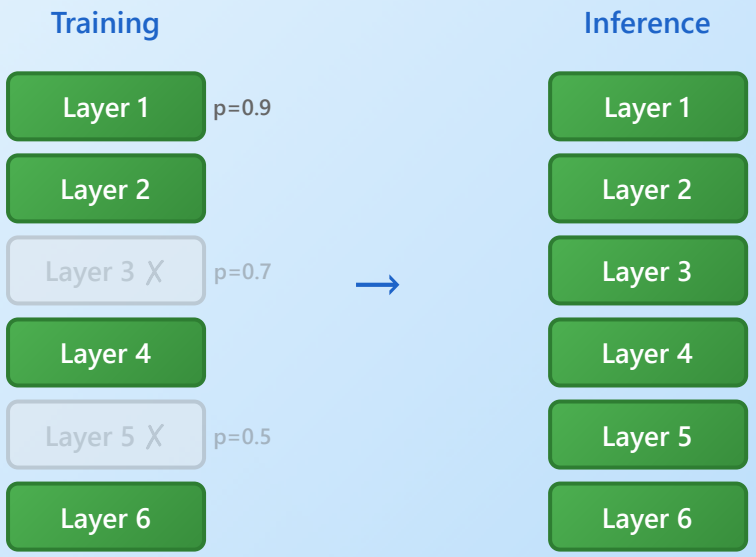
Stochastic Depth

Training vs Inference



Survival Probability

Decreases with layer depth

Early layers (High) — Deep layers (Low)

Key Benefits

- Randomly drops **entire layers** during training
- Creates **shorter networks** during training
- **Improves gradient flow** to early layers
- **Reduces training time** significantly

✓ Enables Extreme Depth

Training of **extremely deep networks** (1000+ layers)