```
------------------------- DeepSpeed Flops Profiler -------------------------
Profile Summary at step 10:
Notations:
data parallel size (dp_size), model parallel size(mp_size),
number of parameters (params), number of multiply-accumulate operations(MACs),
number of floating-point operations (flops), floating-point operations per second
(FLOPS),
fwd latency (forward propagation latency), bwd latency (backward propagation
latency),
step (weights update latency), iter latency (sum of fwd, bwd and step latency)

params per gpu:                                                 11.4 M
params of model = params per GPU * mp_size:                     11.4 M
fwd MACs per GPU:                                               42.59 GMACs
fwd flops per GPU:                                              85.59 G
fwd flops of model = fwd flops per GPU * mp_size:               85.59 G
fwd latency:                                                    99.9 ms
fwd FLOPS per GPU = fwd flops per GPU / fwd latency:            856.74 GFLOPS

----------------------------- Aggregated Profile per GPU
-----------------------------
Top 1 modules in terms of params, MACs or fwd latency at different model depths:
depth 0:
    params      - {'PoolFormerModel': '11.4 M'}
    MACs        - {'PoolFormerModel': '42.59 GMACs'}
    fwd latency - {'PoolFormerModel': '99.9 ms'}
depth 1:
    params      - {'PoolFormerEncoder': '11.4 M'}
    MACs        - {'PoolFormerEncoder': '42.59 GMACs'}
    fwd latency - {'PoolFormerEncoder': '99.74 ms'}
depth 2:
    params      - {'ModuleList': '11.4 M'}
    MACs        - {'ModuleList': '42.59 GMACs'}
    fwd latency - {'ModuleList': '1.8 ms'}
depth 3:
    params      - {'ModuleList': '9.47 M'}
    MACs        - {'ModuleList': '37.14 GMACs'}
    fwd latency - {'ModuleList': '97.18 ms'}
depth 4:
    params      - {'PoolFormerLayer': '9.47 M'}
    MACs        - {'PoolFormerLayer': '37.14 GMACs'}
    fwd latency - {'PoolFormerLayer': '97.18 ms'}
depth 5:
    params      - {'PoolFormerOutput': '9.45 M'}
    MACs        - {'PoolFormerOutput': '37.14 GMACs'}
    fwd latency - {'PoolFormerGroupNorm': '47.05 ms'}

------------------------------ Detailed Profile per GPU
------------------------------
```

```
Each module profile is listed after its name in the following order:
params, percentage of total params, MACs, percentage of total MACs, fwd latency,
percentage of total fwd latency, fwd FLOPS

Note: 1. A module can have torch.nn.module or torch.nn.functional to compute logits
(e.g. CrossEntropyLoss). They are not counted as submodules, thus not to be printed
out. However they make up the difference between a parent's MACs (or latency) and
the sum of its submodules'.
2. Number of floating-point operations is a theoretical estimation, thus FLOPS
computed using that could be larger than the maximum system throughput.
3. The fwd latency listed in the top module's profile is directly captured at the
module forward function in PyTorch, thus it's less than the fwd latency shown above
which is captured in DeepSpeed.

PoolFormerModel(
  11.4 M, 100.00% Params, 42.59 GMACs, 100.00% MACs, 99.9 ms, 100.00% latency,
856.74 GFLOPS,
  (encoder): PoolFormerEncoder(
    11.4 M, 100.00% Params, 42.59 GMACs, 100.00% MACs, 99.74 ms, 99.85% latency,
858.06 GFLOPS,
    (patch_embeddings): ModuleList(
      (0): PoolFormerEmbeddings(
        9.47 k, 0.08% Params, 693.63 MMACs, 1.63% MACs, 480.65 us, 0.48% latency,
2.9 TFLOPS,
        (projection): Conv2d(9.47 k, 0.08% Params, 693.63 MMACs, 1.63% MACs, 334.74
us, 0.34% latency, 4.16 TFLOPS, 3, 64, kernel_size=(7, 7), stride=(4, 4),
padding=(2, 2))
        (norm): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.52 us, 0.04%
latency, 0.0 FLOPS, )
      )
      (1): PoolFormerEmbeddings(
        73.86 k, 0.65% Params, 1.36 GMACs, 3.19% MACs, 407.46 us, 0.41% latency,
6.68 TFLOPS,
        (projection): Conv2d(73.86 k, 0.65% Params, 1.36 GMACs, 3.19% MACs, 258.92
us, 0.26% latency, 10.51 TFLOPS, 64, 128, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1))
        (norm): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.0 us, 0.04%
latency, 0.0 FLOPS, )
      )
      (2): PoolFormerEmbeddings(
        368.96 k, 3.24% Params, 1.7 GMACs, 3.99% MACs, 404.36 us, 0.40% latency,
8.41 TFLOPS,
        (projection): Conv2d(368.96 k, 3.24% Params, 1.7 GMACs, 3.99% MACs, 252.01
us, 0.25% latency, 13.49 TFLOPS, 128, 320, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1))
        (norm): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.95 us, 0.04%
latency, 0.0 FLOPS, )
      )
      (3): PoolFormerEmbeddings(
        1.48 M, 12.94% Params, 1.7 GMACs, 3.99% MACs, 502.82 us, 0.50% latency, 6.76
```

```
TFLOPS,
        (projection): Conv2d(1.48 M, 12.94% Params, 1.7 GMACs, 3.99% MACs, 335.22
us, 0.34% latency, 10.14 TFLOPS, 320, 512, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1))
        (norm): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 37.67 us, 0.04%
latency, 0.0 FLOPS, )
      )
    )
    (block): ModuleList(
      (0): ModuleList(
        (0): PoolFormerLayer(
          33.47 k, 0.29% Params, 2.42 GMACs, 5.67% MACs, 8.18 ms, 8.19% latency,
599.8 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 531.2 us, 0.53% latency, 8.88
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 190.73 us, 0.19%
latency, 24.74 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            33.09 k, 0.29% Params, 2.42 GMACs, 5.67% MACs, 2.16 ms, 2.16% latency,
2.25 TFLOPS,
            (conv1): Conv2d(16.64 k, 0.15% Params, 1.21 GMACs, 2.84% MACs, 252.01
us, 0.25% latency, 9.66 TFLOPS, 64, 256, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(16.45 k, 0.14% Params, 1.21 GMACs, 2.84% MACs, 293.49
us, 0.29% latency, 8.25 TFLOPS, 256, 64, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 75.34
us, 0.08% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.3 ms,
1.30% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(128, 0.00% Params, 0 MACs, 0.00% MACs,
1.93 ms, 1.93% latency, 12.24 GFLOPS, 1, 64, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(128, 0.00% Params, 0 MACs, 0.00% MACs,
1.98 ms, 1.99% latency, 11.89 GFLOPS, 1, 64, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 91.79 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (1): PoolFormerLayer(
          33.47 k, 0.29% Params, 2.42 GMACs, 5.67% MACs, 7.9 ms, 7.91% latency,
621.24 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 588.89 us, 0.59% latency, 8.01
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 187.4 us, 0.19%
latency, 25.18 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            33.09 k, 0.29% Params, 2.42 GMACs, 5.67% MACs, 1.91 ms, 1.91% latency,
2.54 TFLOPS,
```

```
            (conv1): Conv2d(16.64 k, 0.15% Params, 1.21 GMACs, 2.84% MACs, 273.23
us, 0.27% latency, 8.91 TFLOPS, 64, 256, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(16.45 k, 0.14% Params, 1.21 GMACs, 2.84% MACs, 237.94
us, 0.24% latency, 10.17 TFLOPS, 256, 64, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 73.43
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.1 ms,
1.10% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(128, 0.00% Params, 0 MACs, 0.00% MACs,
1.95 ms, 1.95% latency, 12.08 GFLOPS, 1, 64, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(128, 0.00% Params, 0 MACs, 0.00% MACs,
1.9 ms, 1.91% latency, 12.4 GFLOPS, 1, 64, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 88.93 us, 0.09%
latency, 0.0 FLOPS, )
        )
      )
      (1): ModuleList(
        (0): PoolFormerLayer(
          132.48 k, 1.16% Params, 2.42 GMACs, 5.67% MACs, 7.91 ms, 7.92% latency,
615.85 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 591.75 us, 0.59% latency, 3.99
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 206.47 us, 0.21%
latency, 11.43 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            131.71 k, 1.16% Params, 2.42 GMACs, 5.67% MACs, 1.9 ms, 1.90% latency,
2.55 TFLOPS,
            (conv1): Conv2d(66.05 k, 0.58% Params, 1.21 GMACs, 2.84% MACs, 242.71
us, 0.24% latency, 9.99 TFLOPS, 128, 512, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(65.66 k, 0.58% Params, 1.21 GMACs, 2.84% MACs, 241.52
us, 0.24% latency, 10.01 TFLOPS, 512, 128, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 72.0 us,
0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.09 ms,
1.09% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
1.97 ms, 1.97% latency, 5.99 GFLOPS, 1, 128, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
1.88 ms, 1.89% latency, 6.26 GFLOPS, 1, 128, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 91.79 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (1): PoolFormerLayer(
          132.48 k, 1.16% Params, 2.42 GMACs, 5.67% MACs, 8.39 ms, 8.40% latency,
580.59 GFLOPS,
          (pooling): PoolFormerPooling(
```

```
            0, 0.00% Params, 0 MACs, 0.00% MACs, 666.62 us, 0.67% latency, 3.54
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 297.31 us, 0.30%
latency, 7.94 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            131.71 k, 1.16% Params, 2.42 GMACs, 5.67% MACs, 2.09 ms, 2.09% latency,
2.32 TFLOPS,
            (conv1): Conv2d(66.05 k, 0.58% Params, 1.21 GMACs, 2.84% MACs, 240.09
us, 0.24% latency, 10.1 TFLOPS, 128, 512, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(65.66 k, 0.58% Params, 1.21 GMACs, 2.84% MACs, 248.19
us, 0.25% latency, 9.74 TFLOPS, 512, 128, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 78.2 us,
0.08% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.26 ms,
1.26% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
2.01 ms, 2.01% latency, 5.88 GFLOPS, 1, 128, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
1.98 ms, 1.99% latency, 5.95 GFLOPS, 1, 128, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 101.8 us, 0.10%
latency, 0.0 FLOPS, )
        )
      )
      (2): ModuleList(
        (0): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 8.19 ms, 8.20% latency,
925.09 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 555.28 us, 0.56% latency, 2.66
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 192.17 us, 0.19%
latency, 7.67 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 2.03 ms, 2.04% latency,
3.71 TFLOPS,
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 269.65
us, 0.27% latency, 14.02 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 286.34
us, 0.29% latency, 13.19 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 74.15
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.15 ms,
1.15% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.91 ms, 1.91% latency, 3.86 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
```

```
2.15 ms, 2.15% latency, 3.44 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 91.79 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (1): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 7.8 ms, 7.81% latency,
971.2 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 560.05 us, 0.56% latency, 2.63
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 222.21 us, 0.22%
latency, 6.64 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 1.93 ms, 1.93% latency,
3.92 TFLOPS,
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 238.18
us, 0.24% latency, 15.87 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 240.8
us, 0.24% latency, 15.68 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 105.38
us, 0.11% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.11 ms,
1.12% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.86 ms, 1.86% latency, 3.96 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.92 ms, 1.92% latency, 3.85 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 89.17 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (2): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 7.98 ms, 7.99% latency,
949.25 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 595.57 us, 0.60% latency, 2.48
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 236.75 us, 0.24%
latency, 6.23 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 1.98 ms, 1.98% latency,
3.82 TFLOPS,
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 261.78
us, 0.26% latency, 14.44 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 242.71
us, 0.24% latency, 15.56 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 73.19
us, 0.07% latency, 0.0 FLOPS, p=0.0)
```

```
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.15 ms,
1.15% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.88 ms, 1.88% latency, 3.92 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
2.0 ms, 2.01% latency, 3.68 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 91.08 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (3): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 8.19 ms, 8.20% latency,
925.04 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 712.39 us, 0.71% latency, 2.07
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 228.4 us, 0.23%
latency, 6.46 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 2.0 ms, 2.00% latency,
3.78 TFLOPS,
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 239.13
us, 0.24% latency, 15.81 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 289.2
us, 0.29% latency, 13.06 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 75.82
us, 0.08% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.16 ms,
1.16% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.89 ms, 1.89% latency, 3.9 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.94 ms, 1.94% latency, 3.8 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 86.78 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (4): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 8.14 ms, 8.15% latency,
930.48 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 530.48 us, 0.53% latency, 2.78
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 191.93 us, 0.19%
latency, 7.68 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 1.99 ms, 2.00% latency,
3.79 TFLOPS,
```

```
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 254.87
us, 0.26% latency, 14.83 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 242.95
us, 0.24% latency, 15.54 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 71.76
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.19 ms,
1.19% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.95 ms, 1.95% latency, 3.79 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
2.05 ms, 2.06% latency, 3.59 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 99.9 us, 0.10%
latency, 0.0 FLOPS, )
        )
        (5): PoolFormerLayer(
          822.72 k, 7.22% Params, 3.77 GMACs, 8.86% MACs, 8.08 ms, 8.08% latency,
937.85 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 552.65 us, 0.55% latency, 2.67
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 197.65 us, 0.20%
latency, 7.46 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            820.8 k, 7.20% Params, 3.77 GMACs, 8.86% MACs, 2.04 ms, 2.04% latency,
3.71 TFLOPS,
            (conv1): Conv2d(410.88 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 252.72
us, 0.25% latency, 14.96 TFLOPS, 320, 1280, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(409.92 k, 3.60% Params, 1.89 GMACs, 4.43% MACs, 239.85
us, 0.24% latency, 15.74 TFLOPS, 1280, 320, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 71.76
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.2 ms,
1.21% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
1.91 ms, 1.91% latency, 3.86 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(640, 0.01% Params, 0 MACs, 0.00% MACs,
2.01 ms, 2.01% latency, 3.67 GFLOPS, 1, 320, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 118.02 us,
0.12% latency, 0.0 FLOPS, )
        )
      )
      (3): ModuleList(
        (0): PoolFormerLayer(
          2.1 M, 18.44% Params, 2.42 GMACs, 5.67% MACs, 8.13 ms, 8.14% latency,
595.36 GFLOPS,
          (pooling): PoolFormerPooling(
```

```
            0, 0.00% Params, 0 MACs, 0.00% MACs, 555.28 us, 0.56% latency, 1.06
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 221.73 us, 0.22%
latency, 2.66 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            2.1 M, 18.42% Params, 2.42 GMACs, 5.67% MACs, 2.03 ms, 2.03% latency,
2.38 TFLOPS,
            (conv1): Conv2d(1.05 M, 9.22% Params, 1.21 GMACs, 2.84% MACs, 269.41 us,
0.27% latency, 8.98 TFLOPS, 512, 2048, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(1.05 M, 9.20% Params, 1.21 GMACs, 2.84% MACs, 272.51 us,
0.27% latency, 8.87 TFLOPS, 2048, 512, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 72.72
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.17 ms,
1.17% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(1.02 k, 0.01% Params, 0 MACs, 0.00%
MACs, 1.96 ms, 1.96% latency, 1.5 GFLOPS, 1, 512, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(1.02 k, 0.01% Params, 0 MACs, 0.00%
MACs, 1.98 ms, 1.98% latency, 1.49 GFLOPS, 1, 512, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 88.69 us, 0.09%
latency, 0.0 FLOPS, )
        )
        (1): PoolFormerLayer(
          2.1 M, 18.44% Params, 2.42 GMACs, 5.67% MACs, 8.31 ms, 8.31% latency,
582.85 GFLOPS,
          (pooling): PoolFormerPooling(
            0, 0.00% Params, 0 MACs, 0.00% MACs, 528.34 us, 0.53% latency, 1.12
GFLOPS,
            (pool): AvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 190.26 us, 0.19%
latency, 3.1 GFLOPS, kernel_size=3, stride=1, padding=1)
          )
          (output): PoolFormerOutput(
            2.1 M, 18.42% Params, 2.42 GMACs, 5.67% MACs, 2.14 ms, 2.14% latency,
2.26 TFLOPS,
            (conv1): Conv2d(1.05 M, 9.22% Params, 1.21 GMACs, 2.84% MACs, 285.39 us,
0.29% latency, 8.47 TFLOPS, 512, 2048, kernel_size=(1, 1), stride=(1, 1))
            (conv2): Conv2d(1.05 M, 9.20% Params, 1.21 GMACs, 2.84% MACs, 256.06 us,
0.26% latency, 9.44 TFLOPS, 2048, 512, kernel_size=(1, 1), stride=(1, 1))
            (drop): PoolFormerDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 73.67
us, 0.07% latency, 0.0 FLOPS, p=0.0)
            (act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.27 ms,
1.27% latency, 0.0 FLOPS, )
          )
          (before_norm): PoolFormerGroupNorm(1.02 k, 0.01% Params, 0 MACs, 0.00%
MACs, 1.9 ms, 1.90% latency, 1.56 GFLOPS, 1, 512, eps=1e-05, affine=True)
          (after_norm): PoolFormerGroupNorm(1.02 k, 0.01% Params, 0 MACs, 0.00%
MACs, 2.14 ms, 2.14% latency, 1.38 GFLOPS, 1, 512, eps=1e-05, affine=True)
          (drop_path): Identity(0, 0.00% Params, 0 MACs, 0.00% MACs, 93.46 us, 0.09%
```

```
latency, 0.0 FLOPS, )
        )
      )
    )
  )
)
------------------------------------------------------------------------------
```