How to Run the Code

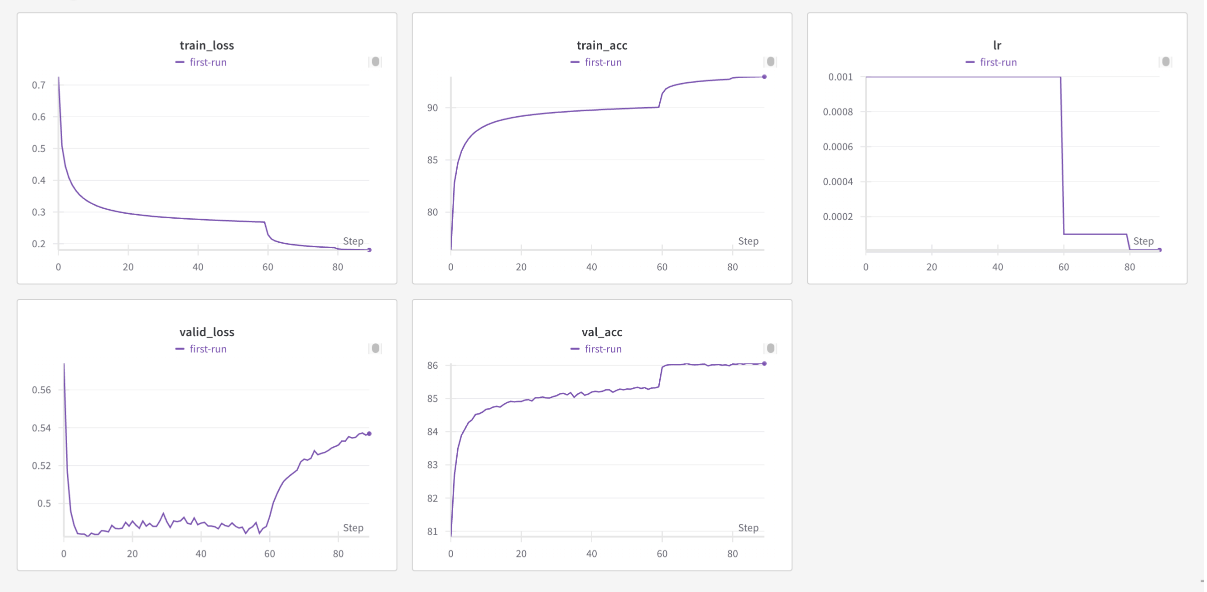
1. Install the required libraries such as PyTorch, torchaudio, wandb, numpy, and others listed in the script.
2. Prepare the dataset by ensuring MFCC and transcript files are in the appropriate directory structure: train-clean-100, dev-clean, and test-clean.
3. To start training, run the training script. The model will be trained, validated, and checkpoints will be saved when validation accuracy improves.
4. After training, run the test script to generate phoneme predictions. The results will be saved in a CSV file.

Model and Hyperparameters

- The model uses 8 fully connected layers with GELU activations, BatchNorm, and Dropout layers.
- Input size: (2*context + 1) * 28
- Output size: 42 phoneme classes.
- Hyperparameters:
  - Optimizer: AdamW with LR of 1e-3, batch size: 4096, context window: 25.
  - The model was trained for 90 epochs, with the best performance at epoch 78.

Key Experiments

- Different architectures were tested; the final architecture with 2048 hidden units and dropout yielded the best accuracy (85%).
- Time and frequency masking on MFCC inputs improved model robustness.


train_loss
first-run
Step
train_acc
first-run
Step
lr
first-run
Step
valid_loss
first-run
Step
val_acc
first-run
Step