AI Model Design

Yu-Chi Chu

07 Mar 2024

Implementation Process:

1. **Package Loading and Initialization:** Load necessary packages, including torch, torchvision, etc. Initialize the learning rate (lr) and set the number of epochs (EPOCH) to 50.
2. **Data Transformation:** Perform data transformation to improve training results. Transformations include normalization and data augmentation, among other methods. Data transformation converts the data into a format that is easier for machine learning algorithms to process and accelerates data loading.
3. **Dataset Splitting and Loading:** Split the dataset into a validation set and a training set. The dataset is divided with 10% allocated to the validation set and the remaining 90% to the training set. Then, the training set, validation set, and testing set are loaded into their respective DataLoaders.
4. **Model Selection:** For MNIST, a smaller and simpler model is used for training. For CIFAR10, due to its higher data complexity, a ResNet-like model is constructed. This model uses multiple convolutional layers and reduces the number of fully-connected layers to decrease the number of model parameters, as the number of parameters is usually determined by the fully-connected layers.
5. **Model Training (Loss Function and Optimizer):** Model training involves selecting a loss function and an optimizer. Cross-entropy is used as the loss function, and Stochastic Gradient Descent (SGD) is used as the optimizer. Unlike Gradient Descent (GD), which calculates the gradient of the loss function using the entire training set at once before updating parameters, SGD uses randomly sampled mini-batches and updates parameters after calculating the gradient of each mini-batch or the average of the gradients of a small batch. Furthermore, a learning rate scheduler, specifically OneCycleLR, is added to adjust the learning rate and improve learning effectiveness.
6. **Testing and Evaluation:** Run the trained model on the testing set to obtain the test accuracy and verify the model's performance.
7. **Training and Validation Curve Visualization:** Plot the training and validation loss and accuracy to observe the model's learning curves. Identify areas for potential optimization from the plots and improve the model through iterative training.

Q: What are the meanings of shuffle and drop_last in DataLoader?

shuffle rearranges the entire dataset (data) in each epoch. The default value is False. drop_last determines whether to discard the last incomplete batch when the total data length is not divisible by the batch_size. The default value is False.

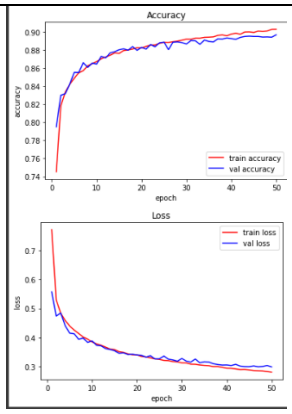
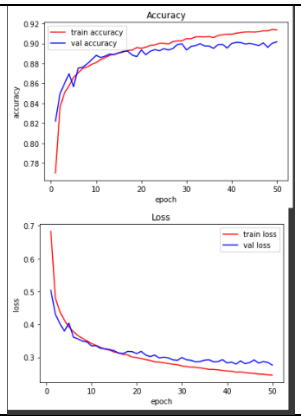
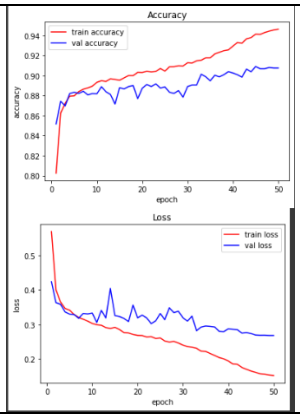
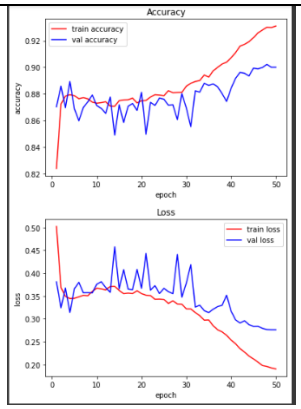
Result:

Model Parameters:

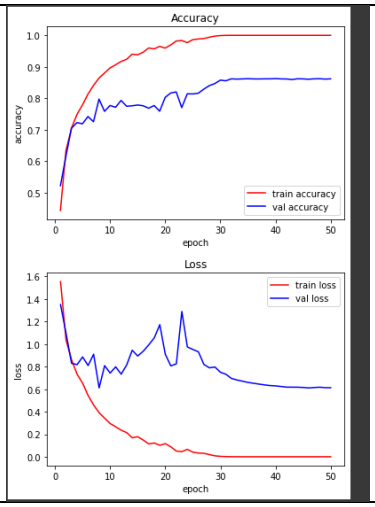
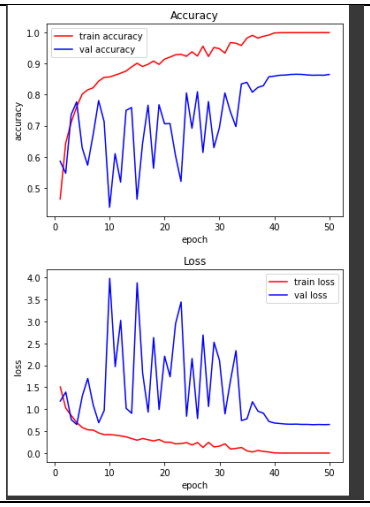
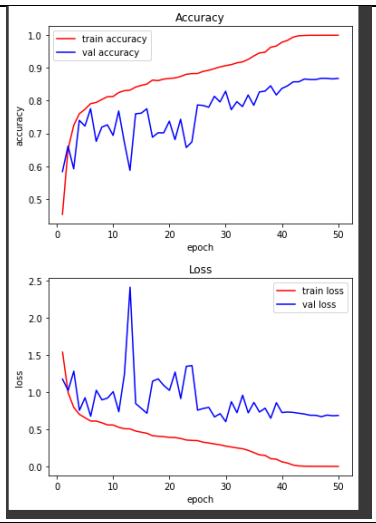
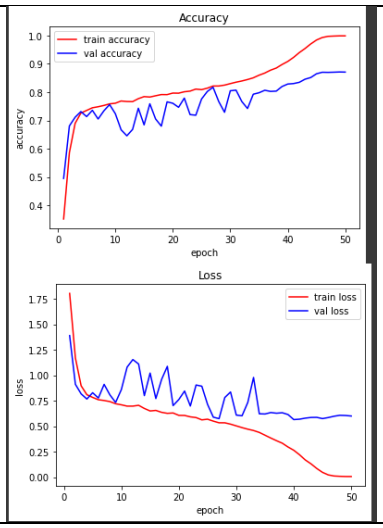
	Parameters	MACs
MNIST	0.05441M	0.000248768G
CIFAR10	0.44073M	0.034514176G

Training Comparison:

MNIST:

BATCH_SIZE	64	32
Without lr_schedualer		
Test accuracy	0.892	0.900
With lr_schedualer		
Test accuracy	0.910	0.903

CIFAR10:

BATCH_SIZE	128	64
With lr_schedualer		
Test accuracy	0.857	0.862
BATCH_SIZE	32	16
With lr_schedualer		
Test accuracy	0.864	0.868