A Appendix

You may include other additional sections here.

A.1 AlexNet Architecture and experimental details

```
model arch DataParallel(
  (module): AlexNet(
    (features): Sequential(
      (0): Conv2d(3, 64, kernel_size=(11, 11), stride=(4, 4), padding=(5, 5))
      (1): ReLU(inplace=True)
      (2): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
      (3): Conv2d(64, 192, kernel_size=(5, 5), stride=(1, 1), padding=(2, 2))
      (4): ReLU(inplace=True)
      (5): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
      (6): Conv2d(192, 384, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (7): ReLU(inplace=True)
      (8): Conv2d(384, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (9): ReLU(inplace=True)
      (10): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
      (11): ReLU(inplace=True)
      (12): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False
    )
    (classifier): Linear(in_features=256, out_features=10, bias=True)
  )
)
    Total params: 2.47M
```

B Experimental details

B.1 AlexNet on CIFAR 10

We train AlexNet using the above architecture on the CIFAR-10 dataset. We train for 164 epochs decaying the learning rate by a factor of 10 at 81 and 122 epochs.

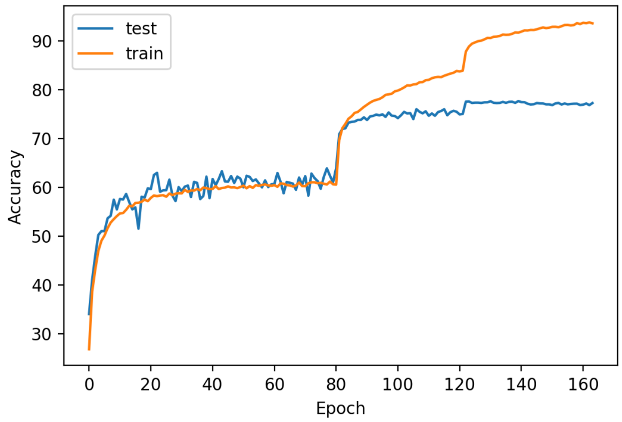

Figure 8: Training and test accuracy for AlexNet trained on CIFAR-10

B.2 ResNet-110 on CIFAR 10

We train ResNet-110 on the CIFAR-10 dataset. We train for 164 epochs decaying the learning rate by a factor of 10 at 81 and 122 epochs. We use a weight decay of 1e-4.

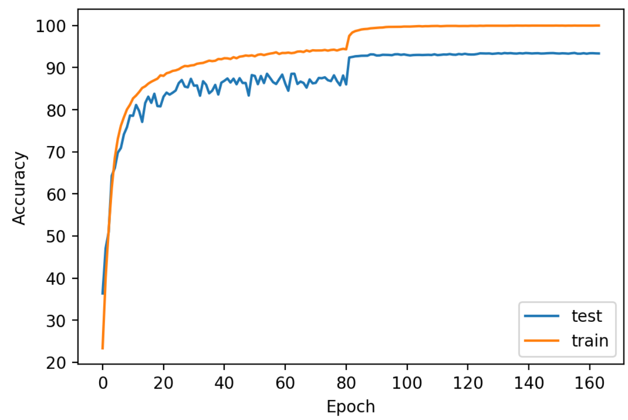

Figure 9: Training and test accuracy for ResNet-110 trained on CIFAR-10

B.3 RESNET-18 ON IMAGENET

We train ResNet-18 on the ImageNet dataset. We train for 90 epochs decaying the learning rate by a factor of 10 at 31 and 61 epochs. We use a weight decay of 1e-4.

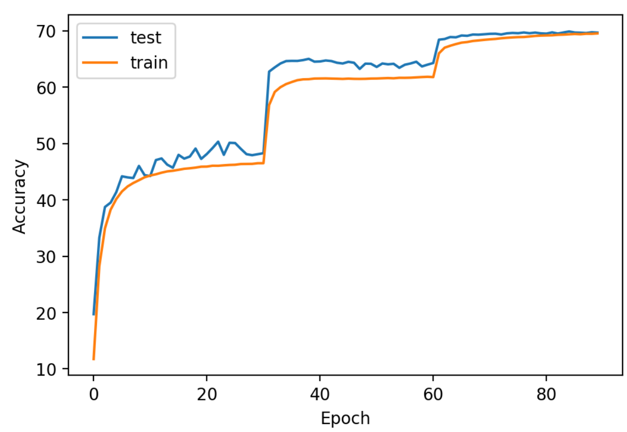

Figure 10: Training and test accuracy for ResNet-18 trained on ImageNet

C EXPERIMENTS USING THE WHOLE f_i^t

We can also analyze self similarity when considering the function $f_i^t([N])$ in isolation (i.e. not applied on top of the current history $\mathcal{H}_t$.) We plot our results in Figure 12.

D RESNET CIFAR 10 PATH DISTANCES

We try to asses whether the overlap between path vectors is related to the distance between original points in data space. We plot our results in Figure 13. We find that there is not a clean linear relationship between distance in original space and overlap in path vectors, supporting the case for considering each image update as its own function.

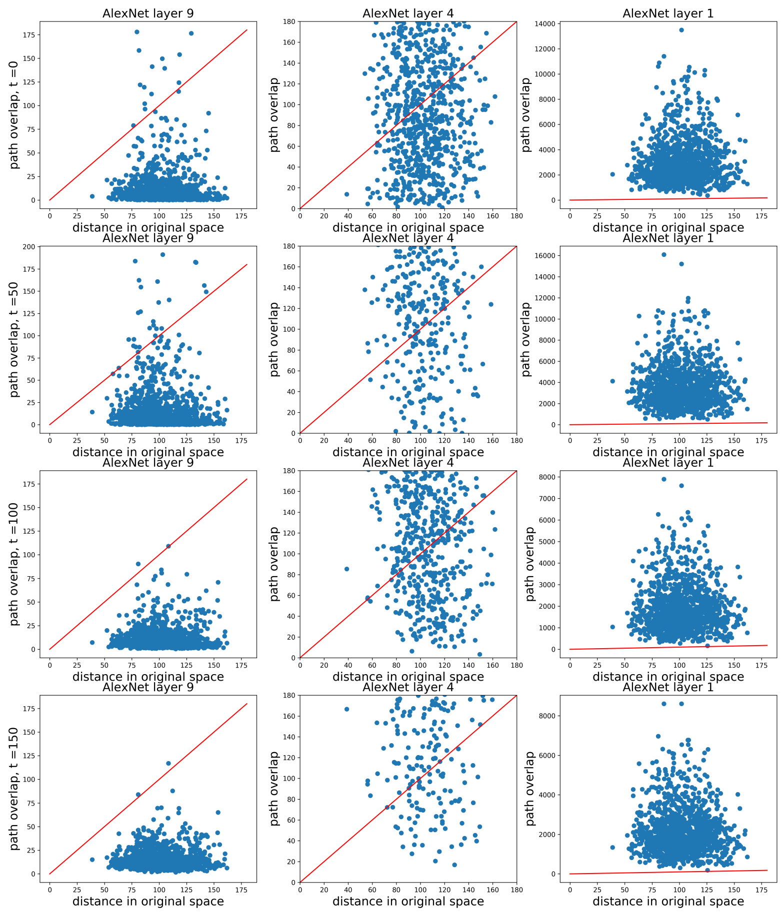

Figure 11: Path overlap $p_i^t \cdot p_j^t$ versus distance in original space $||x_i - x_j||_2$ for AlexNet trained on CIFAR 10 at various timesteps during training. Layer 6 shown on the left and Layer 2 shown on the right. We find that there is not a clean linear relationship between distance in original space and overlap in path vectors, supporting the case for considering each image update as its own function.

E ALEXNET SIMILARITY

F VGG-19 CROSS PATH SIMILARITY

We calculate the cross similarity $\frac{(p_i^t \odot p_j^t)\cdot(p_i^{t-1} \odot p_j^{t-1})}{|(p_i^t \odot p_j^t)||(p_i^{t-1} \odot p_j^{t-1})|}$ for VGG-19 trained on CIFAR 10 data and report our results in Figure 16.

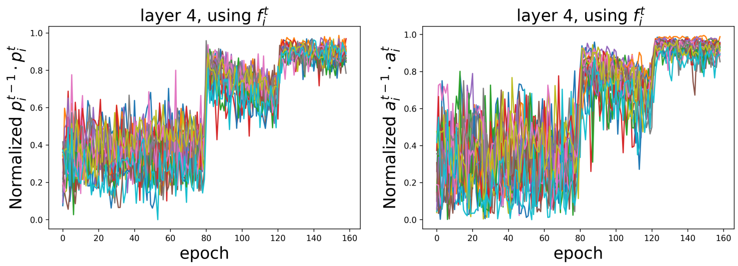

Figure 12: Self similarity $\frac{p_i^{t-1} \cdot p_i^t}{||p_i^{t-1}||||p_i^t||}$ for AlexNet trained on CIFAR-10 data We apply each function f_i^t by itself (without mixing with other functions). The variance is higher than in the mixed case, and perfect self similarity does not exist even towards the end of training.

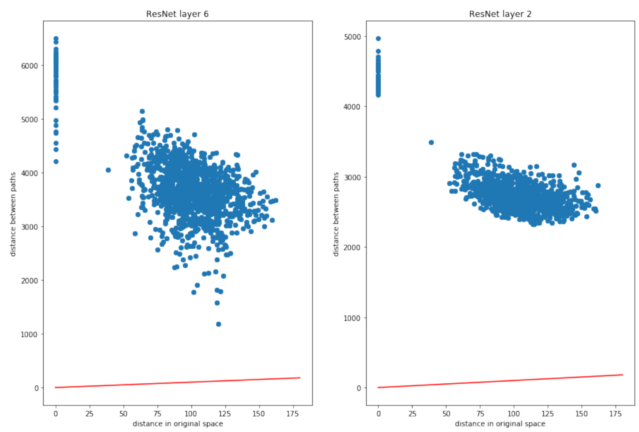

Figure 13: Path overlap $p_i^t \cdot p_j^t$ versus distance in original space $||x_i - x_j||_2$ for ResNet trained on CIFAR 10. Layer 6 shown on the left and Layer 2 shown on the right. We find that there is not a clean linear relationship between distance in original space and overlap in path vectors, supporting the case for considering each image update as its own function.

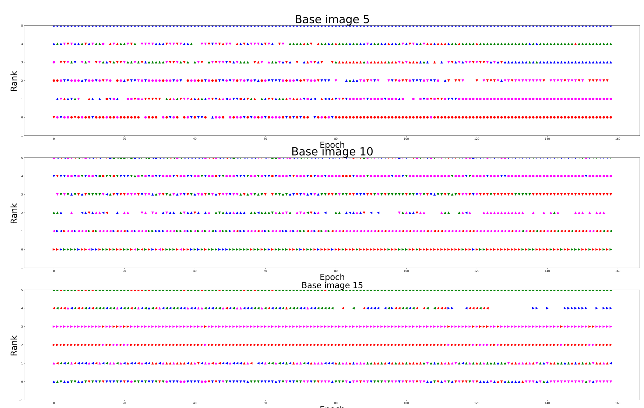

Figure 14: Plotting closeness to baseline image x_j for $j = 5, 10, 15$. Closeness defined as $\frac{a_i^{t-1} \cdot a_i^t}{||a_i^{t-1}||||a_i^t||}$ for AlexNet trained on CIFAR 10 data. We find the identity of the nearest neighbor changes.

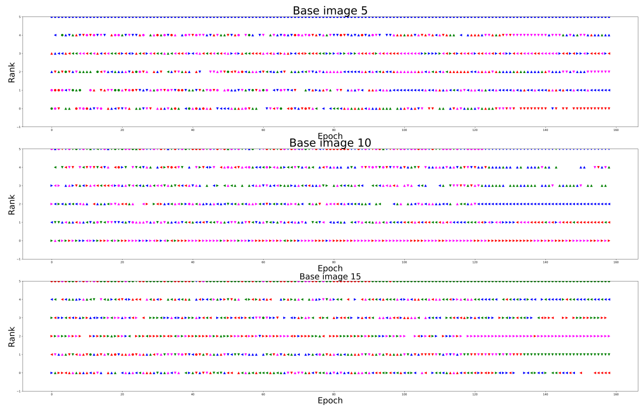

Figure 15: Plotting closeness to baseline image x_j for $j = 5, 10, 15$. Closeness defined as $\frac{a_i^{t-1} \cdot a_i^t}{||a_i^{t-1}||||a_i^t||}$ for AlexNet trained on CIFAR 10 data. We find the identity of the nearest neighbor changes.

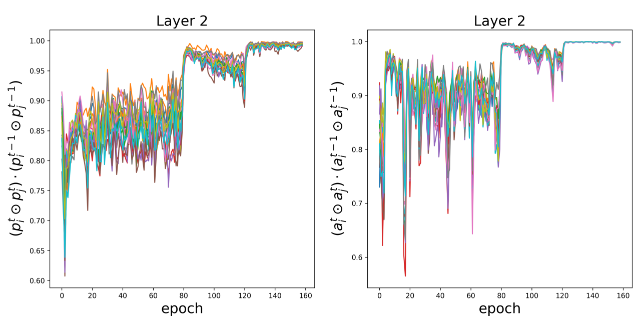

Figure 16: cross path similarity $\frac{(p_i^t \odot p_j^t) \cdot (p_i^{t-1} \odot p_j^{t-1})}{|(p_i^t \odot p_j^t)||(p_i^{t-1} \odot p_j^{t-1})|}$ for VGG-19 trained on CIFAR 10 data.