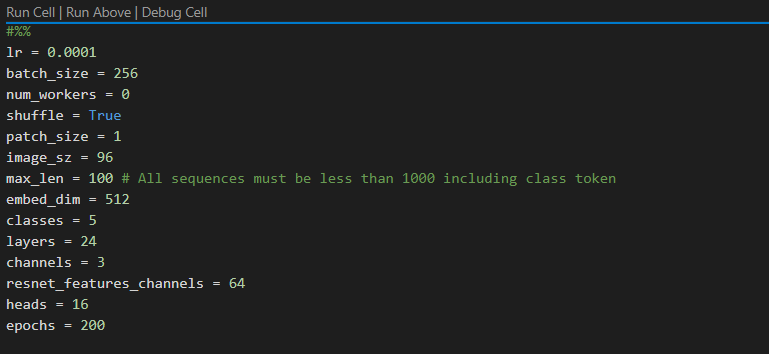
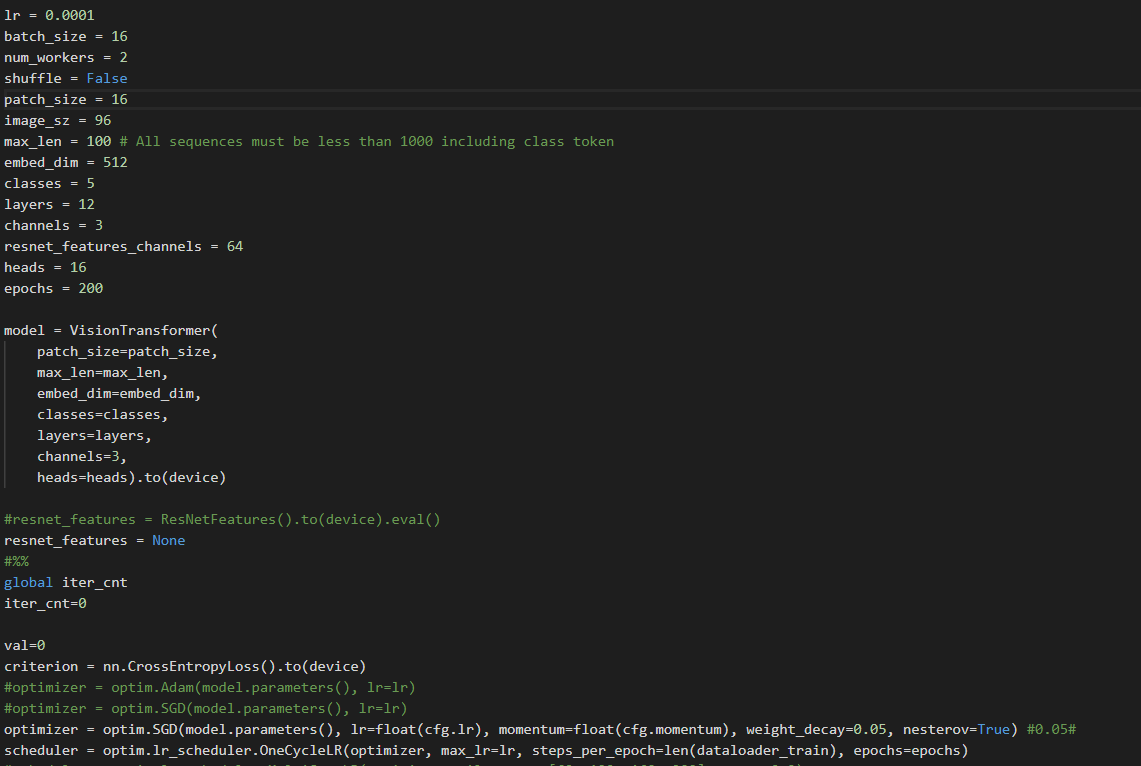
\* I can work with any patch size I want by modifying the max\_len variable here:



But I got the CUDA out of memory error.

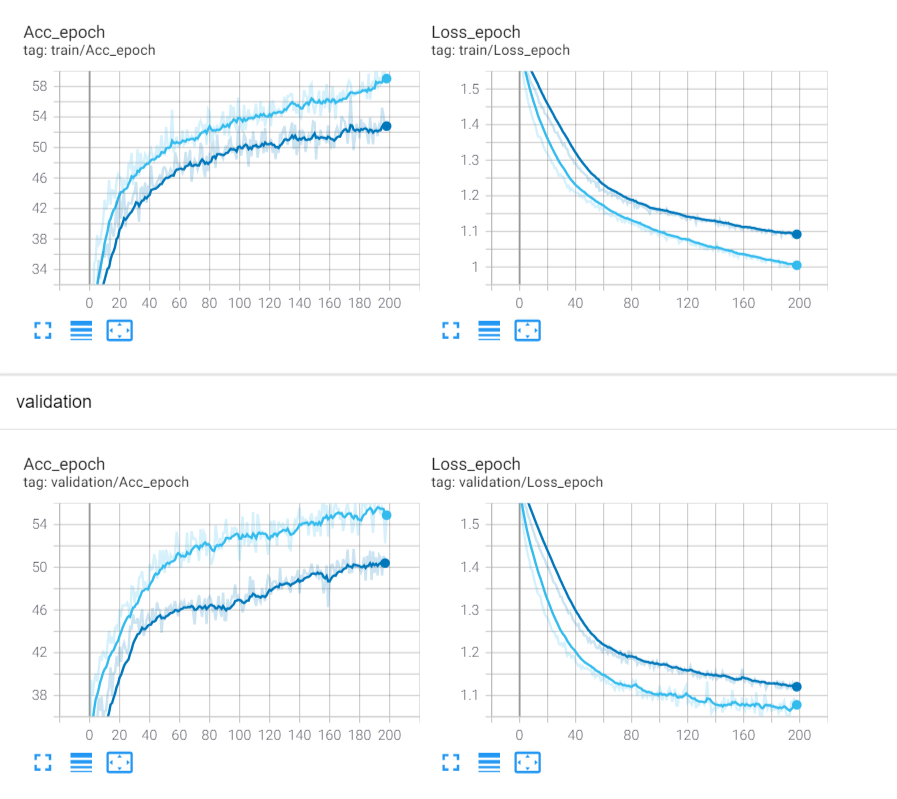
\* For training **VIT only**, those are the best parameters so far:





+LeakyRELU

* Increasing layers from 12 to 24 increases the accuracy of the model.



* Changing heads from 16 to 8 or 32 reduces accuracy. (To DO)