FFN Structure (Layer 0)

Input
[hidden_size]

Feed-Forward Network (Dense)

up_proj
[h → ffn_dim]

gate_proj
[h → ffn_dim]
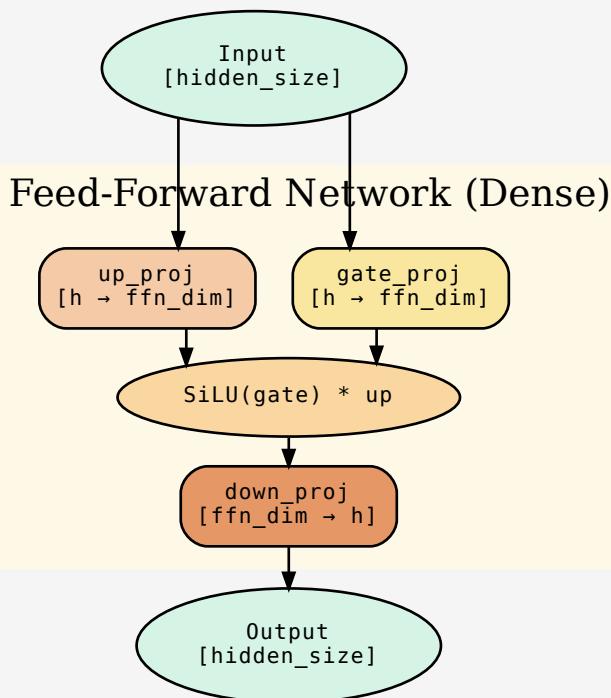
SiLU(gate) * up

down_proj
[ffn_dim → h]

Output
[hidden_size]

Model: qwen2
Layers: 24
Hidden size: 896
Attention heads: 14
Max position: 131072