```
Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3), bias=False) [forward_compute_time=36.842,backward_compute_time=0.039,activation_size=321126400.0,parameter_size=37632.0]
                                                                                                              BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
[forward_compute_time=15.771,backward_compute_time=18.005,activation_size=321126400.0,parameter_size=512.0]
                                                                                                                 ReLU()
[forward_compute_time=96.593,backward_compute_time=77.547,activation_size=321126400.0,parameter_size=0.0]
                                                                                                                 MaxPool2d(kernel_size=3, stride=2, padding=1, dilation=1, ceil_mode=False)
[forward_compute_time=12.293,backward_compute_time=116.823,activation_size=80281600.0,parameter_size=0.0]
                                                  Conv2d(64, 64, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) \\ [forward\_compute\_time=11.785,backward\_compute\_time=28.743, activation\_size=80281600.0, parameter\_size=147456.0]
                                                      BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=4.263,backward_compute_time=4.974,activation_size=80281600.0,parameter_size=512.0]
                                                                                                                            ReLU()
                                                       [forward compute time=3.017,backward compute time=2.763,activation size=80281600.0,parameter size=0.0]
                                                  Conv2d(64, 64, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) \\ [forward\_compute\_time=11.524, backward\_compute\_time=28.147, activation\_size=80281600.0, parameter\_size=147456.0]
                                                         BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
[forward_compute_time=4.250,backward_compute_time=4.985,activation_size=80281600.0,parameter_size=512.0]
                                                                                                                   Add [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=80281600.0,parameter_size=0.0]
                                                                                                                   [forward_compute_time=7.930,backward_compute_time=6.142,activation_size=80281600.0,parameter_size=0.0]
                                                     Conv2d(64, 64, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)\\ [forward\_compute\_time=0.000, backward\_compute\_time=0.000, activation\_size=80281600.0, parameter\_size=147456.0]
                                                     BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=80281600.0,parameter_size=512.0]
                                                       [forward_compute_time=3.005,backward_compute_time=3.084,activation_size=40140800.0,parameter_size=0.0]
                                                 Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) [forward_compute_time=11.740,backward_compute_time=28.462,activation_size=80281600.0,parameter_size=147456.0]
                                                        BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=4.216,backward_compute_time=4.977,activation_size=80281600.0,parameter_size=512.0]
                                                                                                                   [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=80281600.0,parameter_size=0.0]
                                                                                                                   [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=0.0]
Conv2d(64, 128, kernel\_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) \\ [forward\_compute\_time=0.000, backward\_compute\_time=0.000, activation\_size=40140800.0, parameter\_size=294912.0]
                                                                                                                                                                                                                                                                    Conv2d(64, 128, kernel\_size=(1, 1), stride=(2, 2), bias=False)
                                                                                                                                                                                                                                [forward compute time=0.000,backward compute time=0.000,activation size=40140800.0,parameter size=32768.0]
    BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=1024.0]
                                                                                                                                                                                                                               BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=1024.0]
         [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=0.0]
                                                      Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=589824.0]
                                                                                  BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=1024.0]
                                                                                                                                              [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=0.0]
                                                                                                                                              [forward compute time=0.000,backward compute time=0.000,activation size=20070400.0,parameter size=0.0]
                                                                                Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=589824.0]
                                                                                  BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=1024.0]
                                                                                    [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=0.0]
                                                                                Conv2d(128, 128, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) \\ [forward\_compute\_time=0.000,backward\_compute\_time=0.000,activation\_size=40140800.0,parameter\_size=589824.0]
                                                                                  BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=1024.0]
                                                                                                                                              [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=40140800.0,parameter_size=0.0]
                                                                                                                                                                                                                  ReLU()
                                                                                                                                              [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=0.0]
                        Conv2d(128, 256, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=1179648.0]
                                                                                                                                                                                                                                                           Conv2d(128, 256, kernel_size=(1, 1), stride=(2, 2), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=131072.0]
                                  BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2048.0]
                                                                                                                                                                                                                                                          BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2048.0]
                                                                                                              ReLU()
                                                [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]
                                                                                  Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2359296.0]
                                                                                                              BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2048.0]
                                                                                                                                                                          [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=0.0]
                                                                                                                                                                               [forward compute time=0.000,backward compute time=0.000,activation size=0.0,parameter size=0.0]
                                                                                                                                     Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)
                                                                                                          [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2359296.0]
                                                                                                            BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2048.0]
                                                                                                                     [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]
                                                                                                          Conv2d(256, 256, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) \\ [forward\_compute\_time=0.000, backward\_compute\_time=0.000, activation\_size=20070400.0, parameter\_size=2359296.0]
                                                                                                             BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=20070400.0,parameter_size=2048.0]
                                                                                                                                                                          [forward compute time=0.000,backward compute time=0.000,activation size=20070400.0,parameter size=0.0]
                                                                                                                                                                               ReLU() \\ [forward\_compute\_time=0.000,backward\_compute\_time=0.000,activation\_size=0.0,parameter\_size=0.0]
                                                   Conv2d(256, 512, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4718592.0]
                                                                                                                                                                                                                                                                                      Conv2d(256, 512, kernel_size=(1, 1), stride=(2, 2), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=524288.0]
                                                              BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4096.0]
                                                                                                                                                                                                                                                                                      BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4096.0]
                                                                           [forward compute time=0.000,backward compute time=0.000,activation size=0.0,parameter size=0.0]
                                                                                                              Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=9437184.0]
                                                                                                                                         BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4096.0]
                                                                                                                                                                                                     [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=0.0]
                                                                                                                                                                                                           [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]
                                                                                                                                      Conv2d(512, 512, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)\\ [forward\_compute\_time=0.000, backward\_compute\_time=0.000, activation\_size=10035200.0, parameter\_size=9437184.0]
                                                                                                                                       BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4096.0]
                                                                                                                                                                                                               ReLU()
                                                                                                                                                [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]
                                                                                                                                      Conv2d(512, 512, kernel\_size=(3, 3), stride=(1, 1), padding=(1, 1), bias=False)\\ [forward\_compute\_time=0.000, backward\_compute\_time=0.000, activation\_size=10035200.0, parameter\_size=9437184.0]
                                                                                                                                         BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=4096.0]
                                                                                                                                                                                                     [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=10035200.0,parameter_size=0.0]
                                                                                                                                                                                                                                                                         ReLU()
                                                                                                                                                                                                           [forward compute time=0.000,backward compute time=0.000,activation size=0.0,parameter size=0.0]
                                                                                                                                                                                                      AdaptiveAvgPool2d(output_size=(1, 1))
[forward_compute_time=0.000,backward_compute_time=0.000,activation_size=204800.0,parameter_size=0.0]
                                                                                                                                                                                                                                                       Flatten(start_dim=1, end_dim=-1)
                                                                                                                                                                                                      [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=204800.0,parameter_size=0.0]
```

Linear(in_features=512, out_features=1000, bias=True)
[forward_compute_time=0.000,backward_compute_time=0.000,activation_size=400000.0,parameter_size=2052000.0]

 $Input 0 \\ [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]$