

LSUV Initialization

Layer-Sequential Unit-Variance Initialization

Two-Step Process

1 Initialize with **orthogonal matrices**



2 Normalize to **unit variance** layer by layer

Var=1

Layer 1

Var=1

Layer 2

Var=1

Layer N

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