BasicLT: Basic Level Abstraction and Selective Differentiation 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 from highly imbalanced supervision, where rare classes lack sufficient examples for reliable fine- grained discrimination. Most existing methods address class imbalance but still require all categories to compete in a flat fine- grained label space, regardless of their available supervision.

Granularity Mismatch under Scarcity

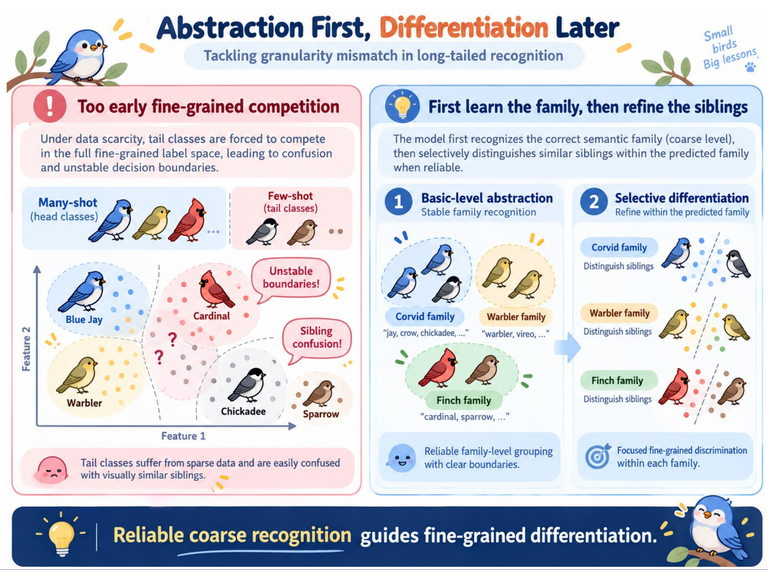
- **Root Cause:** Under severe data scarcity, tail classes are forced into fine-grained discrimination before the model establishes reliable coarse semantic structure. Recognizing the correct semantic family may be feasible even when distinguishing visually similar sibling classes remains difficult.
- **Consequence:** Premature fine- grained competition produces unstable decision boundaries, weak sibling-level discrimination, and poor rare -class generalization.

Limitations of Existing Approaches

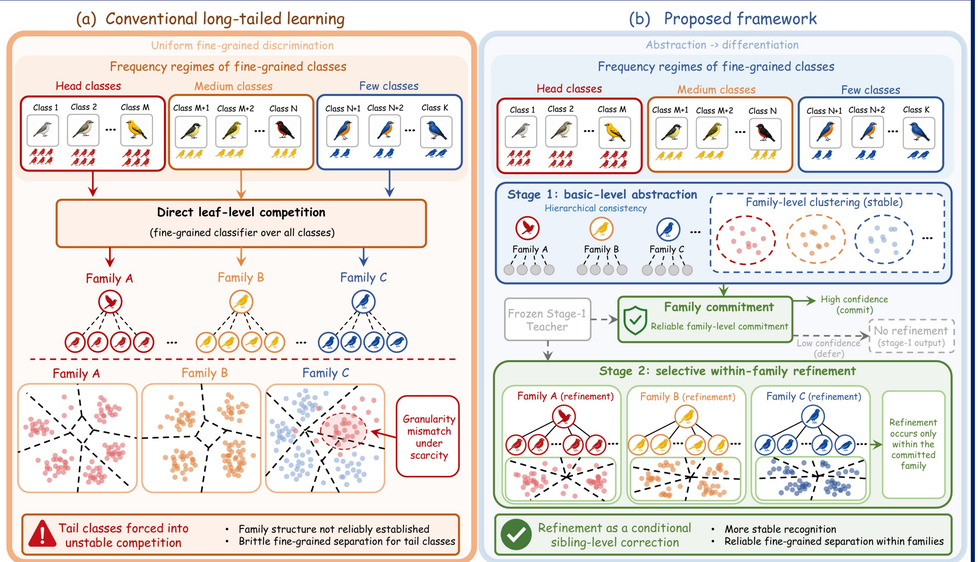
- **Rebalancing and Calibration:** Adjust class contributions or decision scores without explicitly controlling semantic granularity.
- **Multi-Expert Learning:** Improves specialization but generally retains fine- grained competition across all classes.
- **Hierarchy-Aware Recognition:** Exploits semantic structure but does not necessarily use family level reliability to control when additional fine- grained refinement is applied.

Our Key Insight —— Abstraction First, Differentiation Later.

Instead of imposing additional fine-grained discrimination uniformly, the model should first learn stable basic-level semantic structure and then refine sibling classes when family-level commitment is reliable. This progression enables **reliable coarse recognition to guide fine-grained differentiation** , particularly for medium -shot and few-shot categories.



2. Method: BasicLT



Overview: From Abstraction to Differentiation

BasicLT introduces a **two-stage framework combining basic-level abstraction and commitment-guided selective differentiation**. It first establishes stable family level structure, then selectively refines fine grained predictions under reliable semantic commitment.

Stage 1: Basic-Level Abstraction

Stage 1 jointly learns fine-grained recognition and basic-level semantic structure through an explicit family prediction branch.

- **Basic-Level Supervision:** A dedicated branch learns coarse semantic families alongside fine-grained classification.
- **Hierarchical Consistency:** Aligns aggregated fine-grained predictions with the supervised family distribution using a stop -gradient target.
- **Abstraction-First Learning:** Prioritizes family-level supervision early and strengthens coarse -fine consistency later.

Stage 2: Commitment-Guided Selective Differentiation

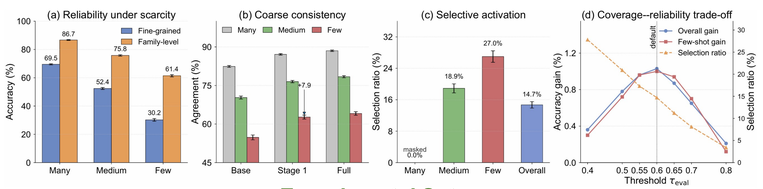
Stage 2 introduces family-conditioned residual refinement, activated only when the learned family structure is sufficiently reliable.

- **Teacher-Guided Commitment:** A frozen Stage 1 teacher selects medium- and few-shot training samples using family confidence and coarse prediction agreement.
- **Within-Family Refinement:** Centered, family-masked residuals improve sibling-level discrimination without modifying predictions outside the selected family.
- **Stable Optimization:** Teacher-student distillation and residual regularization preserve learned semantic structure.

Final Prediction: Conservative Selective Refinement

- **Selective Inference:** Refinement requires reliable student-side family commitment and an eligible medium -or few-shot base prediction.
- **Safe Fallback:** Uncommitted samples retain their original predictions without residual correction.
- **Core Advantage:** BasicLT controls **when and where fine-grained discrimination is introduced** , improving recognition under severe scarcity.

3. Experiments



Experimental Setup

BasicLT is evaluated on five benchmarks: CIFAR-100 LT, CIFAR-10-LT, ImageNet LT, iNaturalist 2018, and Places-LT. Results are reported using Top-1 accuracy over five independent runs.

Mechanism Analysis

- **Reliable Abstraction:** Family-level prediction is substantially more reliable than fine-grained recognition under scarcity.
- **Improved Consistency:** Stage 1 strengthens coarse semantic agreement, while Stage 2 selectively refines committed low-frequency predictions.
- **Controlled Refinement:** Commitment-based activation balances refinement coverage and prediction reliability.

Key Findings

On CIFAR-100 LT (IR=100), BasicLT improves overall accuracy from **52.6% to 60.6% (+8.0 pp)** and outperforms uniform refinement by **4.8 pp**. Consistent improvements across five benchmarks validate the abstraction-to-differentiation strategy.

Method	Source	IR = 100				IR = 50				IR = 10			
		Many	Medium	Few	Overall	Many	Medium	Few	Overall	Many	Medium	Few	Overall
CE-DRW [2]*	*NeurIPS 19*	63.4	41.2	15.7	41.4	60.6	39.0	22.9	45.0	62.5	48.6	-	58.2
LDAM-DRW [2]*	*NeurIPS 19*	61.2	39.5	14.4	42.0	64.9	45.1	28.7	46.6	70.3	57.0	-	58.7
BS [17]*	*NeurIPS 20*	60.8	44.6	20.4	46.1	66.0	47.2	30.1	50.9	68.0	56.0	-	61.0
cRT [3]*	*ICLR 20*	61.8	46.2	27.4	47.3	65.4	52.1	32.5	51.7	73.3	67.5	-	63.2
LADE [18]*	*CVPR 21*	60.2	43.9	19.6	45.6	65.5	46.3	29.7	50.1	69.0	57.8	-	61.6
MiSLAS [19]*	*CVPR 21*	62.1	45.3	21.1	46.8	66.8	48.4	31.2	51.5	70.0	58.5	-	62.5
RIDE (3 experts) [5]*	*ICLR 21*	64.7	46.5	22.3	48.0	67.5	50.8	27.9	51.7	66.4	49.4	-	61.8
BCL [6]*	*CVPR 22*	63.1	42.9	23.9	44.2	61.6	43.1	34.3	49.1	62.2	51.8	-	58.9
SADE [20]*	*NeurIPS 22*	63.4	49.2	25.1	49.8	67.2	53.5	31.8	53.8	72.9	62.5	-	63.6
BalPoE [21]*	*CVPR 23*	65.3	51.1	28.0	52.0	69.2	55.9	34.5	56.3	73.0	63.2	-	64.8
ConCutmix [22]†	*TIP 24*	67.0	52.0	29.2	53.2	69.0	55.6	38.0	57.4	73.0	63.0	-	64.5
LSC [23]*	*ICML 24*	66.0	50.0	27.0	51.8	69.0	55.0	36.0	56.5	73.6	63.8	-	65.0
PRL [24]*	*NeurIPS 24*	68.7	55.3	31.2	52.8	67.5	53.5	39.0	56.6	74.1	64.5	-	65.6
ProCo [25]†	*TPAMI 24*	69.0	52.7	34.1	52.8	69.1	55.2	37.0	57.1	73.9	64.0	-	65.5
ICL [26]†	*IJCAI 25*	70.6	61.1	36.5	57.6	69.1	56.8	38.9	61.3	73.9	66.8	-	69.3
LOS [27]*	*ICLR 25*	68.6	53.8	30.2	54.9	69.1	55.8	38.4	58.8	74.0	64.3	-	69.7
FeatRecon [28]†	*ICLR 25*	68.2	53.3	34.0	52.5	68.8	54.3	37.2	57.0	73.4	63.5	-	65.3
IP-TSL [29]†	*NeurIPS 25*	62.4	59.7	31.9	52.4	68.9	55.0	37.5	55.9	73.5	64.1	-	62.6
BasicLT (Ours)	-	**78.4**±0.3	**63.2**±0.4	**37.0**±0.5	**60.6**±0.4	**77.4**±0.4	**59.6**±0.4	**53.1**±0.5	**65.7**±0.3	**76.2**±0.5	**68.3**±0.3	-	**73.8**±0.5

Method	Source	CIFAR-10-LT (IR=100)				ImageNet-LT				iNat. 2018	Places-LT
		Many	Medium	Few	All	Many	Medium	Few	All	All	All
Cross-Entropy	-	90.5	62.1	37.8	70.4	40.9	10.7	0.4	41.6	61.7	27.4
LDAM-DRW [2]*	*NeurIPS 19*	83.3	73.5	72.5	77.0	60.4	46.9	30.7	48.8	64.6	30.5
BS [17]*	*NeurIPS 20*	88.5	76.2	68.3	79.1	62.2	48.8	38.6	51.4	70.6	39.4
cRT [3]*	*ICLR 20*	88.7	76.5	63.8	78.4	61.8	46.2	27.3	50.3	65.2	33.2
MiSLAS [19]*	*CVPR 21*	89.7	79.2	73.1	82.1	61.7	51.3	35.8	52.7	70.7	33.5
RIDE (3 experts) [5]*	*ICLR 21*	96.1	89.0	76.5	87.2	66.2	52.3	36.5	55.4	68.8	40.3
BCL [6]*	*CVPR 22*	92.9	83.3	55.3	81.2	66.5	53.6	30.7	51.8	71.8	-
SADE [20]*	*NeurIPS 22*	95.8	88.5	76.4	86.9	67.2	54.7	35.1	56.1	72.7	40.9
BalPoE [21]*	*CVPR 23*	96.3	89.2	76.4	87.3	68.3	57.2	40.5	58.5	75.0	40.8
LSC [23]*	*ICML 24*	92.5	82.0	73.5	83.5	66.0	50.0	27.0	56.5	73.9	-
PRL [24]*	*NeurIPS 24*	93.8	84.1	74.5	84.5	**72.5**	58.2	41.8	60.8	75.1	41.6
ProCo [25]†	*TPAMI 24*	96.4	89.1	76.6	87.5	69.1	59.8	42.1	60.2	73.5	-
ICL [26]†	*IJCAI 25*	96.9	88.5	76.3	87.1	68.2	59.1	45.8	60.2	75.9	42.1
LOS [27]*	*ICLR 25*	92.0	83.5	74.2	84.8	-	-	-	54.4	70.8	41.9
FeatRecon [28]*	*ICLR 25*	94.5	86.8	74.3	85.2	-	-	-	56.8	72.9	42.3
IP-TSL [29]†	*NeurIPS 25*	82.1	77.2	66.5	76.4	61.2	50.8	36.5	51.7	74.0	-
BasicLT (Ours)	-	**97.1**±0.3	**89.9**±0.2	**77.2**±0.6	**87.8**±0.4	70.4±0.4	**62.4**±0.2	**49.5**±0.5	**63.2**±0.5	**77.2**±0.5	**44.5**±0.4