

Animal_Classification-transfer-enet-s

June 13, 2024

1 Changing to transfer learning with efficientnet v2 s

From the previous model, we have swapped out the model for efficientnet v2 s that has been pretrained and tweaking it to use our data and classes.

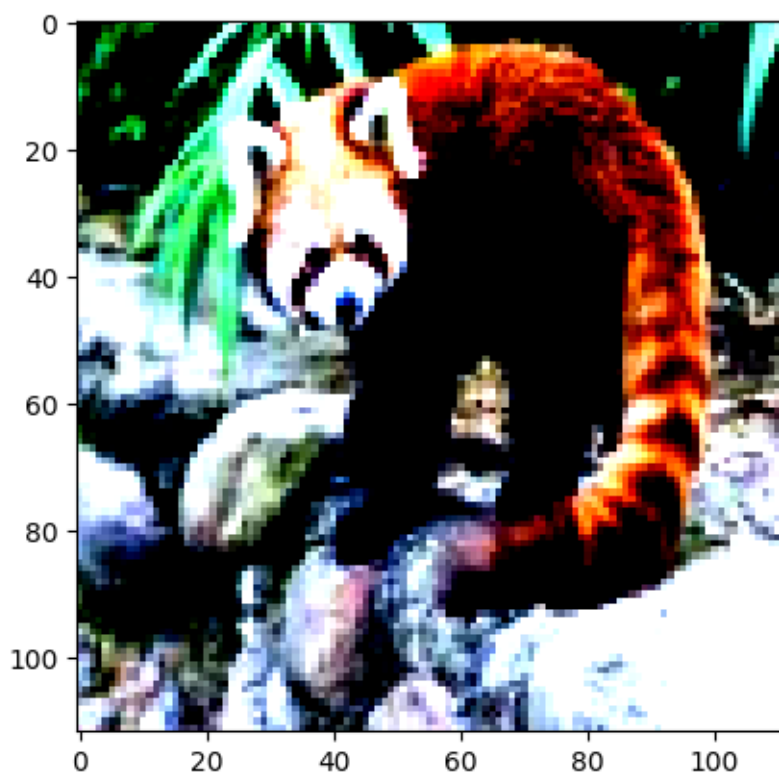
Validation Loss: 0.6571 Validation Accuracy: 93.44% FLOPS: 1.51G

Size of training dataset : 6270

```
torch.Size([3, 112, 112])
```

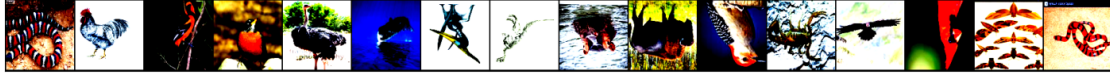
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).

Label: ailurus-fulgens (5)



(5330, 313, 627)

Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).



```
ConvolutionalNetwork(  
  (model): EfficientNet(  
    (features): Sequential(  
      (0): Conv2dNormActivation(  
        (0): Conv2d(3, 24, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),  
bias=False)  
        (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,  
track_running_stats=True)  
        (2): SiLU(inplace=True)  
      )  
      (1): Sequential(  
        (0): FusedMBConv(  
          (block): Sequential(  
            (0): Conv2dNormActivation(  
              (0): Conv2d(24, 24, kernel_size=(3, 3), stride=(1, 1), padding=(1,  
1), bias=False)  
              (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,  
track_running_stats=True)  
              (2): SiLU(inplace=True)  
            )  
          )  
          (stochastic_depth): StochasticDepth(p=0.0, mode=row)  
        )  
        (1): FusedMBConv(  
          (block): Sequential(  
            (0): Conv2dNormActivation(  
              (0): Conv2d(24, 24, kernel_size=(3, 3), stride=(1, 1), padding=(1,  
1), bias=False)  
              (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,  
track_running_stats=True)  
              (2): SiLU(inplace=True)  
            )  
          )  
          (stochastic_depth): StochasticDepth(p=0.005, mode=row)  
        )  
      )  
      (2): Sequential(  

```

```

(0): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(24, 96, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(96, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.01, mode=row)
)
(1): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.015000000000000003, mode=row)
)
(2): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)

```

```

        (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.02, mode=row)
)
(3): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.025, mode=row)
)
)
(3): Sequential(
  (0): FusedMBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), bias=False)
        (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(192, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      )
    )
  )
  (stochastic_depth): StochasticDepth(p=0.030000000000000006, mode=row)
)
(1): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(

```

```

        (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.035, mode=row)
)
(2): FusedMBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
            (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
    )
    (stochastic_depth): StochasticDepth(p=0.04, mode=row)
)
(3): FusedMBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
            (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```

```

        )
    )
    (stochastic_depth): StochasticDepth(p=0.045, mode=row)
)
)
(4): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(256, 256, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=256, bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(256, 16, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(16, 256, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
      )
      (3): Conv2dNormActivation(
        (0): Conv2d(256, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      )
    )
    (stochastic_depth): StochasticDepth(p=0.05, mode=row)
  )
  (1): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
  )
)

```

```

        (1): Conv2dNormActivation(
          (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
          (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.055000000000000001, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
          (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(

```

```

        (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.06000000000000001, mode=row)
)
(3): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.065, mode=row)
)
(4): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```



```

        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
        (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.07, mode=row)
)
(5): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
            (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
        )
    )
)

```

```

        )
        (3): Conv2dNormActivation(
          (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.075, mode=row)
    )
  )
  (5): Sequential(
    (0): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(128, 768, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(768, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(768, 768, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=768, bias=False)
          (1): BatchNorm2d(768, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(768, 32, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(32, 768, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(768, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.08, mode=row)
    )
  )
  (1): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(

```

```

        (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.085, mode=row)
)
(2): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)

```

```

        (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.09, mode=row)
)
(3): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
        )
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.095, mode=row)
)
(4): MBConv(

```

```

        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          )
        )
        (stochastic_depth): StochasticDepth(p=0.1, mode=row)
      )
    (5): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
    )

```

```

        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.10500000000000001, mode=row)
    )
    (6): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
      )
      (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      )
    )
    (stochastic_depth): StochasticDepth(p=0.11000000000000001, mode=row)

```

```

    )
    (7): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.11500000000000002, mode=row)
    )
    (8): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```

```

        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.12000000000000002, mode=row)
)
)
(6): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=960, bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,

```



```

track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.125, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
    )
    (stochastic_depth): StochasticDepth(p=0.13, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(

```

```

        (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
        (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.135, mode=row)
    )
    (3): MBCConv(
    (block): Sequential(
    (0): Conv2dNormActivation(
    (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),

```

```

bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.14, mode=row)
)
(4): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.14500000000000002, mode=row)
)
(5): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)

```

```

    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.15, mode=row)
)
(6): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
  )
)

```

```

        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.155, mode=row)
    )
    (7): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.16, mode=row)
    )
    (8): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,

```

```

track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
        (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.165, mode=row)
)
(9): MBCConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
        )
    )
)

```

```

        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.17, mode=row)
)
(10): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.175, mode=row)
)
(11): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),

```

```

bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.18, mode=row)
    )
    (12): MBCConv(
    (block): Sequential(
    (0): Conv2dNormActivation(
    (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))

```



```

        (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.185, mode=row)
)
(13): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
            (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
    )
    )
    (stochastic_depth): StochasticDepth(p=0.19, mode=row)
)
(14): MBConv(
    (block): Sequential(

```

```

        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.195, mode=row)
    )
  )
  (7): Conv2dNormActivation(
    (0): Conv2d(256, 1280, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(1280, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
  )
)
(avgpool): AdaptiveAvgPool2d(output_size=1)
(classifier): Sequential(
  (0): Dropout(p=0.2, inplace=True)
  (1): Linear(in_features=1280, out_features=151, bias=True)
)
)
)

images.shape: torch.Size([16, 3, 112, 112])
out.shape: torch.Size([16, 151])

```

```

out[0]: tensor([-4.7994, -5.1829, -5.0027, -4.9127, -4.6998, -4.8673, -5.1724,
-4.9495,
-4.7703, -4.9698, -5.1741, -5.1912, -5.0215, -4.8155, -5.0213, -5.2369,
-4.8082, -5.2048, -5.1846, -4.8295, -4.9353, -5.2770, -4.9934, -5.1266,
-4.6548, -4.9796, -5.0709, -5.0723, -4.7300, -5.0411, -5.0543, -4.9714,
-5.1474, -4.8505, -4.9984, -5.0507, -4.8566, -5.0439, -4.9996, -5.1005,
-5.1156, -5.0616, -5.0866, -5.0087, -5.2058, -5.1470, -4.8633, -5.0163,
-4.9701, -4.8954, -4.9969, -5.2226, -5.0340, -5.0524, -5.0602, -4.9395,
-5.2437, -4.9152, -4.7581, -4.9960, -5.0949, -5.0790, -4.9998, -5.0557,
-5.1472, -5.0925, -5.0727, -4.8898, -4.9881, -4.9445, -5.0515, -5.1039,
-4.9463, -4.9599, -4.9802, -5.0435, -5.2297, -4.9317, -5.1567, -4.9077,
-5.1739, -4.7727, -4.9432, -4.9620, -5.1138, -5.1272, -4.9757, -5.2533,
-5.0020, -4.8290, -5.0521, -5.1993, -4.9562, -5.3153, -4.9967, -5.0638,
-4.8776, -5.0088, -4.6869, -5.0928, -5.4245, -5.0838, -5.0627, -5.0189,
-5.1908, -5.1645, -4.9533, -5.1461, -5.1133, -5.0064, -5.1321, -4.9497,
-5.1428, -5.0813, -4.9845, -5.0299, -4.8336, -4.9603, -5.1759, -5.0610,
-4.9300, -4.6662, -4.9472, -4.9394, -5.1029, -4.7909, -5.1264, -5.1141,
-4.9729, -4.8760, -5.1422, -5.1973, -4.8764, -5.1530, -5.2282, -4.9745,
-5.1271, -5.0359, -4.9405, -5.0989, -5.3952, -5.2781, -4.9934, -4.8897,
-5.0299, -5.2391, -5.0535, -4.9889, -5.0309, -4.9962, -4.9986],
device='cuda:0', grad_fn=<SelectBackward0>)

```

```

ConvolutionalNetwork(
  (model): EfficientNet(
    (features): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(3, 24, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),
bias=False)
        (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Sequential(
        (0): FusedMBConv(
          (block): Sequential(
            (0): Conv2dNormActivation(
              (0): Conv2d(24, 24, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
              (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
              (2): SiLU(inplace=True)
            )
          )
          (stochastic_depth): StochasticDepth(p=0.0, mode=row)
        )
        (1): FusedMBConv(
          (block): Sequential(

```

```

        (0): Conv2dNormActivation(
          (0): Conv2d(24, 24, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
          (1): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.005, mode=row)
    )
  (2): Sequential(
    (0): FusedMBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 96, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), bias=False)
          (1): BatchNorm2d(96, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(96, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.01, mode=row)
    )
    (1): FusedMBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
          (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.015000000000000003, mode=row)
    )
  )

```

```

(2): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.02, mode=row)
)
(3): FusedMBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(192, 48, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.025, mode=row)
)
)
(3): Sequential(
  (0): FusedMBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(48, 192, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), bias=False)
        (1): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
  )

```

```

        (1): Conv2dNormActivation(
          (0): Conv2d(192, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.030000000000000006, mode=row)
    )
    (1): FusedMBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.035, mode=row)
    )
    (2): FusedMBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.04, mode=row)
    )
    (3): FusedMBConv(
      (block): Sequential(

```

```

        (0): Conv2dNormActivation(
          (0): Conv2d(64, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.045, mode=row)
    )
  (4): Sequential(
    (0): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(256, 256, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=256, bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(256, 16, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(16, 256, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(256, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
    )
  )

```

```

    )
    (stochastic_depth): StochasticDepth(p=0.05, mode=row)
)
(1): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.055000000000000001, mode=row)
)
(2): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)

```



```

        (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.060000000000000001, mode=row)
    )
    (3): MBConv(
        (block): Sequential(
            (0): Conv2dNormActivation(
                (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
                (2): SiLU(inplace=True)
            )
            (1): Conv2dNormActivation(
                (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
                (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
                (2): SiLU(inplace=True)
            )
            (2): SqueezeExcitation(
                (avgpool): AdaptiveAvgPool2d(output_size=1)
                (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
                (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
                (activation): SiLU(inplace=True)
                (scale_activation): Sigmoid()
            )
            (3): Conv2dNormActivation(
                (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,

```

```

track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.065, mode=row)
    )
    (4): MBConv(
        (block): Sequential(
            (0): Conv2dNormActivation(
                (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
                (2): SiLU(inplace=True)
            )
            (1): Conv2dNormActivation(
                (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
                (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
                (2): SiLU(inplace=True)
            )
            (2): SqueezeExcitation(
                (avgpool): AdaptiveAvgPool2d(output_size=1)
                (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
                (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
                (activation): SiLU(inplace=True)
                (scale_activation): Sigmoid()
            )
            (3): Conv2dNormActivation(
                (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            )
        )
    (stochastic_depth): StochasticDepth(p=0.07, mode=row)
    )
    (5): MBConv(
        (block): Sequential(
            (0): Conv2dNormActivation(
                (0): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
                (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
                (2): SiLU(inplace=True)
            )
            (1): Conv2dNormActivation(

```

```

        (0): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=512, bias=False)
        (1): BatchNorm2d(512, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(32, 512, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.075, mode=row)
)
)
(5): Sequential(
  (0): MBCConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(128, 768, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(768, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(768, 768, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=768, bias=False)
        (1): BatchNorm2d(768, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(768, 32, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(32, 768, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
      )
    )
  )
)

```

```

        (3): Conv2dNormActivation(
          (0): Conv2d(768, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.08, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.085, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,

```

```

track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.09, mode=row)
)
(3): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
    )
  )
)

```

```

        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.095, mode=row)
)
(4): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.1, mode=row)
)
(5): MBCConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),

```

```

bias=False)
    (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
    (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.10500000000000001, mode=row)
    )
    (6): MBConv(
    (block): Sequential(
    (0): Conv2dNormActivation(
    (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
    (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))

```

```

        (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.11000000000000001, mode=row)
)
(7): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
            (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
            (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
    )
    )
    (stochastic_depth): StochasticDepth(p=0.11500000000000002, mode=row)
)
(8): MBConv(
    (block): Sequential(

```



```

        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
    (stochastic_depth): StochasticDepth(p=0.12000000000000002, mode=row)
  )
(6): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(960, 960, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=960, bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
  )
)

```

```

    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.125, mode=row)
)
(1): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
)

```

```

        (stochastic_depth): StochasticDepth(p=0.13, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.135, mode=row)
    )
    (3): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,

```

```

track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.14, mode=row)
  )
  (4): MBCConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
        (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```

```

    )
    )
    (stochastic_depth): StochasticDepth(p=0.14500000000000002, mode=row)
)
(5): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.15, mode=row)
)
(6): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),

```

```

padding=(1, 1), groups=1536, bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.155, mode=row)
    )
    (7): MBConv(
    (block): Sequential(
    (0): Conv2dNormActivation(
    (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
    (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
    (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
    (activation): SiLU(inplace=True)
    (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
    (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)

```

```

        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.16, mode=row)
)
(8): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.165, mode=row)
)
(9): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )

```

```

        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.17, mode=row)
    )
    (10): MBCConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(

```



```

        (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.175, mode=row)
)
(11): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.18, mode=row)
)
(12): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```

```

        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
        (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.185, mode=row)
)
(13): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
            (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
        )
    )
)

```

```

        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.19, mode=row)
    )
    (14): MBCConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(256, 1536, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1536, 1536, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1536, bias=False)
          (1): BatchNorm2d(1536, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1536, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 1536, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1536, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.195, mode=row)
    )
  )
  (7): Conv2dNormActivation(
    (0): Conv2d(256, 1280, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(1280, eps=0.001, momentum=0.1, affine=True,
track_running_stats=True)

```

```

        (2): SiLU(inplace=True)
    )
)
(avgpool): AdaptiveAvgPool2d(output_size=1)
(classifier): Sequential(
  (0): Dropout(p=0.2, inplace=True)
  (1): Linear(in_features=1280, out_features=151, bias=True)
)
)
)

```

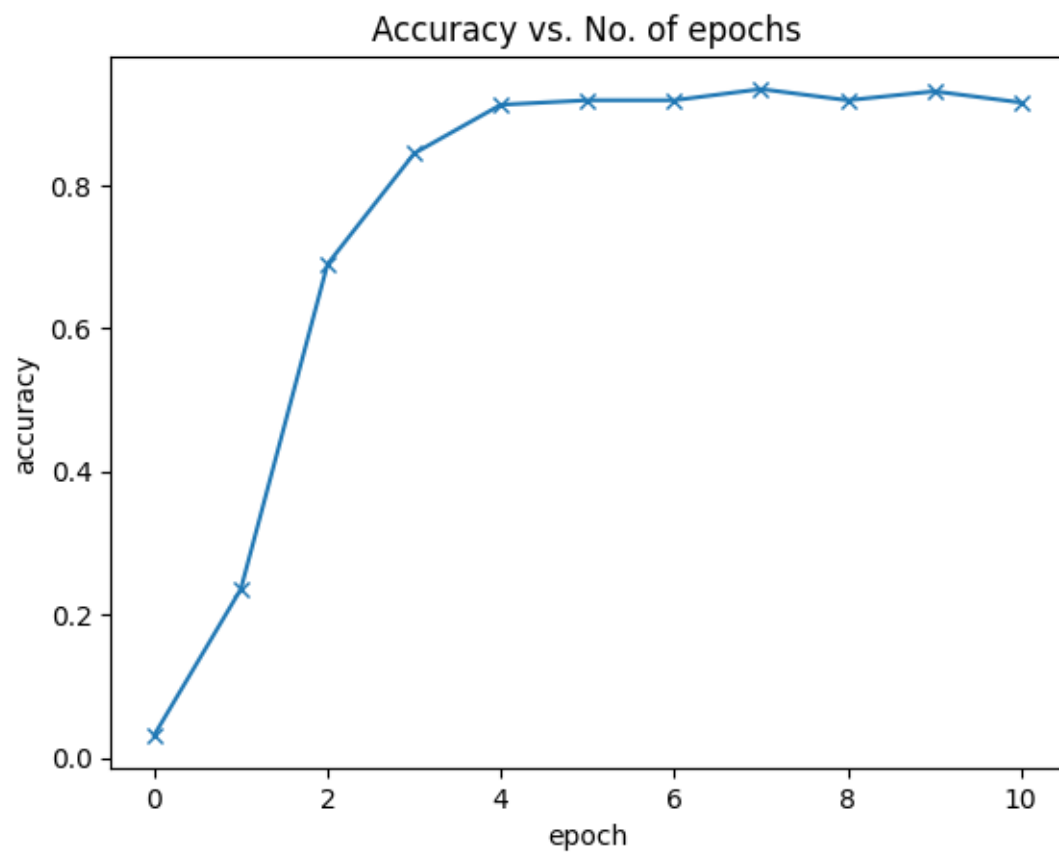
```
[{'val_loss': 5.047817230224609, 'val_acc': 0.03125}]
```

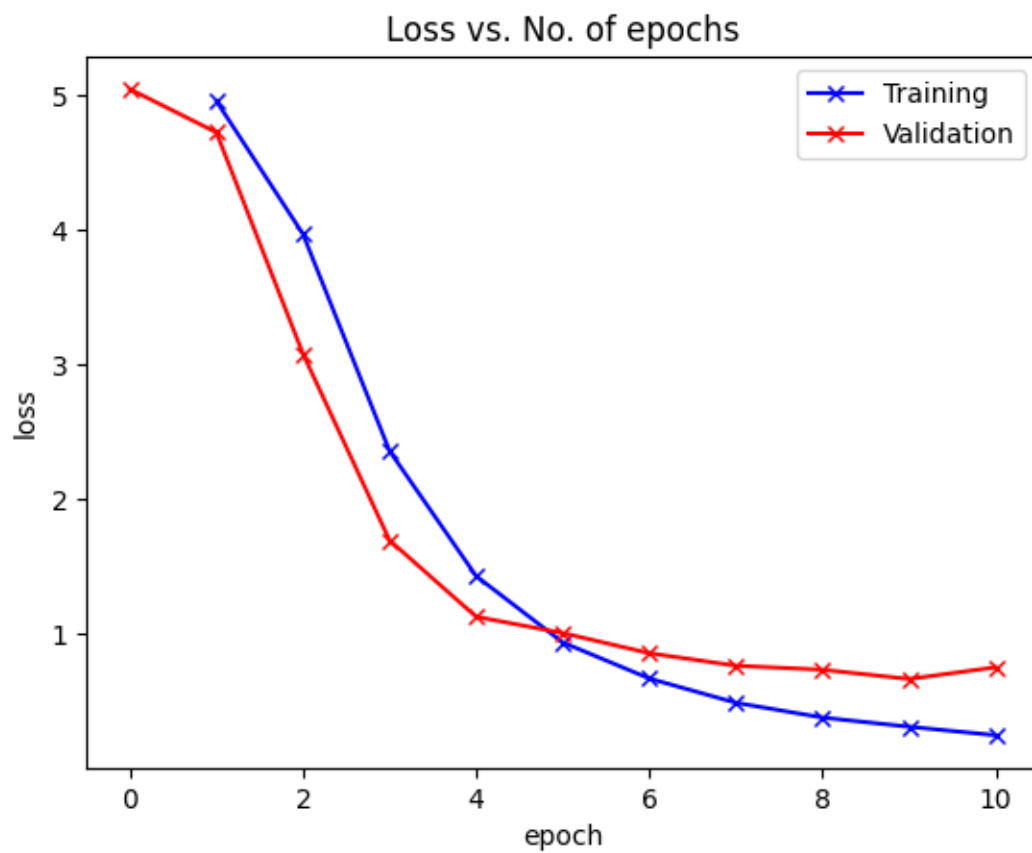
627

```

0%|          | 0/334 [00:00<?, ?it/s]
Epoch [0], train_loss: 4.9595, val_loss: 4.7259, val_acc: 0.2354
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [1], train_loss: 3.9699, val_loss: 3.0797, val_acc: 0.6896
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [2], train_loss: 2.3567, val_loss: 1.6923, val_acc: 0.8444
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [3], train_loss: 1.4267, val_loss: 1.1240, val_acc: 0.9125
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [4], train_loss: 0.9333, val_loss: 1.0034, val_acc: 0.9187
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [5], train_loss: 0.6650, val_loss: 0.8530, val_acc: 0.9187
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [6], train_loss: 0.4843, val_loss: 0.7602, val_acc: 0.9344
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [7], train_loss: 0.3747, val_loss: 0.7302, val_acc: 0.9187
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [8], train_loss: 0.3062, val_loss: 0.6615, val_acc: 0.9313
0%|          | 0/334 [00:00<?, ?it/s]
Epoch [9], train_loss: 0.2433, val_loss: 0.7486, val_acc: 0.9156

```





```
{'val_loss': 0.6571369171142578, 'val_acc': 0.934374988079071}
```

1.1 FLOPs

+ Number of FLOPs: 1.51G