

Animal_Classification-transfer-mobile-v3-small

June 13, 2024

1 Changing to transfer learning with Mobile Net V3 Small

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

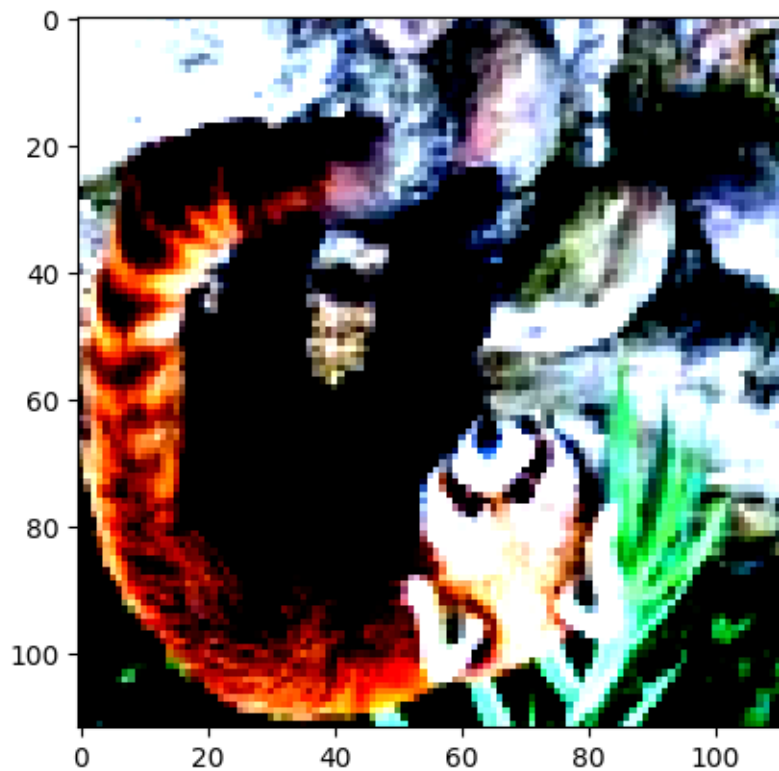
Validation Loss: 1.2289 Validation Accuracy: 85.76% FLOPS: 0.03G

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=(2, 2), 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): SqueezeExcitation(  
            (avgpool): AdaptiveAvgPool2d(output_size=1)  
            (fc1): Conv2d(16, 8, kernel_size=(1, 1), stride=(1, 1))  
            (fc2): Conv2d(8, 16, kernel_size=(1, 1), stride=(1, 1))  
            (activation): ReLU()  
            (scale_activation): Hardsigmoid()  
          )  
          (2): 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, 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=(2, 2), 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)
        )
      )
    )
    (3): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(24, 88, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(88, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(88, 88, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=88, bias=False)
          (1): BatchNorm2d(88, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): ReLU(inplace=True)
        )
        (2): Conv2dNormActivation(
          (0): Conv2d(88, 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, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,

```

```

track_running_stats=True)
    (2): Hardswish()
    )
    (1): Conv2dNormActivation(
        (0): Conv2d(96, 96, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=96, bias=False)
        (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(96, 24, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(24, 96, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(96, 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, 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=(5, 5), stride=(1, 1), padding=(2,
2), groups=240, bias=False)
                (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
                (2): Hardswish()
            )
            (2): SqueezeExcitation(
                (avgpool): AdaptiveAvgPool2d(output_size=1)
                (fc1): Conv2d(240, 64, kernel_size=(1, 1), stride=(1, 1))
                (fc2): Conv2d(64, 240, kernel_size=(1, 1), stride=(1, 1))
                (activation): ReLU()
                (scale_activation): Hardsigmoid()
            )
            (3): Conv2dNormActivation(

```

```

        (0): Conv2d(240, 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, 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=(5, 5), stride=(1, 1), padding=(2,
2), groups=240, bias=False)
      (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(240, 64, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(64, 240, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(240, 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, 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): Hardswish()
    )
    (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): Hardswish()
    )
    (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, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(8): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 144, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(144, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(144, 144, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=144, bias=False)
      (1): BatchNorm2d(144, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(144, 40, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(40, 144, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(144, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(48, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(9): InvertedResidual(

```

```

        (block): Sequential(
          (0): Conv2dNormActivation(
            (0): Conv2d(48, 288, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(288, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): Hardswish()
          )
          (1): Conv2dNormActivation(
            (0): Conv2d(288, 288, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=288, bias=False)
            (1): BatchNorm2d(288, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): Hardswish()
          )
          (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(288, 72, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(72, 288, kernel_size=(1, 1), stride=(1, 1))
            (activation): ReLU()
            (scale_activation): Hardsigmoid()
          )
          (3): Conv2dNormActivation(
            (0): Conv2d(288, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          )
        )
      )
    )
  (10): InvertedResidual(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
      )
      (1): Conv2dNormActivation(
        (0): Conv2d(576, 576, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=576, bias=False)
        (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(576, 144, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(144, 576, kernel_size=(1, 1), stride=(1, 1))

```

```

        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(576, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
  (11): InvertedResidual(
    (block): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(576, 576, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=576, bias=False)
      (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(576, 144, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(144, 576, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(576, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
  (12): Conv2dNormActivation(
    (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(576, 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=576, out_features=1024, bias=True)
  (1): Hardswish()
  (2): Dropout(p=0.2, inplace=True)
  (3): Linear(in_features=1024, out_features=151, bias=True)
)
)
)

images.shape: torch.Size([8, 3, 112, 112])
out.shape: torch.Size([8, 151])
out[0]: tensor([-5.5542, -5.6103, -5.4379, -5.2414, -5.1411, -4.6883, -4.7928,
-4.9843,
-4.9865, -5.5290, -4.4779, -5.4457, -5.3733, -4.3677, -5.6185, -5.5864,
-4.5083, -5.0144, -4.3929, -4.8658, -4.8815, -5.8586, -4.5566, -5.1401,
-4.9144, -5.6852, -5.5693, -5.1761, -5.0683, -5.5963, -5.0061, -4.9665,
-5.8103, -4.6418, -5.4931, -4.1183, -4.8317, -4.8607, -5.1162, -4.8572,
-5.9464, -4.8994, -4.7951, -4.9792, -4.8258, -5.0329, -5.0679, -5.1438,
-5.2102, -5.3552, -4.8719, -5.9381, -5.1105, -5.4190, -5.4374, -5.0009,
-5.0370, -5.5993, -5.2528, -4.5962, -6.0117, -5.0390, -4.9475, -4.4735,
-5.4557, -4.3037, -5.4275, -5.0196, -5.1799, -6.4949, -4.8761, -5.4855,
-4.9137, -4.8761, -5.1473, -5.3224, -4.8004, -4.7817, -5.6682, -6.2449,
-4.9177, -5.0530, -4.7397, -5.5145, -5.3851, -4.7552, -5.6719, -5.3272,
-5.0901, -4.3154, -6.2621, -4.5276, -5.3719, -4.6722, -5.6334, -4.7404,
-5.1805, -4.7256, -4.8417, -4.6307, -5.3082, -5.1648, -4.6274, -4.5309,
-5.7943, -5.1004, -5.0038, -5.3724, -4.9499, -5.3513, -5.0583, -4.0588,
-5.5525, -4.7955, -5.0639, -4.8760, -5.2460, -5.7086, -4.4823, -5.0263,
-5.4103, -5.0922, -5.9388, -4.5565, -5.2087, -5.8740, -5.4014, -4.6751,
-5.0326, -4.9352, -5.4580, -5.4378, -5.6487, -5.2868, -4.9246, -4.8195,
-5.3997, -5.3548, -4.6796, -5.0489, -4.6602, -4.8843, -4.2516, -5.2222,
-5.9748, -4.5397, -4.6895, -4.2863, -5.2250, -4.9740, -4.8543],
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=(2, 2), 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): SqueezeExcitation(
    (avgpool): AdaptiveAvgPool2d(output_size=1)
    (fc1): Conv2d(16, 8, kernel_size=(1, 1), stride=(1, 1))
    (fc2): Conv2d(8, 16, kernel_size=(1, 1), stride=(1, 1))
    (activation): ReLU()
    (scale_activation): Hardsigmoid()
  )
  (2): 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, 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=(2, 2), 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)
    )
  )
)
(3): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(24, 88, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(88, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)

```

```

        (2): ReLU(inplace=True)
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(88, 88, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), groups=88, bias=False)
      (1): BatchNorm2d(88, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): ReLU(inplace=True)
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(88, 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, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(96, 96, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=96, bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
      (2): Hardswish()
    )
    (2): SqueezeExcitation(
      (avgpool): AdaptiveAvgPool2d(output_size=1)
      (fc1): Conv2d(96, 24, kernel_size=(1, 1), stride=(1, 1))
      (fc2): Conv2d(24, 96, kernel_size=(1, 1), stride=(1, 1))
      (activation): ReLU()
      (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(96, 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, 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=(5, 5), stride=(1, 1), padding=(2,
2), groups=240, bias=False)
          (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(240, 64, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(64, 240, kernel_size=(1, 1), stride=(1, 1))
          (activation): ReLU()
          (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(240, 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, 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=(5, 5), stride=(1, 1), padding=(2,
2), groups=240, bias=False)
        (1): BatchNorm2d(240, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
      )
      (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(240, 64, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(64, 240, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()

```

```

        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(240, 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, 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): Hardswish()
  )
  (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): Hardswish()
  )
  (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, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (1): BatchNorm2d(48, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
  )
)
)
(8): InvertedResidual(
  (block): Sequential(
    (0): Conv2dNormActivation(
      (0): Conv2d(48, 144, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(144, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
    (2): Hardswish()
  )
  (1): Conv2dNormActivation(

```

```

        (0): Conv2d(144, 144, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=144, bias=False)
        (1): BatchNorm2d(144, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        (2): Hardswish()
    )
    (2): SqueezeExcitation(
        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(144, 40, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(40, 144, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
        (0): Conv2d(144, 48, kernel_size=(1, 1), stride=(1, 1), bias=False)
        (1): BatchNorm2d(48, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
)
)
(9): InvertedResidual(
    (block): Sequential(
        (0): Conv2dNormActivation(
            (0): Conv2d(48, 288, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(288, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): Hardswish()
        )
        (1): Conv2dNormActivation(
            (0): Conv2d(288, 288, kernel_size=(5, 5), stride=(2, 2), padding=(2,
2), groups=288, bias=False)
            (1): BatchNorm2d(288, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
            (2): Hardswish()
        )
        (2): SqueezeExcitation(
            (avgpool): AdaptiveAvgPool2d(output_size=1)
            (fc1): Conv2d(288, 72, kernel_size=(1, 1), stride=(1, 1))
            (fc2): Conv2d(72, 288, kernel_size=(1, 1), stride=(1, 1))
            (activation): ReLU()
            (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
            (0): Conv2d(288, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
    )
)

```

```

    )
    )
    (10): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(576, 576, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=576, bias=False)
          (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): SqueezeExcitation(
          (avgpool): AdaptiveAvgPool2d(output_size=1)
          (fc1): Conv2d(576, 144, kernel_size=(1, 1), stride=(1, 1))
          (fc2): Conv2d(144, 576, kernel_size=(1, 1), stride=(1, 1))
          (activation): ReLU()
          (scale_activation): Hardsigmoid()
        )
        (3): Conv2dNormActivation(
          (0): Conv2d(576, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
        )
      )
    )
    (11): InvertedResidual(
      (block): Sequential(
        (0): Conv2dNormActivation(
          (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (1): Conv2dNormActivation(
          (0): Conv2d(576, 576, kernel_size=(5, 5), stride=(1, 1), padding=(2,
2), groups=576, bias=False)
          (1): BatchNorm2d(576, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
          (2): Hardswish()
        )
        (2): SqueezeExcitation(

```

```

        (avgpool): AdaptiveAvgPool2d(output_size=1)
        (fc1): Conv2d(576, 144, kernel_size=(1, 1), stride=(1, 1))
        (fc2): Conv2d(144, 576, kernel_size=(1, 1), stride=(1, 1))
        (activation): ReLU()
        (scale_activation): Hardsigmoid()
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(576, 96, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (1): BatchNorm2d(96, eps=0.001, momentum=0.01, affine=True,
track_running_stats=True)
    )
  )
)
(12): Conv2dNormActivation(
  (0): Conv2d(96, 576, kernel_size=(1, 1), stride=(1, 1), bias=False)
  (1): BatchNorm2d(576, 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=576, out_features=1024, bias=True)
  (1): Hardswish()
  (2): Dropout(p=0.2, inplace=True)
  (3): Linear(in_features=1024, out_features=151, bias=True)
)
)
)

```

```
[{'val_loss': 5.08850622177124, 'val_acc': 0.02187499962747097}]
```

627

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

```
Epoch [0], train_loss: 4.5812, val_loss: 3.5535, val_acc: 0.4250
```

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

```
Epoch [1], train_loss: 2.9841, val_loss: 2.3200, val_acc: 0.6219
```

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

```
Epoch [2], train_loss: 2.2071, val_loss: 1.7945, val_acc: 0.7312
```

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

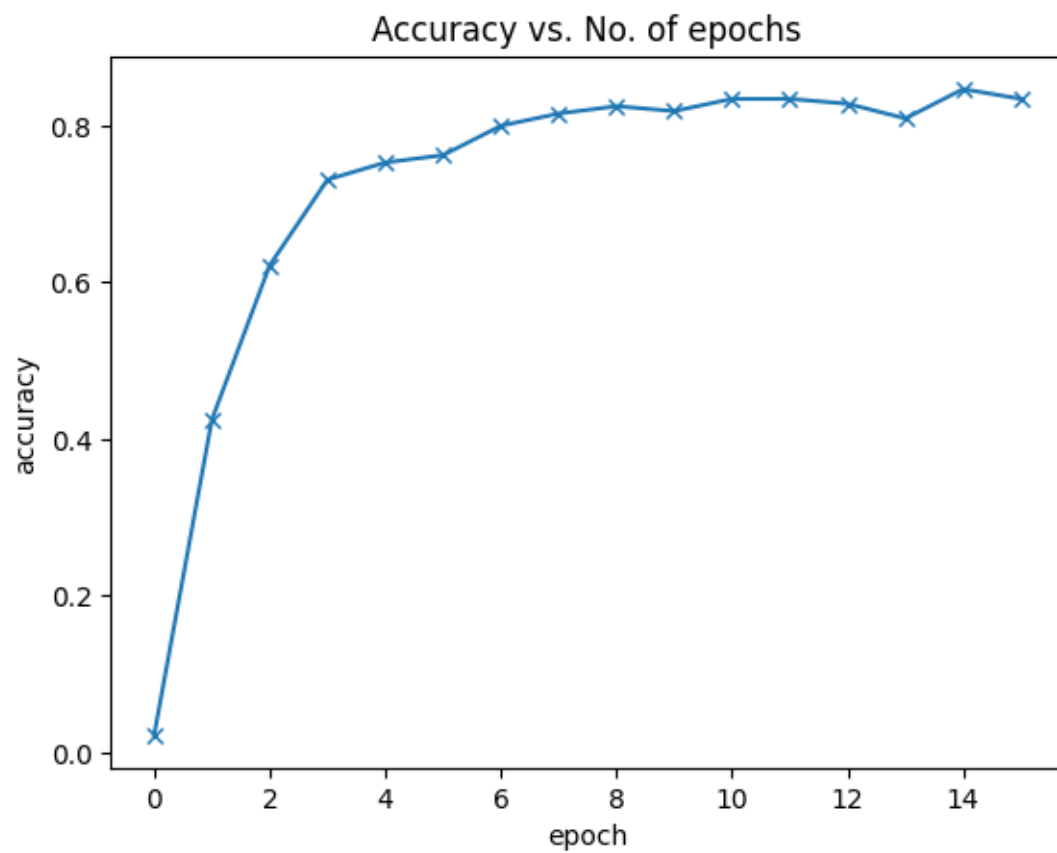
```
Epoch [3], train_loss: 1.7772, val_loss: 1.7229, val_acc: 0.7531
```

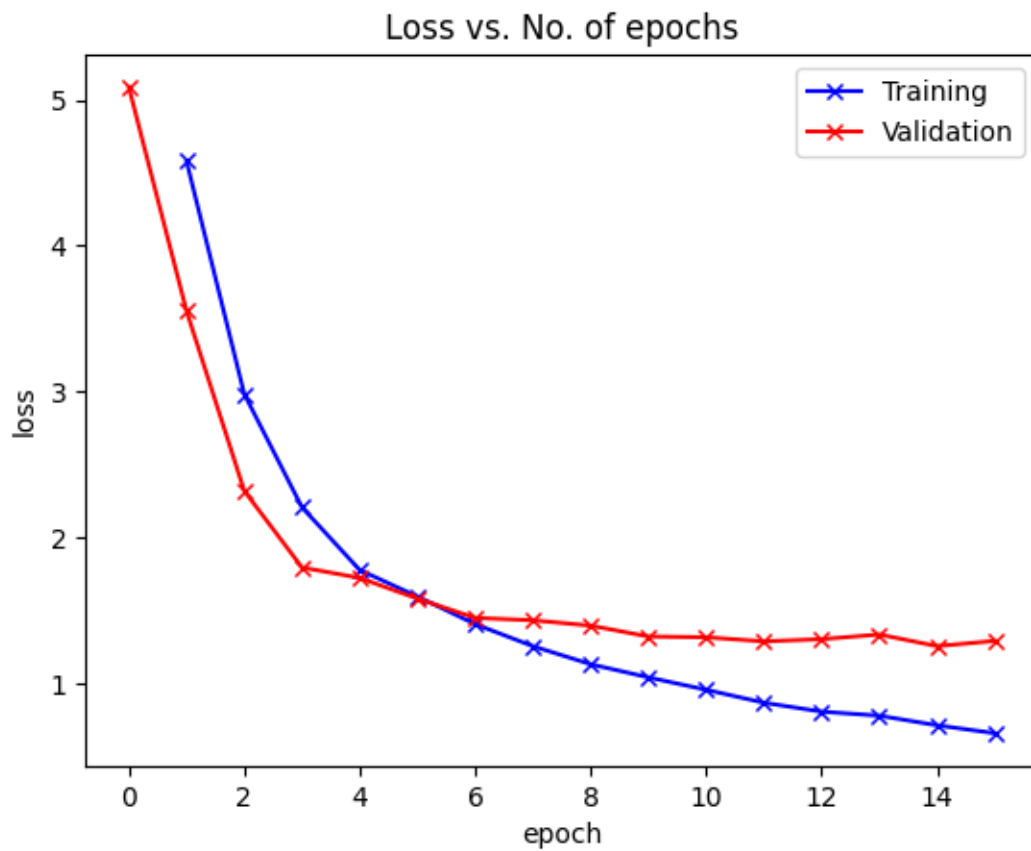
```
0%|          | 0/667 [00:00<?, ?it/s]
```

```
Epoch [4], train_loss: 1.5954, val_loss: 1.5807, val_acc: 0.7625
```



```
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [5], train_loss: 1.4059, val_loss: 1.4487, val_acc: 0.8000
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [6], train_loss: 1.2548, val_loss: 1.4326, val_acc: 0.8156
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [7], train_loss: 1.1310, val_loss: 1.3956, val_acc: 0.8250
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [8], train_loss: 1.0410, val_loss: 1.3212, val_acc: 0.8188
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [9], train_loss: 0.9561, val_loss: 1.3160, val_acc: 0.8344
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [10], train_loss: 0.8672, val_loss: 1.2879, val_acc: 0.8344
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [11], train_loss: 0.8060, val_loss: 1.3030, val_acc: 0.8281
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [12], train_loss: 0.7782, val_loss: 1.3350, val_acc: 0.8094
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [13], train_loss: 0.7123, val_loss: 1.2550, val_acc: 0.8469
0%|          | 0/667 [00:00<?, ?it/s]
Epoch [14], train_loss: 0.6580, val_loss: 1.2911, val_acc: 0.8344
```





```
{'val_loss': 1.2289390563964844, 'val_acc': 0.8575949668884277}
```

1.1 FLOPs

+ Number of FLOPs: 0.03G