I have only trained the models with the given data. I have not split them for testing.

Observations:

- The CNN's weights are usually initialised through *Xavier's Initialisation* in PyTorch. Which means the CNN is filled with certain weights before training it.
- The CNN that was trained on pseudo-labels always reached good training accuracy(~90% by 30 iterations).
- This is because starting from the weights initialised by *Xavier's Initialisation*, the model will easily go to nearly the same predictions it made(i.e., the pseudo labels) in the same number of iterations(i.e., 30).
- For the ground truth model, the rate at which the training accuracy increases is heavily depended on the amount of Mel frequency data that was fed into the model.

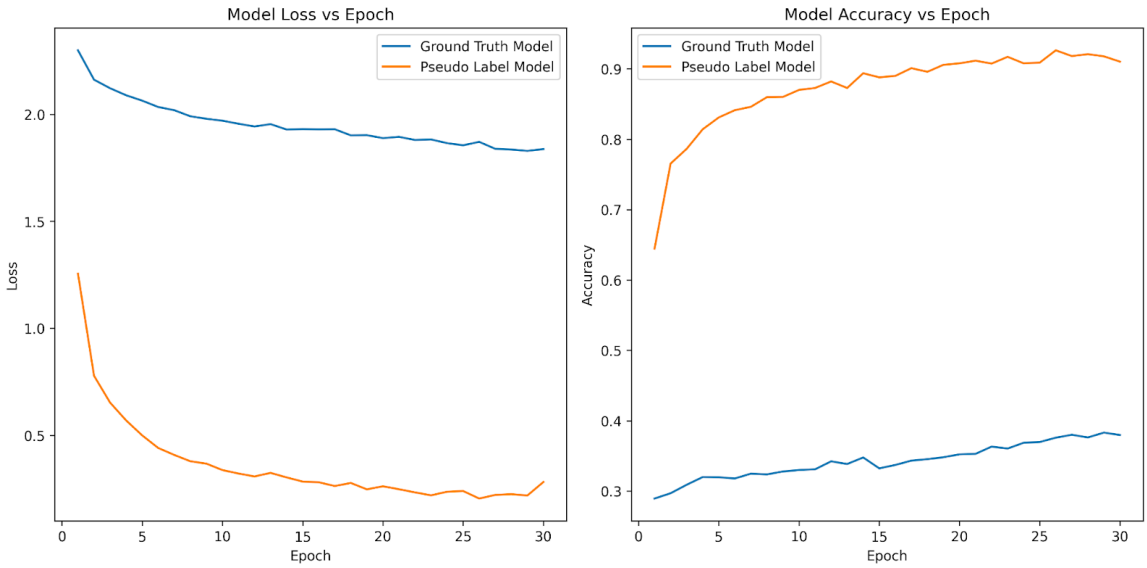


Fig.1: Trained on Mel Spectrogram sliced at [0:10,200:210]
==> Mel frequency 0-10 & Time window 200-210

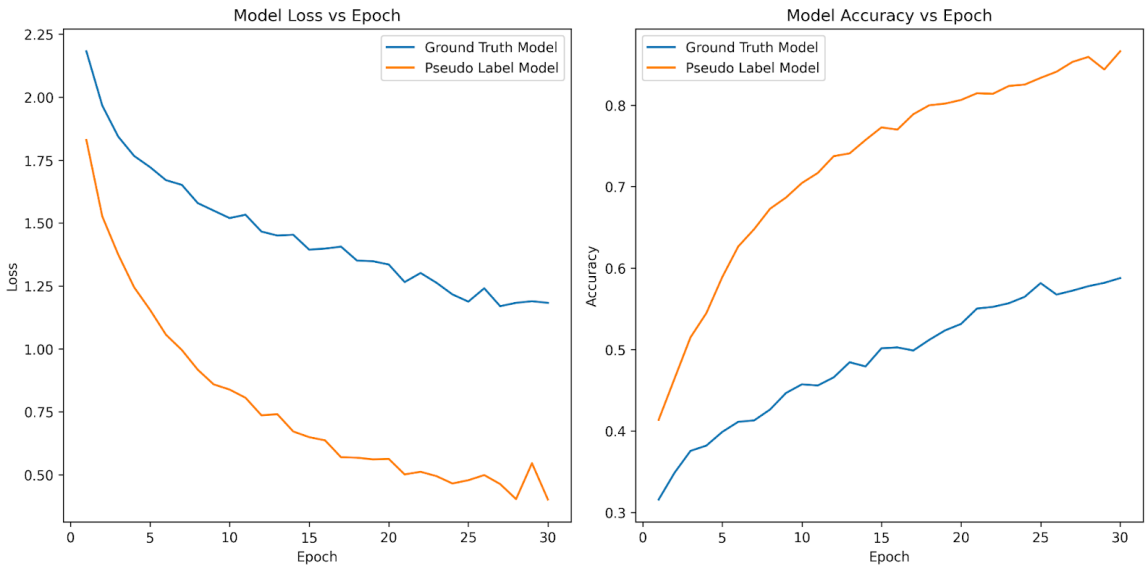


Fig.2: Trained on Mel Spectrogram sliced at [0:30,200:230]

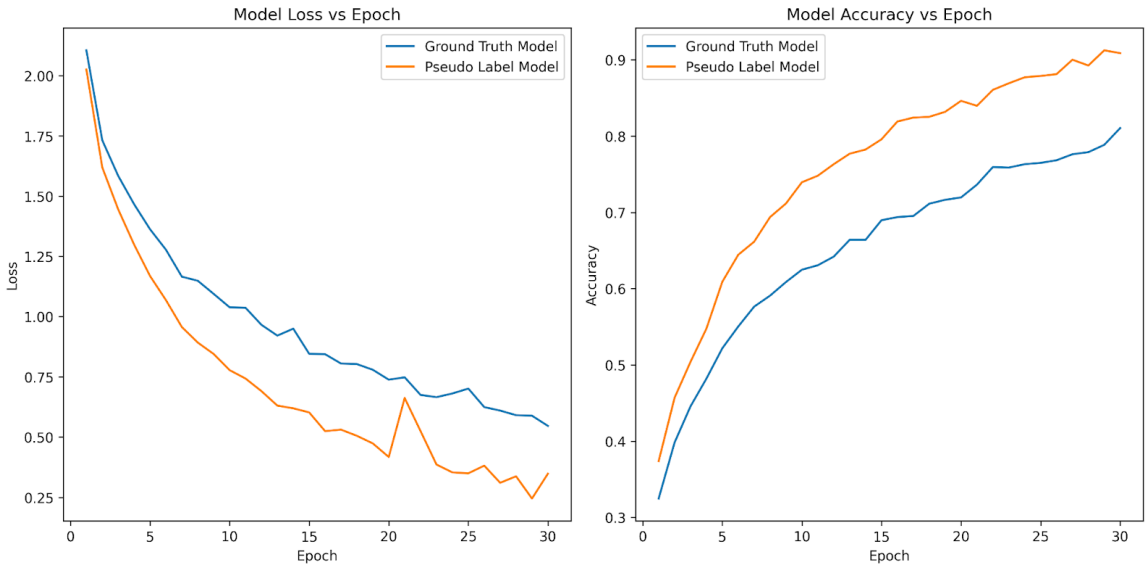


Fig.3: Trained on Mel Spectrogram sliced at [0:50,200:250]

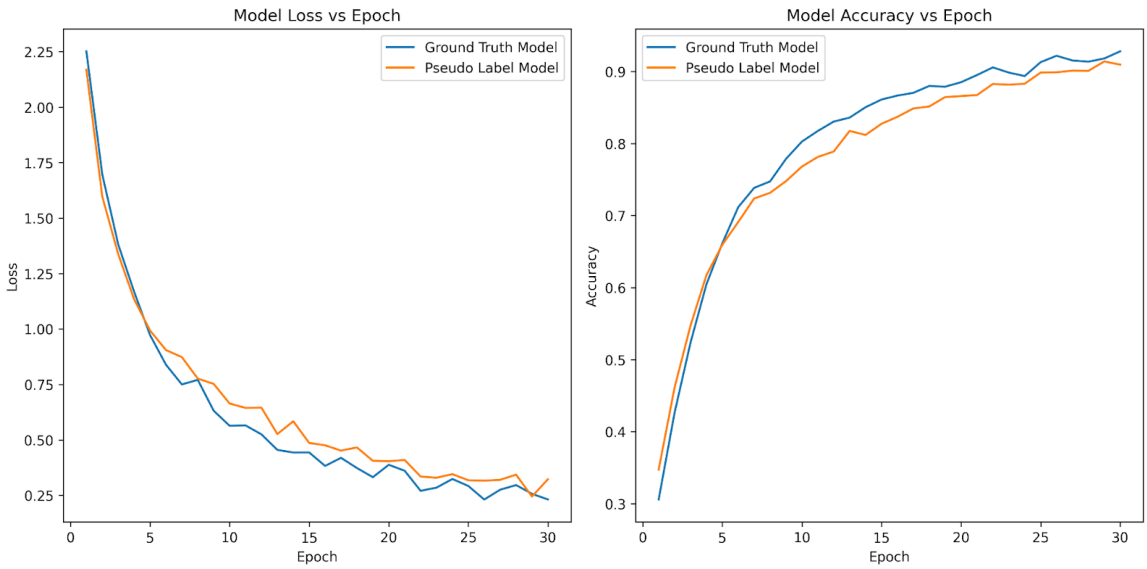


Fig.4: Trained on Mel Spectrogram sliced at [0:100,150:250]