

----- 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: | 86.74 M |
| params of model = params per GPU * mp_size: | 86.74 M |
| fwd MACs per GPU: | 241.82 GMACs |
| fwd flops per GPU: | 483.98 G |
| fwd flops of model = fwd flops per GPU * mp_size: | 483.98 G |
| fwd latency: | 248.71 ms |
| fwd FLOPS per GPU = fwd flops per GPU / fwd latency: | 1.95 TFLOPS |

----- Aggregated Profile per GPU -----

Top 1 modules in terms of params, MACs or fwd latency at different model depths:

depth 0:

| | |
|-------------|---------------------------------|
| params | - {'SwinModel': '86.74 M'} |
| MACs | - {'SwinModel': '241.82 GMACs'} |
| fwd latency | - {'SwinModel': '248.71 ms'} |

depth 1:

| | |
|-------------|-----------------------------------|
| params | - {'SwinEncoder': '86.73 M'} |
| MACs | - {'SwinEncoder': '241.67 GMACs'} |
| fwd latency | - {'SwinEncoder': '245.32 ms'} |

depth 2:

| | |
|-------------|----------------------------------|
| params | - {'ModuleList': '86.73 M'} |
| MACs | - {'ModuleList': '241.67 GMACs'} |
| fwd latency | - {'ModuleList': '245.11 ms'} |

depth 3:

| | |
|-------------|---------------------------------|
| params | - {'SwinStage': '86.73 M'} |
| MACs | - {'SwinStage': '241.67 GMACs'} |
| fwd latency | - {'SwinStage': '245.11 ms'} |

depth 4:

| | |
|-------------|---------------------------------|
| params | - {'ModuleList': '83.97 M'} |
| MACs | - {'ModuleList': '239.2 GMACs'} |
| fwd latency | - {'ModuleList': '239.38 ms'} |

depth 5:

| | |
|-------------|--------------------------------|
| params | - {'SwinLayer': '83.97 M'} |
| MACs | - {'SwinLayer': '239.2 GMACs'} |
| fwd latency | - {'SwinLayer': '239.38 ms'} |

depth 6:

| | |
|--------|------------------------------------|
| params | - {'SwinAttention': '28.03 M'} |
| MACs | - {'SwinAttention': '81.36 GMACs'} |

```

    fwd latency - {'SwinAttention': '106.82 ms'}
depth 7:
    params      - {'Linear': '55.9 M'}
    MACs        - {'Linear': '157.84 GMACs'}
    fwd latency - {'SwinSelfAttention': '87.4 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.

```

SwinModel(
  86.74 M, 100.00% Params, 241.82 GMACs, 100.00% MACs, 248.71 ms, 100.00% latency,
  1.95 TFLOPS,
  (embeddings): SwinEmbeddings(
    6.53 k, 0.01% Params, 154.14 MMACs, 0.06% MACs, 1.64 ms, 0.66% latency, 199.92
    GFLOPS,
    (patch_embeddings): SwinPatchEmbeddings(
      6.27 k, 0.01% Params, 154.14 MMACs, 0.06% MACs, 317.57 us, 0.13% latency,
      980.85 GFLOPS,
      (projection): Conv2d(6.27 k, 0.01% Params, 154.14 MMACs, 0.06% MACs, 229.84
      us, 0.09% latency, 1.36 TFLOPS, 3, 128, kernel_size=(4, 4), stride=(4, 4))
    )
    (norm): LayerNorm(256, 0.00% Params, 0 MACs, 0.00% MACs, 1.15 ms, 0.46% latency,
    13.93 GFLOPS, (128,), eps=1e-05, elementwise_affine=True)
    (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 59.6 us, 0.02% latency,
    0.0 FLOPS, p=0.0, inplace=False)
  )
  (encoder): SwinEncoder(
    86.73 M, 99.99% Params, 241.67 GMACs, 99.94% MACs, 245.32 ms, 98.63% latency,
    1.97 TFLOPS,
    (layers): ModuleList(
      (0): SwinStage(
        529.99 k, 0.61% Params, 21.18 GMACs, 8.76% MACs, 23.28 ms, 9.36% latency,
        1.82 TFLOPS,
        (blocks): ModuleList(
          (0): SwinLayer(
            198.95 k, 0.23% Params, 10.18 GMACs, 4.21% MACs, 10.23 ms, 4.11%
            latency, 1.99 TFLOPS,

```

```

        (layernorm_before): LayerNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
1.22 ms, 0.49% latency, 13.18 GFLOPS, (128,)), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            66.72 k, 0.08% Params, 3.6 GMACs, 1.49% MACs, 5.26 ms, 2.11% latency,
1.37 TFLOPS,
            (self): SwinSelfAttention(
                50.21 k, 0.06% Params, 2.78 GMACs, 1.15% MACs, 4.4 ms, 1.77%
latency, 1.26 TFLOPS,
                (query): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
605.82 us, 0.24% latency, 2.71 TFLOPS, in_features=128, out_features=128, bias=True)
                (key): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
555.04 us, 0.22% latency, 2.96 TFLOPS, in_features=128, out_features=128, bias=True)
                (value): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
637.77 us, 0.26% latency, 2.58 TFLOPS, in_features=128, out_features=128, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.15 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs, 763.18 us, 0.31%
latency, 2.15 TFLOPS,
                (dense): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
654.94 us, 0.26% latency, 2.51 TFLOPS, in_features=128, out_features=128, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.05 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
        )
        (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 27.66 us,
0.01% latency, 0.0 FLOPS, p=0.1)
        (layernorm_after): LayerNorm(256, 0.00% Params, 0 MACs, 0.00% MACs, 1.06
ms, 0.43% latency, 15.19 GFLOPS, (128,)), eps=1e-05, elementwise_affine=True)
        (intermediate): SwinIntermediate(
            66.05 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 1.25 ms, 0.50% latency,
5.26 TFLOPS,
            (dense): Linear(66.05 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 550.99
us, 0.22% latency, 11.94 TFLOPS, in_features=128, out_features=512, bias=True)
            (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 621.32 us, 0.25% latency, 0.0 FLOPS, )
        )
        (output): SwinOutput(
            65.66 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 636.34 us, 0.26%
latency, 10.34 TFLOPS,
            (dense): Linear(65.66 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 534.77
us, 0.22% latency, 12.3 TFLOPS, in_features=512, out_features=128, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.42 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
    )
    (1): SwinLayer(
        198.95 k, 0.23% Params, 10.18 GMACs, 4.21% MACs, 11.34 ms, 4.56%
latency, 1.8 TFLOPS,

```

```

        (layernorm_before): LayerNorm(256, 0.00% Params, 0 MACs, 0.00% MACs,
1.02 ms, 0.41% latency, 15.76 GFLOPS, (128,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            66.72 k, 0.08% Params, 3.6 GMACs, 1.49% MACs, 4.43 ms, 1.78% latency,
1.63 TFLOPS,
            (self): SwinSelfAttention(
                50.21 k, 0.06% Params, 2.78 GMACs, 1.15% MACs, 3.61 ms, 1.45%
latency, 1.54 TFLOPS,
                (query): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
612.02 us, 0.25% latency, 2.69 TFLOPS, in_features=128, out_features=128, bias=True)
                (key): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
582.46 us, 0.23% latency, 2.82 TFLOPS, in_features=128, out_features=128, bias=True)
                (value): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
556.95 us, 0.22% latency, 2.95 TFLOPS, in_features=128, out_features=128, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 42.92 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs, 740.53 us, 0.30%
latency, 2.22 TFLOPS,
                (dense): Linear(16.51 k, 0.02% Params, 822.08 MMACs, 0.34% MACs,
625.61 us, 0.25% latency, 2.63 TFLOPS, in_features=128, out_features=128, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 45.3 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
        )
        (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.23 us,
0.01% latency, 0.0 FLOPS, p=0.1)
        (layernorm_after): LayerNorm(256, 0.00% Params, 0 MACs, 0.00% MACs, 1.13
ms, 0.45% latency, 14.23 GFLOPS, (128,), eps=1e-05, elementwise_affine=True)
        (intermediate): SwinIntermediate(
            66.05 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 1.27 ms, 0.51% latency,
5.18 TFLOPS,
            (dense): Linear(66.05 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 589.85
us, 0.24% latency, 11.15 TFLOPS, in_features=128, out_features=512, bias=True)
            (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 601.53 us, 0.24% latency, 0.0 FLOPS, )
        )
        (output): SwinOutput(
            65.66 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 711.2 us, 0.29%
latency, 9.25 TFLOPS,
            (dense): Linear(65.66 k, 0.08% Params, 3.29 GMACs, 1.36% MACs, 601.53
us, 0.24% latency, 10.93 TFLOPS, in_features=512, out_features=128, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.72 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
    )
    (downsample): SwinPatchMerging(
        132.1 k, 0.15% Params, 822.08 MMACs, 0.34% MACs, 1.61 ms, 0.65% latency,

```

```

1.03 TFLOPS,
    (reduction): Linear(131.07 k, 0.15% Params, 822.08 MMACs, 0.34% MACs,
160.22 us, 0.06% latency, 10.26 TFLOPS, in_features=512, out_features=256,
bias=False)
    (norm): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs, 1.05 ms, 0.42%
latency, 15.36 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    )
    )
    (1): SwinStage(
    2.11 M, 2.43% Params, 20.87 GMACs, 8.63% MACs, 21.05 ms, 8.46% latency, 1.98
TFLOPS,
    (blocks): ModuleList(
    (0): SwinLayer(
    791.11 k, 0.91% Params, 10.02 GMACs, 4.14% MACs, 8.8 ms, 3.54% latency,
2.28 TFLOPS,
    (layernorm_before): LayerNorm(512, 0.00% Params, 0 MACs, 0.00% MACs,
1.04 ms, 0.42% latency, 7.69 GFLOPS, (256,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
    264.52 k, 0.30% Params, 3.45 GMACs, 1.42% MACs, 4.02 ms, 1.62%
latency, 1.71 TFLOPS,
    (self): SwinSelfAttention(
    198.73 k, 0.23% Params, 2.62 GMACs, 1.08% MACs, 3.29 ms, 1.32%
latency, 1.59 TFLOPS,
    (query): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
550.75 us, 0.22% latency, 2.99 TFLOPS, in_features=256, out_features=256, bias=True)
    (key): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
505.92 us, 0.20% latency, 3.25 TFLOPS, in_features=256, out_features=256, bias=True)
    (value): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
535.73 us, 0.22% latency, 3.07 TFLOPS, in_features=256, out_features=256, bias=True)
    (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.39 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (output): SwinSelfOutput(
    65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs, 651.6 us, 0.26%
latency, 2.52 TFLOPS,
    (dense): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
552.18 us, 0.22% latency, 2.98 TFLOPS, in_features=256, out_features=256, bias=True)
    (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.66 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 24.08 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(512, 0.00% Params, 0 MACs, 0.00% MACs, 1.04
ms, 0.42% latency, 7.68 GFLOPS, (256,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
    263.17 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 1.24 ms, 0.50%
latency, 5.32 TFLOPS,
    (dense): Linear(263.17 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 560.28
us, 0.23% latency, 11.74 TFLOPS, in_features=256, out_features=1024, bias=True)

```

```

        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 599.38 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        262.4 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 661.13 us, 0.27%
latency, 9.95 TFLOPS,
        (dense): Linear(262.4 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 558.61
us, 0.22% latency, 11.77 TFLOPS, in_features=1024, out_features=256, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.57 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(1): SwinLayer(
    791.11 k, 0.91% Params, 10.02 GMACs, 4.14% MACs, 10.54 ms, 4.24%
latency, 1.9 TFLOPS,
    (layernorm_before): LayerNorm(512, 0.00% Params, 0 MACs, 0.00% MACs,
1.05 ms, 0.42% latency, 7.66 GFLOPS, (256,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        264.52 k, 0.30% Params, 3.45 GMACs, 1.42% MACs, 4.33 ms, 1.74%
latency, 1.59 TFLOPS,
        (self): SwinSelfAttention(
            198.73 k, 0.23% Params, 2.62 GMACs, 1.08% MACs, 3.54 ms, 1.42%
latency, 1.48 TFLOPS,
            (query): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
581.26 us, 0.23% latency, 2.83 TFLOPS, in_features=256, out_features=256, bias=True)
            (key): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
542.88 us, 0.22% latency, 3.03 TFLOPS, in_features=256, out_features=256, bias=True)
            (value): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
555.52 us, 0.22% latency, 2.96 TFLOPS, in_features=256, out_features=256, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 42.92 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs, 704.77 us, 0.28%
latency, 2.33 TFLOPS,
            (dense): Linear(65.79 k, 0.08% Params, 822.08 MMACs, 0.34% MACs,
598.43 us, 0.24% latency, 2.75 TFLOPS, in_features=256, out_features=256, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.57 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.05 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(512, 0.00% Params, 0 MACs, 0.00% MACs, 1.12
ms, 0.45% latency, 7.16 GFLOPS, (256,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        263.17 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 1.27 ms, 0.51%
latency, 5.18 TFLOPS,
        (dense): Linear(263.17 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 607.97
us, 0.24% latency, 10.82 TFLOPS, in_features=256, out_features=1024, bias=True)
    )
)

```

```

        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 582.46 us, 0.23% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        262.4 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 676.87 us, 0.27%
latency, 9.72 TFLOPS,
        (dense): Linear(262.4 k, 0.30% Params, 3.29 GMACs, 1.36% MACs, 572.44
us, 0.23% latency, 11.49 TFLOPS, in_features=1024, out_features=256, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.09 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
)
    (downsample): SwinPatchMerging(
        526.34 k, 0.61% Params, 822.08 MMACs, 0.34% MACs, 1.61 ms, 0.65% latency,
1.03 TFLOPS,
        (reduction): Linear(524.29 k, 0.60% Params, 822.08 MMACs, 0.34% MACs,
149.97 us, 0.06% latency, 10.96 TFLOPS, in_features=1024, out_features=512,
bias=False)
        (norm): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs, 1.05 ms, 0.42%
latency, 7.66 GFLOPS, (1024,), eps=1e-05, elementwise_affine=True)
    )
)
    (2): SwinStage(
        58.89 M, 67.89% Params, 179.81 GMACs, 74.36% MACs, 181.43 ms, 72.95%
latency, 1.98 TFLOPS,
        (blocks): ModuleList(
            (0): SwinLayer(
                3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.09 ms, 3.65% latency,
2.19 TFLOPS,
                (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.07 ms, 0.43% latency, 3.74 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
                (attention): SwinAttention(
                    1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.19 ms, 1.68% latency,
1.61 TFLOPS,
                    (self): SwinSelfAttention(
                        790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.46 ms, 1.39%
latency, 1.47 TFLOPS,
                        (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
583.89 us, 0.23% latency, 2.82 TFLOPS, in_features=512, out_features=512, bias=True)
                        (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
527.14 us, 0.21% latency, 3.12 TFLOPS, in_features=512, out_features=512, bias=True)
                        (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
564.34 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
                        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 42.92 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
                    )
                    (output): SwinSelfOutput(
                        262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 652.55 us, 0.26%
latency, 2.52 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
547.17 us, 0.22% latency, 3.0 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.33 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 27.18 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.11 ms, 0.44% latency, 3.63 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.24 ms, 0.50% latency,
5.3 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 545.26
us, 0.22% latency, 12.06 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 618.46 us, 0.25% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 687.84 us, 0.28%
latency, 9.56 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 583.41
us, 0.23% latency, 11.27 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.33 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (1): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.19 ms, 4.10% latency,
1.95 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
996.59 us, 0.40% latency, 4.03 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.38 ms, 1.76% latency,
1.54 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.65 ms, 1.47%
latency, 1.4 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
638.25 us, 0.26% latency, 2.58 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
551.22 us, 0.22% latency, 2.98 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
579.83 us, 0.23% latency, 2.84 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 40.53 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 619.65 us, 0.25%
latency, 2.65 TFLOPS,

```



```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
521.18 us, 0.21% latency, 3.15 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.95 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 29.56 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
979.9 us, 0.39% latency, 4.1 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.17 ms, 0.47% latency,
5.62 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 534.53
us, 0.21% latency, 12.3 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 559.33 us, 0.22% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 741.96 us, 0.30%
latency, 8.86 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 613.45
us, 0.25% latency, 10.72 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.81 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(2): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.36 ms, 3.76% latency,
2.13 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.11 ms, 0.44% latency, 3.63 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.35 ms, 1.75% latency,
1.55 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.44 ms, 1.38%
latency, 1.48 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
554.56 us, 0.22% latency, 2.96 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
597.48 us, 0.24% latency, 2.75 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
502.35 us, 0.20% latency, 3.27 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.25 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 821.59 us, 0.33%
latency, 2.0 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
709.53 us, 0.29% latency, 2.32 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.0 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 38.15 us,
0.02% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.06 ms, 0.43% latency, 3.79 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.32 ms, 0.53% latency,
4.99 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 583.65
us, 0.23% latency, 11.27 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 648.02 us, 0.26% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 653.51 us, 0.26%
latency, 10.06 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 548.12
us, 0.22% latency, 12.0 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.05 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (3): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.72 ms, 4.31% latency,
1.86 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.01 ms, 0.41% latency, 3.97 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.33 ms, 1.74% latency,
1.55 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.58 ms, 1.44%
latency, 1.42 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
597.48 us, 0.24% latency, 2.75 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
565.05 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
530.24 us, 0.21% latency, 3.1 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 41.72 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 666.14 us, 0.27%
latency, 2.47 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
565.05 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.9 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.23 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.22 ms, 0.49% latency, 3.29 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.34 ms, 0.54% latency,
4.92 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 543.12
us, 0.22% latency, 12.11 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 715.26 us, 0.29% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 932.46 us, 0.37%
latency, 7.05 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 819.92
us, 0.33% latency, 8.02 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.48 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(4): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 8.86 ms, 3.56% latency,
2.25 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.11 ms, 0.45% latency, 3.62 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.09 ms, 1.64% latency,
1.65 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.35 ms, 1.35%
latency, 1.52 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
571.01 us, 0.23% latency, 2.88 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
564.1 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
582.93 us, 0.23% latency, 2.82 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 40.77 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 658.99 us, 0.26%
latency, 2.49 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
558.38 us, 0.22% latency, 2.94 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.9 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 24.08 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
986.1 us, 0.40% latency, 4.07 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.22 ms, 0.49% latency,
5.4 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 560.76
us, 0.23% latency, 11.73 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 582.46 us, 0.23% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 692.13 us, 0.28%
latency, 9.5 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 590.09
us, 0.24% latency, 11.15 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 33.62 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(5): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.32 ms, 4.15% latency,
1.93 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
989.68 us, 0.40% latency, 4.06 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.48 ms, 1.80% latency,
1.5 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.73 ms, 1.50%
latency, 1.37 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
674.49 us, 0.27% latency, 2.44 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
554.56 us, 0.22% latency, 2.96 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
590.32 us, 0.24% latency, 2.79 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 40.77 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 666.62 us, 0.27%
latency, 2.47 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
564.58 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.19 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 29.8 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
990.63 us, 0.40% latency, 4.05 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.27 ms, 0.51% latency,
5.18 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 610.11
us, 0.25% latency, 10.78 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 586.03 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 663.52 us, 0.27%
latency, 9.91 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 559.81
us, 0.23% latency, 11.75 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 33.86 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (6): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.41 ms, 3.79% latency,
2.11 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.02 ms, 0.41% latency, 3.95 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.45 ms, 1.79% latency,
1.51 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.57 ms, 1.44%
latency, 1.43 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
525.0 us, 0.21% latency, 3.13 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
609.4 us, 0.25% latency, 2.7 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
544.31 us, 0.22% latency, 3.02 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 44.11 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 797.99 us, 0.32%
latency, 2.06 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
688.31 us, 0.28% latency, 2.39 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.0 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 26.46 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.08 ms, 0.44% latency, 3.71 GFLOPS, (512,)), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.4 ms, 0.56% latency,
4.69 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 687.12
us, 0.28% latency, 9.57 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 622.27 us, 0.25% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 670.43 us, 0.27%
latency, 9.81 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 570.54
us, 0.23% latency, 11.53 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.42 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(7): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.06 ms, 4.05% latency,
1.98 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
998.26 us, 0.40% latency, 4.02 GFLOPS, (512,)), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.41 ms, 1.77% latency,
1.53 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.63 ms, 1.46%
latency, 1.4 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
567.91 us, 0.23% latency, 2.9 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
520.23 us, 0.21% latency, 3.16 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
577.21 us, 0.23% latency, 2.85 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 40.77 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 694.27 us, 0.28%
latency, 2.37 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
576.73 us, 0.23% latency, 2.85 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.05 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 29.56 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
983.48 us, 0.40% latency, 4.08 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.18 ms, 0.47% latency,
5.58 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 550.51
us, 0.22% latency, 11.95 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 551.94 us, 0.22% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 647.31 us, 0.26%
latency, 10.16 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 548.6
us, 0.22% latency, 11.99 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.95 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(8): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 8.91 ms, 3.58% latency,
2.23 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.11 ms, 0.44% latency, 3.63 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.07 ms, 1.64% latency,
1.65 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.33 ms, 1.34%
latency, 1.53 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
559.33 us, 0.22% latency, 2.94 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
575.54 us, 0.23% latency, 2.86 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
514.03 us, 0.21% latency, 3.2 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 40.77 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 654.94 us, 0.26%
latency, 2.51 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
522.85 us, 0.21% latency, 3.14 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 47.45 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 25.03 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
985.62 us, 0.40% latency, 4.07 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.24 ms, 0.50% latency,
5.3 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 543.83
us, 0.22% latency, 12.09 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 603.2 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 691.89 us, 0.28%
latency, 9.51 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 576.97
us, 0.23% latency, 11.4 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.01 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (9): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.83 ms, 4.35% latency,
1.84 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
989.44 us, 0.40% latency, 4.06 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.54 ms, 1.83% latency,
1.48 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.79 ms, 1.52%
latency, 1.34 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
622.99 us, 0.25% latency, 2.64 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
506.64 us, 0.20% latency, 3.25 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
536.44 us, 0.22% latency, 3.06 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 56.03 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 667.33 us, 0.27%
latency, 2.46 TFLOPS,

```



```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
552.42 us, 0.22% latency, 2.98 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 42.92 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.47 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.09 ms, 0.44% latency, 3.67 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.31 ms, 0.53% latency,
5.01 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 581.03
us, 0.23% latency, 11.32 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 634.43 us, 0.26% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.03 ms, 0.41% latency,
6.39 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 922.2
us, 0.37% latency, 7.13 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.33 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (10): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 8.7 ms, 3.50% latency,
2.29 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.1 ms, 0.44% latency, 3.65 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 3.92 ms, 1.58% latency,
1.72 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.18 ms, 1.28%
latency, 1.6 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
548.12 us, 0.22% latency, 3.0 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
498.29 us, 0.20% latency, 3.3 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
539.54 us, 0.22% latency, 3.05 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 41.01 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 661.61 us, 0.27%
latency, 2.49 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
540.73 us, 0.22% latency, 3.04 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.9 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 24.32 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.0 ms, 0.40% latency, 4.01 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.23 ms, 0.49% latency,
5.34 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 544.07
us, 0.22% latency, 12.09 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 608.92 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 645.64 us, 0.26%
latency, 10.19 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 542.16
us, 0.22% latency, 12.13 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 33.38 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(11): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.1 ms, 4.06% latency,
1.97 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.03 ms, 0.42% latency, 3.88 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.35 ms, 1.75% latency,
1.55 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.57 ms, 1.43%
latency, 1.43 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
561.71 us, 0.23% latency, 2.93 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
541.45 us, 0.22% latency, 3.04 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
548.36 us, 0.22% latency, 3.0 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.87 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 698.57 us, 0.28%
latency, 2.35 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
594.85 us, 0.24% latency, 2.76 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 32.66 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 29.8 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
983.0 us, 0.40% latency, 4.08 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.19 ms, 0.48% latency,
5.54 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 522.61
us, 0.21% latency, 12.58 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 589.13 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 720.02 us, 0.29%
latency, 9.13 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 575.3
us, 0.23% latency, 11.43 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 55.31 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(12): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.21 ms, 3.70% latency,
2.16 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.04 ms, 0.42% latency, 3.88 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.28 ms, 1.72% latency,
1.57 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.54 ms, 1.43%
latency, 1.44 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
578.4 us, 0.23% latency, 2.84 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
558.14 us, 0.22% latency, 2.95 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
601.29 us, 0.24% latency, 2.73 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.63 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 654.94 us, 0.26%
latency, 2.51 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
549.55 us, 0.22% latency, 2.99 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 33.62 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 25.51 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.14 ms, 0.46% latency, 3.53 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.25 ms, 0.50% latency,
5.24 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 556.95
us, 0.22% latency, 11.81 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 621.32 us, 0.25% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 679.97 us, 0.27%
latency, 9.67 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 575.3
us, 0.23% latency, 11.43 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.81 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(13): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 10.98 ms, 4.41% latency,
1.81 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.17 ms, 0.47% latency, 3.43 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.65 ms, 1.87% latency,
1.45 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.8 ms, 1.53%
latency, 1.34 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
705.24 us, 0.28% latency, 2.33 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
599.62 us, 0.24% latency, 2.74 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
576.26 us, 0.23% latency, 2.85 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.87 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 758.65 us, 0.31%
latency, 2.17 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
646.11 us, 0.26% latency, 2.54 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.52 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 31.71 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.14 ms, 0.46% latency, 3.54 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.35 ms, 0.54% latency,
4.87 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 602.25
us, 0.24% latency, 10.92 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 663.52 us, 0.27% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 721.69 us, 0.29%
latency, 9.11 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 607.97
us, 0.24% latency, 10.82 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.48 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(14): SwinLayer(
    3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.97 ms, 4.01% latency,
2.0 TFLOPS,
    (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.09 ms, 0.44% latency, 3.69 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (attention): SwinAttention(
        1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.79 ms, 1.93% latency,
1.41 TFLOPS,
        (self): SwinSelfAttention(
            790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.86 ms, 1.55%
latency, 1.32 TFLOPS,
            (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
610.83 us, 0.25% latency, 2.69 TFLOPS, in_features=512, out_features=512, bias=True)
            (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
594.14 us, 0.24% latency, 2.77 TFLOPS, in_features=512, out_features=512, bias=True)
            (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
581.98 us, 0.23% latency, 2.83 TFLOPS, in_features=512, out_features=512, bias=True)
            (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.97 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
        )
        (output): SwinSelfOutput(
            262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 836.85 us, 0.34%
latency, 1.96 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
713.35 us, 0.29% latency, 2.3 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 45.06 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 28.37 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.12 ms, 0.45% latency, 3.58 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.4 ms, 0.56% latency,
4.7 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 619.65
us, 0.25% latency, 10.61 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 693.32 us, 0.28% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 730.75 us, 0.29%
latency, 9.0 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 618.7
us, 0.25% latency, 10.63 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.76 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (15): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 11.35 ms, 4.56% latency,
1.75 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.1 ms, 0.44% latency, 3.63 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.8 ms, 1.93% latency,
1.4 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.96 ms, 1.59%
latency, 1.28 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
660.18 us, 0.27% latency, 2.49 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
679.49 us, 0.27% latency, 2.42 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
558.85 us, 0.22% latency, 2.94 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.01 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 749.11 us, 0.30%
latency, 2.19 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
632.52 us, 0.25% latency, 2.6 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 37.19 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 43.63 us,
0.02% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.15 ms, 0.46% latency, 3.48 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.34 ms, 0.54% latency,
4.91 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 607.97
us, 0.24% latency, 10.82 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 645.64 us, 0.26% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 941.04 us, 0.38%
latency, 6.99 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 824.21
us, 0.33% latency, 7.98 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.95 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (16): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 9.85 ms, 3.96% latency,
2.02 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.27 ms, 0.51% latency, 3.16 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.45 ms, 1.79% latency,
1.51 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.64 ms, 1.46%
latency, 1.4 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
636.58 us, 0.26% latency, 2.58 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
594.14 us, 0.24% latency, 2.77 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
565.77 us, 0.23% latency, 2.91 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.49 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 724.08 us, 0.29%
latency, 2.27 TFLOPS,

```

```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
601.05 us, 0.24% latency, 2.74 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.95 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 30.04 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.11 ms, 0.45% latency, 3.61 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.4 ms, 0.56% latency,
4.71 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 635.86
us, 0.26% latency, 10.34 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 677.35 us, 0.27% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 738.14 us, 0.30%
latency, 8.91 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 623.7
us, 0.25% latency, 10.54 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 37.43 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (17): SwinLayer(
        3.16 M, 3.64% Params, 9.94 GMACs, 4.11% MACs, 11.3 ms, 4.54% latency,
1.76 TFLOPS,
        (layernorm_before): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.1 ms, 0.44% latency, 3.65 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(
            1.05 M, 1.21% Params, 3.37 GMACs, 1.39% MACs, 4.92 ms, 1.98% latency,
1.37 TFLOPS,
            (self): SwinSelfAttention(
                790.67 k, 0.91% Params, 2.54 GMACs, 1.05% MACs, 3.86 ms, 1.55%
latency, 1.32 TFLOPS,
                (query): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
645.16 us, 0.26% latency, 2.55 TFLOPS, in_features=512, out_features=512, bias=True)
                (key): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
607.25 us, 0.24% latency, 2.71 TFLOPS, in_features=512, out_features=512, bias=True)
                (value): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
567.44 us, 0.23% latency, 2.9 TFLOPS, in_features=512, out_features=512, bias=True)
                (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 44.58 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
            )
            (output): SwinSelfOutput(
                262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs, 948.67 us, 0.38%
latency, 1.73 TFLOPS,

```



```

        (dense): Linear(262.66 k, 0.30% Params, 822.08 MMACs, 0.34% MACs,
822.07 us, 0.33% latency, 2.0 TFLOPS, in_features=512, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 37.43 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 33.86 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(1.02 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.12 ms, 0.45% latency, 3.57 GFLOPS, (512,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 1.39 ms, 0.56% latency,
4.72 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 617.5
us, 0.25% latency, 10.65 TFLOPS, in_features=512, out_features=2048, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 691.41 us, 0.28% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 714.3 us, 0.29% latency,
9.21 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 3.29 GMACs, 1.36% MACs, 598.19
us, 0.24% latency, 10.99 TFLOPS, in_features=2048, out_features=512, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 37.91 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    )
    )
    (downsample): SwinPatchMerging(
        2.1 M, 2.42% Params, 822.08 MMACs, 0.34% MACs, 1.69 ms, 0.68% latency,
976.0 GFLOPS,
        (reduction): Linear(2.1 M, 2.42% Params, 822.08 MMACs, 0.34% MACs, 161.89
us, 0.07% latency, 10.16 TFLOPS, in_features=2048, out_features=1024, bias=False)
        (norm): LayerNorm(4.1 k, 0.00% Params, 0 MACs, 0.00% MACs, 1.11 ms, 0.44%
latency, 3.63 GFLOPS, (2048,), eps=1e-05, elementwise_affine=True)
    )
    )
    (3): SwinStage(
        25.2 M, 29.06% Params, 19.81 GMACs, 8.19% MACs, 19.35 ms, 7.78% latency,
2.05 TFLOPS,
        (blocks): ModuleList(
            (0): SwinLayer(
                12.6 M, 14.53% Params, 9.9 GMACs, 4.10% MACs, 9.88 ms, 3.97% latency,
2.01 TFLOPS,
                (layernorm_before): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.14 ms, 0.46% latency, 1.76 GFLOPS, (1024,), eps=1e-05, elementwise_affine=True)
                (attention): SwinAttention(
                    4.2 M, 4.85% Params, 3.33 GMACs, 1.38% MACs, 4.9 ms, 1.97% latency,
1.36 TFLOPS,
                    (self): SwinSelfAttention(

```

```

        3.15 M, 3.64% Params, 2.51 GMACs, 1.04% MACs, 4.04 ms, 1.62%
latency, 1.24 TFLOPS,
        (query): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
685.21 us, 0.28% latency, 2.4 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (key): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs, 706.2
us, 0.28% latency, 2.33 TFLOPS, in_features=1024, out_features=1024, bias=True)
        (value): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
680.69 us, 0.27% latency, 2.42 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.49 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (output): SwinSelfOutput(
        1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs, 749.35 us, 0.30%
latency, 2.19 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
596.05 us, 0.24% latency, 2.76 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 45.78 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
    (drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 25.75 us,
0.01% latency, 0.0 FLOPS, p=0.1)
    (layernorm_after): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.16 ms, 0.46% latency, 1.74 GFLOPS, (1024,), eps=1e-05, elementwise_affine=True)
    (intermediate): SwinIntermediate(
        4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 1.32 ms, 0.53% latency,
4.99 TFLOPS,
        (dense): Linear(4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 626.56
us, 0.25% latency, 10.5 TFLOPS, in_features=1024, out_features=4096, bias=True)
        (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 608.68 us, 0.24% latency, 0.0 FLOPS, )
    )
    (output): SwinOutput(
        4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 682.12 us, 0.27% latency,
9.64 TFLOPS,
        (dense): Linear(4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 577.69
us, 0.23% latency, 11.38 TFLOPS, in_features=4096, out_features=1024, bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 34.33 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
    (1): SwinLayer(
        12.6 M, 14.53% Params, 9.9 GMACs, 4.10% MACs, 9.38 ms, 3.77% latency,
2.11 TFLOPS,
        (layernorm_before): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.04 ms, 0.42% latency, 1.94 GFLOPS, (1024,), eps=1e-05, elementwise_affine=True)
        (attention): SwinAttention(

```

```

4.2 M, 4.85% Params, 3.33 GMACs, 1.38% MACs, 4.42 ms, 1.78% latency,
1.51 TFLOPS,
    (self): SwinSelfAttention(
        3.15 M, 3.64% Params, 2.51 GMACs, 1.04% MACs, 3.59 ms, 1.44%
latency, 1.4 TFLOPS,
        (query): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
597.72 us, 0.24% latency, 2.75 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (key): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs, 601.53
us, 0.24% latency, 2.73 TFLOPS, in_features=1024, out_features=1024, bias=True)
        (value): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
564.34 us, 0.23% latency, 2.91 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 46.25 us,
0.02% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
    (output): SwinSelfOutput(
        1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs, 737.43 us, 0.30%
latency, 2.23 TFLOPS,
        (dense): Linear(1.05 M, 1.21% Params, 822.08 MMACs, 0.34% MACs,
623.46 us, 0.25% latency, 2.64 TFLOPS, in_features=1024, out_features=1024,
bias=True)
        (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 35.76 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
    )
)
(drop_path): SwinDropPath(0, 0.00% Params, 0 MACs, 0.00% MACs, 25.03 us,
0.01% latency, 0.0 FLOPS, p=0.1)
(layer_norm_after): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs,
1.1 ms, 0.44% latency, 1.83 GFLOPS, (1024,), eps=1e-05, elementwise_affine=True)
(intermediate): SwinIntermediate(
    4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 1.38 ms, 0.55% latency,
4.78 TFLOPS,
    (dense): Linear(4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 632.52
us, 0.25% latency, 10.4 TFLOPS, in_features=1024, out_features=4096, bias=True)
    (intermediate_act_fn): GELUActivation(0, 0.00% Params, 0 MACs, 0.00%
MACs, 661.61 us, 0.27% latency, 0.0 FLOPS, )
)
(output): SwinOutput(
    4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 690.94 us, 0.28% latency,
9.52 TFLOPS,
    (dense): Linear(4.2 M, 4.84% Params, 3.29 GMACs, 1.36% MACs, 579.6 us,
0.23% latency, 11.35 TFLOPS, in_features=4096, out_features=1024, bias=True)
    (dropout): Dropout(0, 0.00% Params, 0 MACs, 0.00% MACs, 36.48 us,
0.01% latency, 0.0 FLOPS, p=0.0, inplace=False)
)
)
)
)
)

```

```
)  
(layernorm): LayerNorm(2.05 k, 0.00% Params, 0 MACs, 0.00% MACs, 1.13 ms, 0.46%  
latency, 1.77 GFLOPS, (1024,)), eps=1e-05, elementwise_affine=True)  
(pooler): AdaptiveAvgPool1d(0, 0.00% Params, 0 MACs, 0.00% MACs, 348.57 us, 0.14%  
latency, 1.15 GFLOPS, output_size=1)  
)  
-----
```