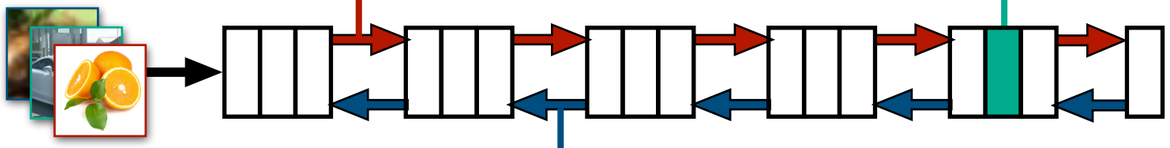
ZEST Dataset Construction

Training Input Collection:

Zero-shot **Forward** metrics
- Activation

Zero-shot **Static** metrics
- Weight

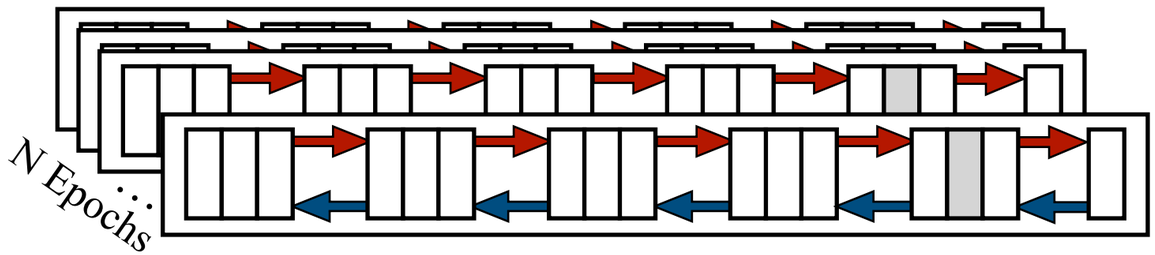


Zero-shot **Backward** metrics
- Gradient
- Fisher
- Snip
- Synflow

Collect $\text{feat}_{fwd}[i]$, $\text{feat}_{bwd}[i]$, $\text{feat}_{s}[i]$ for each layer i

Training Label Collection:

For each layer #i, fine-tune until convergence

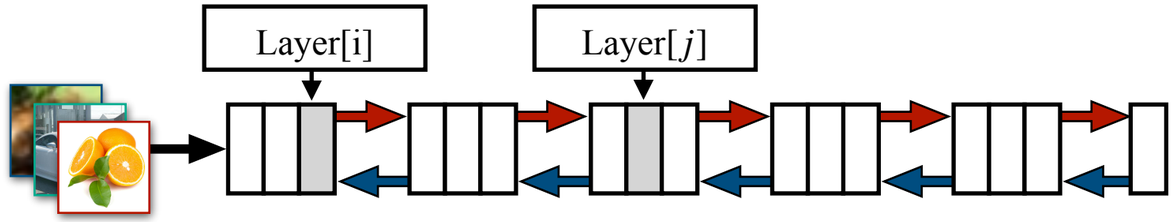


Obtain final accuracy and record as $\text{Score}_{gt}[i]$

Cost: train #layers × N epochs (expensive)

ZEST Train and Inference

Train ZEST Predictor:



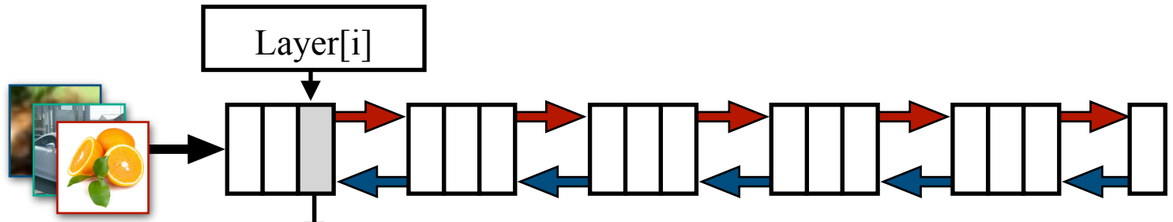
Input: $\text{feat}_{fwd}[i]$, $\text{feat}_{bwd}[i]$, $\text{feat}_{s}[i]$
Label: $\text{Score}_{gt}[i]$

Input: $\text{feat}_{fwd}[j]$, $\text{feat}_{bwd}[j]$, $\text{feat}_{s}[j]$
Label: $\text{Score}_{gt}[j]$

ZEST

Pair-wise Ranking Loss

ZEST Inference:



Input: $\text{feat}_{fwd}[i]$, $\text{feat}_{bwd}[i]$, $\text{feat}_{s}[i]$

ZEST Predictor → **Score**$_{pred}[i]$

Cost: one ZEST forward (cheap)