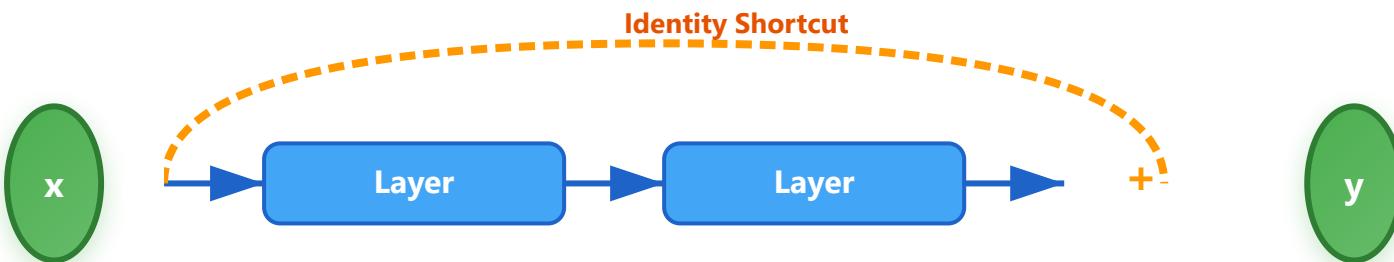


Skip Connection (ResNet): Breakthrough Architecture

CORE FORMULA

$$y = F(x) + x$$



Extreme Depth

Enabled training 152+ layer networks



Gradient Highway

Direct path for gradients to early layers



Residual Learning

Learns $F(x)$ instead of full $H(x)$



Math Benefit

Gradient always has component of 1