

TinyLlama-1.1B

```

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

+ Code + Text
[8]
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1736, in _wrapped_call_impl
return self._call_impl(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1747, in _call_impl
return forward_call(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/litgpt/model.py, line 161, in forward
x = block(x, cos, sin, mask, input_pos, input_pos_maxp1)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1736, in _wrapped_call_impl
return self._call_impl(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1747, in _call_impl
return forward_call(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/litgpt/model.py, line 322, in forward
return self.post_mlp_norm(self.mlp(x_normed)) + x
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1736, in _wrapped_call_impl
return self._call_impl(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1747, in _call_impl
return forward_call(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/litgpt/model.py, line 547, in forward
x_fc_1 = self.fc_1(x)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1736, in _wrapped_call_impl
return self._call_impl(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/module.py, line 1747, in _call_impl
return forward_call(*args, **kwargs)
File ~/opt/conda/lib/python3.12/site-packages/torch/nn/modules/linear.py, line 125, in forward
return F.linear(input, self.weight, self.bias)
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 352.00 MiB. GPU 0 has a total capacity of 79.25 GiB of which 44.75 MiB is free. Process 36263 has 79.20 GiB memory in use. Of the allocated memory 78.81
This will fail because the training job won't fit in our 80GB GPU memory.

```

The baseline

```

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

+ Code + Text
Epoch 1 | iter 108 step 108 | loss train: 1.674, val: 1.621 | iter time: 349.12 ms (step)
Epoch 1 | iter 109 step 109 | loss train: 1.797, val: 1.621 | iter time: 438.05 ms (step)
Epoch 1 | iter 110 step 110 | loss train: 1.435, val: 1.621 | iter time: 386.25 ms (step)
Epoch 1 | iter 111 step 111 | loss train: 1.865, val: 1.621 | iter time: 353.28 ms (step)
Epoch 1 | iter 112 step 112 | loss train: 1.701, val: 1.621 | iter time: 429.57 ms (step)
Epoch 1 | iter 113 step 113 | loss train: 1.377, val: 1.621 | iter time: 426.41 ms (step)
Epoch 1 | iter 114 step 114 | loss train: 1.685, val: 1.621 | iter time: 358.22 ms (step)
Epoch 1 | iter 115 step 115 | loss train: 1.355, val: 1.621 | iter time: 431.38 ms (step)
Epoch 1 | iter 116 step 116 | loss train: 1.719, val: 1.621 | iter time: 428.24 ms (step)
Epoch 1 | iter 117 step 117 | loss train: 1.518, val: 1.621 | iter time: 353.78 ms (step)
Epoch 1 | iter 118 step 118 | loss train: 1.512, val: 1.621 | iter time: 481.58 ms (step)
Epoch 1 | iter 119 step 119 | loss train: 1.539, val: 1.621 | iter time: 428.00 ms (step)
Epoch 1 | iter 120 step 120 | loss train: 1.580, val: 1.621 | iter time: 434.18 ms (step)
Epoch 1 | iter 121 step 121 | loss train: 1.342, val: 1.621 | iter time: 289.34 ms (step)
Epoch 1 | iter 122 step 122 | loss train: 1.669, val: 1.621 | iter time: 427.28 ms (step)
Epoch 1 | iter 123 step 123 | loss train: 1.746, val: 1.621 | iter time: 273.45 ms (step)
Epoch 1 | iter 124 step 124 | loss train: 1.522, val: 1.621 | iter time: 428.85 ms (step)
Epoch 1 | iter 125 step 125 | loss train: 1.484, val: 1.621 | iter time: 354.64 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.

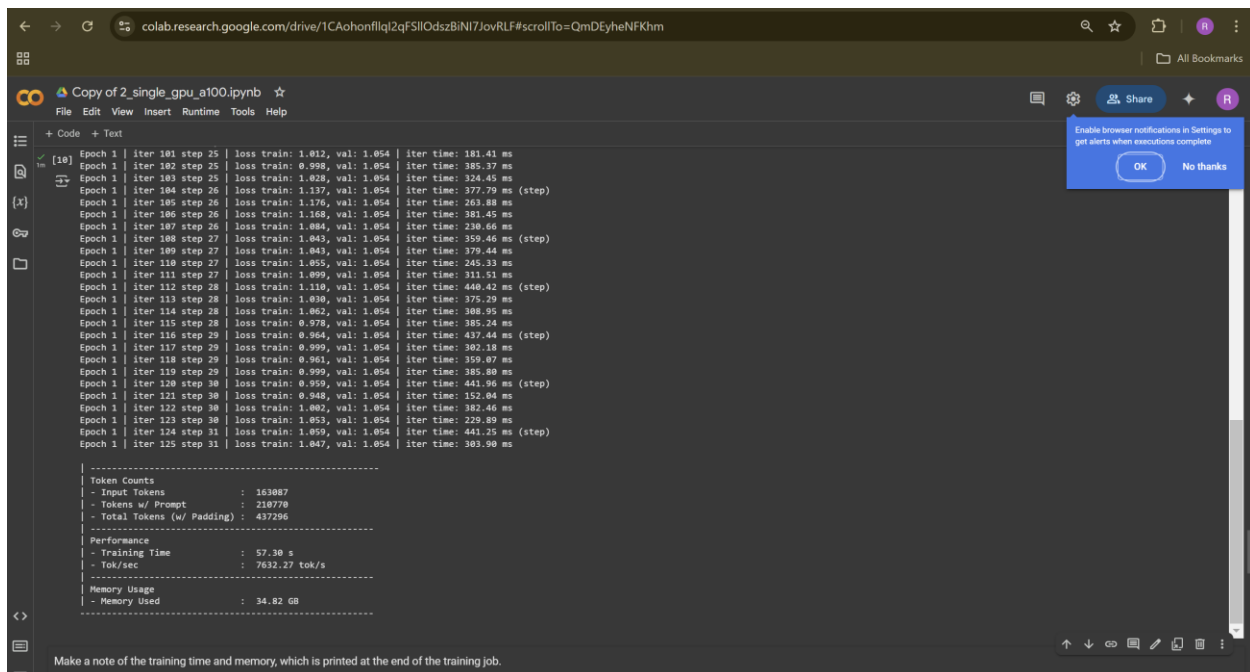
### Response:
"In this photo, all the way up. The caption reads "Through the looking glass guises the world". Just makes it hard to see the colors outside."

iter 125: val loss 1.5937, val time: 12876.29 ms

-----
Token Counts
- Input Tokens      : 163887
- Tokens w/ Prompt  : 210778
- Total Tokens (w/ Padding) : 437296
-----
Performance
- Training Time     : 165.70 s
- Tok/sec           : 2639.04 tok/s
-----
Memory Usage
- Memory Used       : 30.41 GB
-----

```

Reduced batch size

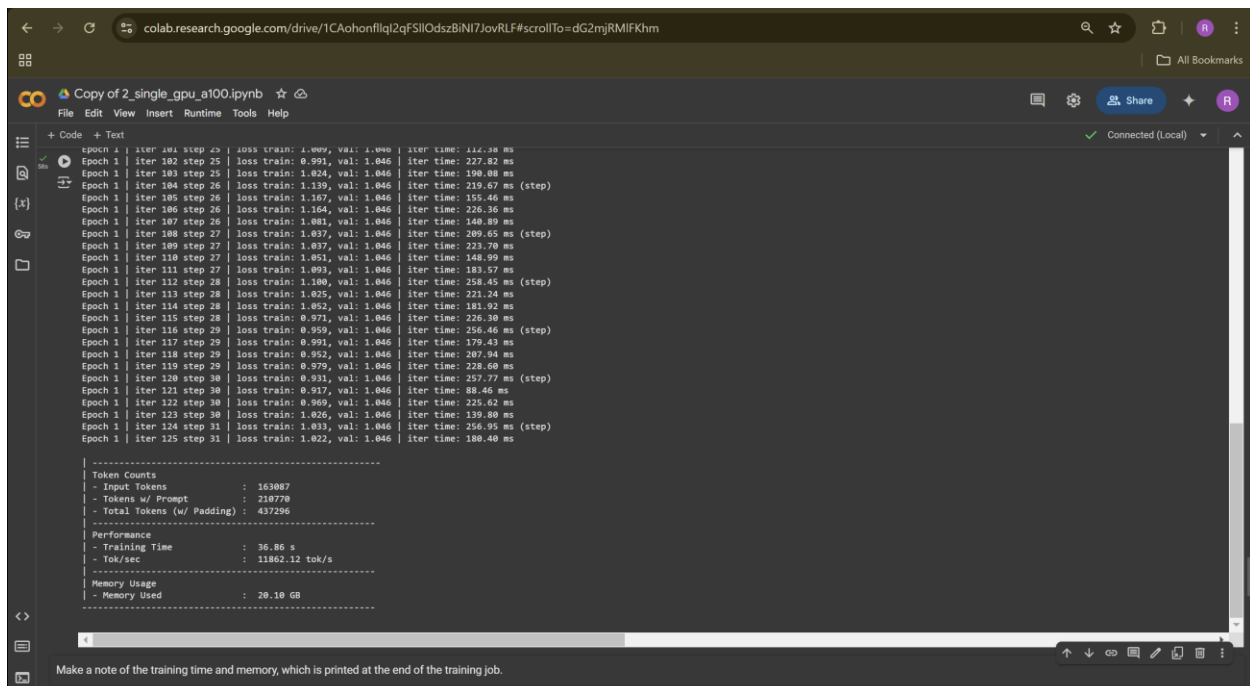


```
Epoch 1 | iter 101 step 25 | loss train: 1.012, val: 1.054 | iter time: 181.41 ms
Epoch 1 | iter 102 step 25 | loss train: 0.998, val: 1.054 | iter time: 385.37 ms
Epoch 1 | iter 103 step 25 | loss train: 1.028, val: 1.054 | iter time: 324.45 ms
Epoch 1 | iter 104 step 26 | loss train: 1.137, val: 1.054 | iter time: 377.79 ms (step)
Epoch 1 | iter 105 step 26 | loss train: 1.176, val: 1.054 | iter time: 263.88 ms
Epoch 1 | iter 106 step 26 | loss train: 1.168, val: 1.054 | iter time: 381.45 ms
Epoch 1 | iter 107 step 26 | loss train: 1.084, val: 1.054 | iter time: 230.66 ms
Epoch 1 | iter 108 step 27 | loss train: 1.043, val: 1.054 | iter time: 359.46 ms (step)
Epoch 1 | iter 109 step 27 | loss train: 1.043, val: 1.054 | iter time: 379.44 ms
Epoch 1 | iter 110 step 27 | loss train: 1.055, val: 1.054 | iter time: 245.33 ms
Epoch 1 | iter 111 step 27 | loss train: 1.099, val: 1.054 | iter time: 311.51 ms
Epoch 1 | iter 112 step 28 | loss train: 1.110, val: 1.054 | iter time: 440.42 ms (step)
Epoch 1 | iter 113 step 28 | loss train: 1.019, val: 1.054 | iter time: 375.29 ms
Epoch 1 | iter 114 step 28 | loss train: 1.002, val: 1.054 | iter time: 308.95 ms
Epoch 1 | iter 115 step 28 | loss train: 0.978, val: 1.054 | iter time: 385.24 ms
Epoch 1 | iter 116 step 29 | loss train: 0.964, val: 1.054 | iter time: 437.44 ms (step)
Epoch 1 | iter 117 step 29 | loss train: 0.999, val: 1.054 | iter time: 302.18 ms
Epoch 1 | iter 118 step 29 | loss train: 0.963, val: 1.054 | iter time: 359.97 ms
Epoch 1 | iter 119 step 29 | loss train: 0.999, val: 1.054 | iter time: 385.80 ms
Epoch 1 | iter 120 step 30 | loss train: 0.959, val: 1.054 | iter time: 441.96 ms (step)
Epoch 1 | iter 121 step 30 | loss train: 0.948, val: 1.054 | iter time: 152.04 ms
Epoch 1 | iter 122 step 30 | loss train: 1.002, val: 1.054 | iter time: 302.46 ms
Epoch 1 | iter 123 step 30 | loss train: 1.053, val: 1.054 | iter time: 229.89 ms
Epoch 1 | iter 124 step 31 | loss train: 1.059, val: 1.054 | iter time: 441.25 ms (step)
Epoch 1 | iter 125 step 31 | loss train: 1.047, val: 1.054 | iter time: 303.90 ms

-----
| Token Counts
| - Input Tokens      : 163087
| - Tokens w/ Prompt  : 210770
| - Total Tokens (w/ Padding) : 437296
|-----
| Performance
| - Training Time     : 57.30 s
| - Tok/sec           : 7632.27 tok/s
|-----
| Memory Usage
| - Memory Used       : 34.82 GB
|-----

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

Gradient accumulation



```
Epoch 1 | iter 101 step 25 | loss train: 1.099, val: 1.040 | iter time: 112.28 ms
Epoch 1 | iter 102 step 25 | loss train: 0.991, val: 1.046 | iter time: 227.82 ms
Epoch 1 | iter 103 step 25 | loss train: 1.024, val: 1.046 | iter time: 190.08 ms
Epoch 1 | iter 104 step 26 | loss train: 1.139, val: 1.046 | iter time: 19.67 ms (step)
Epoch 1 | iter 105 step 26 | loss train: 1.167, val: 1.046 | iter time: 155.46 ms
Epoch 1 | iter 106 step 26 | loss train: 1.164, val: 1.046 | iter time: 226.36 ms
Epoch 1 | iter 107 step 26 | loss train: 1.001, val: 1.046 | iter time: 140.89 ms
Epoch 1 | iter 108 step 27 | loss train: 1.037, val: 1.046 | iter time: 209.65 ms (step)
Epoch 1 | iter 109 step 27 | loss train: 1.037, val: 1.046 | iter time: 223.70 ms
Epoch 1 | iter 110 step 27 | loss train: 1.051, val: 1.046 | iter time: 148.99 ms
Epoch 1 | iter 111 step 27 | loss train: 1.093, val: 1.046 | iter time: 183.57 ms
Epoch 1 | iter 112 step 28 | loss train: 1.100, val: 1.046 | iter time: 258.45 ms (step)
Epoch 1 | iter 113 step 28 | loss train: 1.025, val: 1.046 | iter time: 221.24 ms
Epoch 1 | iter 114 step 28 | loss train: 1.052, val: 1.046 | iter time: 181.92 ms
Epoch 1 | iter 115 step 28 | loss train: 0.971, val: 1.046 | iter time: 226.38 ms
Epoch 1 | iter 116 step 29 | loss train: 0.959, val: 1.046 | iter time: 256.46 ms (step)
Epoch 1 | iter 117 step 29 | loss train: 0.991, val: 1.046 | iter time: 179.43 ms
Epoch 1 | iter 118 step 29 | loss train: 0.952, val: 1.046 | iter time: 207.94 ms
Epoch 1 | iter 119 step 29 | loss train: 0.979, val: 1.046 | iter time: 228.08 ms
Epoch 1 | iter 120 step 30 | loss train: 0.931, val: 1.046 | iter time: 257.77 ms (step)
Epoch 1 | iter 121 step 30 | loss train: 0.917, val: 1.046 | iter time: 88.46 ms
Epoch 1 | iter 122 step 30 | loss train: 0.969, val: 1.046 | iter time: 225.62 ms
Epoch 1 | iter 123 step 30 | loss train: 1.026, val: 1.046 | iter time: 139.80 ms
Epoch 1 | iter 124 step 31 | loss train: 1.033, val: 1.046 | iter time: 256.95 ms (step)
Epoch 1 | iter 125 step 31 | loss train: 1.022, val: 1.046 | iter time: 180.40 ms

-----
| Token Counts
| - Input Tokens      : 163087
| - Tokens w/ Prompt  : 210770
| - Total Tokens (w/ Padding) : 437296
|-----
| Performance
| - Training Time     : 36.86 s
| - Tok/sec           : 11862.12 tok/s
|-----
| Memory Usage
| - Memory Used       : 20.10 GB
|-----

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

Reduced precision

```
Copy of 2_single_gpu_0100.ipynb
File Edit View Insert Runtime Tools Help
+ Code + Text
Connected (Local)

iter 100: val loss 1.0542, val time: 8054.90 ms
Epoch 1 | iter 101 step 25 | loss train: 1.012, val: 1.054 | iter time: 126.80 ms
Epoch 1 | iter 102 step 25 | loss train: 0.997, val: 1.054 | iter time: 245.14 ms
Epoch 1 | iter 103 step 25 | loss train: 1.027, val: 1.054 | iter time: 207.83 ms
Epoch 1 | iter 104 step 26 | loss train: 1.137, val: 1.054 | iter time: 260.69 ms (step)
Epoch 1 | iter 105 step 26 | loss train: 1.176, val: 1.054 | iter time: 169.95 ms
Epoch 1 | iter 106 step 26 | loss train: 1.168, val: 1.054 | iter time: 242.69 ms
Epoch 1 | iter 107 step 26 | loss train: 1.083, val: 1.054 | iter time: 158.64 ms
Epoch 1 | iter 108 step 27 | loss train: 1.043, val: 1.054 | iter time: 251.46 ms (step)
Epoch 1 | iter 109 step 27 | loss train: 1.042, val: 1.054 | iter time: 239.30 ms
Epoch 1 | iter 110 step 27 | loss train: 1.055, val: 1.054 | iter time: 167.56 ms
Epoch 1 | iter 111 step 27 | loss train: 1.099, val: 1.054 | iter time: 201.66 ms
Epoch 1 | iter 112 step 28 | loss train: 1.110, val: 1.054 | iter time: 298.92 ms (step)
Epoch 1 | iter 113 step 28 | loss train: 1.030, val: 1.054 | iter time: 238.62 ms
Epoch 1 | iter 114 step 28 | loss train: 1.062, val: 1.054 | iter time: 199.84 ms
Epoch 1 | iter 115 step 28 | loss train: 0.979, val: 1.054 | iter time: 245.01 ms
Epoch 1 | iter 116 step 29 | loss train: 0.964, val: 1.054 | iter time: 297.97 ms (step)
Epoch 1 | iter 117 step 29 | loss train: 0.999, val: 1.054 | iter time: 193.72 ms
Epoch 1 | iter 118 step 29 | loss train: 0.961, val: 1.054 | iter time: 227.25 ms
Epoch 1 | iter 119 step 29 | loss train: 0.999, val: 1.054 | iter time: 245.28 ms
Epoch 1 | iter 120 step 30 | loss train: 0.959, val: 1.054 | iter time: 299.09 ms (step)
Epoch 1 | iter 121 step 30 | loss train: 0.948, val: 1.054 | iter time: 102.06 ms
Epoch 1 | iter 122 step 30 | loss train: 1.003, val: 1.054 | iter time: 243.16 ms
Epoch 1 | iter 123 step 30 | loss train: 1.054, val: 1.054 | iter time: 158.03 ms
Epoch 1 | iter 124 step 31 | loss train: 1.059, val: 1.054 | iter time: 298.37 ms (step)
Epoch 1 | iter 125 step 31 | loss train: 1.047, val: 1.054 | iter time: 195.37 ms

-----
| Token Counts
| - Input Tokens      : 163087
| - Tokens w/ Prompt  : 210770
| - Total Tokens (w/ Padding) : 437296
|-----
| Performance
| - Training Time     : 38.67 s
| - Tok/sec           : 11307.95 tok/s
|-----
| Memory Usage
| - Memory Used       : 31.32 GB
|-----

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

Mixed precision