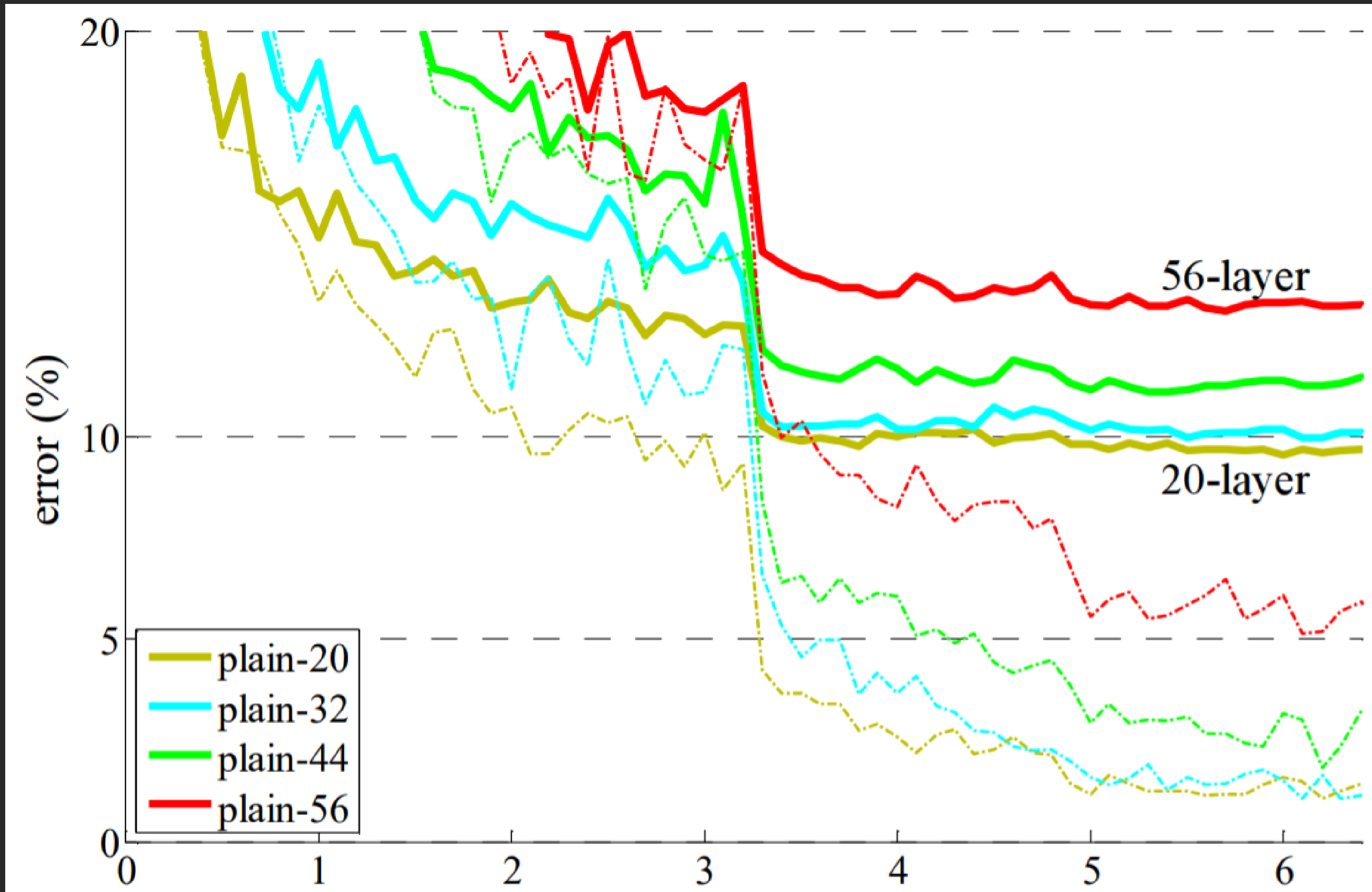
關於 ResNet 的三兩事

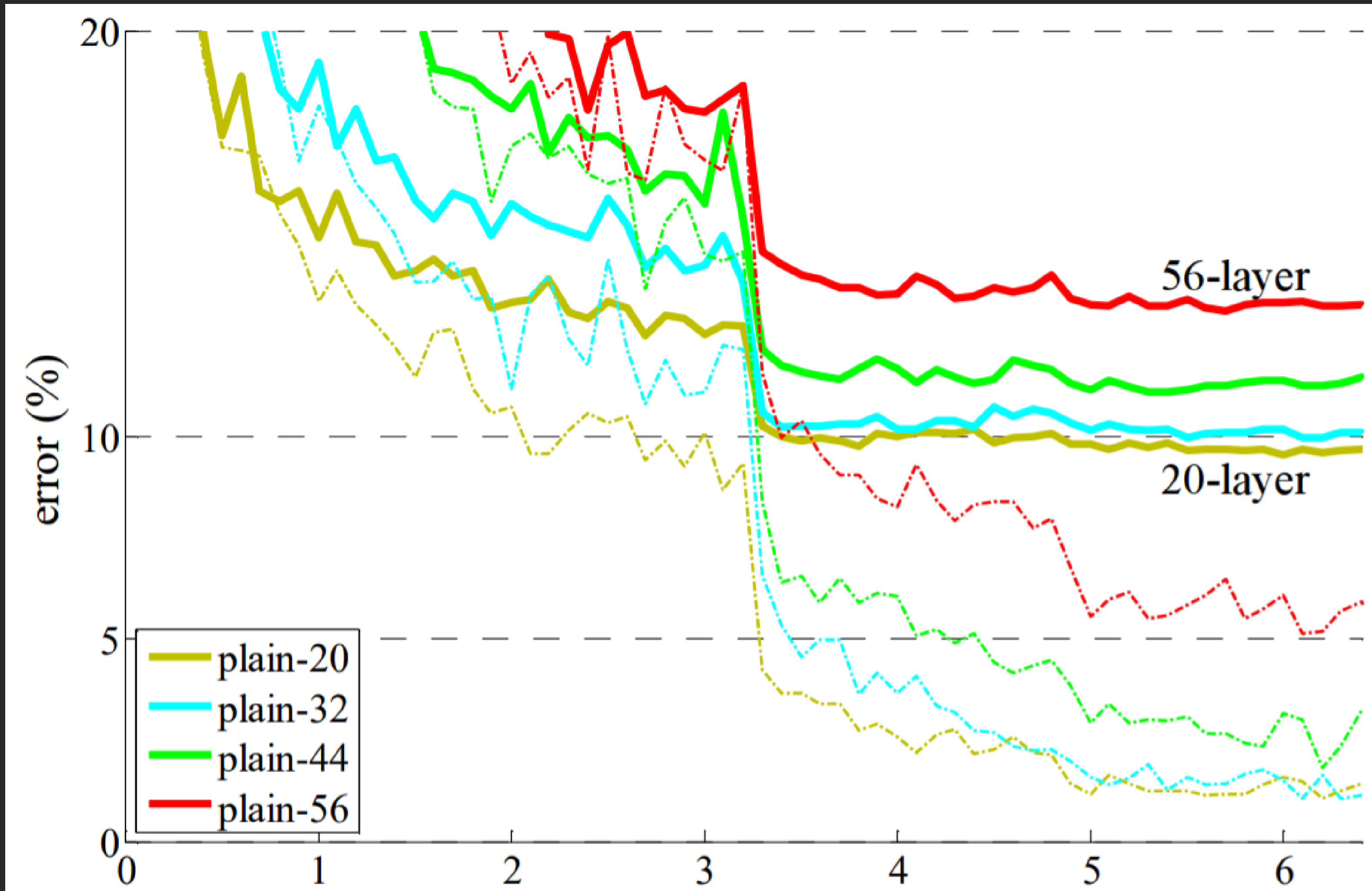
問題：模型是不是越深越好

答案是否定的



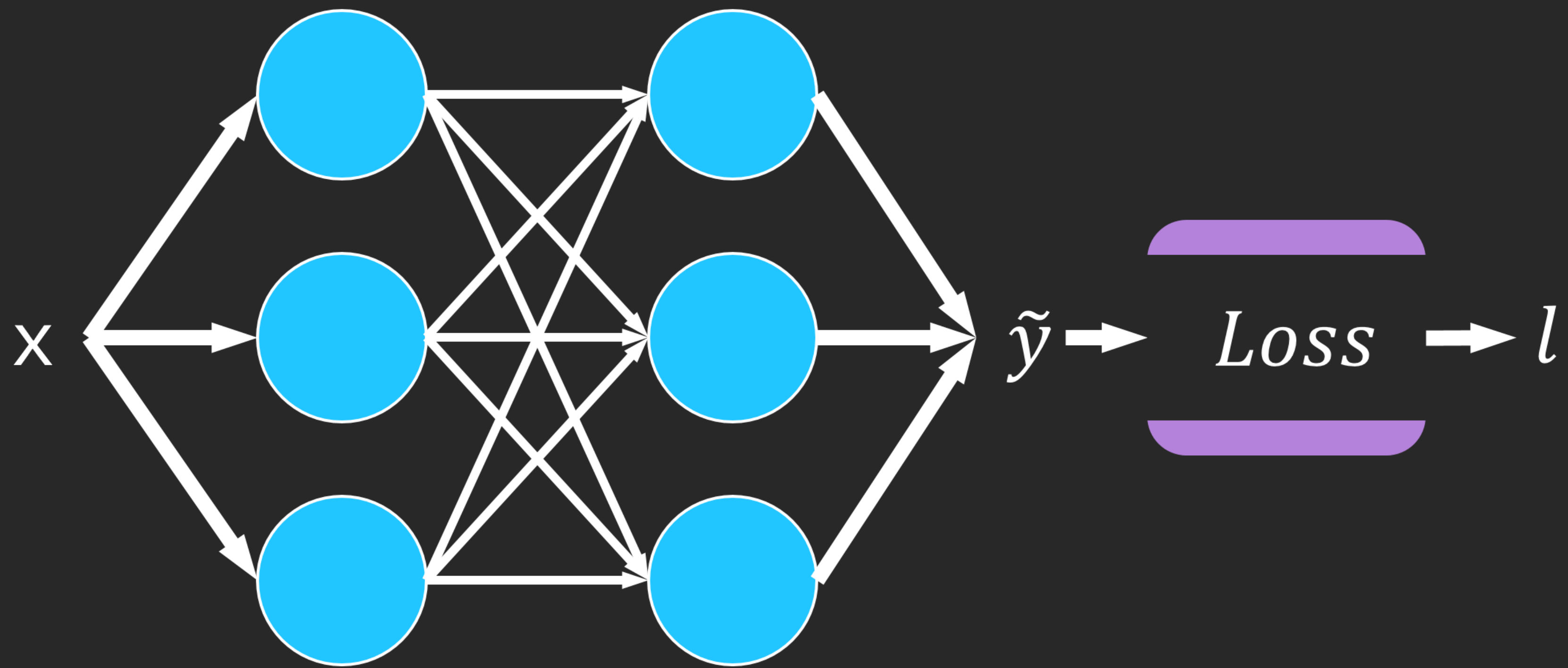
20
10
5
0
error (%)
0
1
2
3
4
5
6
56-layer
20-layer
plain-20
plain-32
plain-44
plain-56

可以看到，增加層數反而使 error 增加了



這個現象是源於

梯度消失
Vanishing Gradient


x
$\tilde{y}$
$Loss$
l

$$x \rightarrow f_1 \rightarrow h_1 \rightarrow f_2 \rightarrow h_2 \rightarrow \ldots$$

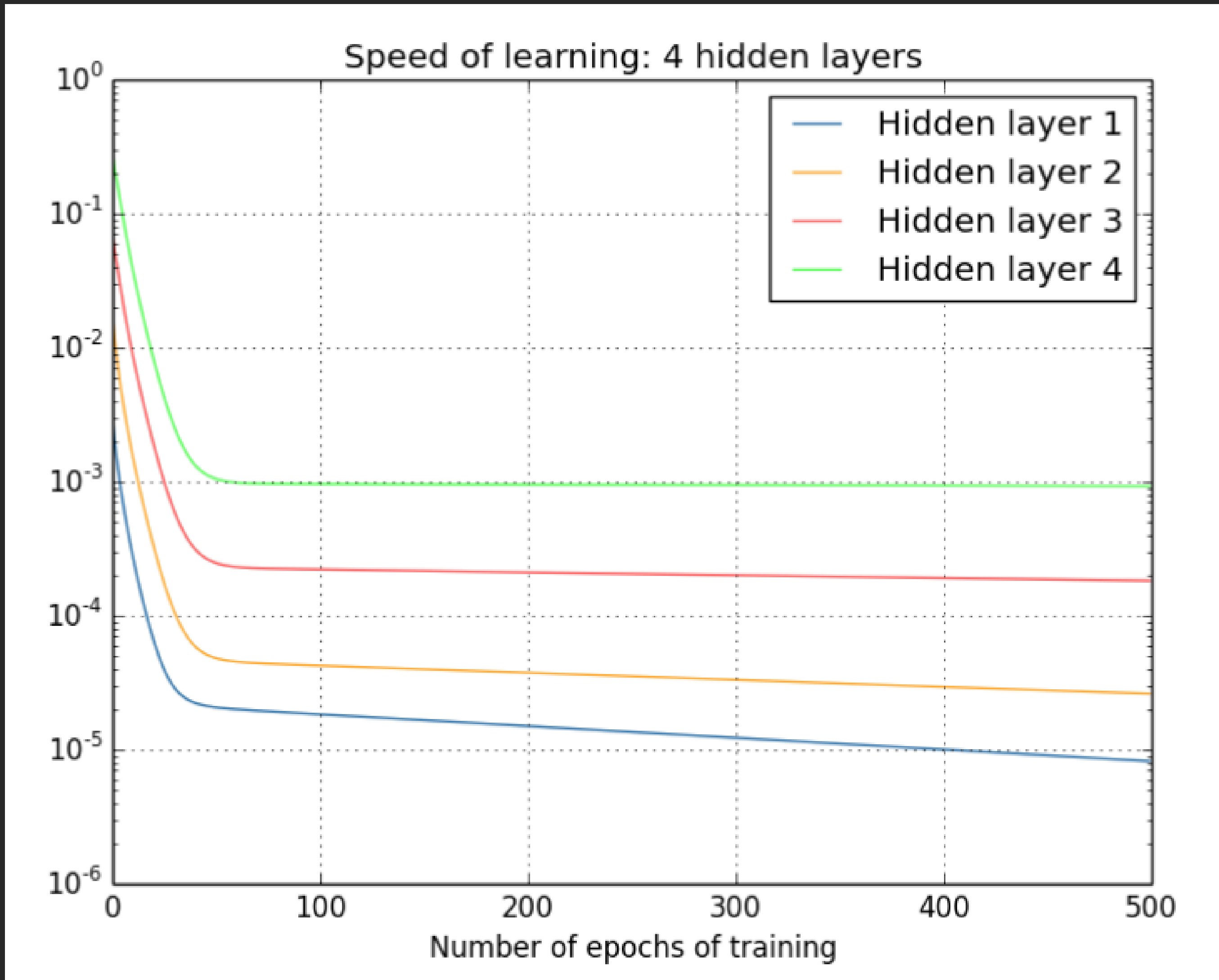
$$\ldots \rightarrow f_N \rightarrow h_N \rightarrow Loss \rightarrow l$$

$$h_i = f_i(h_{i-1}) = \alpha(w_i h_{i-1} + b_i)$$

$$\nabla w_i = \frac{\partial l}{\partial w_i} = \frac{\partial l}{\partial h_N} \frac{\partial h_N}{\partial h_{N-1}} \cdots \frac{\partial h_{i+1}}{\partial h_i} \frac{\partial h_i}{\partial w_i}$$

$$\nabla w_i = \frac{\partial l}{\partial w_i} = \frac{\partial l}{\partial h_N} \underbrace{\frac{\partial h_N}{\partial h_{N-1}} \cdots \frac{\partial h_{i+1}}{\partial h_i}}_{\text{越多項越容易出現梯度消失or爆炸}} \frac{\partial h_i}{\partial w_i}$$

越靠前面的 layer 更新越慢

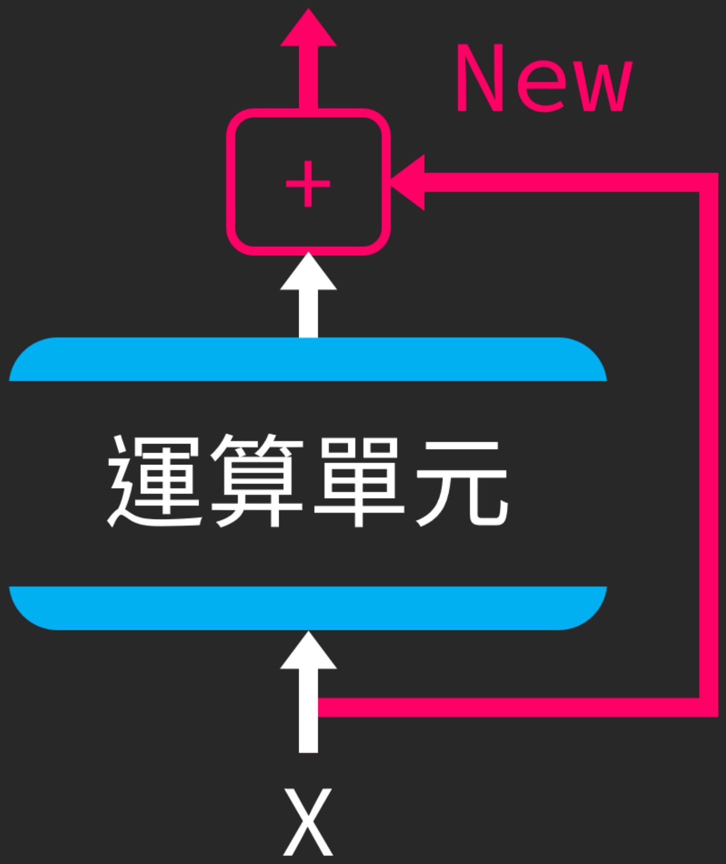


超有效又簡單的解決方案

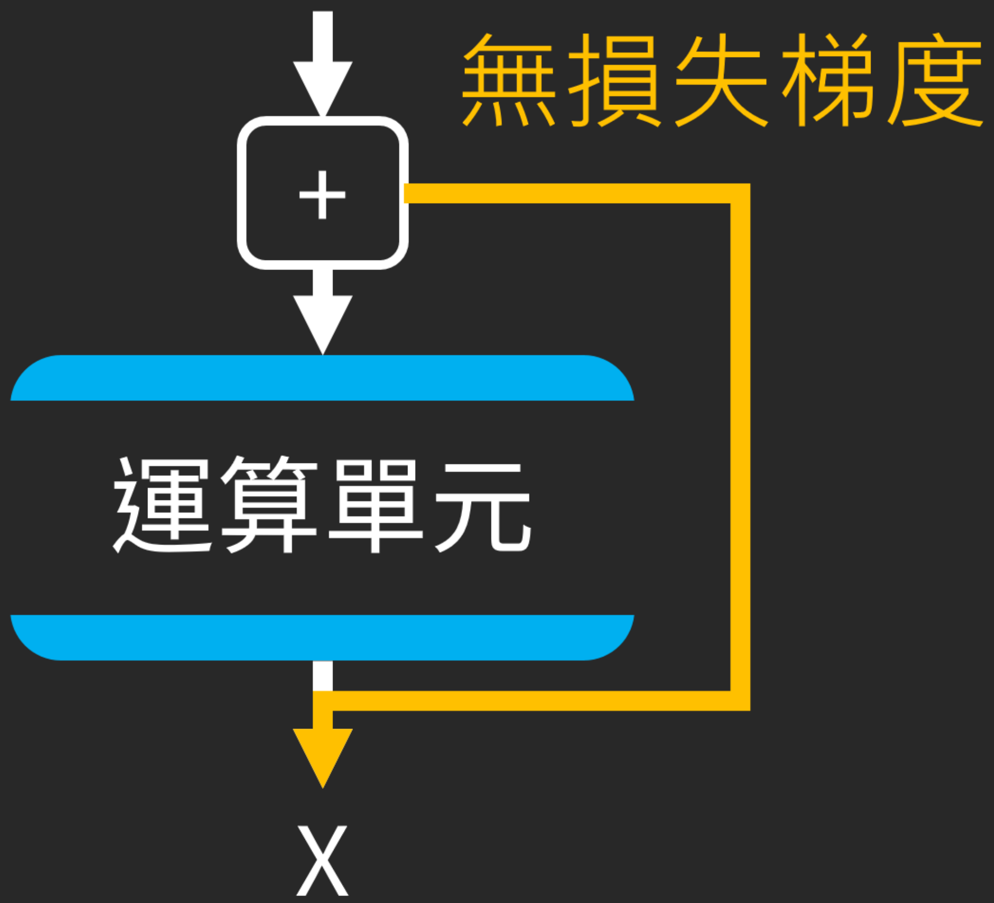
Residual Block

白話文：加一個 shortcut

Shortcut Forward



Shortcut Backward

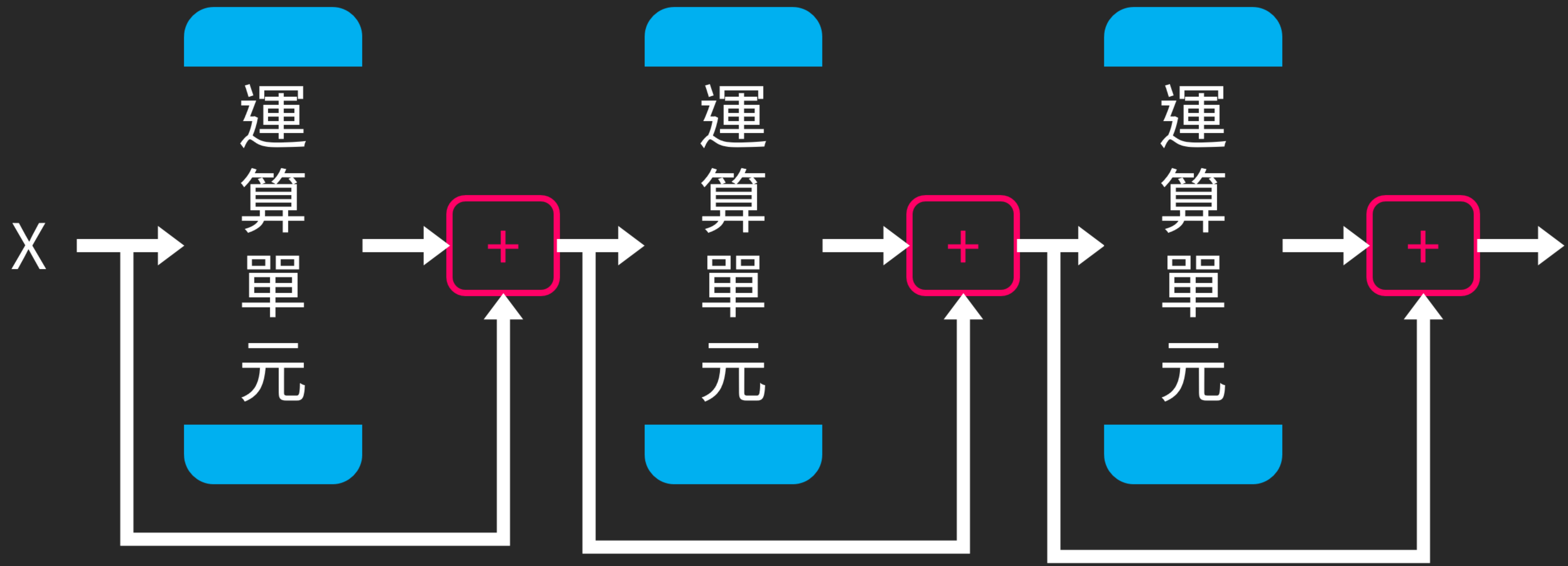


shortcut 達成了

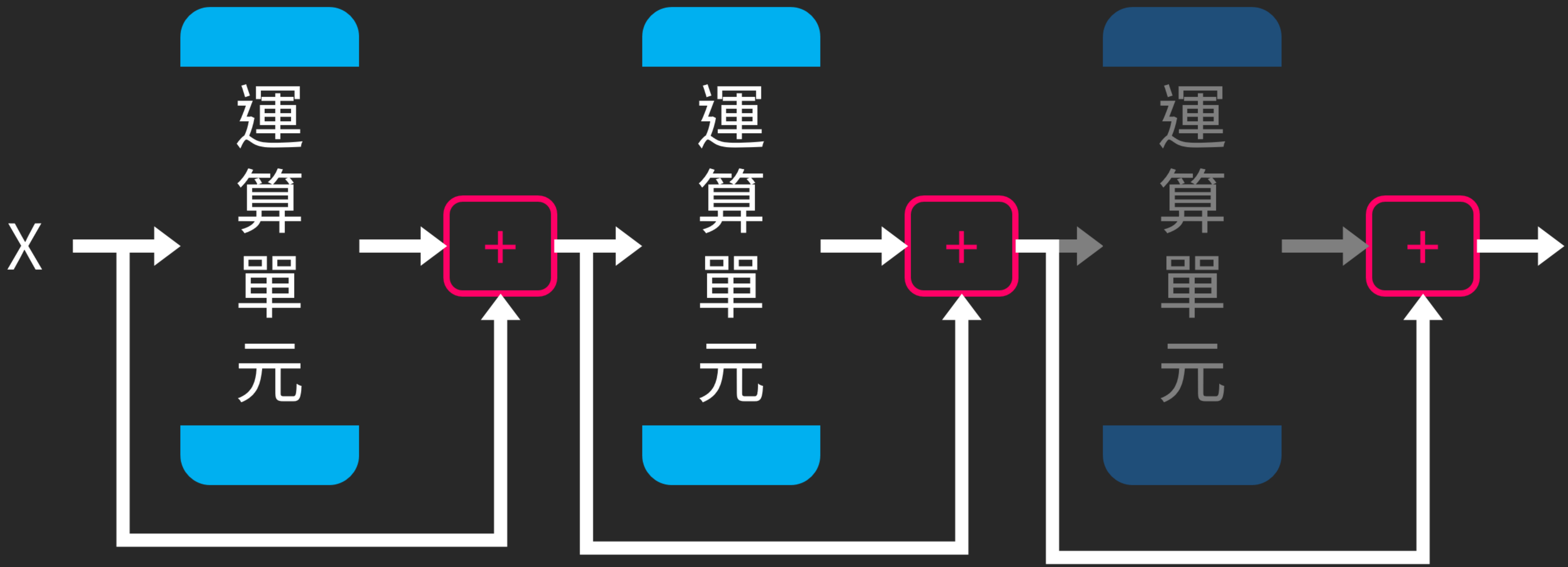
恆等映射
Identity Mappings

消除了冗於結構的負面影響

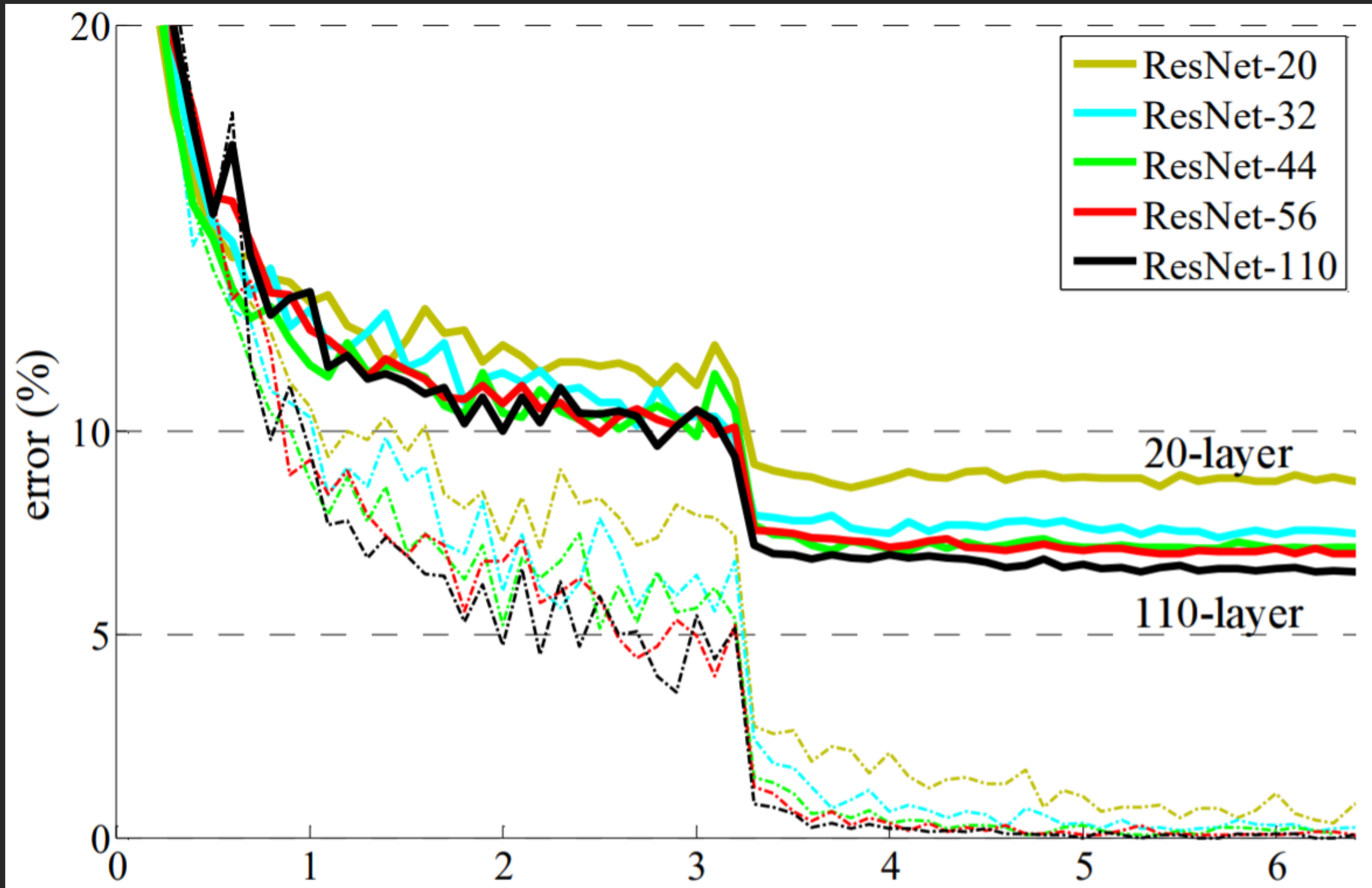
人為設計的多層結構，不一定每一層都有用處



模型只要將冗於的部分歸 0 就不會受到拖累

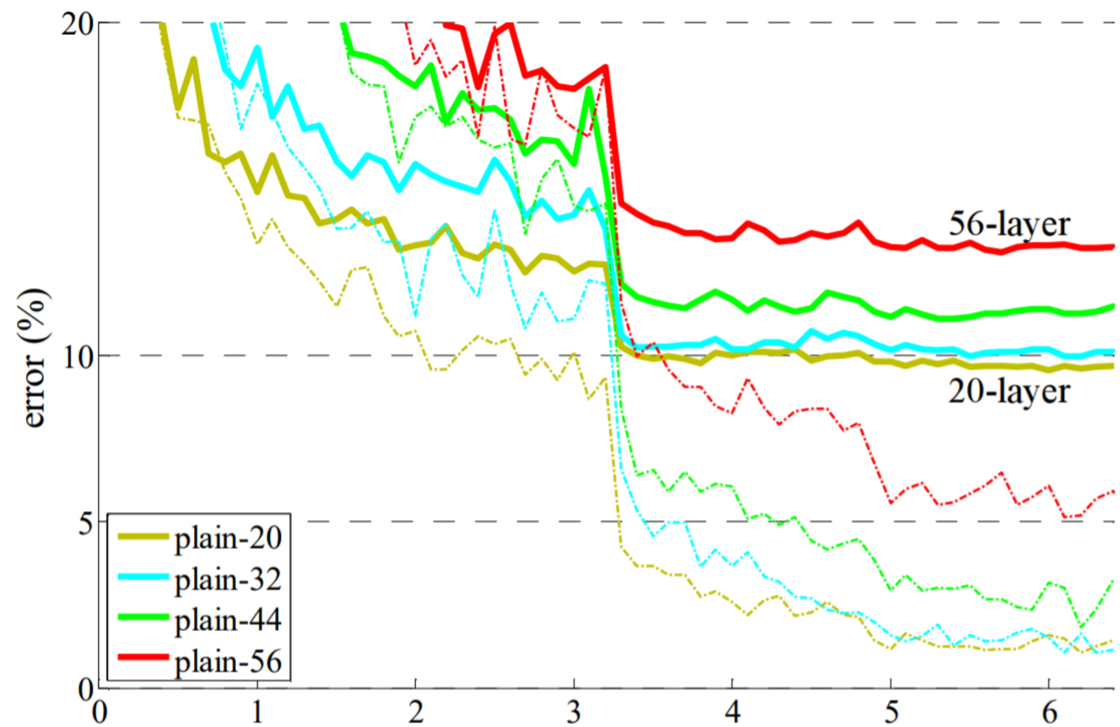


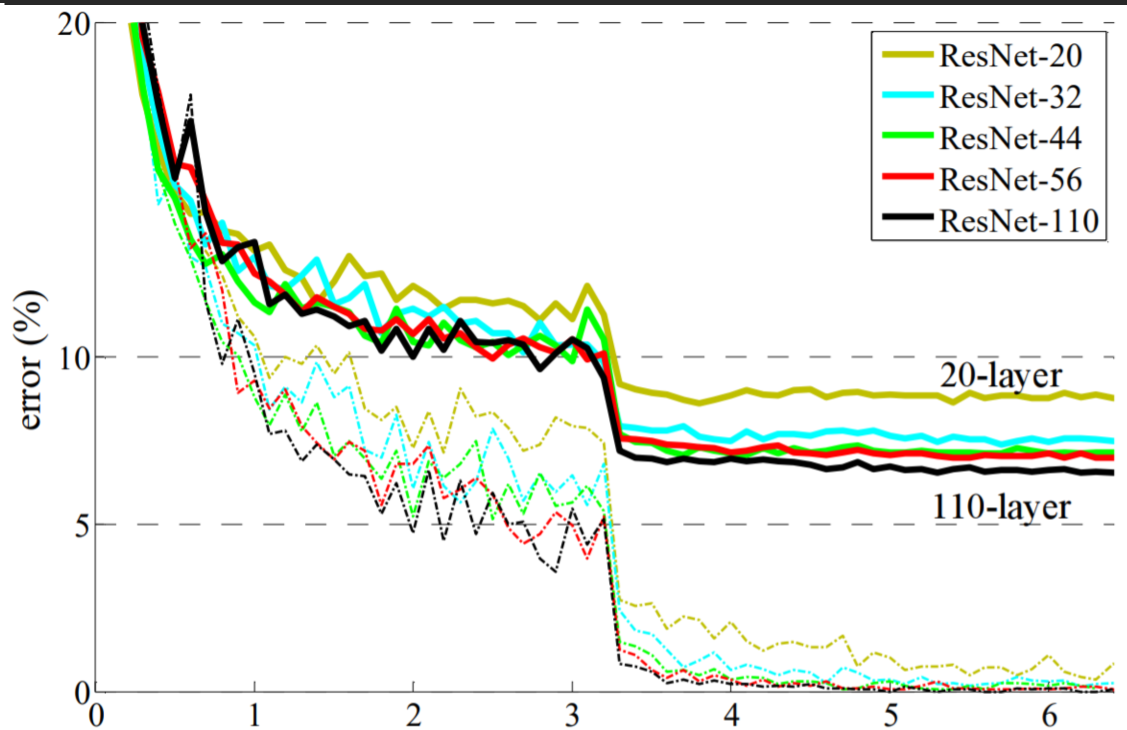
ResNet 效果



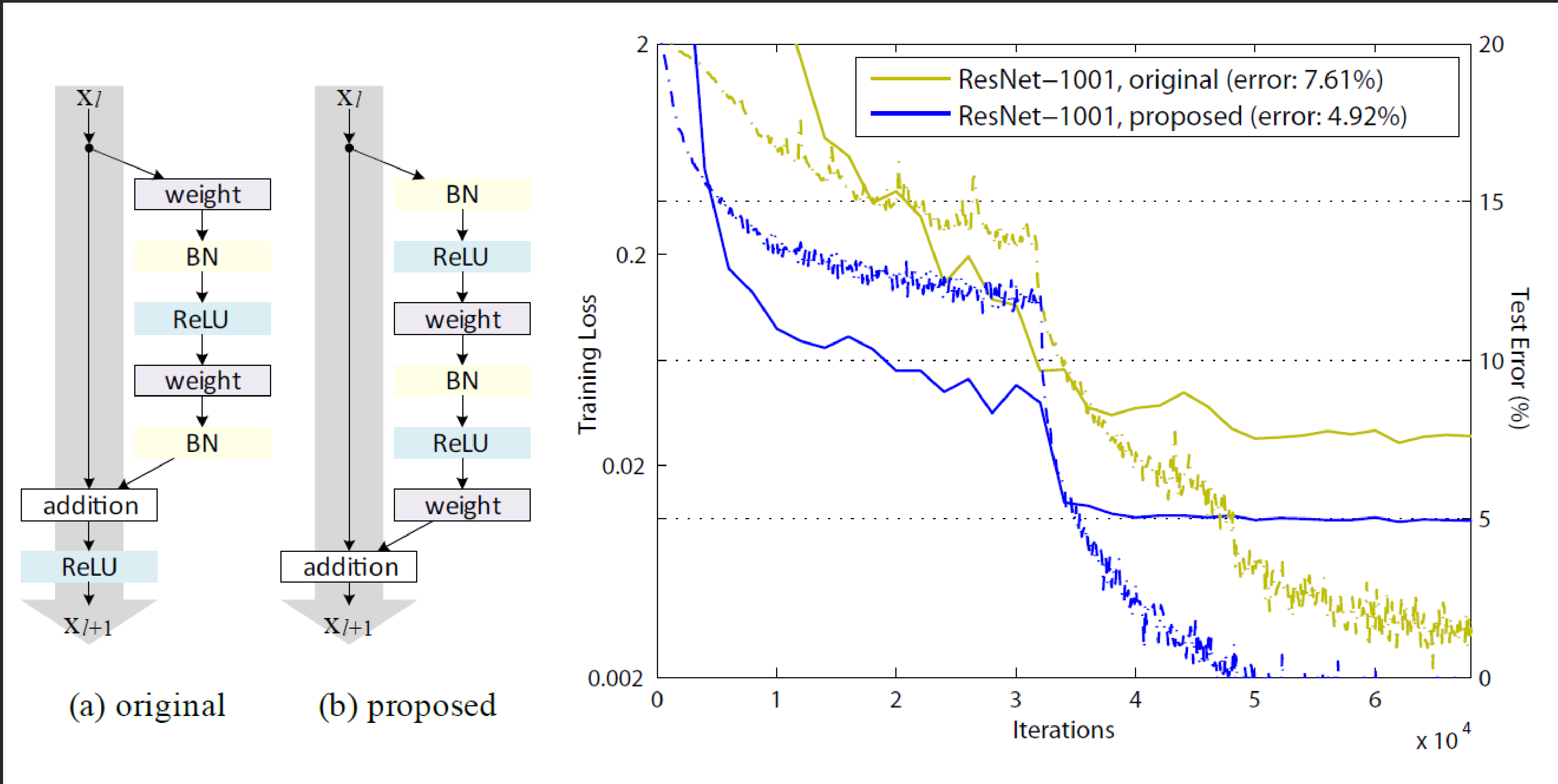
ResNet-20
ResNet-32
ResNet-44
ResNet-56
ResNet-110
20-layer
110-layer
error (%)
20
10
5
0
0
1
2
3
4
5
6

PlainCNN vs ResNet





Shortcut 越乾淨越好



更多詳細的實驗請看

Deep Residual Learning for Image Recognition

與

Identity Mappings in Deep Residual Networks

參考

[Identity Mappings in Deep Residual Networks 論文介紹]