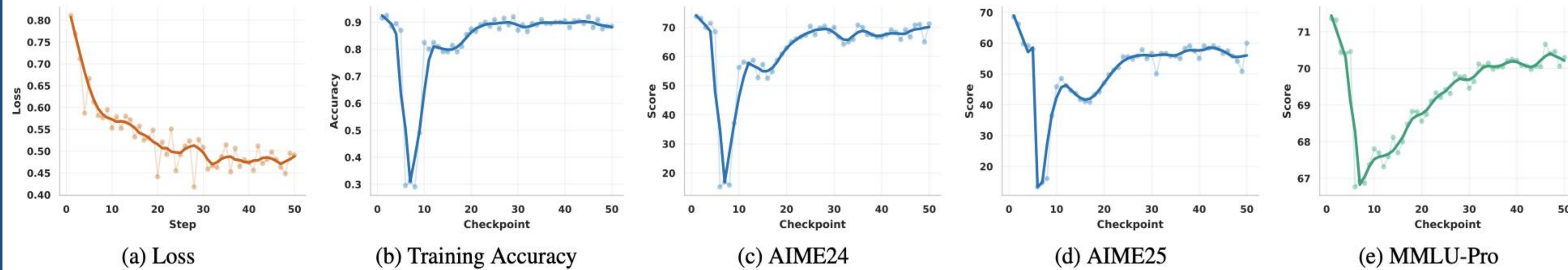


Takeaway 1. In continual distillation, loss can fall while performance crashes then recovers; We term this phenomenon **Imitation Shock**.



Takeaway 3. Direct continual distillation fails not because reasoning is absent, but because it is systematically delayed by imitation anchors.

Takeaway 4. Effective continual distillation requires clearing the optimization path for yet-to-learn tokens.

Takeaway 2. Pre-bottleneck updates are not necessary for effective continual distillation; keeping only post-bottleneck updates works better.

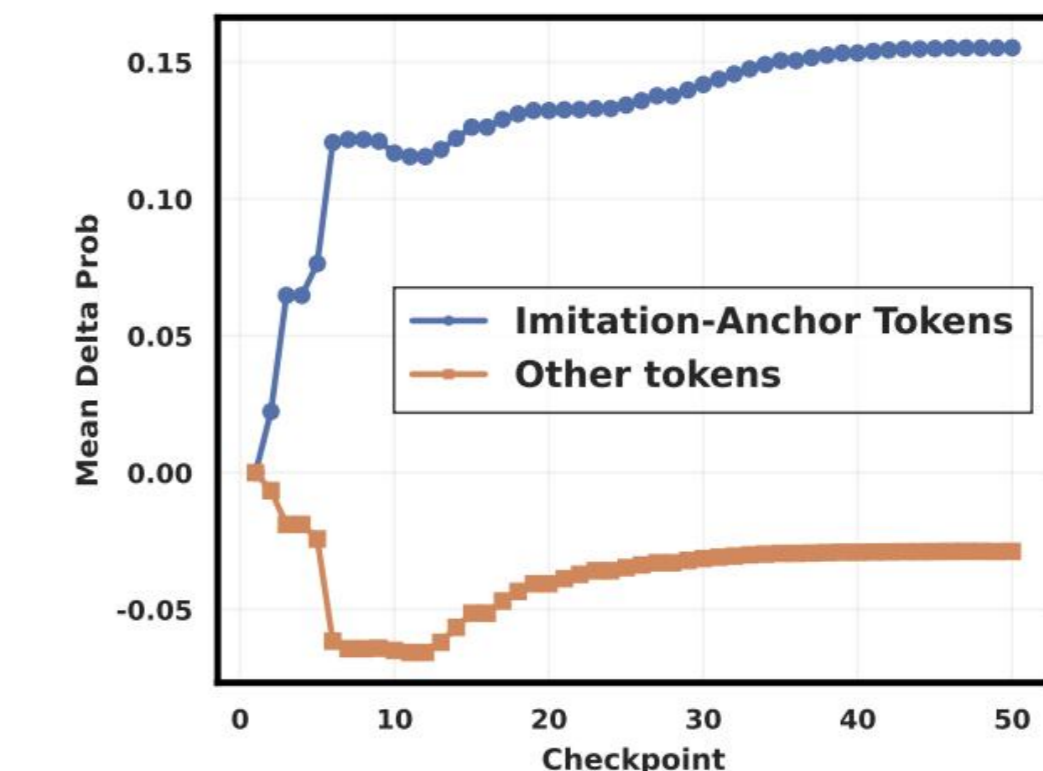
Table 1. Recovering-residual transfer extracted from the Imitation Bottleneck. Δ Avg denotes the average performance change relative to the BASE model. Standard distillation leads to degradation, while retaining only the post-bottleneck updates converts the drop into consistent gains.

Method	BOBA-200			S1K-200		
	AIME24	AIME25	Δ Avg	AIME24	AIME25	Δ Avg
BASE	75.83	67.08	–	75.83	67.08	–
SFT (R1)	71.25	55.00	↓8.33	72.50	55.84	↓7.29
RRT (R1)	76.67	68.54	↑1.15	76.67	70.63	↑2.20
SFT (QWQ)	73.33	63.33	↓3.13	73.33	64.17	↓2.70
RRT (QWQ)	76.04	69.17	↑1.15	75.00	71.67	↑1.88

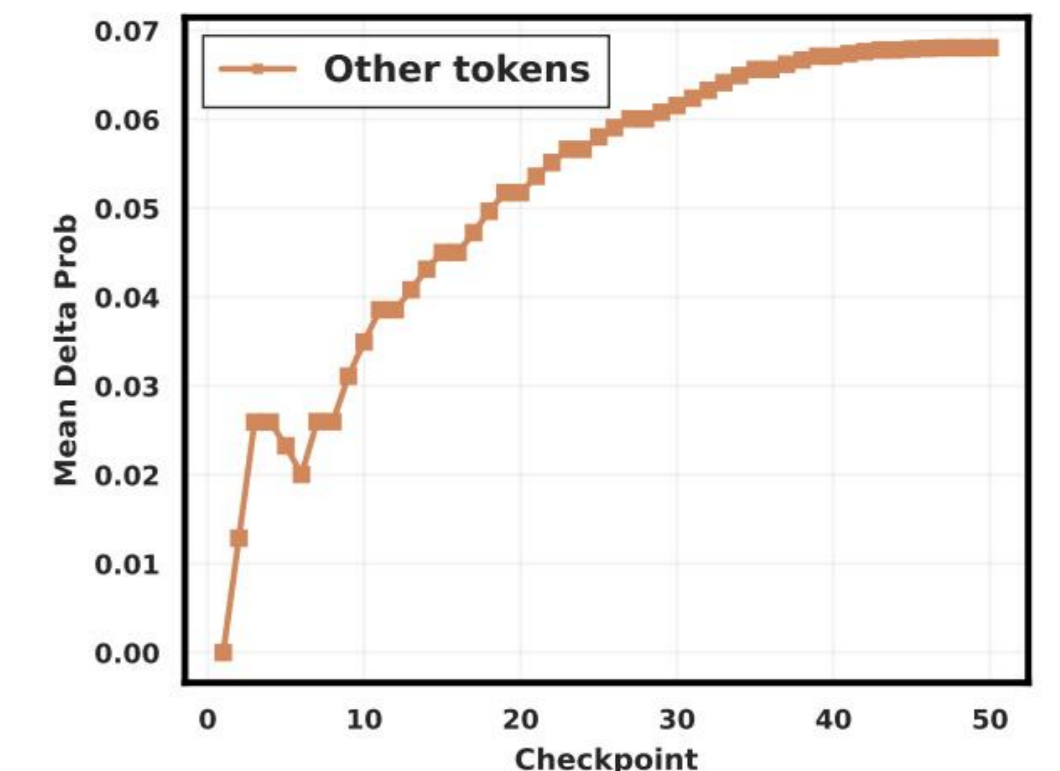
4. Main Results

Method	BOBA-200			S1K-200		
	AM24	AM25	AVG	AM24	AM25	AVG
BASE	75.83	67.08	71.46	75.83	67.08	71.46
SFT (R1)	71.25	55.00	63.13	72.50	55.83	64.17
RRT (R1)	76.67	68.54	72.61	76.67	70.63	73.65
-T3S (R1)	30.63	25.63	28.13	26.67	26.67	26.67
T3S (R1)	80.63	73.96	77.30	80.00	73.13	76.57
SFT (QWQ)	73.33	68.33	70.83	73.33	63.33	68.33
RRT (QWQ)	76.04	69.17	72.61	75.00	71.67	73.33
-T3S (QWQ)	50.00	33.33	41.67	46.67	40.00	43.33
T3S (QWQ)	77.50	70.83	74.17	76.67	72.50	74.59

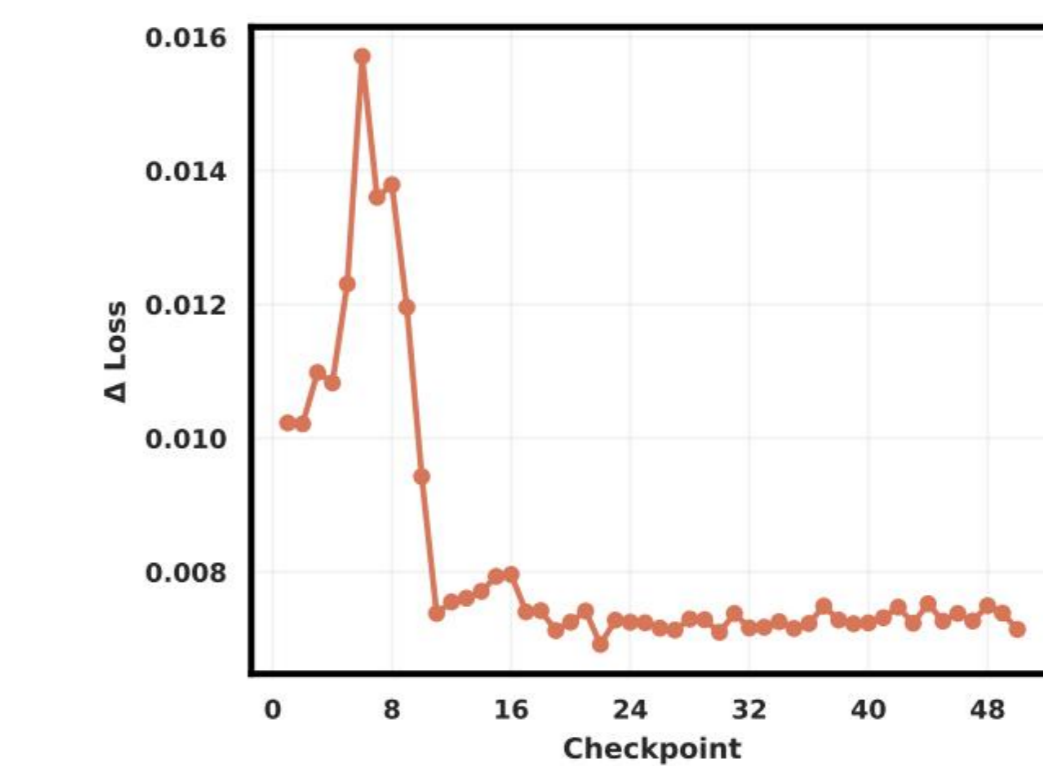
Model	AIME25	MATH500	TheoremQA	AVG
LLaDA-2.0-Mini (Base)	30.00	91.98	45.88	55.95
LLaDA-2.0-Mini (SFT)	30.00	91.58	47.00	56.19
LLaDA-2.0-Mini (T3S)	53.33	93.19	51.50	66.01
Qwen3-14B (no-think)	26.51	88.18	55.88	56.86
Ministral-14B-Instruct-2512	48.59	91.38	56.13	65.37
Ling-Mini	46.25	94.79	50.38	63.81
Qwen3-30B-A3B-Instruct-2507	62.50	96.74	50.12	69.79



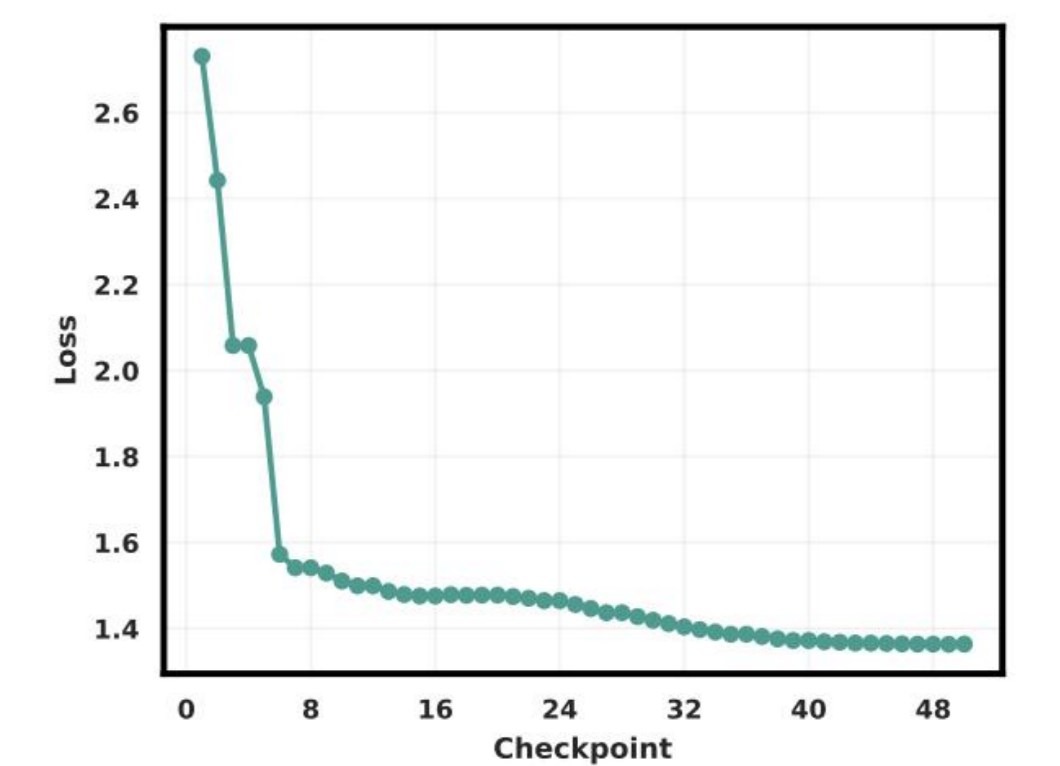
(a) Delta Prob Trends for Two Token Groups (**full objective**)



(b) Delta Prob Trends for Two Token Groups (**train other tokens only**)



(a) Δ Loss on other tokens



(b) Anchor-token loss

Figure 5. One-step intervention on Imitation-Anchor Tokens. Panel (a) reports $\Delta\mathcal{L}_{\text{other}}$ after one gradient step that optimizes only anchor tokens at each checkpoint. Panel (b) shows the corresponding anchor-token loss $\mathcal{L}_{\text{anchor}}$ at that checkpoint.