

# Animal\_Classification-transfer-mobile-v3-large

June 13, 2024

## 1 Changing to transfer learning with Mobile Net V3 Large

From the previous model, we have swapped out the model for Mobile Net V3 Large that has been pretrained and tweaking it to use our data and classes.

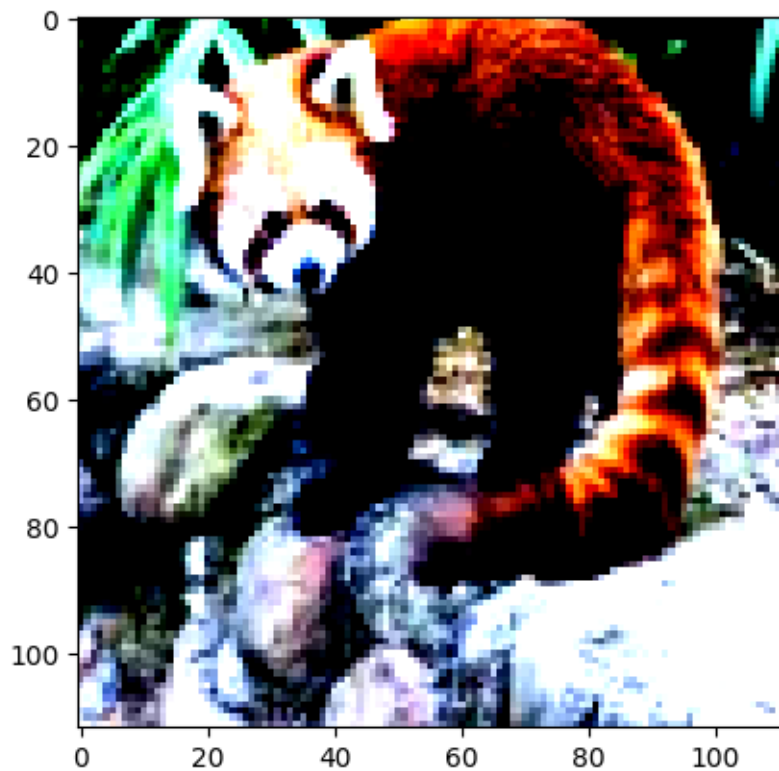
Validation Loss: 1.1079 Validation Accuracy: 89.72% FLOPS: 0.12G

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): MobileNetV3(  
    (features): Sequential(  
      (0): Conv2dNormActivation(  
        (0): Conv2d(3, 16, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),  
bias=False)  
        (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,  
track_running_stats=True)  
        (2): Hardswish()  
      )  
      (1): InvertedResidual(  
        (block): Sequential(  
          (0): Conv2dNormActivation(  
            (0): Conv2d(16, 16, kernel_size=(3, 3), stride=(1, 1), padding=(1,  
1), groups=16, bias=False)  
            (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,  
track_running_stats=True)  
            (2): ReLU(inplace=True)  
          )  
          (1): Conv2dNormActivation(  
            (0): Conv2d(16, 16, kernel_size=(1, 1), stride=(1, 1), bias=False)  
            (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,  
track_running_stats=True)  
          )  
        )  
      )  
      (2): InvertedResidual(  
        (block): Sequential(  
          (0): Conv2dNormActivation(  
            (0): Conv2d(16, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)  
            (1): BatchNorm2d(64, eps=0.001, momentum=0.01, affine=True,  
track_running_stats=True)  
            (2): ReLU(inplace=True)  
          )  
          (1): Conv2dNormActivation(  

```

```

        (0): Conv2d(64, 64, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=64, bias=False)
        (1): BatchNorm2d(64, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): ReLU(inplace=True)
    )
    (2): Conv2dNormActivation(
        (0): Conv2d(64, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(24, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
)
(3): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(72, 72, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=72, bias=False)
      (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(72, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(24, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(4): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(72, 72, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=72, bias=False)
      (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    (2): ReLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(72, 24, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(72, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    )
    )
    (5): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(40, 120, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
        (2): ReLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(120, 120, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=120, bias=False)
          (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
        (2): ReLU(inplace=True)
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(120, 32, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(32, 120, kernel_size=(1, 1), stride=(1, 1))
          (activation): ReLU()
          (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(120, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    )
    (6): InvertedResidual(

```

```

        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(40, 120, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): ReLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(120, 120, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=120, bias=False)
            (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): ReLU(inplace=True)
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(120, 32, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(32, 120, kernel_size=(1, 1), stride=(1, 1))
            (activation): ReLU()
            (scale_activation): Hardsigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(120, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          )
        )
      )
    (7): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(240, 240, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=240, bias=False)
          (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(240, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)

```

```

    )
    )
    )
    (8): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(80, 200, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(200, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(200, 200, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=200, bias=False)
          (1): BatchNorm2d(200, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(200, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    )
    (9): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(80, 184, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(184, 184, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=184, bias=False)
          (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(184, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    )

```

```

(10): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 184, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(184, 184, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=184, bias=False)
      (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(184, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(11): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 480, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(480, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(480, 480, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=480, bias=False)
      (1): BatchNorm2d(480, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(480, 120, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(120, 480, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(480, 112, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(112, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    )
    )
    )
    (12): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(672, 672, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=672, bias=False)
          (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(672, 168, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(168, 672, kernel_size=(1, 1), stride=(1, 1))
          (activation): ReLU()
          (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(112, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    )
    (13): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=672, bias=False)
          (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()

```



```

    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(672, 168, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(168, 672, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(14): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
  )
  (1): Conv2dNormActivation(
    (0): Conv2d(960, 960, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=960, bias=False)
    (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
  )
  (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(960, 240, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(240, 960, kernel_size=(1, 1), stride=(1, 1))
    (activation): ReLU()
    (scale_activation): Hardsigmoid()
  )
  (3): Conv2dNormActivation(
    (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
  )
)
)
(15): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(

```

```

        (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(960, 960, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=960, bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(960, 240, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(240, 960, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    )
    (16): Conv2dNormActivation(
        (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
    )
    )
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (classifier): Sequential(
        (0): Linear(in_features=960, out_features=1280, bias=True)
        (1): Hardswish()
        (2): Dropout(p=0.2, inplace=True)
        (3): 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([-4.7812, -4.4831, -5.3164, -5.1257, -4.5402, -5.5535, -4.8652,  
-5.8874,  
-5.0361, -4.3467, -5.1253, -4.2328, -5.1439, -5.7179, -4.6263, -5.0127,

```

-4.7240, -4.5692, -5.9683, -5.2047, -5.8117, -5.1798, -5.3363, -5.0413,
-5.1740, -4.7090, -5.5426, -4.7905, -5.1183, -5.6845, -5.2732, -4.7975,
-5.8939, -5.3481, -5.0668, -5.2674, -5.8295, -5.5214, -5.5249, -4.5756,
-4.9349, -4.3446, -5.0958, -5.7592, -5.0068, -4.8165, -5.7177, -5.6110,
-5.9156, -5.3247, -5.0197, -4.9382, -5.1333, -5.6891, -5.0468, -4.6456,
-5.6181, -5.3155, -5.4917, -4.9752, -4.8466, -5.4030, -5.6401, -4.9278,
-4.8315, -5.3491, -5.7371, -4.6582, -4.7203, -5.2830, -4.7965, -5.4060,
-6.0145, -5.0599, -4.6089, -4.8897, -5.7236, -4.0639, -5.2115, -5.1297,
-5.0582, -5.1172, -4.7781, -5.1480, -4.8607, -4.8927, -4.0642, -4.6462,
-5.7286, -5.5381, -5.2825, -4.1856, -5.3057, -5.7258, -4.8991, -4.7629,
-4.9660, -4.9021, -5.8596, -4.8411, -5.3522, -4.5745, -5.0632, -5.3313,
-5.3129, -5.0528, -4.5861, -5.0233, -5.0781, -5.6057, -5.0709, -5.6290,
-4.9182, -4.3021, -5.2826, -4.6825, -4.5921, -5.1634, -5.2189, -5.2730,
-4.8037, -4.9847, -5.3225, -5.4399, -4.4496, -5.4737, -5.4246, -4.5623,
-5.0154, -5.2249, -5.0212, -4.2939, -4.9801, -5.2743, -4.5334, -6.1248,
-5.0221, -4.8258, -4.4920, -5.4247, -5.3156, -5.4661, -5.5334, -4.4338,
-4.9518, -5.1836, -5.2404, -4.6790, -5.5185, -4.9575, -5.4492],
device='cuda:0', grad_fn=<SelectBackward0>)

```

```

ConvolutionalNetwork(
  (model): MobileNetV3(
    (features): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(3, 16, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1),
bias=False)
        (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
      )
      (1): InvertedResidual(
        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(16, 16, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=16, bias=False)
            (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): ReLU(inplace=True)
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(16, 16, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(16, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          )
        )
      )
      (2): InvertedResidual(
        (block): Sequential(

```

```

        (0): Conv2dNormActivation(
          (0): Conv2d(16, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(64, 64, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=64, bias=False)
          (1): BatchNorm2d(64, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(64, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(24, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    (3): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(72, 72, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=72, bias=False)
          (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(72, 24, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(24, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    (4): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    (2): ReLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(72, 72, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=72, bias=False)
      (1): BatchNorm2d(72, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(72, 24, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(24, 72, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(72, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(5): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(40, 120, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(120, 120, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=120, bias=False)
      (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(120, 32, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(32, 120, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(

```

```

        (0): Conv2d(120, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
)
)
(6): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(40, 120, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): ReLU(inplace=True)
  )
  (1): Conv2dNormActivation(
    (0): Conv2d(120, 120, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=120, bias=False)
    (1): BatchNorm2d(120, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): ReLU(inplace=True)
  )
  (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(120, 32, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(32, 120, kernel_size=(1, 1), stride=(1, 1))
    (activation): ReLU()
    (scale_activation): Hardsigmoid()
  )
  (3): Conv2dNormActivation(
    (0): Conv2d(120, 40, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(40, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
  )
)
)
(7): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(40, 240, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
  )
  (1): Conv2dNormActivation(
    (0): Conv2d(240, 240, kernel_size=(3, 3), stride=(2, 2), padding=(1,
1), groups=240, bias=False)
    (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    (2): Hardswish()
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(240, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(8): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 200, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(200, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(200, 200, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=200, bias=False)
      (1): BatchNorm2d(200, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(200, 80, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(80, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(9): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(80, 184, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(184, 184, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=184, bias=False)
      (1): BatchNorm2d(184, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
    )
  )
)

```





```

        (fc2): Conv2d(120, 480, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(480, 112, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(112, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(12): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
  )
  (1): Conv2dNormActivation(
    (0): Conv2d(672, 672, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=672, bias=False)
    (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    (2): Hardswish()
  )
  (2): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(672, 168, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(168, 672, kernel_size=(1, 1), stride=(1, 1))
    (activation): ReLU()
    (scale_activation): Hardsigmoid()
  )
  (3): Conv2dNormActivation(
    (0): Conv2d(672, 112, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(112, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
  )
)
)
(13): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(112, 672, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
  )
)

```

```

    )
    (1): Conv2dNormActivation(
      (0): Conv2d(672, 672, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=672, bias=False)
      (1): BatchNorm2d(672, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(672, 168, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(168, 672, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(672, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(14): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(960, 960, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=960, bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(960, 240, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(240, 960, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    )
    )
    )
    (15): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(960, 960, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=960, bias=False)
          (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(960, 240, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(240, 960, kernel_size=(1, 1), stride=(1, 1))
          (activation): ReLU()
          (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(960, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(160, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    (16): Conv2dNormActivation(
      (0): Conv2d(160, 960, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(960, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    )
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (classifier): Sequential(
      (0): Linear(in_features=960, out_features=1280, bias=True)
      (1): Hardswish()
      (2): Dropout(p=0.2, inplace=True)
      (3): Linear(in_features=1280, out_features=151, bias=True)
    )

```

```
)  
)
```

```
[{'val_loss': 5.10817813873291, 'val_acc': 0.02500000037252903}]
```

```
627
```

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

```
Epoch [0], train_loss: 4.1468, val_loss: 2.5995, val_acc: 0.5875
```

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

```
Epoch [1], train_loss: 2.1886, val_loss: 1.7783, val_acc: 0.7875
```

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

```
Epoch [2], train_loss: 1.5442, val_loss: 1.3930, val_acc: 0.8188
```

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

```
Epoch [3], train_loss: 1.1747, val_loss: 1.2415, val_acc: 0.8219
```

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

```
Epoch [4], train_loss: 0.9517, val_loss: 1.1572, val_acc: 0.8531
```

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

```
Epoch [5], train_loss: 0.7862, val_loss: 1.1514, val_acc: 0.8688
```

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

```
Epoch [6], train_loss: 0.6645, val_loss: 1.1776, val_acc: 0.8688
```

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

```
Epoch [7], train_loss: 0.5927, val_loss: 1.1007, val_acc: 0.8750
```

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

```
Epoch [8], train_loss: 0.5039, val_loss: 1.0720, val_acc: 0.8813
```

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

```
Epoch [9], train_loss: 0.4561, val_loss: 0.9861, val_acc: 0.9094
```

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

```
Epoch [10], train_loss: 0.3608, val_loss: 1.1562, val_acc: 0.8750
```

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

```
Epoch [11], train_loss: 0.3625, val_loss: 1.1689, val_acc: 0.8813
```

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

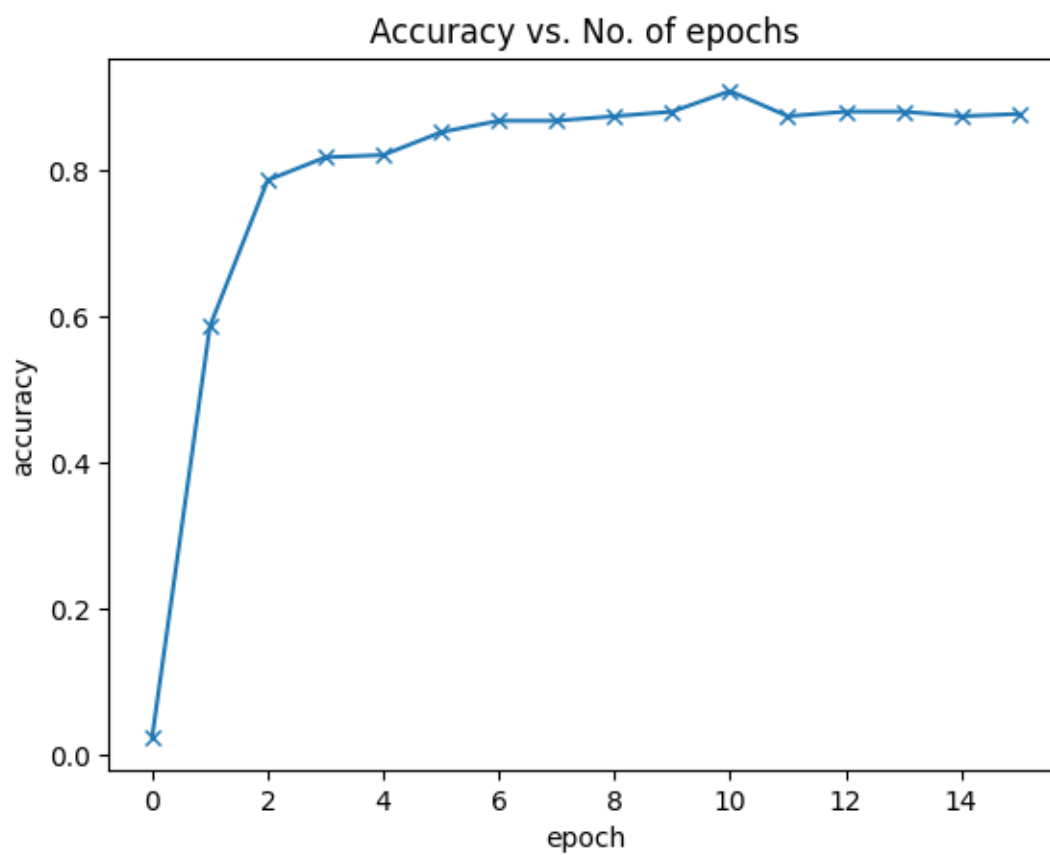
```
Epoch [12], train_loss: 0.3294, val_loss: 1.1434, val_acc: 0.8813
```

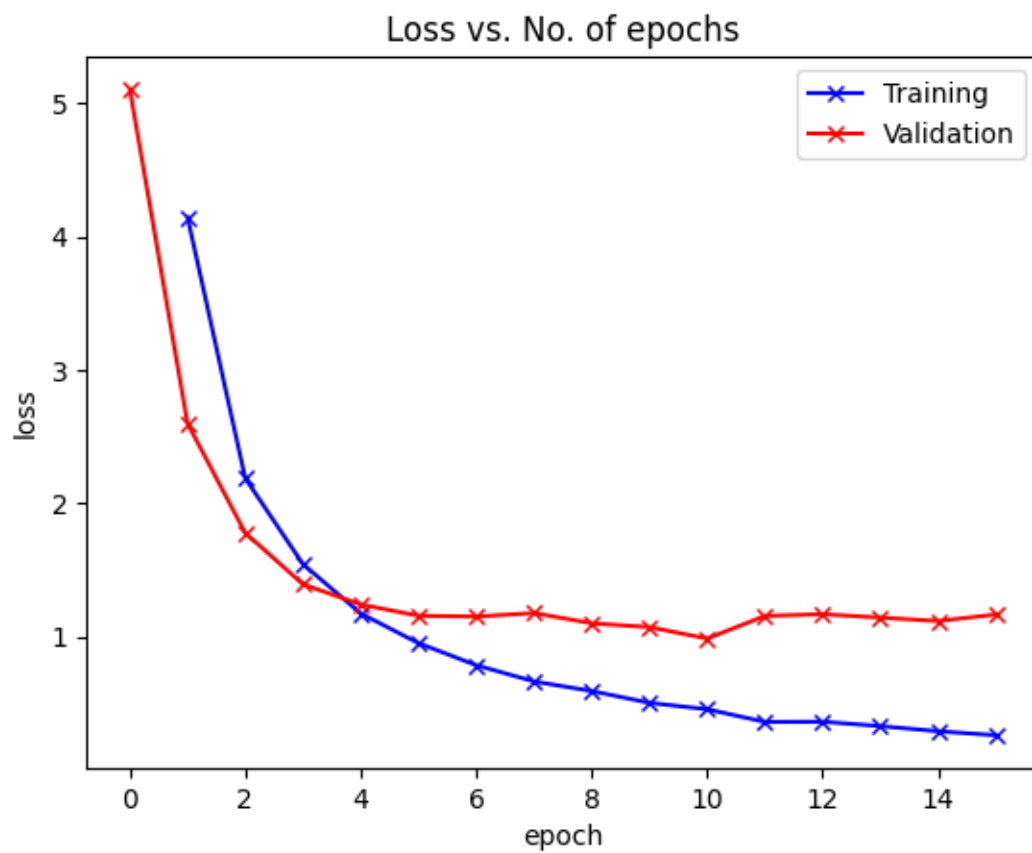
```
0%|          | 0/667 [00:00<?, ?it/s]
```

```
Epoch [13], train_loss: 0.2910, val_loss: 1.1179, val_acc: 0.8750
```

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

Epoch [14], train\_loss: 0.2611, val\_loss: 1.1665, val\_acc: 0.8781





```
{'val_loss': 1.1078855991363525, 'val_acc': 0.8971519470214844}
```

## 1.1 FLOPs

+ Number of FLOPs: 0.12G