Token Depths

Tokens remain: 197

Layer 1

Tokens remain: 190

Layer 2

Tokens remain: 10

Layer K

Embedding

Halting probability f

Transformer Block

Token halting

Token removal & reorg.

Adaptive Halting

Transformer Block

Adaptive Halting

Transformer Block

Adaptive Halting

Class token memory

Mean-field aggregation

Task head

ImageNet1K Examples for Adaptive Tokens