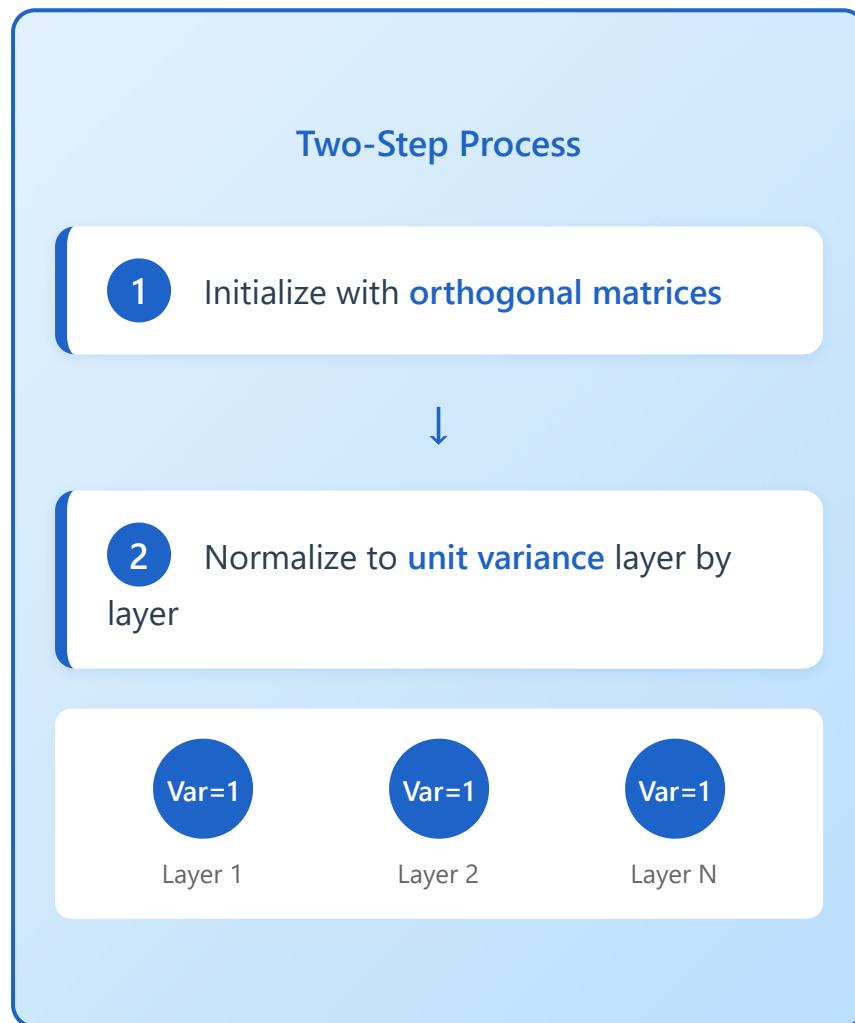


LSUV Initialization

Layer-Sequential Unit-Variance Initialization



Key Characteristics

Sequential process: Initializes and normalizes one layer at a time

Orthogonal initialization: Provides better starting point than random

Data-driven: Uses actual data to estimate variance

More sophisticated: Advanced compared to Xavier/He methods

✓ **Particularly Effective**
Excels in **very deep networks** with **100+ layers**

Uses **small batch of data** for variance estimation

 Additional Resource: DeepLearning.AI - Weight Initialization Guide