

nvidia apex

Scarecrow 1123 wrote a trainer for allennlp that uses nvidia's apex package to enable mixed precision training.

The full gist is available [here](#).

This is a copy of the trainer provided..

I find that my models are more often successful if I specify "O1" instead of "O2" for amp. This uses only a set of whitelisted operations in half precision mode.

[This trainer](#) has the change already made.

To use this during training include a snippet like this in your training json config.

```
{
  // ....
  "trainer": {
    "type": "fp16-trainer",
    "mixed_precision": true,
    // other options
  }
  // ....
}
```

and make sure the trainer is in a directory that you are including using `--include-package` .

For a bert model I was training, it ran out of VRAM on a single GTX 1070 without apex configured. However with apex configured the model was only using 4.5GB. There was no discernable penalty with regard to the number of epochs required though I haven't investigated a ton.

Revision #1

Created Thu, Oct 24, 2019 2:06 AM by [kenneth](#)

Updated Thu, Oct 24, 2019 2:15 AM by [kenneth](#)