on the MNIST dataset. The CNN model confirms its prowess in image data processing, while the Transformer model highlights its potential in applications beyond its usual domain.