Limitations of Shallow Networks

UNIVERSAL APPROXIMATION THEOREM

Single hidden layer networks can theoretically approximate any function
BUT exponentially many hidden units may be required

↓

No Hierarchical Patterns (1)

Cannot efficiently capture hierarchical structure in images and text

Compositional Struggle (2)

Must learn all combinations from scratch without reusing patterns

Poor Generalization (3)

Weak performance on high-dimensional data with limited samples

Limited Abstraction (4)

Manual feature engineering still required for complex tasks