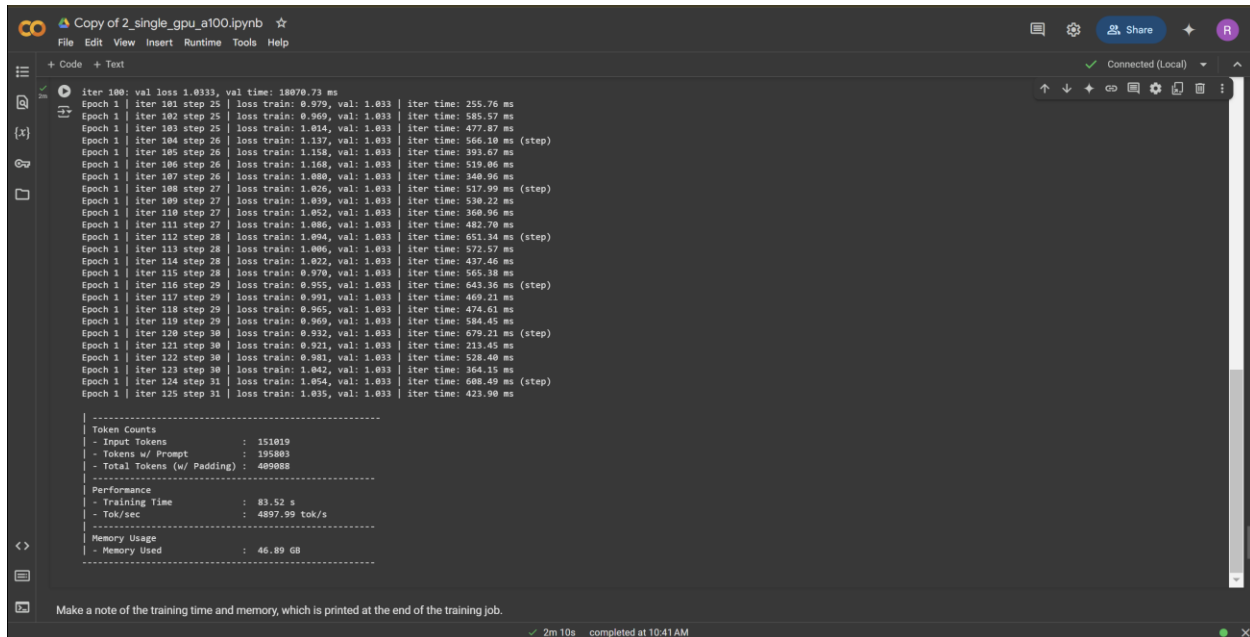


Larger models



```
Copy of 2_single_gpu_a100.ipynb
File Edit View Insert Runtime Tools Help

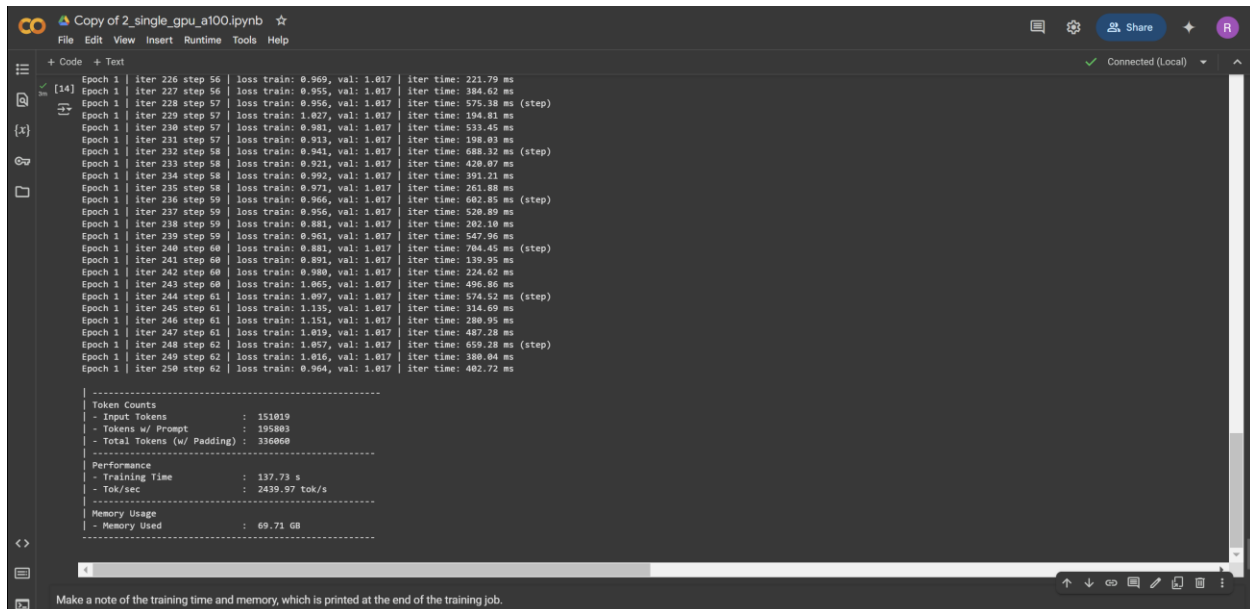
+ Code + Text

iter 100: val loss 1.0333, val time: 18070.73 ms
Epoch 1 | iter 101 step 25 | loss train: 0.979, val: 1.033 | iter time: 255.76 ms
Epoch 1 | iter 102 step 25 | loss train: 0.969, val: 1.033 | iter time: 585.57 ms
Epoch 1 | iter 103 step 25 | loss train: 1.014, val: 1.033 | iter time: 477.87 ms
Epoch 1 | iter 104 step 26 | loss train: 1.137, val: 1.033 | iter time: 566.10 ms (step)
Epoch 1 | iter 105 step 26 | loss train: 1.158, val: 1.033 | iter time: 393.67 ms
Epoch 1 | iter 106 step 26 | loss train: 1.168, val: 1.033 | iter time: 519.06 ms
Epoch 1 | iter 107 step 26 | loss train: 1.088, val: 1.033 | iter time: 340.96 ms
Epoch 1 | iter 108 step 27 | loss train: 1.026, val: 1.033 | iter time: 517.99 ms (step)
Epoch 1 | iter 109 step 27 | loss train: 1.039, val: 1.033 | iter time: 530.22 ms
Epoch 1 | iter 110 step 27 | loss train: 1.052, val: 1.033 | iter time: 360.96 ms
Epoch 1 | iter 111 step 27 | loss train: 1.086, val: 1.033 | iter time: 482.70 ms
Epoch 1 | iter 112 step 28 | loss train: 1.094, val: 1.033 | iter time: 551.34 ms (step)
Epoch 1 | iter 113 step 28 | loss train: 1.006, val: 1.033 | iter time: 572.57 ms
Epoch 1 | iter 114 step 28 | loss train: 1.022, val: 1.033 | iter time: 437.46 ms
Epoch 1 | iter 115 step 28 | loss train: 0.978, val: 1.033 | iter time: 565.38 ms
Epoch 1 | iter 116 step 29 | loss train: 0.955, val: 1.033 | iter time: 643.36 ms (step)
Epoch 1 | iter 117 step 29 | loss train: 0.991, val: 1.033 | iter time: 469.21 ms
Epoch 1 | iter 118 step 29 | loss train: 0.965, val: 1.033 | iter time: 474.61 ms
Epoch 1 | iter 119 step 29 | loss train: 0.969, val: 1.033 | iter time: 584.45 ms
Epoch 1 | iter 120 step 30 | loss train: 0.932, val: 1.033 | iter time: 679.21 ms (step)
Epoch 1 | iter 121 step 30 | loss train: 0.921, val: 1.033 | iter time: 213.45 ms
Epoch 1 | iter 122 step 30 | loss train: 0.981, val: 1.033 | iter time: 528.40 ms
Epoch 1 | iter 123 step 30 | loss train: 1.042, val: 1.033 | iter time: 364.15 ms
Epoch 1 | iter 124 step 31 | loss train: 1.054, val: 1.033 | iter time: 688.49 ms (step)
Epoch 1 | iter 125 step 31 | loss train: 1.035, val: 1.033 | iter time: 423.90 ms

-----
| Token Counts
| - Input Tokens      : 151019
| - Tokens w/ Prompt  : 195003
| - Total Tokens (w/ Padding) : 409088
| -----
| Performance
| - Training Time     : 83.52 s
| - Tok/sec           : 4897.99 tok/s
| -----
| Memory Usage
| - Memory Used       : 46.89 GB
| -----

Make a note of the training time and memory, which is printed at the end of the training job.
```

Larger model -3b



```
Copy of 2_single_gpu_a100.ipynb
File Edit View Insert Runtime Tools Help

+ Code + Text

Epoch 1 | iter 226 step 56 | loss train: 0.969, val: 1.017 | iter time: 221.79 ms
Epoch 1 | iter 227 step 56 | loss train: 0.955, val: 1.017 | iter time: 384.62 ms
Epoch 1 | iter 228 step 57 | loss train: 0.956, val: 1.017 | iter time: 575.38 ms (step)
Epoch 1 | iter 229 step 57 | loss train: 1.027, val: 1.017 | iter time: 104.81 ms
Epoch 1 | iter 230 step 57 | loss train: 0.981, val: 1.017 | iter time: 533.45 ms
Epoch 1 | iter 231 step 57 | loss train: 0.913, val: 1.017 | iter time: 108.03 ms
Epoch 1 | iter 232 step 58 | loss train: 0.941, val: 1.017 | iter time: 688.32 ms (step)
Epoch 1 | iter 233 step 58 | loss train: 0.921, val: 1.017 | iter time: 428.07 ms
Epoch 1 | iter 234 step 58 | loss train: 0.992, val: 1.017 | iter time: 391.21 ms
Epoch 1 | iter 235 step 58 | loss train: 0.971, val: 1.017 | iter time: 261.88 ms
Epoch 1 | iter 236 step 59 | loss train: 0.966, val: 1.017 | iter time: 602.85 ms (step)
Epoch 1 | iter 237 step 59 | loss train: 0.956, val: 1.017 | iter time: 520.89 ms
Epoch 1 | iter 238 step 59 | loss train: 0.881, val: 1.017 | iter time: 202.10 ms
Epoch 1 | iter 239 step 59 | loss train: 0.961, val: 1.017 | iter time: 547.96 ms
Epoch 1 | iter 240 step 60 | loss train: 0.881, val: 1.017 | iter time: 704.45 ms (step)
Epoch 1 | iter 241 step 60 | loss train: 0.891, val: 1.017 | iter time: 139.95 ms
Epoch 1 | iter 242 step 60 | loss train: 0.900, val: 1.017 | iter time: 224.62 ms
Epoch 1 | iter 243 step 60 | loss train: 1.005, val: 1.017 | iter time: 496.86 ms
Epoch 1 | iter 244 step 61 | loss train: 1.097, val: 1.017 | iter time: 574.52 ms (step)
Epoch 1 | iter 245 step 61 | loss train: 1.135, val: 1.017 | iter time: 314.69 ms
Epoch 1 | iter 246 step 61 | loss train: 1.151, val: 1.017 | iter time: 280.95 ms
Epoch 1 | iter 247 step 61 | loss train: 1.019, val: 1.017 | iter time: 487.28 ms
Epoch 1 | iter 248 step 62 | loss train: 1.057, val: 1.017 | iter time: 659.28 ms (step)
Epoch 1 | iter 249 step 62 | loss train: 1.016, val: 1.017 | iter time: 380.04 ms
Epoch 1 | iter 250 step 62 | loss train: 0.964, val: 1.017 | iter time: 402.72 ms

-----
| Token Counts
| - Input Tokens      : 151019
| - Tokens w/ Prompt  : 195003
| - Total Tokens (w/ Padding) : 336068
| -----
| Performance
| - Training Time     : 137.73 s
| - Tok/sec           : 2439.97 tok/s
| -----
| Memory Usage
| - Memory Used       : 69.71 GB
| -----

Make a note of the training time and memory, which is printed at the end of the training job.
```

Larger model -7b

```

return component(**cfg)
.....
File ~/opt/conda/lib/python3.12/site-packages/litgpt/finetune/full.py, line 122, in setup
fabric.launch(main, devices, resume, seed, config, data, checkpoint_dir, out_dir, train, eval, optimizer)
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric.py, line 837, in launch
return self._wrap_and_launch(function, self, *args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric/fabric.py, line 923, in _wrap_and_launch
return to_run(*args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric/fabric.py, line 928, in _wrap_with_setup
return to_run(*args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/litgpt/finetune/full.py, line 171, in main
token_counts = fit(fabric, state, train_dataloader, val_dataloader, devices, resume, checkpoint_dir, out_dir, train, eval, data)
.....
File ~/opt/conda/lib/python3.12/site-packages/litgpt/finetune/full.py, line 272, in fit
optimizer.step()
File ~/opt/conda/lib/python3.12/site-packages/torch/optim/lr_scheduler.py, line 137, in wrapper
return func._get_opt(opt, opt._class_)(*args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric/wrappers.py, line 87, in step
output = self._strategy.optimizer_step(*args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric/strategies/strategy.py, line 204, in optimizer_step
return self.precision.optimizer_step(optimizer, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/lightning/fabric/plugins/precision/precision.py, line 124, in optimizer_step
return optimizer.step(**kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/torch/optim/optimizer.py, line 487, in wrapper
out = func(*args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/torch/optim/optimizer.py, line 91, in _use_grad
ret = func(self, *args, **kwargs)
.....
File ~/opt/conda/lib/python3.12/site-packages/torch/optim/adamw.py, line 209, in step
has_complex = self._init_group(
.....
File ~/opt/conda/lib/python3.12/site-packages/torch/optim/adamw.py, line 152, in _init_group
state["exp_avg_sq"] = torch.zeros_like(
.....
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 150.00 MiB. GPU 0 has a total capacity of 79.25 GiB of which 62.75 MiB is free. Process 52339 has 79.18 GiB memory in use. Of the allocated memory 78.64

```

Larger model -13b (Adam optimizer) (fail)

```

epoch 1 | iter 20 step 20 | loss train: 2.402, val: n/a | iter time: 158.55 ms (step)
epoch 1 | iter 21 step 21 | loss train: 2.402, val: n/a | iter time: 158.54 ms (step)
epoch 1 | iter 22 step 22 | loss train: 2.401, val: n/a | iter time: 158.51 ms (step)
epoch 1 | iter 23 step 23 | loss train: 2.401, val: n/a | iter time: 159.05 ms (step)
epoch 1 | iter 24 step 24 | loss train: 1.637, val: n/a | iter time: 159.01 ms (step)
epoch 1 | iter 25 step 25 | loss train: 1.446, val: n/a | iter time: 243.68 ms (step)
Validating ...
Generate a funny caption for the following photo.
Below is an instruction that describes a task. Write a response that appropriately completes the request.

### Instruction:
Generate a funny caption for the following photo.

### Response:
A man in a tie and coat walks into a bar carrying a small dog wearing a coat.

### Verify:

### Pass
You have passed this challenge.

### Fail
### Try again

iter 25: val loss 1.6281, val time: 7315.42 ms

-----
| Token Counts
| - Input Tokens      : 2592
| - Tokens w/ Prompt  : 3739
| - Total Tokens (w/ Padding) : 3739
|-----
| Performance
| - Training Time     : 13.77 s
| - Tok/sec           : 271.51 tok/s
|-----
| Memory Usage
| - Memory Used       : 52.38 GB
|-----

```

Larger model -13b (SGD optimizer) (25 steps)