Gradient Flow Analysis

GOAL

Understand how gradients propagate through different activations

ReLU

⚠ Risky

Binary gradient (0 or 1)

Can block gradients completely when inactive

Sigmoid / Tanh

✗ Poor

Max derivative: 0.25

Gradients shrink exponentially

Leaky ReLU

✓ Good

Always allows gradient

Small slope in negative region

Smooth (ELU, GELU, Swish)

★ Best

Continuous gradients

Better for optimization

🎯 Ideal Property

Gradients neither vanish nor explode across depth

Modern Approach

Good Activations + **Normalization** + **Skip Connections**