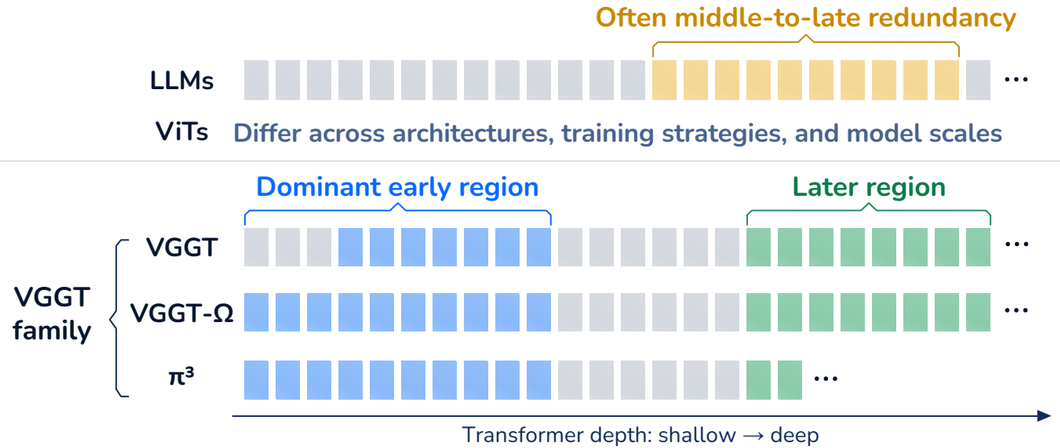
(a) Layer Redundancy Patterns

§3 · Finding 1



1 **Two redundancy regions *consistently* observed in VGGT Family**

3 models, 3,018 pruning settings | *exhaustively evaluated across 4 datasets and 7 geometric metrics*

(b) Structured Pruning

§4 · Findings 2–3

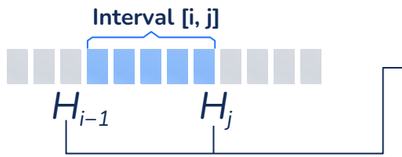
Approximate Additivity



$D(A \cup B) \approx D(A) + D(B)$

2 **Approximate additivity enables $O(L^4) \rightarrow O(L^2)$ two-interval pruning search**

Efficient CKA Proxy

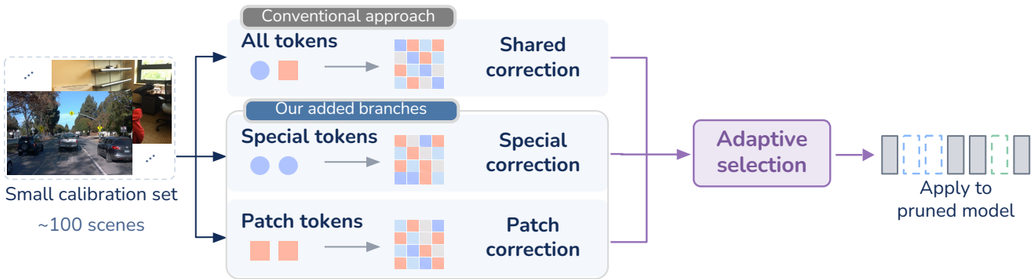


1 – CKA → Approximate Interval Degradation

3 **CKA provides an effective one-pass proxy for two-interval pruning**

(c) Calibration-based Recovery

§5 · Finding 4



4
- **Linear calibration restores accuracy after pruning**
- **Token type-specific mapping further improves**

(d) Accuracy and Efficiency

§6 · Evaluation

