TailCon: Mitigating Tail Signal Erosion through Memory Consolidation for Long-Tailed Recognition

Yuan Dong[1], Di Wu[1], Zhe Zhao[1], Liheng Yu[1], Xiaofeng Cao[3], Pengkun Wang[1,2*] Yang Wang[1,2*]
[1]University of Science and Technology of China [2]Suzhou Institute for Advanced Research [3]Tongji University

1. Background & Motivation

The Challenge of Long-Tailed Recognition

Long- tailed recognition (LTR) suffers not only from insufficient rare class supervision but also from the difficulty of accumulating sparse tail evidence during stochastic training. Frequent head- class updates can dominate optimization, preventing rare -class knowledge from being effectively reinforced. Even informative tail observations may lose their influence before related rare examples reappear.

Tail Signal Erosion

- **Root Cause:** In conflict -dominant regimes, head-class gradients can interfere with tail-class optimization. Since rare examples appear intermittently, their learning signals may be progressively weakened by intervening head dominated updates.
- **Consequence:** Rare- class evidence is diluted before sufficient reinforcement, resulting in biased representations, weak tail discrimination, and poor generalization.

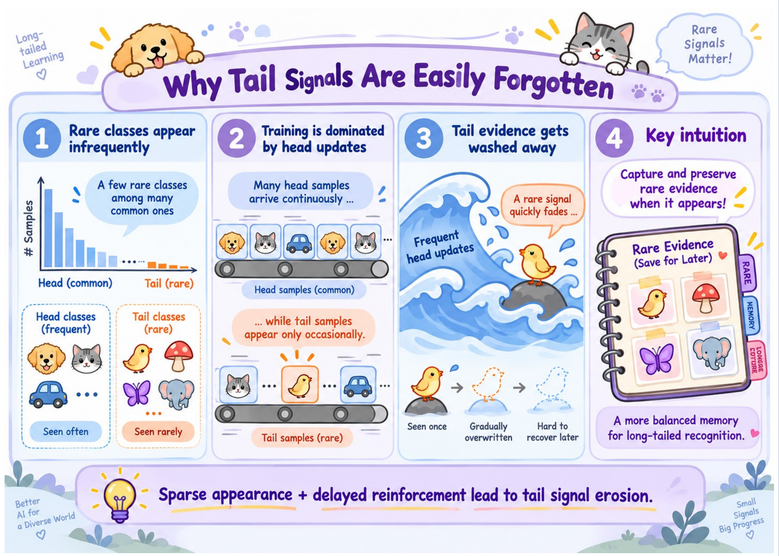
Limitations of Existing Approaches

- **Rebalancing and Calibration:** Correct class -prior bias but do not explicitly preserve sparse evidence during training.
- **Multi-Expert Learning:** Improves specialization but still relies primarily on parametric gradient updates.

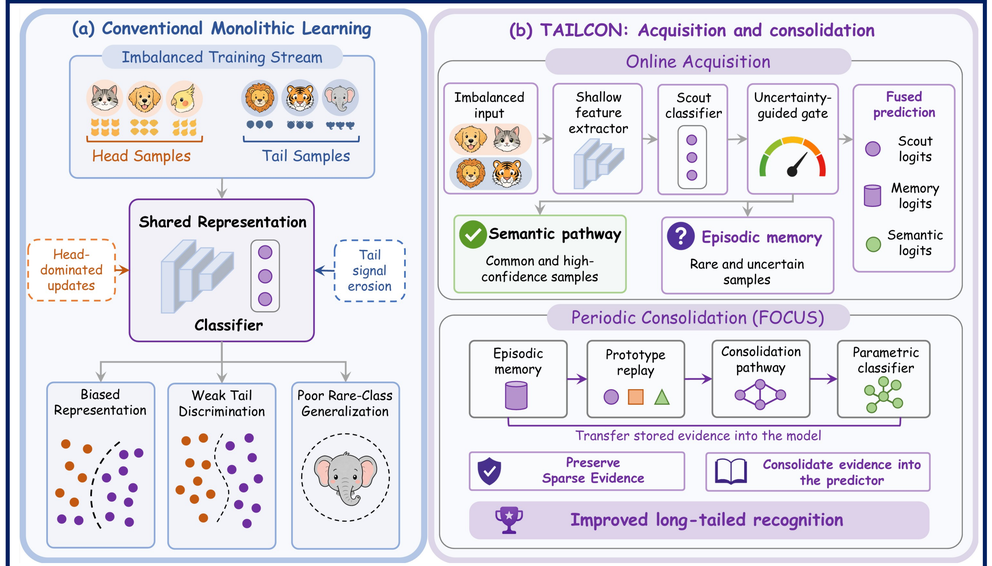
Memory-Based Retrieval: Preserves local evidence, but stored knowledge may remain external to the parametric classifier.

Our Key Insight

Preserve First, Consolidate Later.
Rare-class evidence should be captured when observed and subsequently integrated into the parametric predictor through controlled consolidation, rather than relying solely on immediate gradient updates or test-time retrieval.



2. Method: TAILCON



Overview: Separating Acquisition from Consolidation

TAILCON combines **confidence-guided online learning, detached episodic memory, and periodic FOCUS consolidation** to preserve sparse rare-class evidence and progressively integrate stored knowledge into the parametric classifier.

Stage 1: Confidence-Guided Online Learning

A lightweight scout classifier estimates prediction confidence, while a deeper semantic pathway learns global decision structure.

- **Adaptive Routing:** Low-confidence samples receive stronger semantic-pathway contributions through a confidence gate.
- **Budget Control:** Adaptive threshold feedback stabilizes semantic activation and prevents uncontrolled routing.

Stage 2: Detached Episodic Memory Acquisition

A non-parametric episodic memory preserves uncertain or rare representations through detached feature storage.

- **Selective Writing:** Samples are stored when confidence is low or class frequency falls below a rarity threshold.
- **Prototype Memory:** Detached shallow features form bounded class-specific prototypes for similarity based retrieval.
- **Evidence Preservation:** Stop-gradient writes protect sparse evidence without introducing additional gradient interference through the memory branch.

Stage 3: FOCUS — Periodic Memory Consolidation

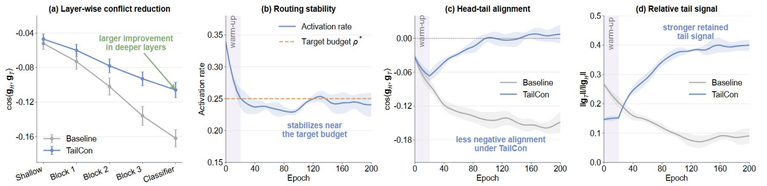
Beyond retrieval-only memory, FOCUS periodically transfers stored evidence into the parametric classifier.

- **Priority Replay:** Rare and uncertain prototypes receive higher sampling probabilities during consolidation.
- **Knowledge Alignment:** A lightweight linear adapter maps stored prototypes into the classifier-head feature space.
- **Memory Distillation**: KL distillation transfers memory-derived knowledge to the classifier head with frozen feature extractors.

Final Prediction

TAILCON combines scout predictions, episodic memory evidence, and confidence-gated semantic predictions into a unified output.
Core Advantage: Fast non parametric preservation protects sparse evidence, while slower consolidation transforms stored knowledge into lasting classifier improvements.

3. Experiments



Experimental Setup

TAILCON is evaluated on CIFAR-100 LT, CIFAR 10-LT, ImageNet-LT, and iNaturalist 2018 using Top -1 accuracy. All results are averaged over five independent runs.

Mechanism Analysis

- **Reduced Gradient Conflict:** TAILCON alleviates negative head-tail gradient alignment, particularly in deeper layers.
- **Stable Routing:** Adaptive confidence gating maintains the target activation budget.
- **Stronger Tail Signals:** Rare class gradients retain greater relative strength throughout training.

Key Findings

TAILCON achieves consistent improvements across four benchmarks. **FOCUS improves overall accuracy by 2.6 pp** on CIFAR-100-LT (IR=100), while **memory-disabled inference also gains 2.6 pp**, demonstrating effective consolidation beyond test-time retrieval.

Method	Type	IR=100				IR=50				IR=10			
		Many	Medium	Few	All	Many	Medium	Few	All	Many	Medium	Few	All
CE-DRW* [3]	*Re-bal.*	63.4	41.2	15.7	41.4	60.6	39.0	22.9	45.0	62.5	48.6	–	58.2
LDAM-DRW* [3]	*Re-bal.*	61.2	39.5	14.4	42.0	64.9	45.1	28.7	46.6	70.3	57.0	–	58.7
Balanced Softmax* [4]	*Re-bal.*	60.8	44.6	20.4	46.1	66.0	47.2	30.1	50.9	68.0	56.0	–	61.0
LADE* [5]	*Re-bal.*	60.2	43.9	19.6	45.6	65.5	46.3	29.7	50.1	69.0	57.8	–	61.6
MiSLAS* [11]	*Re-bal.*	62.1	45.3	21.1	46.8	66.8	48.4	31.2	51.5	70.0	58.5	–	62.5
BCL* [6]	*Re-bal.*	63.1	42.9	23.9	44.2	61.6	43.1	34.3	49.1	62.2	51.8	–	58.9
LSC* [13]	*Decoup.*	66.0	50.0	27.0	51.8	69.0	55.0	36.0	56.5	73.6	63.8	–	65.0
LOS† [12]	*Decoup.*	68.6	53.8	30.2	54.9	69.1	55.8	38.4	58.8	74.0	64.3	–	69.7
IP-TSL* [30]	*Decoup.*	62.4	59.7	31.9	52.4	–	–	–	55.9	–	–	–	62.6
RIDE* [17]	*Expert*	64.7	46.5	22.3	48.0	67.5	50.8	27.9	51.7	66.4	49.4	–	61.8
SADE* [18]	*Expert*	63.4	49.2	25.1	49.8	67.2	53.5	31.8	53.8	72.9	62.5	–	63.6
BalPoE* [39]	*Expert*	65.3	51.1	28.0	52.0	69.2	55.9	34.5	56.3	73.0	63.2	–	64.8
PRL* [19]	*Expert*	68.7	55.3	31.2	52.8	67.5	53.5	39.0	56.6	74.1	64.5	–	65.6
ICL* [20]	*Expert*	–	–	–	57.6	–	–	–	61.3	–	–	–	69.3
ProCo† [40]	*Proto.*	69.0	52.7	34.1	52.8	69.1	55.2	37.0	57.1	73.9	64.0	–	65.5
ConCutMix† [23]	*Aug.*	67.0	52.0	29.2	53.2	69.0	55.6	38.0	57.4	73.0	63.0	–	64.5
FeatRecon† [41]	*Recon.*	68.2	53.3	34.0	52.5	68.8	54.3	37.2	57.0	73.4	63.5	–	65.3
TailCon	*Memory consol.*	**76.1**±0.3	**60.1**±0.4	**35.4**±0.5	**58.5**±0.4	**74.8**±0.4	**57.7**±0.3	**47.2**±0.5	**62.8**±0.3	**74.5**±0.2	**66.7**±0.4	–	**72.1**±0.3

Method	CIFAR-10-LT	ImageNet-LT	iNat 2018
LDAM-DRW* [3]	77.0	48.8	64.6
RIDE* [17]	78.1	56.3	71.8
SADE* [18]	78.8	58.8	72.7
ConCutMix† [23]	81.7	58.5	72.1
FeatRecon† [41]	85.2	56.8	72.9
PRL* [19]	80.5	60.8	75.1
BalPoE* [39]	85.2	58.5	75.0
IP-TSL* [30]	76.4	51.7	74.0
ProCo† [40]	85.9	60.2	73.5
NCL++* [42]	87.2	59.6	75.2
ICL* [20]	87.9	60.2	75.9
TailCon	**88.4**±0.3	**63.2**±0.4	**76.9**±0.3

Variant	Many	Medium	Few	All
Baseline	61.2±0.4	39.5±0.5	14.4±0.4	42.0±0.3
Memory only	65.1±0.5	45.3±0.4	25.8±0.6	46.5±0.4
Semantic only	69.3±0.3	48.2±0.5	19.5±0.5	48.8±0.4
Dual path w/o FOCUS	74.9±0.4	56.5±0.3	31.3±0.5	55.9±0.4
Full w/o gate	74.5±0.3	58.6±0.4	34.4±0.4	57.2±0.3
TailCon	**76.1**±0.3	**60.1**±0.4	**35.4**±0.5	**58.5**±0.4

Training setting	Many	Medium	Few	All
Without FOCUS	71.8±0.4	51.0±0.5	24.6±0.6	50.7±0.4
With FOCUS	**73.9**±0.3	**55.4**±0.4	**28.7**±0.5	**53.3**±0.3