deeplearning.ai

Case Studies

Residual Networks (ResNets)

Residual block

$a^{[l]}$ $a^{[l+1]}$ $a^{[l+2]}$

"shortcut"/skip connection

$a^{[l]}$ → Linear → ReLU $a^{[l+1]}$ → Linear → ⊕ → Relu → $a^{[l+2]}$

"main path"
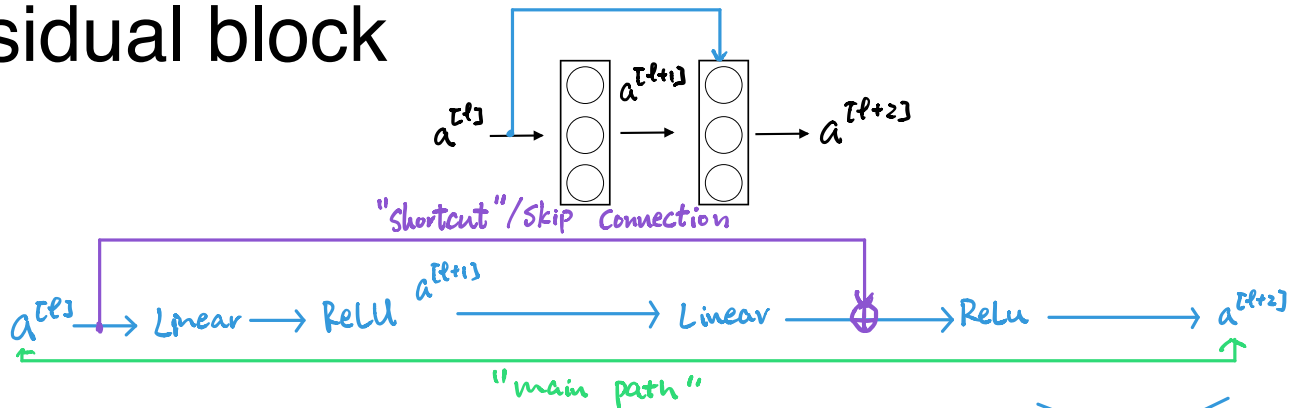
$$z^{[l+1]} = W^{[l+1]} a^{[l]} + b^{[l+1]} \qquad a^{[l+1]} = g(z^{[l+1]}) \qquad z^{[l+2]} = W^{[l+2]} \cdot a^{[l+1]} + b^{[l+2]} \qquad \cancel{a^{[l+2]} = g(z^{[l+2]})}$$

$$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$$

Andrew Ng

Residual Network

"Plain network"

$$x \rightarrow a^{[l]}$$

Plain

training error vs # layers — "Reality", Theory

ResNet

training error vs # layers

[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng