Building a CNN for MNIST Handwritten Digit Classification

Introduction

Welcome! In this assignment, you will build a Convolutional Neural Network (CNN) to classify handwritten digits from the famous MNIST dataset. This dataset is a classic in the field of computer vision and provides a great starting point for understanding image classification with deep learning.

This notebook is structured to guide you step-by-step through the process. You will load the data, preprocess it, define a CNN model, train it, and evaluate its performance. Throughout the assignment, you will have opportunities to experiment and deepen your understanding of the concepts.

Remember to:

- **Read all instructions carefully.**
- **Execute the code cells in order.**
- **Fill in the missing code sections marked as "Students: Fill in the blanks".**
- **Answer the reflection questions in the designated Markdown cell.**
- **Experiment and explore!** Change parameters, layers, and observe the effects.

Let's get started and build our MNIST digit classifier!

Section 1: Setting Up - Imports

Before we dive into building our CNN, we need to import the necessary libraries. These libraries provide pre-built tools and functions that will make our work much easier.

Instructions:

1. **Carefully review the code cell below.** It imports libraries from TensorFlow and Keras, which are powerful frameworks for building and training neural networks.
2. **Execute the code cell by selecting it and pressing [Shift + Enter] (or the "Run" button).**
3. **Ensure there are no error messages after running the cell.** If you encounter errors, double-check that you have TensorFlow and Keras installed in your environment.

```
# Cell 1: Imports
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
```

```
from tensorflow.keras.datasets import mnist
from tensorflow.keras.utils import to_categorical
```

Explanation of Imports:

- **`tensorflow as tf` and `keras`:** TensorFlow is the main deep learning framework, and Keras is its high-level API that simplifies building and training models. We import TensorFlow as `tf` and Keras directly for easy access to their functionalities.
- **`from tensorflow.keras import layers`:** This imports the `layers` module from Keras, which provides various layers for building neural networks (like convolutional layers, dense layers, etc.).
- **`from tensorflow.keras.datasets import mnist`:** This imports the MNIST dataset directly from Keras datasets. This is very convenient for loading and using the MNIST data.
- **`from tensorflow.keras.utils import to_categorical`:** This imports the `to_categorical` function, which we will use to perform one-hot encoding of our labels.

Section 2: Data Loading and Preprocessing

In this section, we will load the MNIST dataset and prepare it for training our CNN model. Preprocessing steps are crucial to ensure our data is in the right format for the model to learn effectively.

Instructions:

1. **Read through the code in the cell below.** Understand how it loads the MNIST dataset and what preprocessing steps are applied.
2. **Execute the code cell.**
3. **Examine the comments in the code** to understand each preprocessing step in detail.

```
# Cell 2: Data Loading and Preprocessing
# Load the MNIST dataset
(x_train, y_train), (x_test, y_test) = mnist.load_data()

# Normalize pixel values to be between 0 and 1
x_train = x_train.astype("float32") / 255.0
x_test = x_test.astype("float32") / 255.0

# Add a channel dimension (for grayscale images, it's 1)
x_train = x_train.reshape(-1, 28, 28, 1)
x_test = x_test.reshape(-1, 28, 28, 1)

```

```
# One-hot encode the labels
num_classes = 10
y_train = to_categorical(y_train, num_classes)
y_test = to_categorical(y_test, num_classes)
```

Explanation of Data Preprocessing:

- **Loading the MNIST dataset:** `mnist.load_data()` loads the MNIST dataset, which is already split into training and testing sets (`(x_train, y_train), (x_test, y_test)`). `x_train` and `x_test` contain the images (pixel data), and `y_train` and `y_test` contain the corresponding labels (digits 0-9).
- **Normalization:** `x_train = x_train.astype("float32") / 255.0` and `x_test = x_test.astype("float32") / 255.0` normalize the pixel values. Pixel values in images are typically in the range 0-255. Dividing by 255 scales them to the range 0-1. This normalization helps the neural network train faster and more effectively.
- **Adding Channel Dimension:** `x_train = x_train.reshape(-1, 28, 28, 1)` and `x_test = x_test.reshape(-1, 28, 28, 1)` reshape the data to add a channel dimension. Even though MNIST images are grayscale (single channel), CNNs in Keras expect input data to have a channel dimension. We reshape from `(number_of_images, 28, 28)` to `(number_of_images, 28, 28, 1)`. The `-1` in `reshape` means "infer the dimension based on the size of the array."
- **One-Hot Encoding:** `y_train = to_categorical(y_train, num_classes)` and `y_test = to_categorical(y_test, num_classes)` perform one-hot encoding on the labels. Instead of representing the digit '3' as a single number, one-hot encoding converts it into a vector `[0, 0, 0, 1, 0, 0, 0, 0, 0, 0]`, where the 4th position (index 3) is 'hot' (value 1), and all other positions are 'cold' (value 0). This is a standard way to represent categorical labels for neural networks in multi-class classification problems. `num_classes = 10` specifies that we have 10 classes (digits 0-9).

Section 3: Model Definition - Building the CNN

Now we will define the architecture of our Convolutional Neural Network (CNN). You will be building a sequential model using Keras layers.

Instructions:

1. **Carefully examine the code in the cell below.** Notice the structure of the `keras.Sequential` model.

2. **Fill in the missing parts** marked with `# Students: Fill in the blanks` to complete the model definition.
3. **Experiment!** You are encouraged to try different configurations for the layers, such as changing the number of filters in the convolutional layers, or adding more layers.

```
# Cell 3: Model Definition
# Build the CNN model.  Students: Fill in the missing parts!
model = keras.Sequential(
    [
        keras.Input(shape=(28, 28, 1)),  # Input layer
        layers.Conv2D(32, kernel_size=(3, 3), activation="relu"), # Conv
        layers.MaxPooling2D(pool_size=(2, 2)), # Max pooling layer 1
        # Students: Add another Conv2D layer here.  Experiment with the
        # layers.Conv2D(____, kernel_size=(____, ____), activation="____
        layers.Conv2D(64, kernel_size=(3, 3), activation="relu"), # Conv
        layers.MaxPooling2D(pool_size=(2, 2)), # Max pooling layer 2
        # Students: Add another MaxPooling2D layer here if needed.
        # layers.MaxPooling2D(pool_size=(____, ____)),  # Max pooling la
        layers.Flatten(),  # Flatten layer
        layers.Dropout(0.5),  # Dropout layer
        layers.Dense(num_classes, activation="softmax"),  # Output layer
    ]
)
```

```
# Cell 3: Model Definition
# Build the CNN model. Students: Fill in the missing parts!
model = keras.Sequential(
    [
        keras.Input(shape=(28, 28, 1)),  # Input layer
        layers.Conv2D(32, kernel_size=(3, 3), activation="relu"), # Conv
        layers.MaxPooling2D(pool_size=(2, 2)), # Max pooling layer 1

        # Students: Add another Conv2D layer here. Experiment with the n
        # This is the standard choice for the second convolutional layer
        layers.Conv2D(64, kernel_size=(3, 3), activation="relu"),  # Con

        # Students: Add another MaxPooling2D layer here if needed.
        # This is the correct placement for the second pooling layer.
        layers.MaxPooling2D(pool_size=(2, 2)),  # Max pooling layer 2

        layers.Flatten(),  # Flatten layer
        layers.Dropout(0.5),  # Dropout layer
        layers.Dense(num_classes, activation="softmax"),  # Output layer
    ]
)

model.summary() # It's always a good idea to print the model summary to
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 26, 26, 32)	320
max_pooling2d_2 (MaxPooling2D)	(None, 13, 13, 32)	0
conv2d_3 (Conv2D)	(None, 11, 11, 64)	18,496
max_pooling2d_3 (MaxPooling2D)	(None, 5, 5, 64)	0
flatten_1 (Flatten)	(None, 1600)	0
dropout_1 (Dropout)	(None, 1600)	0
dense_1 (Dense)	(None, 10)	16,010

Total params: 34,826 (136.04 KB)
Trainable params: 34,826 (136.04 KB)
Non-trainable params: 0 (0.00 B)

Explanation of Layers:

- **`keras.Input(shape=(28, 28, 1))`:** This is the input layer of our model. It specifies the shape of the input images, which are 28x28 pixels with 1 channel (grayscale).
- **`layers.Conv2D(32, kernel_size=(3, 3), activation="relu")`:** This is a 2D Convolutional layer.
  - `32`: This is the number of filters (also called kernels). Each filter learns to detect specific features in the input image.
  - `kernel_size=(3, 3)`: This defines the size of the convolutional filter as 3x3 pixels.
  - `activation="relu"`: ReLU (Rectified Linear Unit) is the activation function. It introduces non-linearity into the model, allowing it to learn complex patterns.
- **`layers.MaxPooling2D(pool_size=(2, 2))`:** This is a Max Pooling layer.
  - `pool_size=(2, 2)`: It reduces the spatial dimensions of the feature maps by taking the maximum value within each 2x2 window. This helps to reduce the number of parameters, control overfitting, and make the model more robust to small shifts and distortions in the input.
- **`layers.Flatten()`:** This layer flattens the 2D feature maps from the convolutional and pooling layers into a 1D vector. This is necessary to connect the convolutional part of the network to the fully connected (Dense) layers.
- **`layers.Dropout(0.5)`:** This is a Dropout layer.

    - `0.5`: This sets the dropout rate to 50%. During training, this layer randomly sets 50% of the input units to 0 at each update. This is a regularization technique that helps to prevent overfitting.
- **`layers.Dense(num_classes, activation="softmax")`**: This is the output Dense (fully connected) layer.
    - `num_classes`: This is set to 10 because we have 10 classes (digits 0-9).
    - `activation="softmax"`: Softmax activation ensures that the output values are probabilities, and they sum up to 1 across all classes. The output will be a vector of 10 probabilities, where each probability represents the model's confidence that the input image belongs to that specific digit class.

Section 4: Model Compilation - Choosing Loss and Optimizer

Before we can train our model, we need to compile it. Compilation involves choosing an optimizer, a loss function, and metrics to evaluate the model's performance.

Instructions:

1. **Examine the code cell below.** You need to fill in the blanks for the `loss` and `optimizer` parameters in `model.compile()`.
2. **Choose an appropriate loss function and optimizer** for this multi-class classification problem.
3. **In the Markdown cell after the code, explain your choices.** Why are these choices suitable for this task?

```
# Cell 4: Model Compilation
# Students: Choose an appropriate loss function and optimizer.  Why did
model.compile(loss="categorical_crossentropy", optimizer="adam", metrics
```

```
model.compile(loss="categorical_crossentropy", optimizer="adam", metrics
```

Explanation of Choices (To be filled by students in the reflection section):

- **Loss Function:** You need to choose a loss function that is appropriate for multi-class classification. Think about what kind of error we are trying to minimize when classifying digits into 10 categories.
- **Optimizer:** You need to choose an optimizer that will efficiently update the model's weights to minimize the loss function. Consider common optimizers used in deep learning.

- **Metrics:** We are using "accuracy" as a metric to evaluate the model's performance. Accuracy is a common metric for classification tasks, representing the percentage of correctly classified images.

Section 5: Model Training - Fitting the Model to the Data

Now it's time to train our CNN model using the training data. Training involves feeding the training data to the model and adjusting its weights to minimize the loss function.

Instructions:

1. **Examine the code cell below.** You need to fill in the blanks for `batch_size` and `epochs` in `model.fit()`.
2. **Choose appropriate values for `batch_size` and `epochs`.**
3. **Run the code cell to start training.** Observe the training progress, especially the loss and accuracy on both the training and validation sets.
4. **Experiment!** Change the `batch_size` and `epochs` and see how it affects the training process and the final performance.

```
# Cell 5: Model Training
# Students: Adjust the batch size and number of epochs.  What happens if
model.fit(x_train, y_train, batch_size=128, epochs=15, validation_split=
```

```
Epoch 1/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 42s 97ms/step - accuracy: 0.7591 - loss: 0.7678
Epoch 2/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 82s 97ms/step - accuracy: 0.9633 - loss: 0.1226
Epoch 3/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 81s 96ms/step - accuracy: 0.9719 - loss: 0.0879
Epoch 4/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 96ms/step - accuracy: 0.9787 - loss: 0.0683
Epoch 5/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 96ms/step - accuracy: 0.9806 - loss: 0.0612
Epoch 6/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 97ms/step - accuracy: 0.9827 - loss: 0.0551
Epoch 7/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 40s 96ms/step - accuracy: 0.9843 - loss: 0.0496
Epoch 8/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 96ms/step - accuracy: 0.9842 - loss: 0.0498
Epoch 9/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 40s 95ms/step - accuracy: 0.9859 - loss: 0.0465
Epoch 10/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 40s 95ms/step - accuracy: 0.9877 - loss: 0.0405
Epoch 11/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 40s 95ms/step - accuracy: 0.9886 - loss: 0.0371
Epoch 12/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 40s 96ms/step - accuracy: 0.9878 - loss: 0.0364
Epoch 13/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 96ms/step - accuracy: 0.9894 - loss: 0.0334
```

```
Epoch 14/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 41s 96ms/step - accuracy: 0.9885 - loss: 0.0345
Epoch 15/15
422/422 ━━━━━━━━━━━━━━━━━━━━ 42s 98ms/step - accuracy: 0.9892 - loss: 0.0331
<keras.src.callbacks.history.History at 0x7af23aaf6510>
```

Explanation of Training Parameters:

- `batch_size`: This determines the number of training samples processed in each mini-batch during training. A larger batch size can speed up training but might require more memory. A smaller batch size can lead to more noisy updates but might generalize better.
- `epochs`: One epoch represents one complete pass through the entire training dataset. More epochs can potentially lead to better training but also increase the risk of overfitting, where the model learns the training data too well and performs poorly on unseen data.
- `validation_split=0.1`: This reserves 10% of the training data as a validation set. During training, the model's performance is evaluated on this validation set after each epoch. This helps to monitor for overfitting and tune hyperparameters.

Section 6: Model Evaluation - Assessing Performance on Test Data

After training, we need to evaluate our model's performance on the test dataset. This gives us an estimate of how well the model generalizes to unseen data.

Instructions:

1. **Run the code cell below.**
2. **Observe the output.** It will print the test loss and test accuracy.
3. **Think about the results.** Is the test accuracy satisfactory? How does it compare to the training and validation accuracy you observed during training?

```
# Cell 6: Model Evaluation
loss, accuracy = model.evaluate(x_test, y_test, verbose=0)
print(f"Test loss: {loss:.4f}")
print(f"Test accuracy: {accuracy:.4f}")
```

```
Test loss: 0.0239
Test accuracy: 0.9915
```

Section 7: Reflection and Answers to Questions

This is an important section! Take some time to reflect on what you have learned and answer the following questions in detail. Your thoughtful answers will demonstrate your understanding of the concepts covered in this assignment.

Reflection Questions:

1. **Conv2D Layer:** What is the role of the Conv2D layer? How do the `kernel_size` and the number of filters affect the learning process? *Hint: Experiment by changing these values in Cell 3.*

- The primary role for a Conv2D layer is to be a feature detector. Its like a specialized scanner that slides accross the input image to find specific patterns. The kernel size defines the dimenstions of a "scanner" window that slides accross the image. A kernel_size=(3, 3) means the scanner is a 3x3 pixel window.

2. **MaxPooling2D Layer:** What is the purpose of the MaxPooling2D layer? How does it contribute to the model's performance? *Hint: Try removing or adding a MaxPooling2D layer and see what happens.*

- The main purpose of the MaxPooling2D layer is to downsample the feature map it receives from a convolutional layer. It shrinks the height and width of the feature maps while trying to keep the most important information. It is an essential layer for making a CNN mor efficient, faster, and more robust by summarizing feature maps and focusing only on the most important information. They reduce computational cost, provide transalation invariance, and help prevent overfitting.

3. **One-Hot Encoding:** Why do we use one-hot encoding for the labels?

- We use one-hot encoding to convert categorical labels like "cat", "dog", "bird" into numerical format that the neural network can understand without creating false relationships between the categories.

4. **Flatten Layer:** Why do we need the Flatten layer before the Dense layer?

- It is a bridge between the feature extraction part of the network and the classification part.

5. **Optimizer and Loss Function:** What optimizer and loss function did you choose in Cell 4? Explain your choices. Why is categorical cross-entropy a suitable loss function for this task? Why is Adam a good choice of optimiser?

- The categorical_crosstropy loss function is the ideal choice for a multi-class classfication problem. This is any task where you need to classify an input into one of

three or more distinct categories. This loss function is suitable because it is designed for probability distributions, it measures the distance between predictions, and it penalizes confident but wrong answers. It gives a very high loss when the model is highly confident about the wrong class.

6. **Batch Size and Epochs:** How did you choose the batch size and number of epochs in Cell 5? What are the effects of changing these parameters? *Hint: Experiment!*

- I chose the batch size to be 128 and the epochs to be 15. These values are common and effective starting points for training a model on a dataset like MNIST or CIFAR-10. The batch size is the number of training samples the model processes before making one update to its internal weights. Instead of showing the model one imahe at a time, or all 60,000 images at once, I showed. batch of 128 images, let it make predictions, calculate the average error for that batch, and then make a single adjustment. 128 is widely used as a sweet spot that balances two competing needs. Using a smaller batch size would cause the model to update its weights more frequently which can sometimes help the mdoel learn faster in terms of updates and avoid getting stuck in poor solutions. The process becomes less stable and each update is based on only a small sample of the data. When using a larger sample size the training process is increased in terms of time per epoch and the learning process is more stable because each update is based on a larger sample of the data. Using a larger batch size does require more RAM though.

7. **Dropout:** Why is the Dropout layer included in the model?

- It is included to prevent overfitting.

8. **Model Architecture:** Describe the overall architecture of your CNN. How many convolutional layers did you use? How many max pooling layers? What is the final dense layer doing?

- A standard and effective architecture for image classification tasks. I used 2 convolutional layers in my model (Conv2D) and 2 maxpooling (MaxPooling2D) layers. The final dense layers is the output layer of my network. It has a crucial and specific job: to perform the final classification.

9. **Performance:** What accuracy did you achieve on the test set? Are you happy with the result? Why or why not? If you're not happy, what could you try to improve the performance?

- I achieved a final test accuracy of 99.15% and a final test loss of 0.0239. This is an outstanding result and I am happy with it as it indicates a highly successful and effective model. To improve the perfomance, I could expand the training dataset by creating slightly modified copies of existing images. I could also implement a learning rate

scheduler that gradually decreases the learning rate during training. I could also train several different high performing models with slightly different architechtures and then average their predictions.

Tips and Explanations:

- **Normalization:** Dividing the pixel values by 255 normalizes them to the range [0, 1]. This is important for training neural networks.
- **Reshaping:** The `reshape` operation adds a channel dimension to the images. For grayscale images, the channel dimension is 1.
- **One-Hot Encoding:** `to_categorical` converts the class labels (0-9) into one-hot encoded vectors.
- **Conv2D Parameters:** The `kernel_size` determines the size of the convolutional filter (e.g., 3x3). The number of filters determines how many different features are learned.
- **MaxPooling2D Parameters:** The `pool_size` determines the size of the pooling window (e.g., 2x2).
- **Optimizer:** The optimizer is the algorithm used to update the model's weights during training.
- **Loss Function:** The loss function measures the error between the model's predictions and the true labels.
- **Batch Size:** The batch size is the number of samples processed in each training iteration.
- **Epochs:** An epoch is one complete pass through the entire training dataset.
- **Dropout:** Dropout is a regularization technique that helps prevent overfitting.

Remember to run each cell to see its output. Experiment with the code and try to understand how different parameters affect the model's performance. Good luck! """

Conclusion and Submission

Congratulations on completing this notebook assignment! You have successfully built and trained a Convolutional Neural Network to classify handwritten digits from the MNIST dataset. You've explored key concepts like convolutional layers, pooling layers, activation functions, optimizers, loss functions, and training procedures. To further solidify your understanding, consider the following:

- **Review your notebook:** Go back through each section, reread the explanations, and make sure you understand the code and the concepts.
- **Experiment further:** Try different CNN architectures, add more layers, change hyperparameters, and see how it affects the performance. Explore other optimizers or loss functions.
- **Reflect on your learning:** Think about the challenges you faced and how you overcame them. What were the most important takeaways for you from this assignment?

Submission Instructions

To submit your assignment:

1. **Save your notebook:** Ensure all your work, including code cells, outputs, and answers to reflection questions, is saved in the notebook.
2. **Print the notebook as a `.pdf` file** and submit it to Canvas.

Deadline: February, 12th