Training Loss (4 models)

- with_pooling | no_softmax
- with_pooling | with_softmax
- no_pooling | no_softmax
- no_pooling | with_softmax