

----- 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:	1.82 M
params of model = params per GPU * mp_size:	1.82 M
fwd MACs per GPU:	2.6 GMACs
fwd flops per GPU:	5.25 G
fwd flops of model = fwd flops per GPU * mp_size:	5.25 G
fwd latency:	252.43 ms
fwd FLOPS per GPU = fwd flops per GPU / fwd latency:	20.82 GFLOPS

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

Top 1 modules in terms of params, MACs or fwd latency at different model depths:
depth 0:

params	- {'MobileNetV1Model': '1.82 M'}
MACs	- {'MobileNetV1Model': '2.6 GMACs'}
fwd latency	- {'MobileNetV1Model': '252.43 ms'}

depth 1:

params	- {'ModuleList': '1.82 M'}
MACs	- {'ModuleList': '2.53 GMACs'}
fwd latency	- {'ModuleList': '241.8 ms'}

depth 2:

params	- {'MobileNetV1ConvLayer': '1.82 M'}
MACs	- {'MobileNetV1ConvLayer': '2.53 GMACs'}
fwd latency	- {'MobileNetV1ConvLayer': '241.8 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.

```
MobileNetV1Model(
  1.82 M, 100.00% Params, 2.6 GMACs, 100.00% MACs, 252.43 ms, 100.00% latency, 20.82 GFLOPS,
  (conv_stem): MobileNetV1ConvLayer(
    696, 0.04% Params, 65.03 MMACs, 2.50% MACs, 9.32 ms, 3.69% latency, 14.47 GFLOPS,
    (convolution): Conv2d(648, 0.04% Params, 65.03 MMACs, 2.50% MACs, 4.44 ms, 1.76% latency, 29.32 GFLOPS, 3, 24, kernel_size=(3, 3), stride=(2, 2), bias=False)
    (normalization): BatchNorm2d(48, 0.00% Params, 0 MACs, 0.00% MACs, 1.61 ms, 0.64% latency, 3.0 GFLOPS, 24, eps=0.001, momentum=0.9997, affine=True, track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.49 ms, 0.59% latency, 0.0 FLOPS, )
  )
  (layer): ModuleList(
    (0): MobileNetV1ConvLayer(
      264, 0.01% Params, 21.68 MMACs, 0.83% MACs, 18.14 ms, 7.19% latency, 2.66 GFLOPS,
      (convolution): Conv2d(216, 0.01% Params, 21.68 MMACs, 0.83% MACs, 10.14 ms, 4.02% latency, 4.28 GFLOPS, 24, 24, kernel_size=(3, 3), stride=(1, 1), groups=24, bias=False)
      (normalization): BatchNorm2d(48, 0.00% Params, 0 MACs, 0.00% MACs, 1.48 ms, 0.59% latency, 3.26 GFLOPS, 24, eps=0.001, momentum=0.9997, affine=True, track_running_stats=True)
      (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.29 ms, 0.51% latency, 0.0 FLOPS, )
    )
    (1): MobileNetV1ConvLayer(
      1.25 k, 0.07% Params, 115.61 MMACs, 4.45% MACs, 25.62 ms, 10.15% latency, 9.4 GFLOPS,
      (convolution): Conv2d(1.15 k, 0.06% Params, 115.61 MMACs, 4.45% MACs, 9.75 ms, 3.86% latency, 23.72 GFLOPS, 24, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (normalization): BatchNorm2d(96, 0.01% Params, 0 MACs, 0.00% MACs, 6.85 ms, 2.71% latency, 1.41 GFLOPS, 48, eps=0.001, momentum=0.9997, affine=True, track_running_stats=True)
      (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 6.08 ms, 2.41% latency, 0.0 FLOPS, )
    )
    (2): MobileNetV1ConvLayer(
      528, 0.03% Params, 10.84 MMACs, 0.42% MACs, 21.23 ms, 8.41% latency, 1.13 GFLOPS,
      (convolution): Conv2d(432, 0.02% Params, 10.84 MMACs, 0.42% MACs, 9.91 ms, 3.93% latency, 2.19 GFLOPS, 48, 48, kernel_size=(3, 3), stride=(2, 2), groups=48, bias=False)
      (normalization): BatchNorm2d(96, 0.01% Params, 0 MACs, 0.00% MACs, 673.77 us, 0.27% latency, 3.57 GFLOPS, 48, eps=0.001, momentum=0.9997, affine=True, track_running_stats=True)
      (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 811.82 us, 0.32%
```

```

latency, 0.0 FLOPS, )
)
(3): MobileNetV1ConvLayer(
  4.8 k, 0.26% Params, 115.61 MMACs, 4.45% MACs, 7.61 ms, 3.01% latency, 31.02
GFLOPS,
  (convolution): Conv2d(4.61 k, 0.25% Params, 115.61 MMACs, 4.45% MACs, 2.53 ms,
1.00% latency, 91.21 GFLOPS, 48, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
  (normalization): BatchNorm2d(192, 0.01% Params, 0 MACs, 0.00% MACs, 1.52 ms,
0.60% latency, 3.18 GFLOPS, 96, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.59 ms, 0.63%
latency, 0.0 FLOPS, )
)
(4): MobileNetV1ConvLayer(
  1.06 k, 0.06% Params, 21.68 MMACs, 0.83% MACs, 16.72 ms, 6.62% latency, 2.88
GFLOPS,
  (convolution): Conv2d(864, 0.05% Params, 21.68 MMACs, 0.83% MACs, 8.78 ms,
3.48% latency, 4.94 GFLOPS, 96, 96, kernel_size=(3, 3), stride=(1, 1), groups=96,
bias=False)
  (normalization): BatchNorm2d(192, 0.01% Params, 0 MACs, 0.00% MACs, 1.46 ms,
0.58% latency, 3.31 GFLOPS, 96, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 3.21 ms, 1.27%
latency, 0.0 FLOPS, )
)
(5): MobileNetV1ConvLayer(
  9.41 k, 0.52% Params, 231.21 MMACs, 8.90% MACs, 19.53 ms, 7.73% latency, 23.93
GFLOPS,
  (convolution): Conv2d(9.22 k, 0.51% Params, 231.21 MMACs, 8.90% MACs, 8.53 ms,
3.38% latency, 54.24 GFLOPS, 96, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
  (normalization): BatchNorm2d(192, 0.01% Params, 0 MACs, 0.00% MACs, 3.76 ms,
1.49% latency, 1.28 GFLOPS, 96, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 3.52 ms, 1.39%
latency, 0.0 FLOPS, )
)
(6): MobileNetV1ConvLayer(
  1.06 k, 0.06% Params, 5.42 MMACs, 0.21% MACs, 11.46 ms, 4.54% latency, 1.05
GFLOPS,
  (convolution): Conv2d(864, 0.05% Params, 5.42 MMACs, 0.21% MACs, 5.59 ms,
2.21% latency, 1.94 GFLOPS, 96, 96, kernel_size=(3, 3), stride=(2, 2), groups=96,
bias=False)
  (normalization): BatchNorm2d(192, 0.01% Params, 0 MACs, 0.00% MACs, 435.83 us,
0.17% latency, 2.76 GFLOPS, 96, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 468.97 us, 0.19%
latency, 0.0 FLOPS, )
)
(7): MobileNetV1ConvLayer(
  18.82 k, 1.04% Params, 115.61 MMACs, 4.45% MACs, 6.82 ms, 2.70% latency, 34.23

```

```

GFLOPS,
    (convolution): Conv2d(18.43 k, 1.01% Params, 115.61 MMACs, 4.45% MACs, 3.25
ms, 1.29% latency, 71.03 GFLOPS, 96, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (normalization): BatchNorm2d(384, 0.02% Params, 0 MACs, 0.00% MACs, 1.66 ms,
0.66% latency, 1.45 GFLOPS, 192, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 790.12 us, 0.31%
latency, 0.0 FLOPS, )
    )
    (8): MobileNetV1ConvLayer(
        2.11 k, 0.12% Params, 10.84 MMACs, 0.42% MACs, 6.94 ms, 2.75% latency, 3.47
GFLOPS,
        (convolution): Conv2d(1.73 k, 0.10% Params, 10.84 MMACs, 0.42% MACs, 2.88 ms,
1.14% latency, 7.53 GFLOPS, 192, 192, kernel_size=(3, 3), stride=(1, 1), groups=192,
bias=False)
        (normalization): BatchNorm2d(384, 0.02% Params, 0 MACs, 0.00% MACs, 622.03 us,
0.25% latency, 3.87 GFLOPS, 192, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
        (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.83 ms, 0.72%
latency, 0.0 FLOPS, )
        )
    (9): MobileNetV1ConvLayer(
        37.25 k, 2.05% Params, 231.21 MMACs, 8.90% MACs, 11.31 ms, 4.48% latency, 41.1
GFLOPS,
        (convolution): Conv2d(36.86 k, 2.03% Params, 231.21 MMACs, 8.90% MACs, 6.58
ms, 2.61% latency, 70.32 GFLOPS, 192, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (normalization): BatchNorm2d(384, 0.02% Params, 0 MACs, 0.00% MACs, 2.05 ms,
0.81% latency, 1.18 GFLOPS, 192, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
        (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 1.8 ms, 0.71%
latency, 0.0 FLOPS, )
        )
    (10): MobileNetV1ConvLayer(
        2.11 k, 0.12% Params, 2.71 MMACs, 0.10% MACs, 5.3 ms, 2.10% latency, 1.14
GFLOPS,
        (convolution): Conv2d(1.73 k, 0.10% Params, 2.71 MMACs, 0.10% MACs, 3.11 ms,
1.23% latency, 1.74 GFLOPS, 192, 192, kernel_size=(3, 3), stride=(2, 2), groups=192,
bias=False)
        (normalization): BatchNorm2d(384, 0.02% Params, 0 MACs, 0.00% MACs, 292.06 us,
0.12% latency, 2.06 GFLOPS, 192, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
        (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 262.5 us, 0.10%
latency, 0.0 FLOPS, )
        )
    (11): MobileNetV1ConvLayer(
        74.5 k, 4.10% Params, 115.61 MMACs, 4.45% MACs, 5.2 ms, 2.06% latency, 44.7
GFLOPS,
        (convolution): Conv2d(73.73 k, 4.06% Params, 115.61 MMACs, 4.45% MACs, 3.24

```

```

ms, 1.29% latency, 71.27 GFLOPS, 192, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 581.5 us,
0.23% latency, 2.07 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 947.24 us, 0.38%
latency, 0.0 FLOPS, )
    )
    (12): MobileNetV1ConvLayer(
    4.22 k, 0.23% Params, 5.42 MMACs, 0.21% MACs, 4.69 ms, 1.86% latency, 2.57
GFLOPS,
    (convolution): Conv2d(3.46 k, 0.19% Params, 5.42 MMACs, 0.21% MACs, 2.24 ms,
0.89% latency, 4.83 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(1, 1), groups=384,
bias=False)
    (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 371.46 us,
0.15% latency, 3.24 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 937.94 us, 0.37%
latency, 0.0 FLOPS, )
    )
    (13): MobileNetV1ConvLayer(
    148.22 k, 8.16% Params, 231.21 MMACs, 8.90% MACs, 7.86 ms, 3.11% latency,
58.98 GFLOPS,
    (convolution): Conv2d(147.46 k, 8.12% Params, 231.21 MMACs, 8.90% MACs, 5.77
ms, 2.28% latency, 80.18 GFLOPS, 384, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 629.19 us,
0.25% latency, 1.91 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 955.82 us, 0.38%
latency, 0.0 FLOPS, )
    )
    (14): MobileNetV1ConvLayer(
    4.22 k, 0.23% Params, 5.42 MMACs, 0.21% MACs, 3.7 ms, 1.47% latency, 3.25
GFLOPS,
    (convolution): Conv2d(3.46 k, 0.19% Params, 5.42 MMACs, 0.21% MACs, 1.73 ms,
0.69% latency, 6.25 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(1, 1), groups=384,
bias=False)
    (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 371.69 us,
0.15% latency, 3.24 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 447.51 us, 0.18%
latency, 0.0 FLOPS, )
    )
    (15): MobileNetV1ConvLayer(
    148.22 k, 8.16% Params, 231.21 MMACs, 8.90% MACs, 7.87 ms, 3.12% latency,
58.89 GFLOPS,
    (convolution): Conv2d(147.46 k, 8.12% Params, 231.21 MMACs, 8.90% MACs, 5.78
ms, 2.29% latency, 80.07 GFLOPS, 384, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)

```

```

(normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 1.12 ms,
0.45% latency, 1.07 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
(activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 494.24 us, 0.20%
latency, 0.0 FLOPS, )
)
(16): MobileNetV1ConvLayer(
4.22 k, 0.23% Params, 5.42 MMACs, 0.21% MACs, 5.39 ms, 2.14% latency, 2.23
GFLOPS,
(convolution): Conv2d(3.46 k, 0.19% Params, 5.42 MMACs, 0.21% MACs, 2.87 ms,
1.14% latency, 3.78 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(1, 1), groups=384,
bias=False)
(normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 418.9 us,
0.17% latency, 2.87 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
(activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 908.61 us, 0.36%
latency, 0.0 FLOPS, )
)
(17): MobileNetV1ConvLayer(
148.22 k, 8.16% Params, 231.21 MMACs, 8.90% MACs, 9.32 ms, 3.69% latency,
49.75 GFLOPS,
(convolution): Conv2d(147.46 k, 8.12% Params, 231.21 MMACs, 8.90% MACs, 6.26
ms, 2.48% latency, 73.92 GFLOPS, 384, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)
(normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 1.07 ms,
0.42% latency, 1.13 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
(activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 911.0 us, 0.36%
latency, 0.0 FLOPS, )
)
(18): MobileNetV1ConvLayer(
4.22 k, 0.23% Params, 5.42 MMACs, 0.21% MACs, 5.03 ms, 1.99% latency, 2.4
GFLOPS,
(convolution): Conv2d(3.46 k, 0.19% Params, 5.42 MMACs, 0.21% MACs, 2.64 ms,
1.05% latency, 4.11 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(1, 1), groups=384,
bias=False)
(normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 394.58 us,
0.16% latency, 3.05 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
(activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 391.48 us, 0.16%
latency, 0.0 FLOPS, )
)
(19): MobileNetV1ConvLayer(
148.22 k, 8.16% Params, 231.21 MMACs, 8.90% MACs, 6.74 ms, 2.67% latency, 68.8
GFLOPS,
(convolution): Conv2d(147.46 k, 8.12% Params, 231.21 MMACs, 8.90% MACs, 5.16
ms, 2.04% latency, 89.59 GFLOPS, 384, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)
(normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 598.67 us,
0.24% latency, 2.01 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,

```

```

track_running_stats=True)
    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 494.72 us, 0.20%
latency, 0.0 FLOPS, )
    )
    (20): MobileNetV1ConvLayer(
        4.22 k, 0.23% Params, 5.42 MMACs, 0.21% MACs, 5.36 ms, 2.12% latency, 2.25
GFLOPS,
        (convolution): Conv2d(3.46 k, 0.19% Params, 5.42 MMACs, 0.21% MACs, 2.57 ms,
1.02% latency, 4.22 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(1, 1), groups=384,
bias=False)
        (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 432.73 us,
0.17% latency, 2.78 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
        (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 863.31 us, 0.34%
latency, 0.0 FLOPS, )
        )
        (21): MobileNetV1ConvLayer(
            148.22 k, 8.16% Params, 231.21 MMACs, 8.90% MACs, 8.65 ms, 3.43% latency,
53.58 GFLOPS,
            (convolution): Conv2d(147.46 k, 8.12% Params, 231.21 MMACs, 8.90% MACs, 5.77
ms, 2.29% latency, 80.15 GFLOPS, 384, 384, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 1.03 ms,
0.41% latency, 1.17 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
            (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 896.22 us, 0.36%
latency, 0.0 FLOPS, )
            )
            (22): MobileNetV1ConvLayer(
                4.22 k, 0.23% Params, 1.35 MMACs, 0.05% MACs, 3.44 ms, 1.36% latency, 875.43
MFLOPS,
                (convolution): Conv2d(3.46 k, 0.19% Params, 1.35 MMACs, 0.05% MACs, 1.55 ms,
0.61% latency, 1.75 GFLOPS, 384, 384, kernel_size=(3, 3), stride=(2, 2), groups=384,
bias=False)
                (normalization): BatchNorm2d(768, 0.04% Params, 0 MACs, 0.00% MACs, 232.7 us,
0.09% latency, 1.29 GFLOPS, 384, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
                (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 180.72 us, 0.07%
latency, 0.0 FLOPS, )
                )
                (23): MobileNetV1ConvLayer(
                    296.45 k, 16.32% Params, 115.61 MMACs, 4.45% MACs, 5.39 ms, 2.14% latency,
43.01 GFLOPS,
                    (convolution): Conv2d(294.91 k, 16.23% Params, 115.61 MMACs, 4.45% MACs, 4.35
ms, 1.72% latency, 53.19 GFLOPS, 384, 768, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                    (normalization): BatchNorm2d(1.54 k, 0.08% Params, 0 MACs, 0.00% MACs, 419.14
us, 0.17% latency, 1.44 GFLOPS, 768, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
                    (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 369.07 us, 0.15%

```

```

latency, 0.0 FLOPS, )
)
(24): MobileNetV1ConvLayer(
  8.45 k, 0.47% Params, 2.71 MMACs, 0.10% MACs, 3.69 ms, 1.46% latency, 1.63
GFLOPS,
  (convolution): Conv2d(6.91 k, 0.38% Params, 2.71 MMACs, 0.10% MACs, 1.83 ms,
0.73% latency, 2.96 GFLOPS, 768, 768, kernel_size=(3, 3), stride=(1, 1), groups=768,
bias=False)
  (normalization): BatchNorm2d(1.54 k, 0.08% Params, 0 MACs, 0.00% MACs, 322.58
us, 0.13% latency, 1.87 GFLOPS, 768, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 250.1 us, 0.10%
latency, 0.0 FLOPS, )
)
(25): MobileNetV1ConvLayer(
  591.36 k, 32.55% Params, 231.21 MMACs, 8.90% MACs, 8.79 ms, 3.48% latency,
52.67 GFLOPS,
  (convolution): Conv2d(589.82 k, 32.47% Params, 231.21 MMACs, 8.90% MACs, 7.45
ms, 2.95% latency, 62.05 GFLOPS, 768, 768, kernel_size=(1, 1), stride=(1, 1),
bias=False)
  (normalization): BatchNorm2d(1.54 k, 0.08% Params, 0 MACs, 0.00% MACs, 643.25
us, 0.25% latency, 936.04 MFLOPS, 768, eps=0.001, momentum=0.9997, affine=True,
track_running_stats=True)
  (activation): ReLU6(0, 0.00% Params, 0 MACs, 0.00% MACs, 343.8 us, 0.14%
latency, 0.0 FLOPS, )
)
)
(pooler): AdaptiveAvgPool2d(0, 0.00% Params, 0 MACs, 0.00% MACs, 271.56 us, 0.11%
latency, 1.11 GFLOPS, output_size=(1, 1))
)
-----

```