Mish

A Self Regularized Non-Linear Non-Monotonous Adaptive Neural Activation Function.

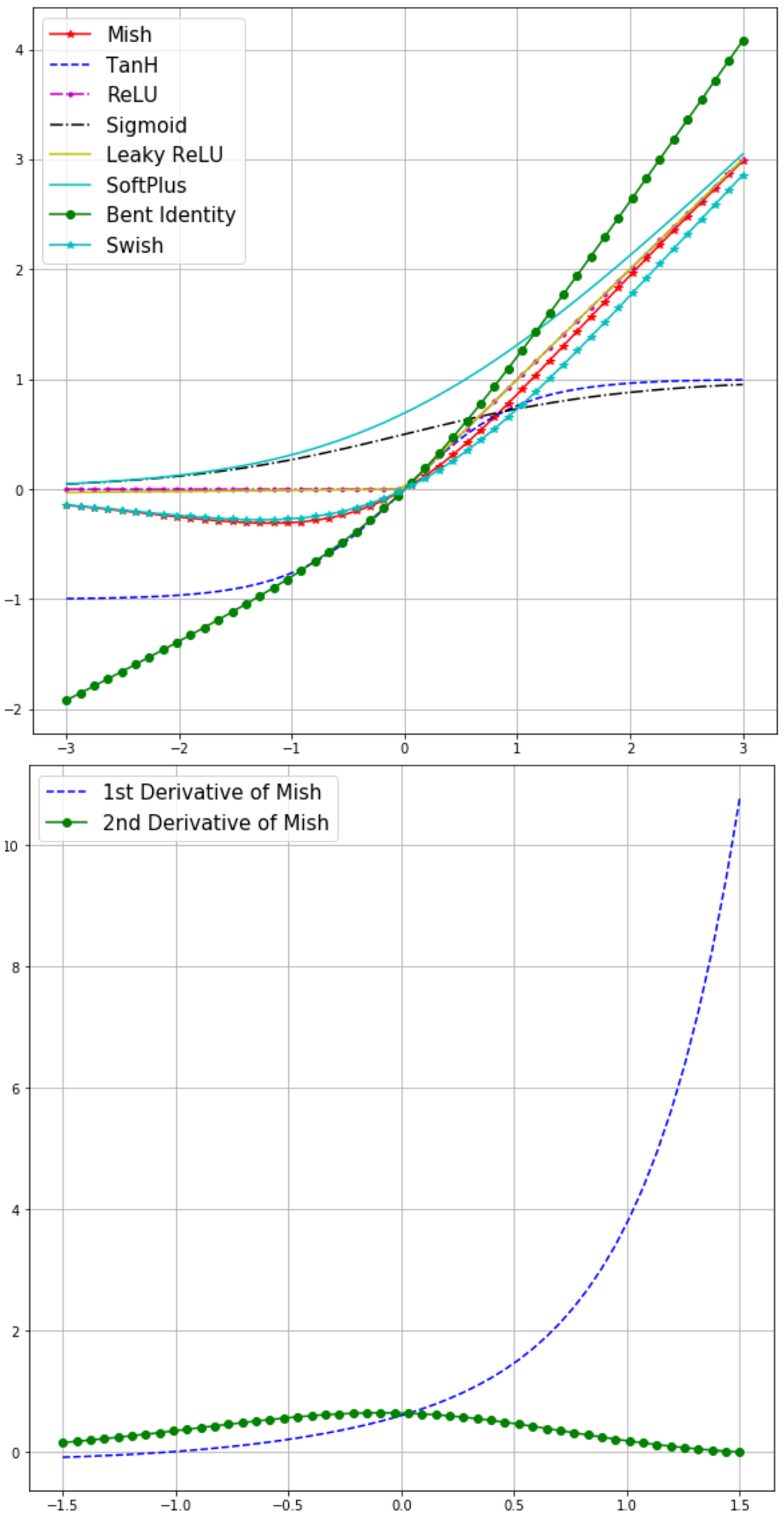
- DIGANTA MISRA

Mish

Introduction

MOTIVATION

Inspired by Swish Activation Function, Mish activation function is a self regularized non-monotonous non-linear adaptive neural activation function. It ranges from approximately -0.31 to infinity. Graphically, it looks nearly the same as Swish, it has a small decay or preserves the gradients in the negative half which results in it having a very sharp global minima.



Mish

Math under the Hood

$$f(x) = x\tanh(\ln(1+e^x))$$

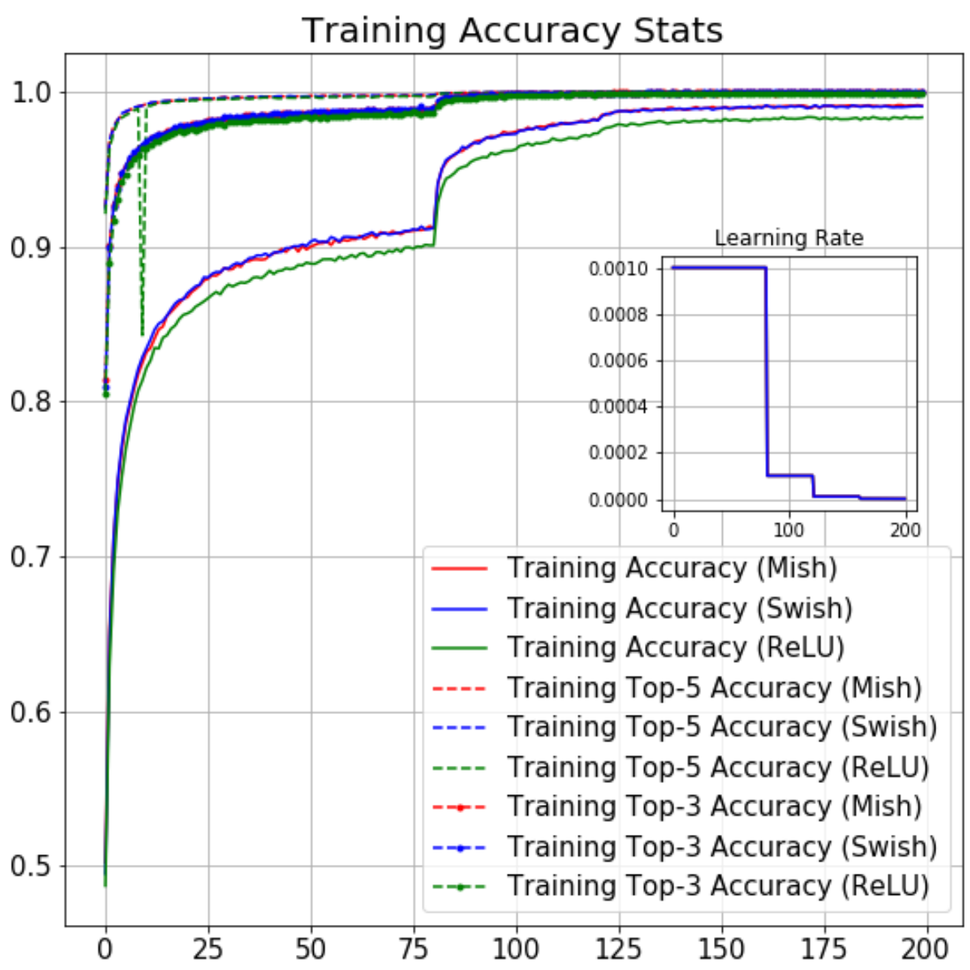
$$f'(x) = \frac{f(x)}{x} + xg'(x)$$

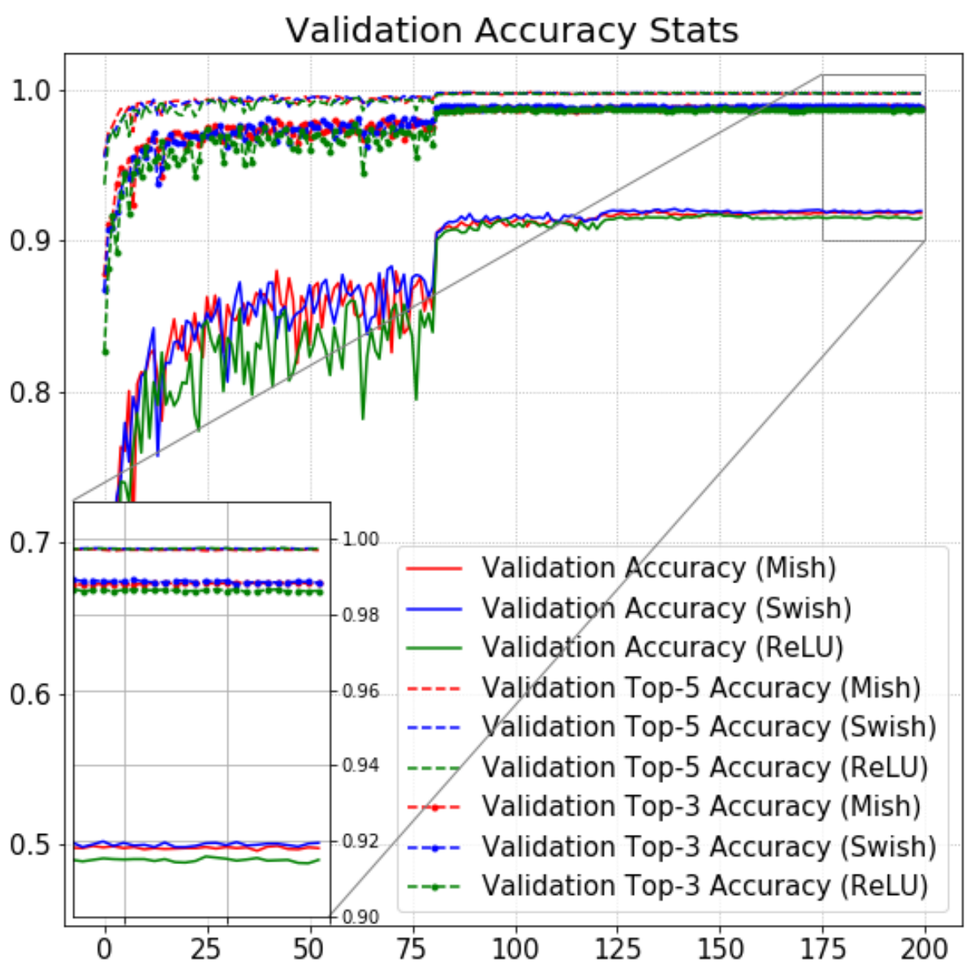
$$f''(x) = \frac{-g'(x)(2f(x) - 2e^x - x - 2)}{(e^x+1)}$$

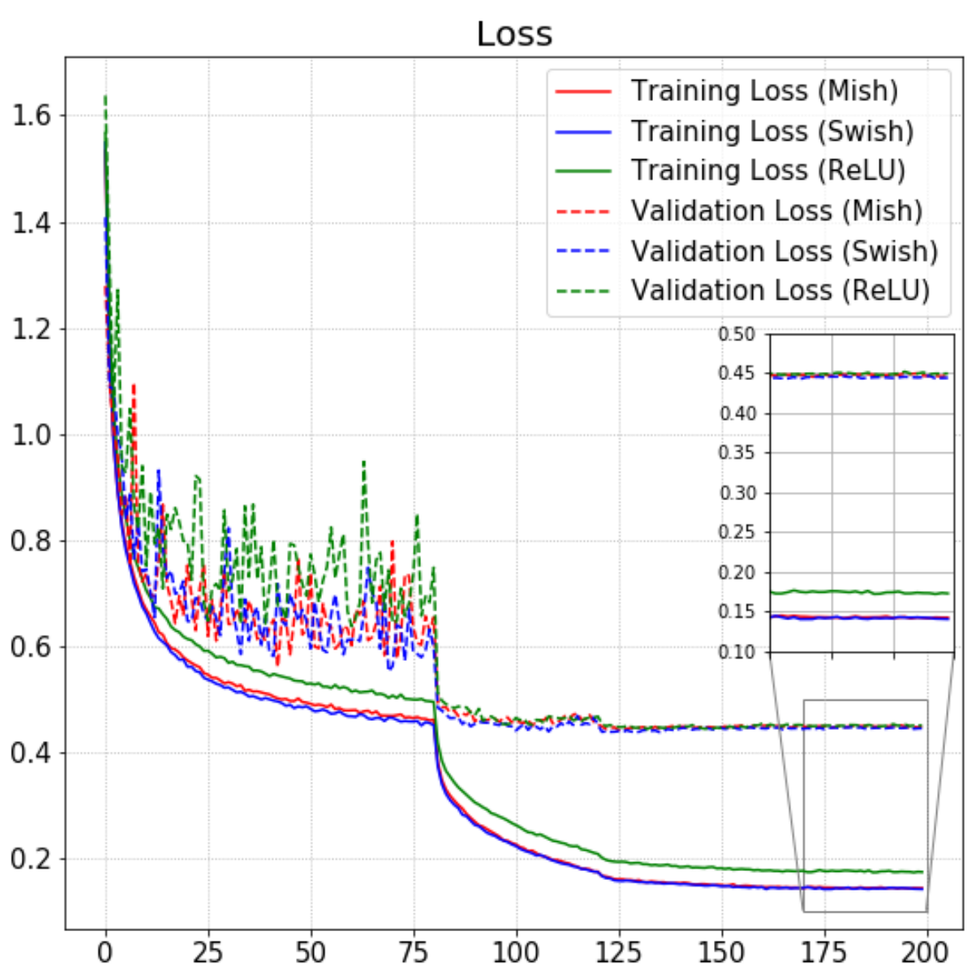
$$g(x) = \tanh(ln(1+e^x))$$

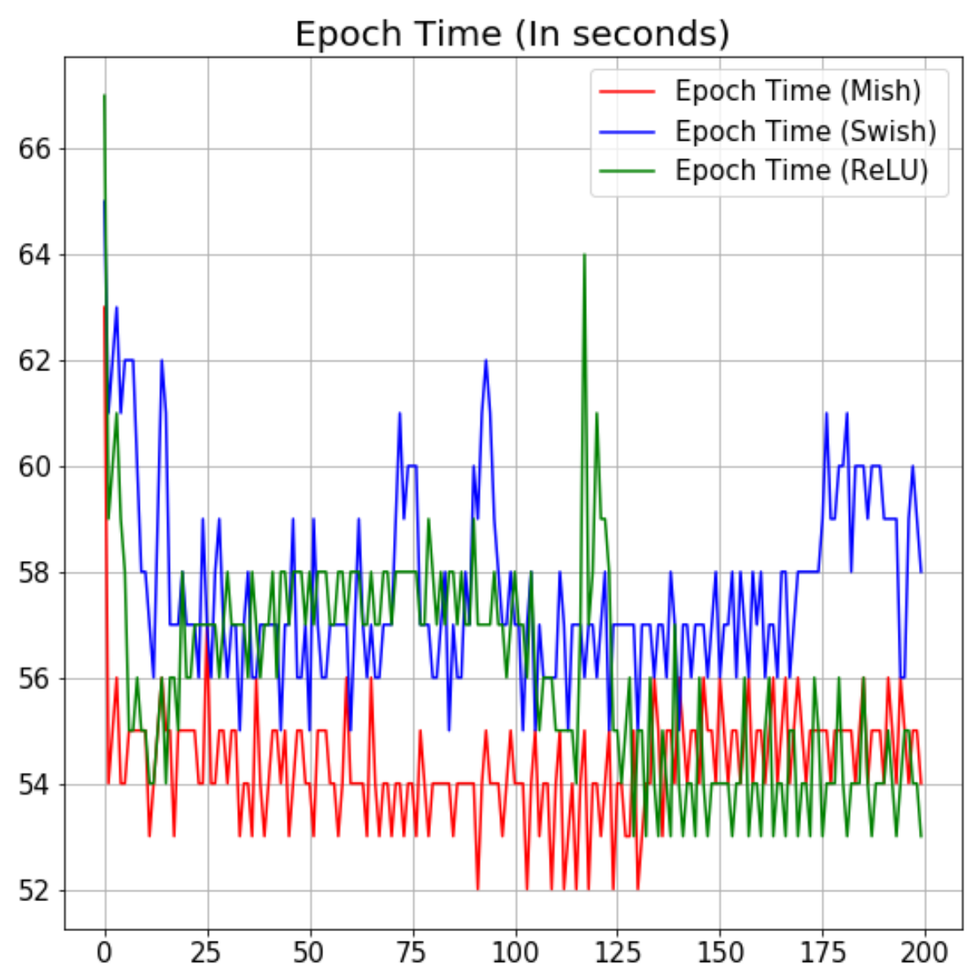
$$g'(x) = \frac{e^x}{\cosh^2(\ln(e^x+1))(e^x+1)}$$

CIFAR-10



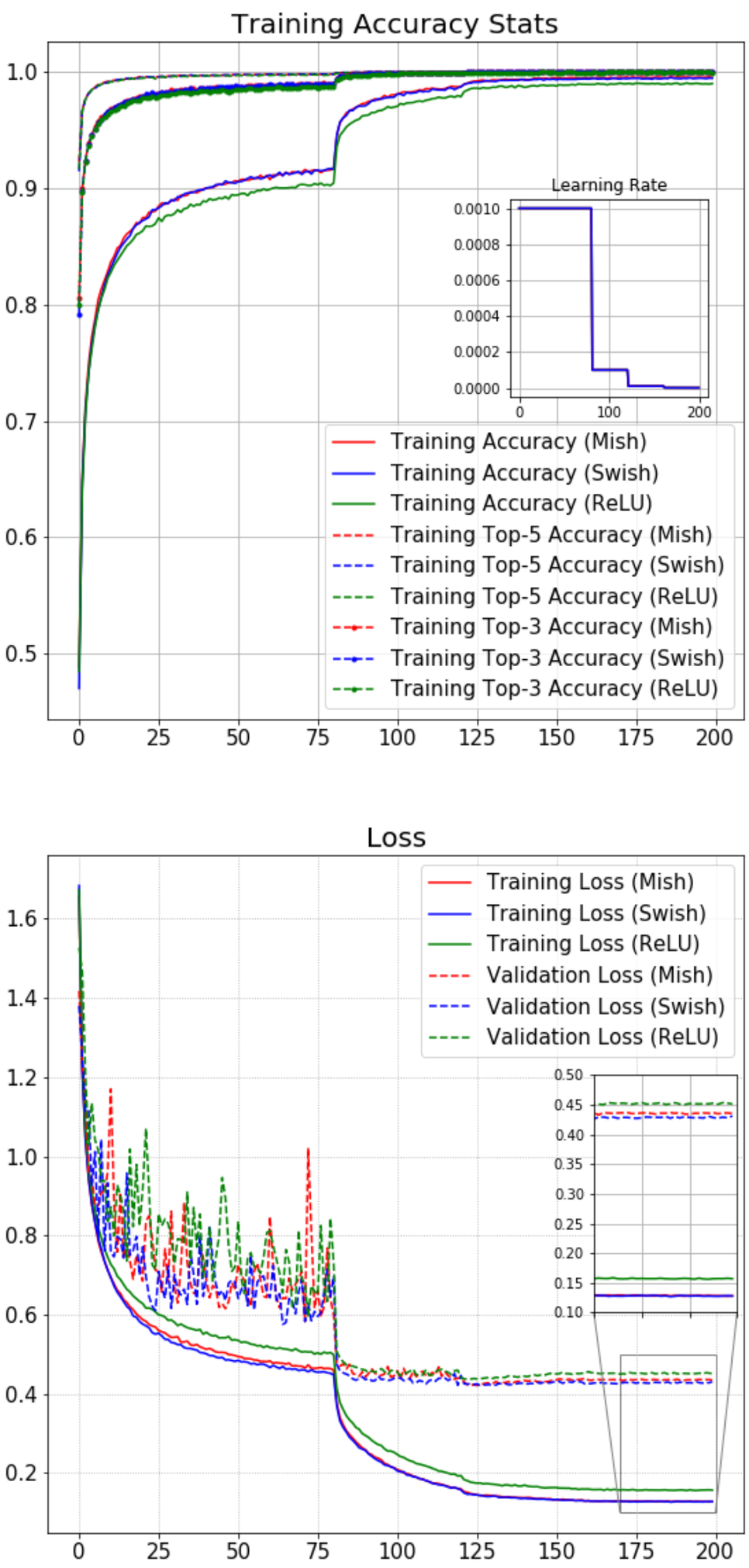


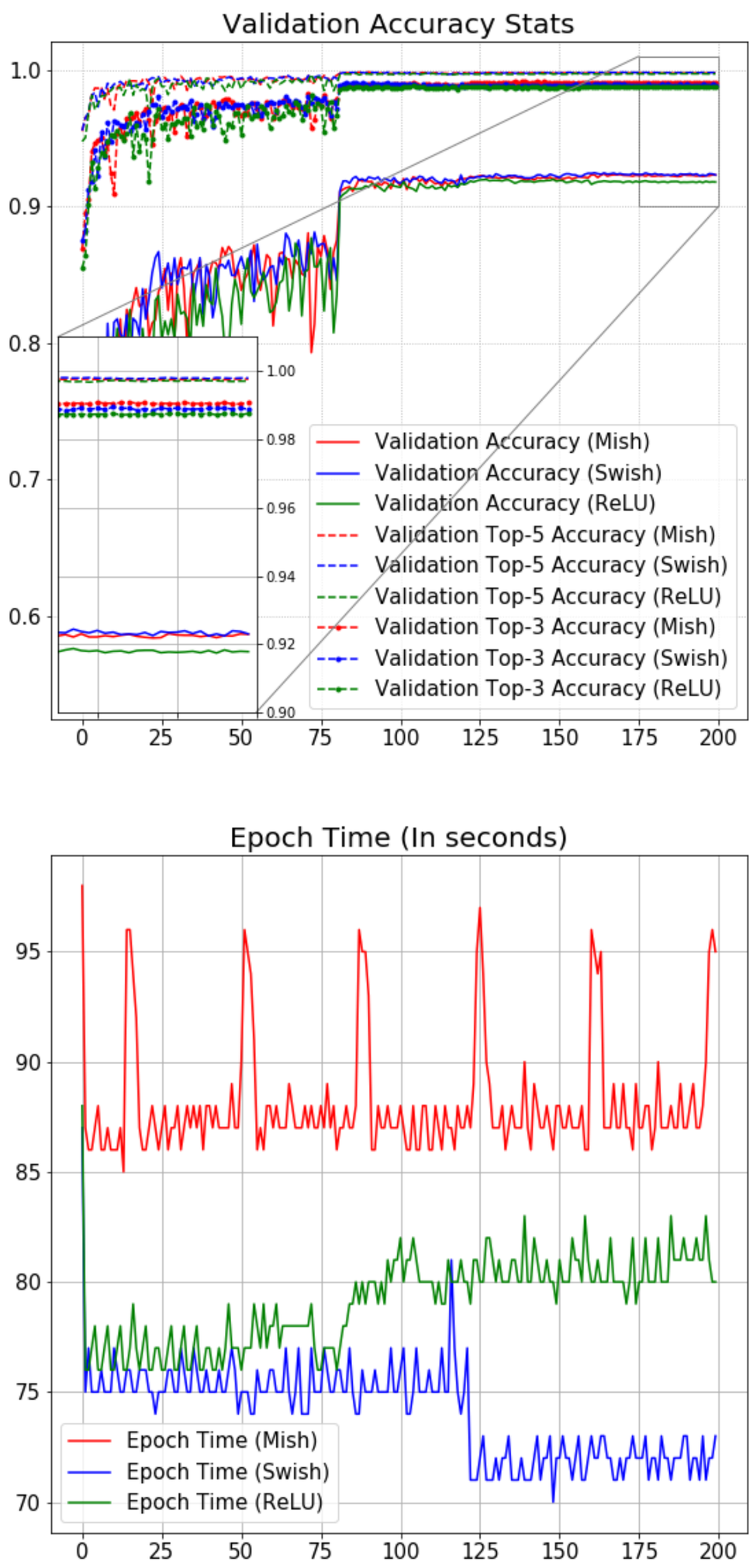




The above result is for ResNet-20 version-1 training on CIFAR-10.

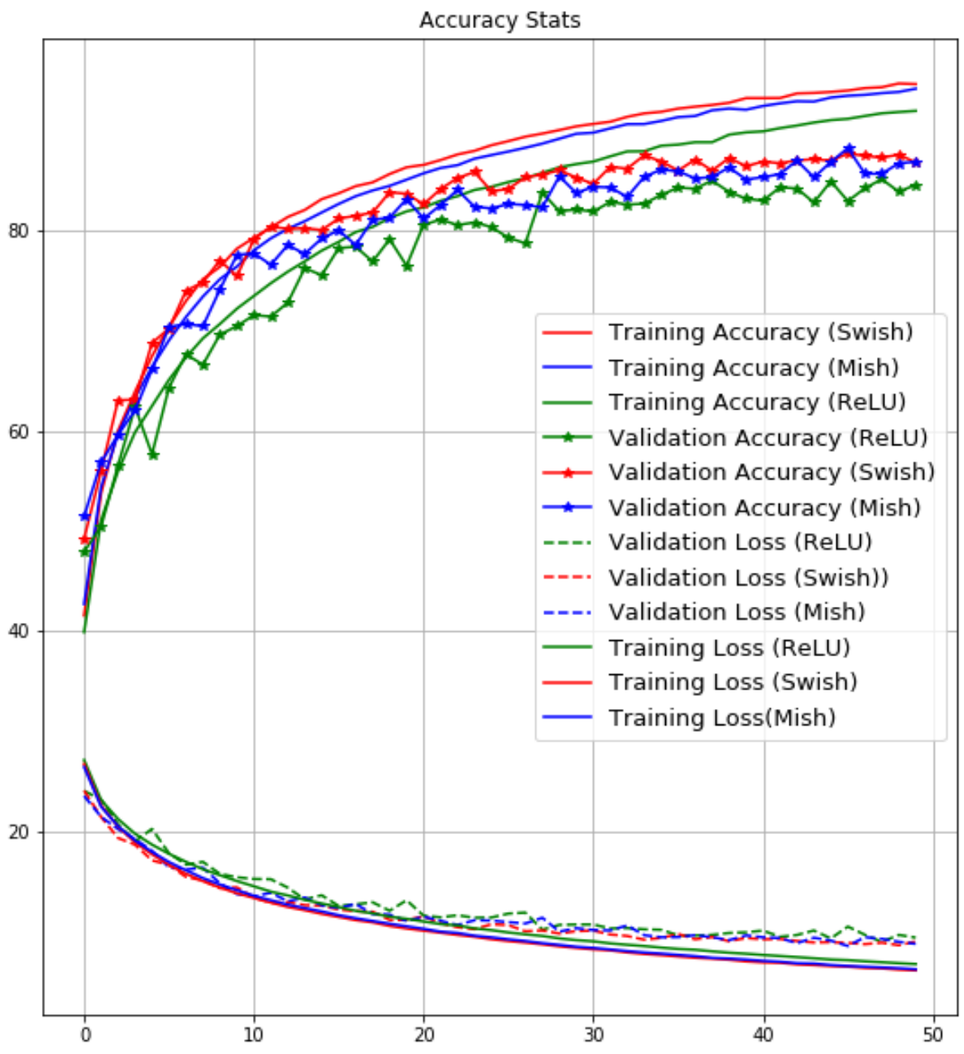
CIFAR-10

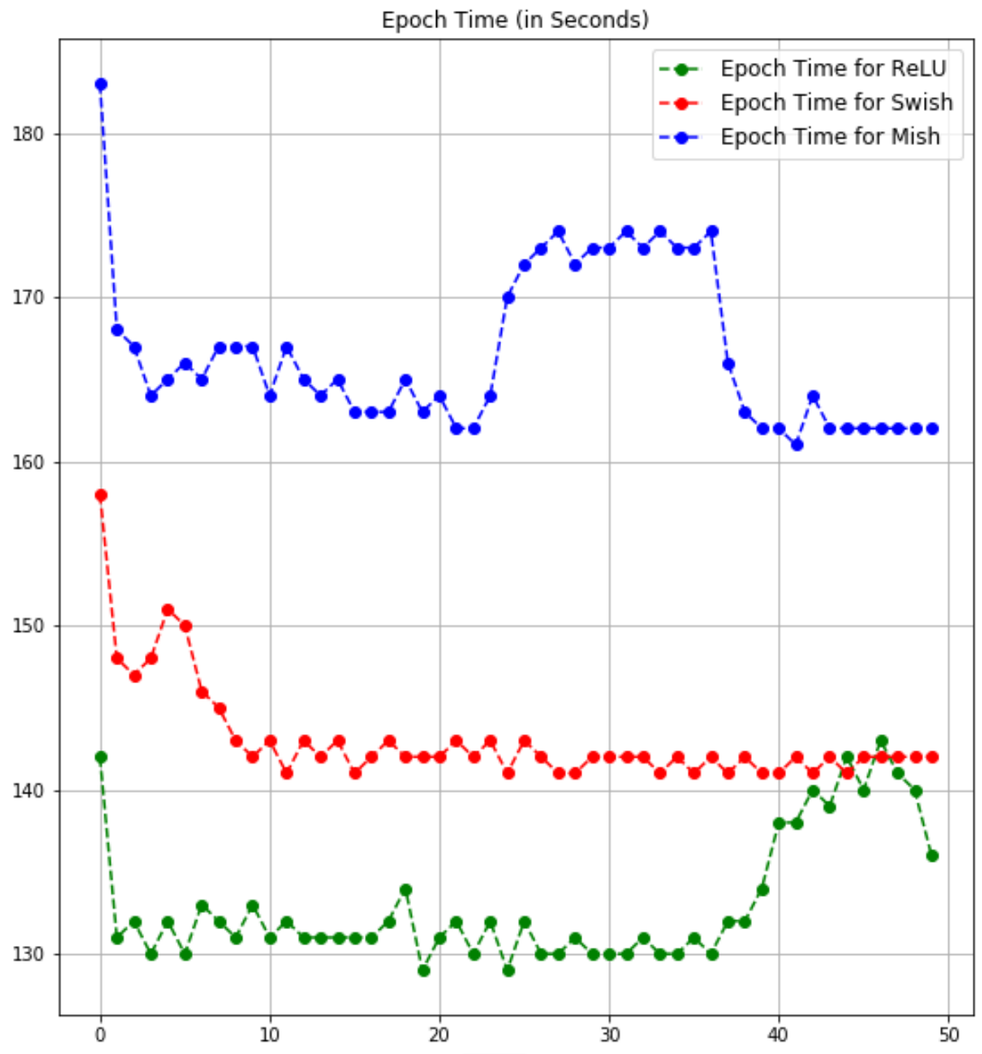


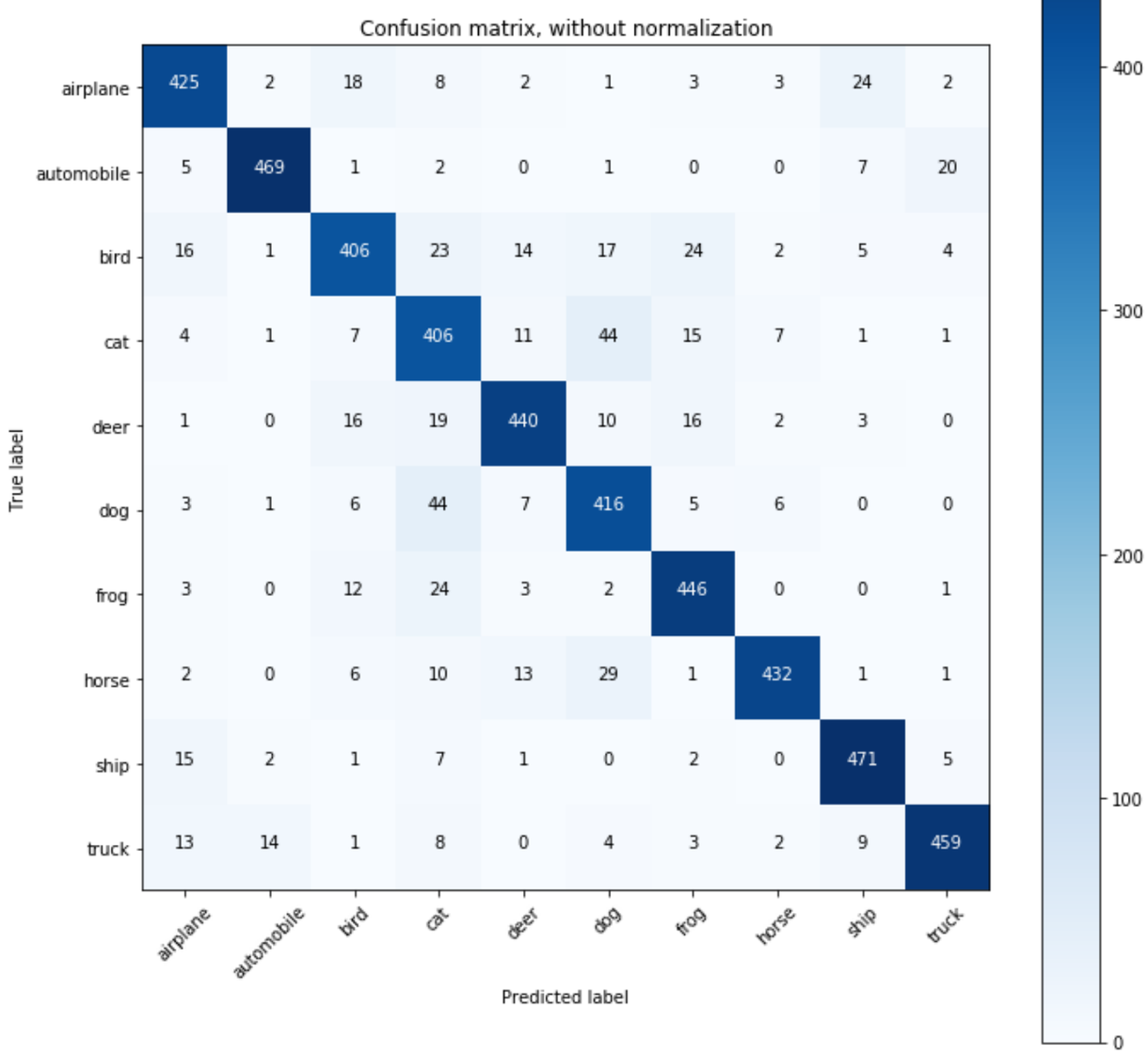


The above result is for ResNet-32 version-1 training on CIFAR-10.

CIFAR-10





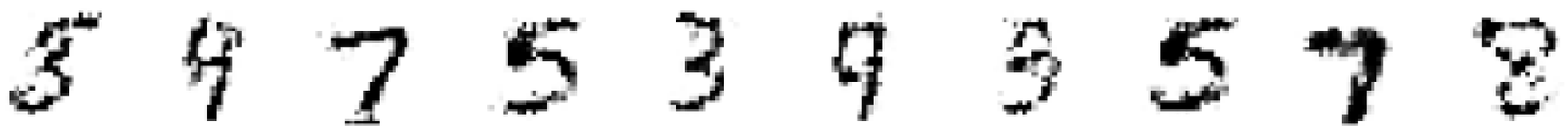


The above result is for ResNet-56 version-2 training on CIFAR-10.

Mish

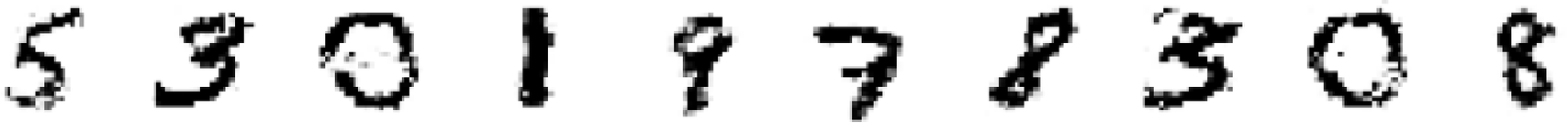
GANs

Epoch 20



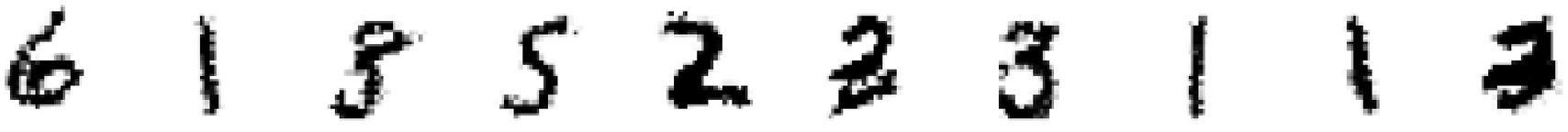
Epoch 30

Epoch 40



Epoch 50

Epoch 60



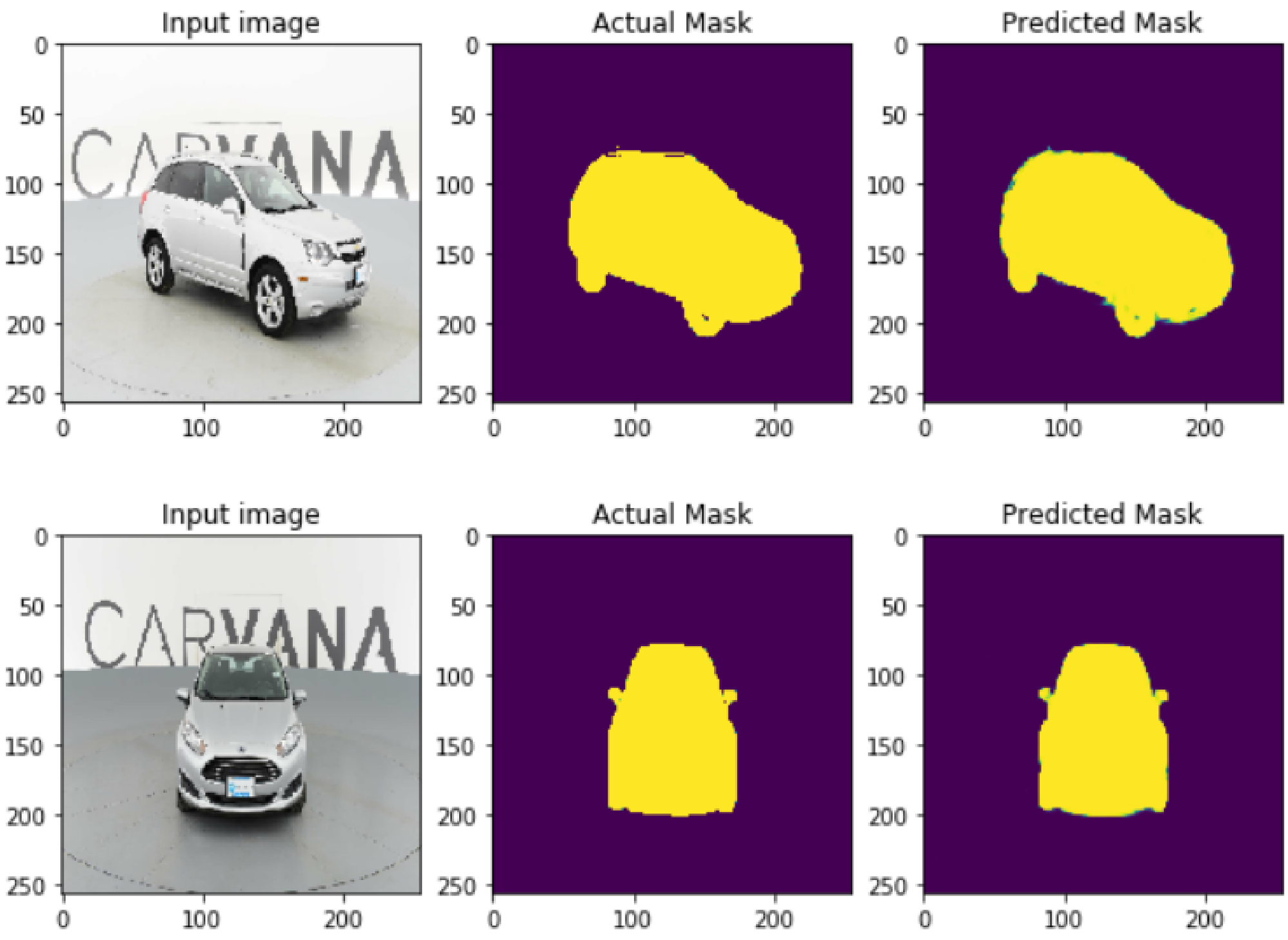
Epoch 70

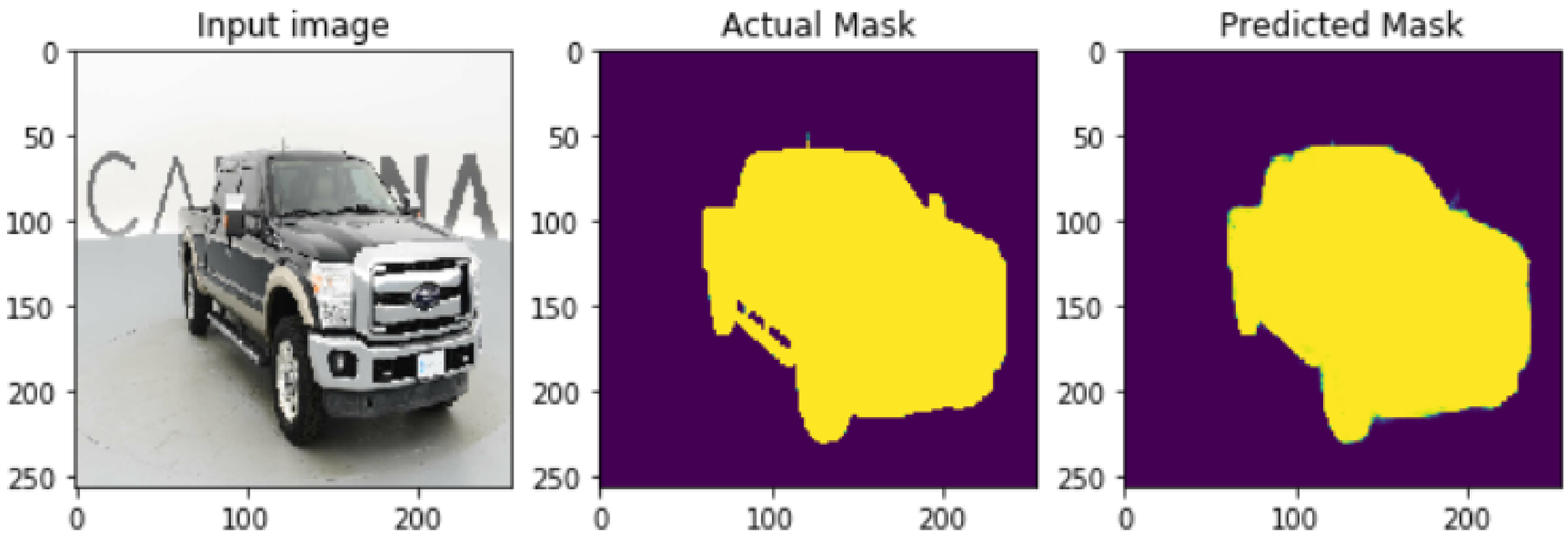
Epoch 80

Epoch 90

Mish

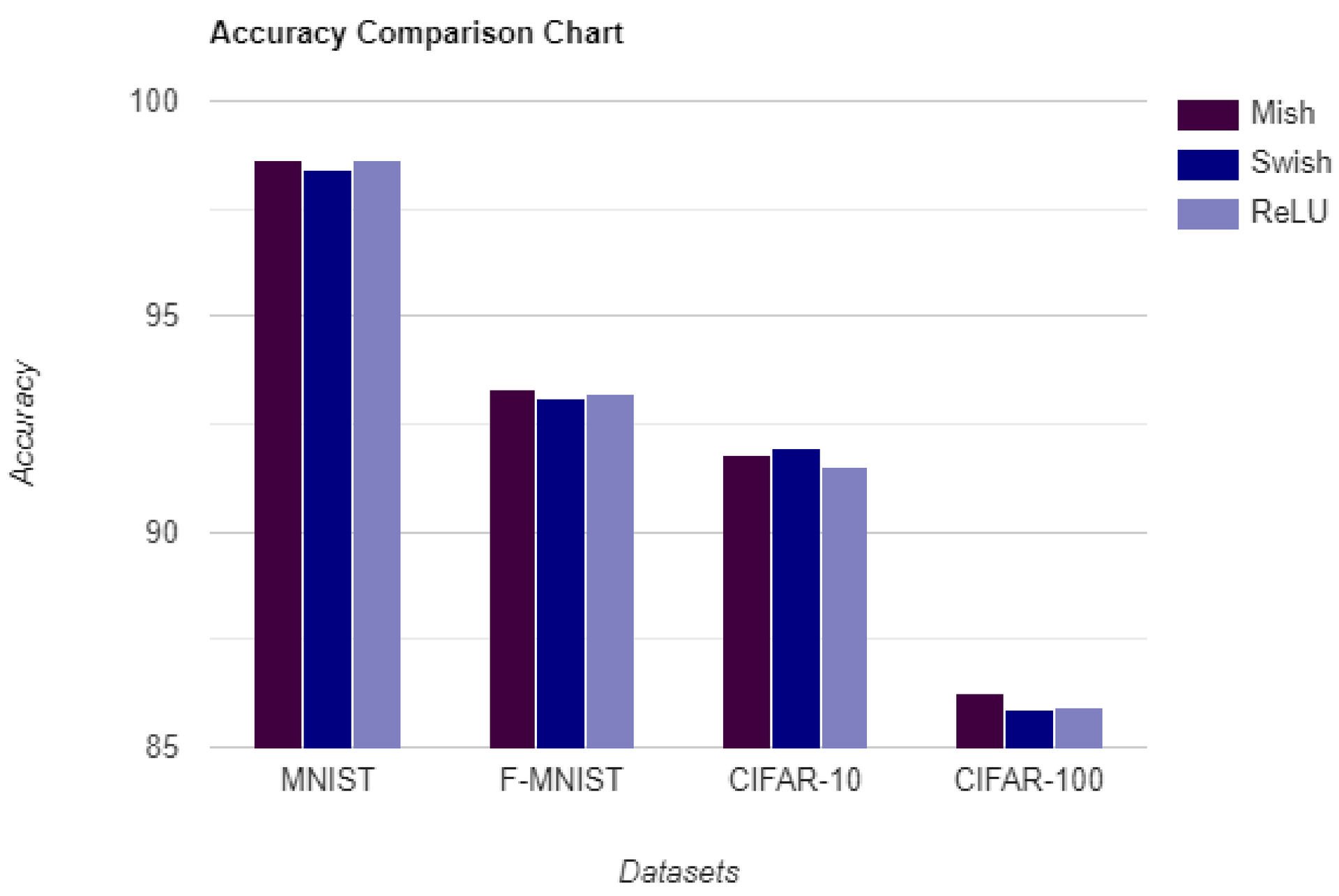
Object Segmentation





Mish

Results



MNIST was trained on a LeNet-5, F-MNIST was trained on a VGG-Net, CIFAR-10 was trained on ResNet-20 version 1 and CIFAR-100 was trained on ResNet-32 version-1. All the accuracies depicted above are Top-1 Testing Accuracy except for CIFAR-100 which shows the Testing Top-3 Accuracy.

Thank You!

Additional Details shall be available on github.com/digantamisra98/Mish