19/12/20:

1. Pytorch checkpointing tutorial: <https://github.com/prigoyal/pytorch_memonger/blob/master/tutorial/Checkpointing_for_PyTorch_models.ipynb>;
2. <https://pytorch.org/docs/stable/cuda.html#torch.cuda.memory_allocated> (memory api’s), <https://pytorch.org/docs/stable/checkpoint.html> (checkpointing api’s) in pytorch;
3. Pytorch generic profiler: <https://pytorch.org/tutorials/recipes/recipes/profiler.html>;
4. Time profile forward and backward passes in pytorch using proper gpu synchronizations: <https://discuss.pytorch.org/t/how-can-i-profile-the-backward-pass-of-model/76310/3>;
5. <https://stackoverflow.com/questions/21285885/remove-line-through-marker-in-matplotlib-legend> - matplotlib legend with good non-line scatter plot (+plt.legen();plt.show());
6. <https://qywu.github.io/2019/05/22/explore-gradient-checkpointing.html#:~:text=By%20applying%20gradient%20checkpointing%20or,on%20longer%20sequences%20of%20tokens.> - Gradient checkpointing with BERT;
7. <https://github.com/maknotavailable/pytorch-pretrained-BERT> - Google’s BERT models;
8. How to release GPU cache: <https://discuss.pytorch.org/t/how-can-we-release-gpu-memory-cache/14530>;