

# Animal\_Classification-transfer-enet

June 13, 2024

## 1 Changing to transfer learning with efficientnet b0

From the previous model, we have swapped out the model for efficientnet b0 that has been pre-trained and tweaking it to use our data and classes.

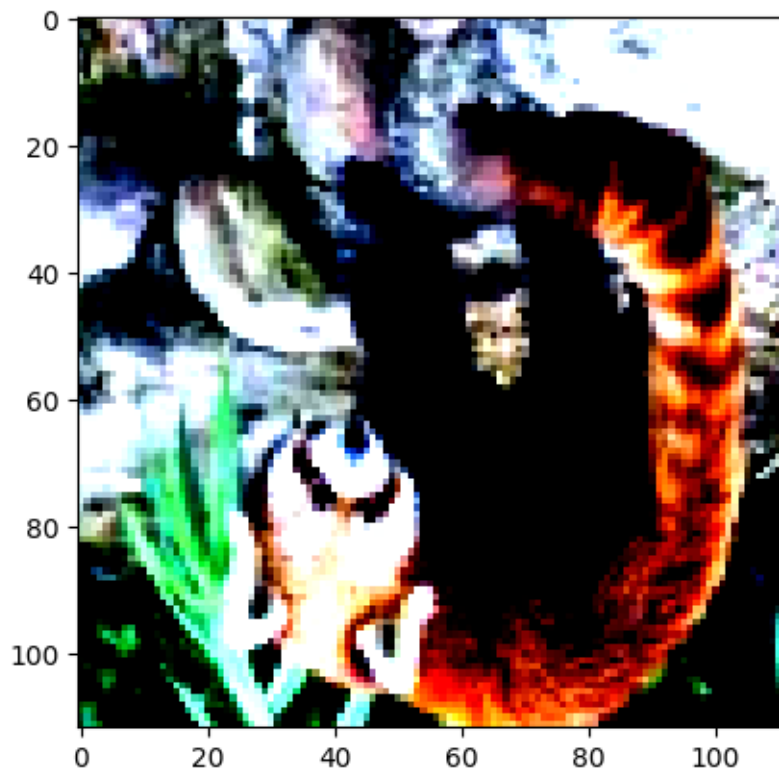
Validation Loss: 0.8605 Validation Accuracy: 91.30% FLOPS: 0.21G

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, 32, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),  
bias=False)  
        (1): BatchNorm2d(32, eps=1e-05, momentum=0.1, affine=True,  
track_running_stats=True)  
        (2): SiLU(inplace=True)  
      )  
      (1): Sequential(  
        (0): MBConv(  
          (block): Sequential(  
            (0): Conv2dNormActivation(  
              (0): Conv2d(32, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1,  
1), groups=32, bias=False)  
              (1): BatchNorm2d(32, eps=1e-05, momentum=0.1, affine=True,  
track_running_stats=True)  
              (2): SiLU(inplace=True)  
            )  
            (1): SqueezeExcitation(  
              (avgpool): AdaptiveAvgPool2d(output_size=1)  
              (fc1): Conv2d(32, 8, kernel_size=(1, 1), stride=(1, 1))  
              (fc2): Conv2d(8, 32, kernel_size=(1, 1), stride=(1, 1))  
              (activation): SiLU(inplace=True)  
              (scale_activation): Sigmoid()  
            )  
            (2): Conv2dNormActivation(  
              (0): Conv2d(32, 16, kernel_size=(1, 1), stride=(1, 1), bias=False)  
              (1): BatchNorm2d(16, eps=1e-05, momentum=0.1, affine=True,  
track_running_stats=True)  
            )  
          )  
        )  
        (stochastic_depth): StochasticDepth(p=0.0, mode=row)  
      )  
    )  
  )  
)
```

```

    )
    (2): Sequential(
      (0): MBConv(
        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(16, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(96, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(96, 96, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=96, bias=False)
            (1): BatchNorm2d(96, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(96, 4, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(4, 96, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(96, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(24, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          )
        )
      )
      (stochastic_depth): StochasticDepth(p=0.0125, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 144, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(144, 144, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=144, bias=False)
          (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
    )
  )
)

```

```

    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(144, 6, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(6, 144, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(144, 24, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(24, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.025, mode=row)
)
)
(3): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(24, 144, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(144, 144, kernel_size=(5, 5), stride=(2, 2),
padding=(2, 2), groups=144, bias=False)
        (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(144, 6, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(6, 144, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(144, 40, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(40, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)

```

```

        )
    )
    (stochastic_depth): StochasticDepth(p=0.037500000000000006, mode=row)
)
(1): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(240, 240, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=240, bias=False)
      (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(240, 10, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(10, 240, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(240, 40, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(40, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.05, mode=row)
)
(4): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )

```

```

        (1): Conv2dNormActivation(
          (0): Conv2d(240, 240, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=240, bias=False)
          (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(240, 10, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(10, 240, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(240, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.0625, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(480, 480, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=480, bias=False)
          (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(

```

```

        (0): Conv2d(480, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.07500000000000001, mode=row)
)
(2): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(480, 480, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=480, bias=False)
      (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(480, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.08750000000000001, mode=row)
)
(5): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)

```

```

        (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(480, 480, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=480, bias=False)
        (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(480, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.1, mode=row)
)
(1): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=672, bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))

```



```

        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.1125, mode=row)
)
(2): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=672, bias=False)
      (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.125, mode=row)
)
(6): Sequential(
  (0): MBConv(

```

```

        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(2, 2),
padding=(2, 2), groups=672, bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(672, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          )
        )
        (stochastic_depth): StochasticDepth(p=0.1375, mode=row)
      )
      (1): MBConv(
        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
            (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
        )
      )

```

```

        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.15000000000000002, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
      )
      (3): Conv2dNormActivation(
        (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      )
    )
    (stochastic_depth): StochasticDepth(p=0.1625, mode=row)

```

```

    )
    (3): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
    )
    (stochastic_depth): StochasticDepth(p=0.17500000000000002, mode=row)
  )
)
(7): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(1152, 1152, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1152, bias=False)

```

```

        (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1152, 320, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(320, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.1875, mode=row)
)
)
(8): Conv2dNormActivation(
  (0): Conv2d(320, 1280, kernel_size=(1, 1), stride=(1, 1), bias=False)
  (1): BatchNorm2d(1280, eps=1e-05, 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([8, 3, 112, 112])
```

```
out.shape: torch.Size([8, 151])
```

```
out[0]: tensor([-5.3306, -5.1478, -4.6661, -4.6892, -4.8743, -5.4414, -5.1203,
-5.2616,
-5.0763, -5.2150, -4.8569, -4.2328, -4.2651, -4.9740, -5.8967, -4.8240,
-4.7153, -5.2438, -5.9391, -4.8895, -4.1349, -5.4026, -4.4439, -5.4200,
-4.7755, -5.4014, -4.2853, -5.0213, -5.3200, -5.0363, -4.8848, -5.0683,
-4.9752, -4.8050, -5.2937, -4.3433, -5.4060, -4.4473, -4.7193, -5.4616,
-4.9635, -5.3621, -5.7129, -5.0883, -4.8706, -5.0065, -4.6918, -4.7496,
-4.8032, -5.0112, -4.9375, -5.1954, -5.3248, -4.8421, -5.0990, -5.6386,
-5.3078, -5.0376, -5.0106, -5.0876, -5.7190, -5.3046, -5.2521, -4.9910,
-5.1990, -5.4892, -4.5919, -5.4890, -5.5808, -5.3988, -5.6106, -5.2343,
```

```

-5.4663, -4.3562, -5.1467, -5.4533, -5.1481, -5.3110, -4.9424, -5.7410,
-5.0600, -4.8960, -4.8630, -4.1390, -4.9727, -4.9667, -4.8241, -5.3833,
-4.6636, -4.5892, -5.3594, -5.4599, -5.1801, -5.2848, -5.1190, -4.6956,
-4.9888, -5.2416, -4.7839, -5.1666, -5.3740, -4.7411, -5.1656, -4.9770,
-6.0877, -4.8717, -5.1582, -4.1725, -4.7421, -4.4821, -5.4239, -5.0960,
-5.3757, -4.8542, -4.7736, -5.5021, -5.3699, -5.4960, -4.9947, -4.6696,
-5.9240, -4.9215, -4.9380, -5.1339, -5.5515, -5.0422, -5.4240, -4.8855,
-4.8572, -5.9668, -4.8704, -5.6482, -5.2565, -5.0988, -5.1251, -5.1164,
-5.1848, -4.5638, -5.3515, -5.2576, -5.5184, -5.0843, -5.3026, -4.8263,
-4.7056, -4.9838, -5.7830, -5.6214, -5.4560, -4.9965, -4.8960],
device='cuda:0', grad_fn=<SelectBackward0>)

```

```

ConvolutionalNetwork(
  (model): EfficientNet(
    (features): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(3, 32, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),
bias=False)
        (1): BatchNorm2d(32, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Sequential(
        (0): MBConv(
          (block): Sequential(
            (0): Conv2dNormActivation(
              (0): Conv2d(32, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=32, bias=False)
              (1): BatchNorm2d(32, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
              (2): SiLU(inplace=True)
            )
            (1): SqueezeExcitation(
              (avgpool): AdaptiveAvgPool2d(output_size=1)
              (fc1): Conv2d(32, 8, kernel_size=(1, 1), stride=(1, 1))
              (fc2): Conv2d(8, 32, kernel_size=(1, 1), stride=(1, 1))
              (activation): SiLU(inplace=True)
              (scale_activation): Sigmoid()
            )
            (2): Conv2dNormActivation(
              (0): Conv2d(32, 16, kernel_size=(1, 1), stride=(1, 1), bias=False)
              (1): BatchNorm2d(16, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            )
          )
        )
        (stochastic_depth): StochasticDepth(p=0.0, mode=row)
      )
    )
  )

```

```

    )
    (2): Sequential(
      (0): MBConv(
        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(16, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(96, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(96, 96, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=96, bias=False)
            (1): BatchNorm2d(96, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(96, 4, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(4, 96, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(96, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(24, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          )
        )
      )
      (stochastic_depth): StochasticDepth(p=0.0125, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 144, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(144, 144, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=144, bias=False)
          (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
    )
  )
)

```

```

    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(144, 6, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(6, 144, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(144, 24, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(24, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.025, mode=row)
)
)
(3): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(24, 144, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(144, 144, kernel_size=(5, 5), stride=(2, 2),
padding=(2, 2), groups=144, bias=False)
        (1): BatchNorm2d(144, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(144, 6, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(6, 144, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(144, 40, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(40, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)

```



```

    )
    )
    (stochastic_depth): StochasticDepth(p=0.037500000000000006, mode=row)
)
(1): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(240, 240, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=240, bias=False)
      (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(240, 10, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(10, 240, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(240, 40, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(40, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.05, mode=row)
)
(4): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )

```

```

        (1): Conv2dNormActivation(
          (0): Conv2d(240, 240, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), groups=240, bias=False)
          (1): BatchNorm2d(240, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(240, 10, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(10, 240, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(240, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.0625, mode=row)
    )
    (1): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(480, 480, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=480, bias=False)
          (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(

```

```

        (0): Conv2d(480, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.07500000000000001, mode=row)
)
(2): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(480, 480, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=480, bias=False)
      (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(480, 80, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(80, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.08750000000000001, mode=row)
)
(5): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1),
bias=False)

```

```

        (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(480, 480, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=480, bias=False)
        (1): BatchNorm2d(480, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(480, 20, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(20, 480, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(480, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
    )
    (stochastic_depth): StochasticDepth(p=0.1, mode=row)
)
(1): MBConv(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=672, bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
        )
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))

```

```

        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.1125, mode=row)
)
(2): MBConv(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=672, bias=False)
      (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(112, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.125, mode=row)
)
(6): Sequential(
  (0): MBConv(

```

```

        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(2, 2),
padding=(2, 2), groups=672, bias=False)
            (1): BatchNorm2d(672, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(672, 28, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(28, 672, kernel_size=(1, 1), stride=(1, 1))
            (activation): SiLU(inplace=True)
            (scale_activation): Sigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(672, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          )
        )
        (stochastic_depth): StochasticDepth(p=0.1375, mode=row)
      )
      (1): MBConv(
        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
            (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
            (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
            (2): SiLU(inplace=True)
          )
        )
      )

```

```

        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
      (stochastic_depth): StochasticDepth(p=0.15000000000000002, mode=row)
    )
    (2): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
        (activation): SiLU(inplace=True)
        (scale_activation): Sigmoid()
      )
      (3): Conv2dNormActivation(
        (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
      )
    )
    (stochastic_depth): StochasticDepth(p=0.1625, mode=row)

```

```

    )
    (3): MBConv(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(1152, 1152, kernel_size=(5, 5), stride=(1, 1),
padding=(2, 2), groups=1152, bias=False)
          (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
          (2): SiLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
          (activation): SiLU(inplace=True)
          (scale_activation): Sigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(1152, 192, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (1): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        )
      )
    )
    (stochastic_depth): StochasticDepth(p=0.17500000000000002, mode=row)
  )
)
(7): Sequential(
  (0): MBConv(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(192, 1152, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(1152, 1152, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), groups=1152, bias=False)

```



```

        (1): BatchNorm2d(1152, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
        (2): SiLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(1152, 48, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(48, 1152, kernel_size=(1, 1), stride=(1, 1))
      (activation): SiLU(inplace=True)
      (scale_activation): Sigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(1152, 320, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (1): BatchNorm2d(320, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    )
  )
  (stochastic_depth): StochasticDepth(p=0.1875, mode=row)
)
)
(8): Conv2dNormActivation(
  (0): Conv2d(320, 1280, kernel_size=(1, 1), stride=(1, 1), bias=False)
  (1): BatchNorm2d(1280, eps=1e-05, 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': 5.034629821777344, 'val_acc': 0.01875000074505806}]
```

627

```
0%|          | 0/667 [00:00<?, ?it/s]
```

Epoch [0], train\_loss: 4.4882, val\_loss: 3.5946, val\_acc: 0.5719

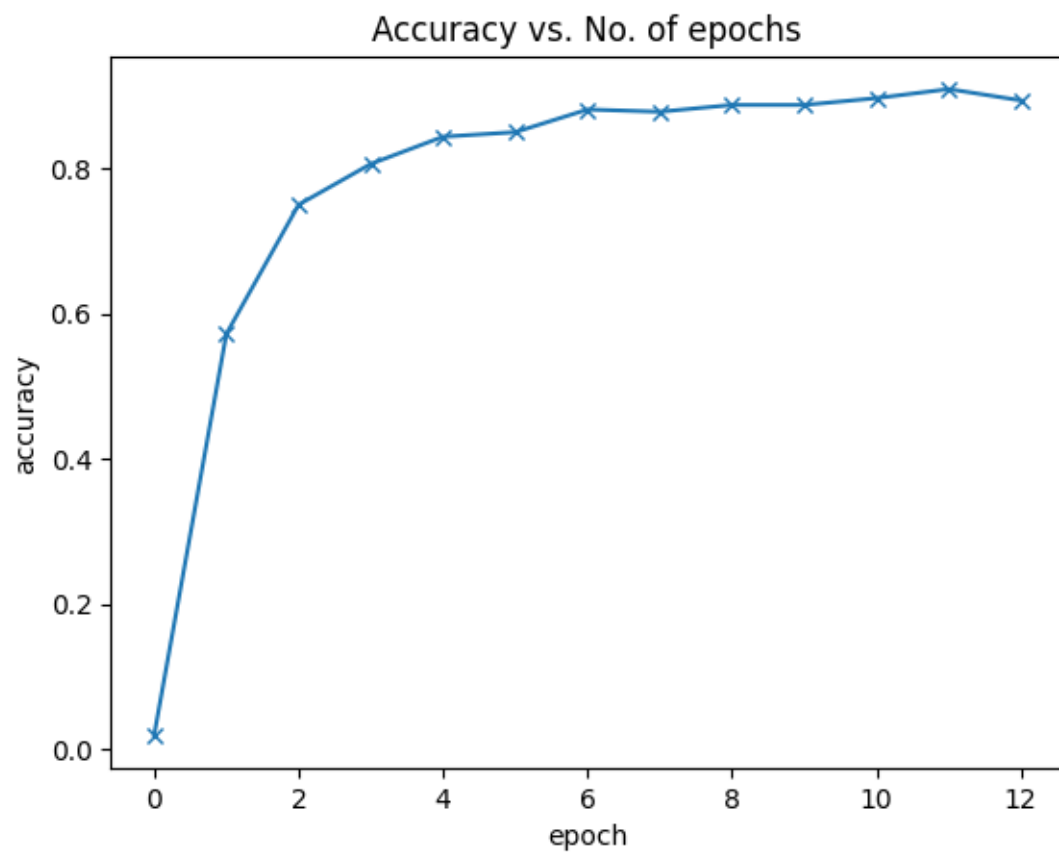
```
0%|          | 0/667 [00:00<?, ?it/s]
```

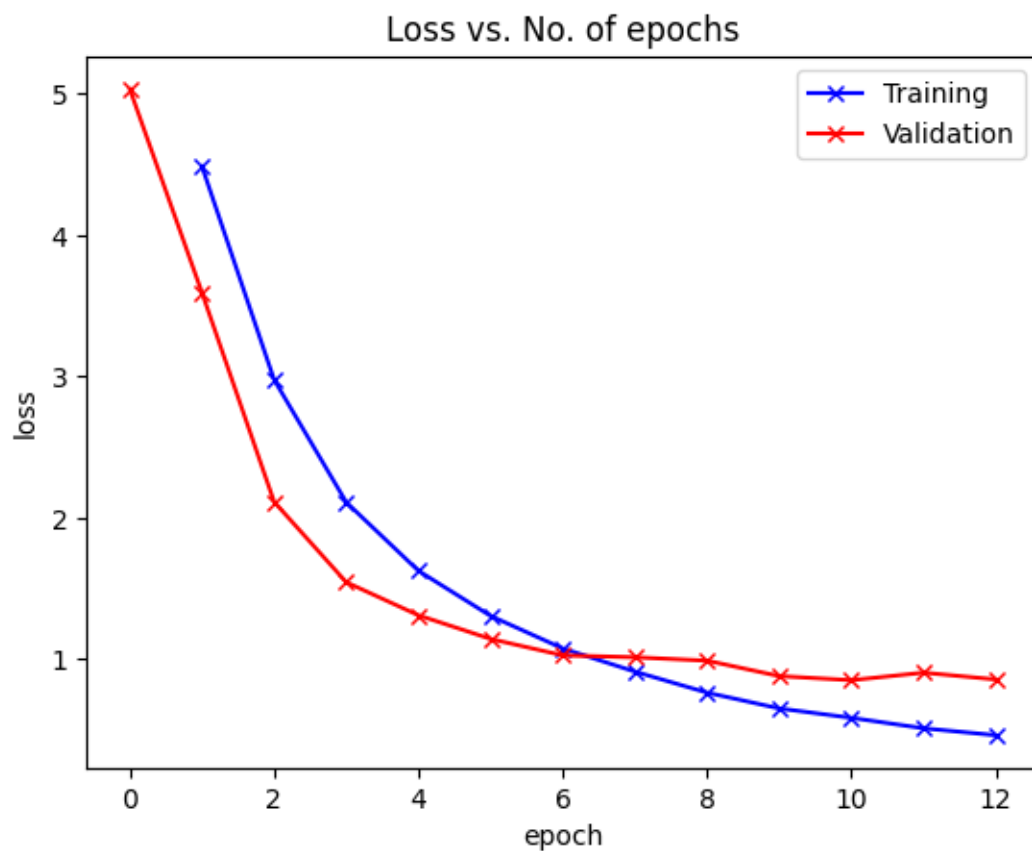
Epoch [1], train\_loss: 2.9778, val\_loss: 2.1140, val\_acc: 0.7500

```
0%|          | 0/667 [00:00<?, ?it/s]
```

Epoch [2], train\_loss: 2.1107, val\_loss: 1.5449, val\_acc: 0.8063

```
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [3], train_loss: 1.6258, val_loss: 1.3110, val_acc: 0.8438
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [4], train_loss: 1.3064, val_loss: 1.1435, val_acc: 0.8500
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [5], train_loss: 1.0760, val_loss: 1.0273, val_acc: 0.8813
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [6], train_loss: 0.9121, val_loss: 1.0137, val_acc: 0.8781
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [7], train_loss: 0.7634, val_loss: 0.9898, val_acc: 0.8875
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [8], train_loss: 0.6517, val_loss: 0.8805, val_acc: 0.8875
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [9], train_loss: 0.5850, val_loss: 0.8516, val_acc: 0.8969
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [10], train_loss: 0.5112, val_loss: 0.9048, val_acc: 0.9094
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [11], train_loss: 0.4621, val_loss: 0.8567, val_acc: 0.8938
```





```
{'val_loss': 0.8605183959007263, 'val_acc': 0.9129747152328491}
```

## 1.1 FLOPs

+ Number of FLOPs: 0.21G