Comparison of Initialization Strategies

Method	Best Use Case	Recommendation
Zero Init	**Never use** - causes symmetry problem	✗ Avoid
Random Small Values	Basic approach, often **insufficient** for deep networks	⚠ Basic
Xavier/Glorot	Best for **tanh/sigmoid** activations	✓ Recommended
He Initialization	Best for **ReLU** and variants (LeakyReLU, PReLU)	✓ Recommended
LSUV	Best for **extremely deep networks** (100+ layers)	✓✓ Advanced
Pre-trained Weights	Best when **transfer learning** is applicable	✓✓ Preferred

Key Takeaway: Choice depends on network architecture and activation functions. Modern practice favors He

initialization for ReLU networks and pre-trained weights when available.