```
ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                      layer1.0.downsample.0.weight (256, 64, 1, 1)
                                                                                                          layer1.0.conv1.weight (64, 64, 1, 1)
                                                                                     ReluBackward0
layer1.0.downsample.1.weight (256)
                             layer1.0.downsample.1.bias (256)
                                                                                                                               layer1.0.bn1.weight (64) layer1.0.bn1.bias
                                                              AccumulateGrad MaxPool2DWithIndicesBackward0 AccumulateGrad
                                                                                                           ConvolutionBackward0
                                                                                                                                  AccumulateGrad
                                         AccumulateGrad
                AccumulateGrad
                                                                                                                  CudnnBatchNormBackward0
                                              CudnnBatchNormBackward0
                                                                                                                                            layer1.0.conv2.weight (64, 64, 3, 3)
                                                                                                                                                         layer1.0.bn2.weight (64) (64)
                                                                                                                                         AccumulateGrad
                                                                                                                                                            AccumulateGrad
                                                                                                                                              CudnnBatchNormBackward0 layer1.0.conv3.weight (256, 64, 1, 1)
                                                                                                                                                                                  layer1.0.bn3.weight (256)
                                                                                                                                                                   AccumulateGrad
                                                                                                                                                                 ConvolutionBackward0 AccumulateGrad
                                                                                                                AddBackward0 layer1.1.conv1.weight (64, 256, 1, 1)
                                                                                                                 ReluBackward0 | layer1.1.bn1.weight (64) | layer1.1.bn1.bias (64)
                                                                                                                               ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                              CudnnBatchNormBackward0 layer1.1.conv2.weight (64, 64, 3, 3)
                                                                                                                                  ReluBackward0 | layer1.1.bn2.weight (64) | layer1.1.bn2.bias (64)
                                                                                                                                                ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                         CudnnBatchNormBackward0 layer1.1.conv3.weight (256, 64, 1, 1)
                                                                                                                                                              ReluBackward0 | layer1.1.bn3.weight | layer1.1.bn3.bias (256) | (256)
                                                                                                                                                                            ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                   AccumulateGrad
                                                                                                                                          AddBackward0 layer1.2.conv1.weight (64, 256, 1, 1)
                                                                                                                                                                               layer1.2.bn1.weight (64) (64)
                                                                                                                                             ReluBackward0 AccumulateGrad
                                                                                                                                                                                                      (64)
                                                                                                                                                          ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                  CudnnBatchNormBackward0 layer1.2.conv2.weight (64, 64, 3, 3)
                                                                                                                                                                      ReluBackward0 | layer1.2.bn2.weight (64) | layer1.2.bn2.bias (64)
                                                                                                                                                                                        ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                              CudnnBatchNormBackward0 layer1.2.conv3.weight (256, 64, 1, 1)
                                                                                                                                                                                                                                    layer1.2.bn3.weight (256) (256)
                                                                                                                                                                                                                    AccumulateGrad
                                                                                                                                                                                                                ConvolutionBackward0 AccumulateGrad
                                                                                                                                                        layer2.0.downsample.0.weight (512, 256, 1, 1)
                                                                                                                                                                                      AddBackward0 layer2.0.conv1.weight (128, 256, 1, 1)
                                                                                                           layer2.0.downsample.1.weight layer2.0.downsample.1.bias (512)
                                                                                                                                                                                                                        layer2.0.bn1.weight layer2.0.bn1.bias (128)
                                                                                                                                                                                      ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                          AccumulateGrad
                                                                                                                                                                                                                                     layer2.0.conv2.weight (128, 128, 3, 3)
                                                                                                                                                                                                            CudnnBatchNormBackward0
                                                                                                                                                                                                                                 AccumulateGrad layer2.0.bn2.weight (128) layer2.0.bn2.bias (128)
                                                                                                                                                                                                                              ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                      CudnnBatchNormBackward0 layer2.0.conv3.weight (512, 128, 1, 1)
                                                                                                                                                                                                                                                                                                    layer2.0.bn3.bias (512)
                                                                                                                                                                                                                                                                               layer2.0.bn3.weight (512)
                                                                                                                                                                                                                                            ReluBackward0
                                                                                                                                                                                                                                                            AccumulateGrad
                                                                                                                                                                                                                                                          ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                 CudnnBatchNormBackwardU
                                                                                                                                                                                                                 AddBackward0 layer2.1.conv1.weight (128, 512, 1, 1)
                                                                                                                                                                                                                                AccumulateGrad layer2.1.bn1.weight (128) layer2.1.bn1.bias (128)
                                                                                                                                                                                                                  ReluBackward0
                                                                                                                                                                                                                               ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                        AccumulateGrad
                                                                                                                                                                                                                              CudnnBatchNormBackward0 layer2.1.conv2.weight (128, 128, 3, 3)
                                                                                                                                                                                                                                  ReluBackward0 | layer2.1.bn2.weight (128) | layer2.1.bn2.bias (128)
                                                                                                                                                                                                                                                   ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                        CudnnBatchNormBackward0 layer2.1.conv3.weight (512, 128, 1, 1)
                                                                                                                                                                                                                                                                               AccumulateGrad layer2.1.bn3.weight (512) layer2.1.bn3.bias (512)
                                                                                                                                                                                                                                                                            ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                           layer2.2.conv1.weight (128, 512, 1, 1)
                                                                                                                                                                                                                                            AddBackward0
                                                                                                                                                                                                                                                                               layer2.2.bn1.weight (128) layer2.2.bn1.bias
                                                                                                                                                                                                                                             ReluBackward0
                                                                                                                                                                                                                                                              AccumulateGrad
                                                                                                                                                                                                                                                           ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                  CudnnBatchNormBackward0 layer2.2.conv2.weight (128, 128, 3, 3)
                                                                                                                                                                                                                                                                       ReluBackward0 layer2.2.bn2.weight (128) layer2.2.bn2.bias (128)
                                                                                                                                                                                                                                                                                     ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                              CudnnBatchNormBackward0 layer2.2.conv3.weight (512, 128, 1, 1)
                                                                                                                                                                                                                                                                                                 ReluBackward0 | layer2.2.bn3.weight | layer2.2.bn3.bias (512) | (512)
                                                                                                                                                                                                                                                                                                                ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                      layer2.3.conv1.weight (128, 512, 1, 1)
                                                                                                                                                                                                                                                                        ReluBackward0 AccumulateGrad layer2.3.bn1.weight (128) layer2.3.bn1.bias (128)
                                                                                                                                                                                                                                                                                      ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                     CudnnBatchNormBackward0 layer2.3.conv2.weight (128, 128, 3, 3)
                                                                                                                                                                                                                                                                                          ReluBackward0 | layer2.3.bn2.weight | layer2.3.bn2.bias (128)
                                                                                                                                                                                                                                                                                                       ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                                                CudnnBatchNormBackward0 layer2.3.conv3.weight (512, 128, 1, 1)
                                                                                                                                                                                                                                                                                                                                                      layer2.3.bn3.weight (512) (512)
                                                                                                                                                                                                                                                                                                                     ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                               layer3.0.downsample.0.weight (1024, 512, 1, 1)
                                                                                                                                                                                                                                                                                                            AddBackward0 layer3.0.conv1.weight (256, 512, 1, 1)
                                                                                                                                                                                                                                                                                                                                              layer3.0.bn1.weight (256) (256)
                                                                                                                                                                                                                                 layer3.0.downsample.1.weight layer3.0.downsample.1.bias (1024)
                                                                                                                                                                                                                                                                                            AccumulateGrad
                                                                                                                                                                                                                                                                                                            ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                          ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                     AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                    AccumulateGrad
                                                                                                                                                                                                                                                                                                                                   CudnnBatchNormBackward0 layer3.0.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                       AccumulateGrad layer3.0.bn2.weight (256) layer3.0.bn2.bias
                                                                                                                                                                                                                                                                                                                                                                            AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                             CudnnBatchNormBackward0 layer3.0.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                   ReluBackward0 layer3.0.bn3.weight (1024) layer3.0.bn3.bias (1024)
                                                                                                                                                                                                                                                                                                                                             layer3.1.conv1.weight (256, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                         AddBackward0
                                                                                                                                                                                                                                                                                                                                                                          layer3.1.bn1.weight (256) (256)
                                                                                                                                                                                                                                                                                                                                         ReluBackward0
                                                                                                                                                                                                                                                                                                                                                         AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                    ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                              CudnnBatchNormBackward0 layer3.1.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                 ReluBackward0 | AccumulateGrad | layer3.1.bn2.weight (256) | layer3.1.bn2.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                 ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                         AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                         CudnnBatchNormBackward0 layer3.1.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                             ReluBackward0 | layer3.1.bn3.weight (1024) | layer3.1.bn3.bias (1024)
                                                                                                                                                                                                                                                                                                                                                                                                          ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                 AddBackward0 layer3.2.conv1.weight (256, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                   ReluBackward0 | layer3.2.bn1.weight (256) | layer3.2.bn1.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                   ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                CudnnBatchNormBackward0 layer3.2.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                     ReluBackward0 | layer3.2.bn2.weight | layer3.2.bn2.bias (256) | (256)
                                                                                                                                                                                                                                                                                                                                                                                                   ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                           AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                           CudnnBatchNormBackward0 layer3.2.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                ReluBackward0 | layer3.2.bn3.weight (1024) | layer3.2.bn3.bias (1024)
                                                                                                                                                                                                                                                                                                                                                                                                                             ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                             layer3.3.conv1.weight (256, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                  layer3.3.bn1.weight (256) (256)
                                                                                                                                                                                                                                                                                                                                                                                                            ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                    CudnnBatchNormBackward0 layer3.3.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                         AccumulateGrad layer3.3.bn2.weight (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                           onvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                          layer3.3.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                    ReluBackward0 | layer3.3.bn3.weight (1024) | layer3.3.bn3.bias (1024)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            CudnnBatchNormBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                           layer3.4.conv1.weight (256, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                         AddBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                         ReluBackward0 | layer3.4.bn1.weight (256) | layer3.4.bn1.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                       CudnnBatchNormBackward0 layer3.4.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                          ReluBackward0 | layer3.4.bn2.weight (256) | layer3.4.bn2.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                                          ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                CudnnBatchNormBackward0 layer3.4.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ReluBackward0 AccumulateGrad layer3.4.bn3.weight (1024) layer3.4.bn3.bias (1024)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CudnnBatchNormBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                    AddBackward0 layer3.5.conv1.weight (256, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                    ReluBackward0 AccumulateGrad layer3.5.bn1.weight (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              layer3.5.bn1.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                    ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CudnnBatchNormBackward0 layer3.5.conv2.weight (256, 256, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       AccumulateGrad layer3.5.bn2.weight (256) layer3.5.bn2.bias (256)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ReluBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ConvolutionBackward0 AccumulateGrad AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CudnnBatchNormBackward0 layer3.5.conv3.weight (1024, 256, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    layer3.5.bn3.weight (1024) (1024)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                             layer4.0.downsample.0.weight (2048, 1024, 1, 1)

AddBackward0 layer4.0.conv1.weight (512, 1024, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                              layer4.0.downsample.1.weight (2048) layer4.0.downsample.1.bias
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           AccumulateGrad layer4.0.bn1.weight (512) layer4.0.bn1.bias (512)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                          AccumulateGrad ReluBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                  AccumulateGrad ConvolutionBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                CudnnBatchNormBackward0 layer4.0.conv2.weight (512, 512, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CudnnBatchNormBackward0 layer4.0.conv3.weight (2048, 512, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ReluBackward0 | layer4.0.bn3.weight (2048) | layer4.0.bn3.bias (2048)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      AddBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ReluBackward0 | layer4.1.bn1.weight (512) | layer4.1.bn1.bias (512)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   CudnnBatchNormBackward0 layer4.1.conv2.weight (512, 512, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         layer4.1.bn2.weight (512) layer4.1.bn2.bias
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        ReluBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             CudnnBatchNormBackward0 layer4.1.conv3.weight (2048, 512, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  AccumulateGrad layer4.1.bn3.weight (2048) layer4.1.bn3.bias (2048)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         CudnnBatchNormBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 layer4.2.conv1.weight (512, 2048, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    layer4.2.bn1.weight (512) layer4.2.bn1.bias
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ReluBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ReluBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CudnnBatchNormBackward0 layer4.2.conv2.weight (512, 512, 3, 3)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              AccumulateGrad layer4.2.bn2.weight (512) layer4.2.bn2.bias (512)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ReluBackward0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         (512)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        ConvolutionBackward0 AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   AccumulateGrad
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  CudnnBatchNormBackward0 layer4.2.conv3.weight (2048, 512, 1, 1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       layer4.2.bn3.weight (2048) (2048)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ReluBackward0 AccumulateGrad
```

conv1.weight (64, 3, 7, 7)