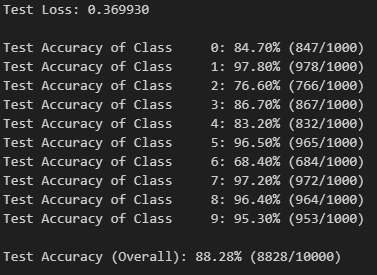
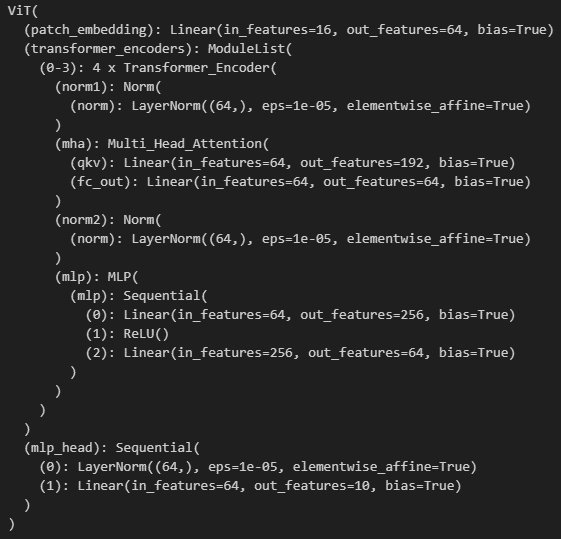
HW7 Report

STuser19 賴昱凱

**Training result:**



**Model:**



**參數量 (a single transformer encoder layer)**

Transformer Encoder

Multi-Head Attention:

* + **norm1**: 每個channel中皆有gamma, beta兩個參數。64個channels
  + **qkv linear layers**: 64 input channels, 192 output channels
  + **fc\_out linear layers**: 64 input channels, 64 output channels
  + **norm2**: 每個channel中皆有gamma, beta兩個參數。64個channels

MLP:

* + **Linear Layer 1**: 64 input channels, 256 output channels
  + **Linear Layer 2**: 256 input channels, 64 output channels

**計算量 (calculate linear layer only for a single transformer encoder layer)**

**一個矩陣與一個矩陣做矩陣乘法**，共需要個乘法以及個加法，共**個乘加計算(MACs)**。

Transformer Encoder

Multi-Head Attention:

* + **qkv linear layers**: 64 input channels, 192 output channels
  + **Calculate attention score alpha = Q @ KT**
  + **Calculate output out = attention @ V**
  + **fc\_out linear layers**: 64 input channels, 64 output channels

MLP:

* + **Linear Layer 1**: 64 input channels, 256 output channels
  + **Linear Layer 2**: 256 input channels, 64 output channels