

Parameter Efficiency: Deep vs. Shallow Networks

Shallow Network

1000 units

Layers:

Total Units:

Parameters:

1
1,000
1M+

Deep Network

100 units

100 units

100 units

Layers:

Total Units:

Parameters:

3
300
30K

Computational Complexity (n inputs)

Shallow

$O(2^n)$

vs

Deep

$O(n^2)$

✓ Better generalization

✓ Faster training

✓ Reduced redundancy

✓ Faster inference