Adapter Architecture: Small Trainable Modules
Inserted Between Frozen Layers

Output

Transformer Layer 6
(FROZEN)

Adapt

Transformer Layer 5
(FROZEN)

Transformer Layer 4
(FROZEN)

Legend:

Frozen
(0 train)

Adapt

Transformer Layer 3
(FROZEN)

Trainable
(adapt)

Parameters:
Base: 99%
Adapters: 1%

Transformer Layer 2
(FROZEN)

Adapt

Transformer Layer 1
(FROZEN)

Input

Key: Freeze expensive pre-trained weights, train cheap adapters (1-2% parameters)