

Gradient flow

Average Gradient

0.008  
0.006  
0.004  
0.002  
0.000

layer\_dict.input\_conv.layer\_dict.conv\_0.weight  
layer\_dict.input\_conv.layer\_dict.bn\_0.weight  
layer\_dict.input\_conv.layer\_dict.bn\_0.bias  
layer\_dict.block\_0\_0.layer\_dict.conv\_0.weight  
layer\_dict.block\_0\_0.layer\_dict.conv\_1.weight  
layer\_dict.block\_0\_1.layer\_dict.conv\_0.weight  
layer\_dict.block\_0\_1.layer\_dict.conv\_1.weight  
layer\_dict.block\_0\_2.layer\_dict.conv\_0.weight  
layer\_dict.block\_0\_2.layer\_dict.conv\_1.weight  
layer\_dict.block\_0\_3.layer\_dict.conv\_0.weight  
layer\_dict.block\_0\_3.layer\_dict.conv\_1.weight  
layer\_dict.block\_0\_4.layer\_dict.conv\_0.weight  
layer\_dict.block\_0\_4.layer\_dict.conv\_1.weight  
layer\_dict.block\_0\_4.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_0.layer\_dict.conv\_0.weight  
layer\_dict.reduction\_block\_0.layer\_dict.conv\_0.bias  
layer\_dict.reduction\_block\_0.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_0.layer\_dict.conv\_1.bias  
layer\_dict.block\_1\_0.layer\_dict.conv\_0.weight  
layer\_dict.block\_1\_0.layer\_dict.conv\_1.weight  
layer\_dict.block\_1\_1.layer\_dict.conv\_0.weight  
layer\_dict.block\_1\_1.layer\_dict.conv\_1.weight  
layer\_dict.block\_1\_2.layer\_dict.conv\_0.weight  
layer\_dict.block\_1\_2.layer\_dict.conv\_1.weight  
layer\_dict.block\_1\_3.layer\_dict.conv\_0.weight  
layer\_dict.block\_1\_3.layer\_dict.conv\_1.weight  
layer\_dict.block\_1\_4.layer\_dict.conv\_0.weight  
layer\_dict.block\_1\_4.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_1.layer\_dict.conv\_0.weight  
layer\_dict.reduction\_block\_1.layer\_dict.conv\_0.bias  
layer\_dict.reduction\_block\_1.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_1.layer\_dict.conv\_1.bias  
layer\_dict.block\_2\_0.layer\_dict.conv\_0.weight  
layer\_dict.block\_2\_0.layer\_dict.conv\_1.weight  
layer\_dict.block\_2\_1.layer\_dict.conv\_0.weight  
layer\_dict.block\_2\_1.layer\_dict.conv\_1.weight  
layer\_dict.block\_2\_2.layer\_dict.conv\_0.weight  
layer\_dict.block\_2\_2.layer\_dict.conv\_1.weight  
layer\_dict.block\_2\_3.layer\_dict.conv\_0.weight  
layer\_dict.block\_2\_3.layer\_dict.conv\_1.weight  
layer\_dict.block\_2\_4.layer\_dict.conv\_0.weight  
layer\_dict.block\_2\_4.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_2.layer\_dict.conv\_0.weight  
layer\_dict.reduction\_block\_2.layer\_dict.conv\_0.bias  
layer\_dict.reduction\_block\_2.layer\_dict.conv\_1.weight  
layer\_dict.reduction\_block\_2.layer\_dict.conv\_1.bias  
logit\_linear\_layer.weight  
logit\_linear\_layer.bias

Layers