Balanced Product of Calibrated Experts for Long-Tailed Recognition

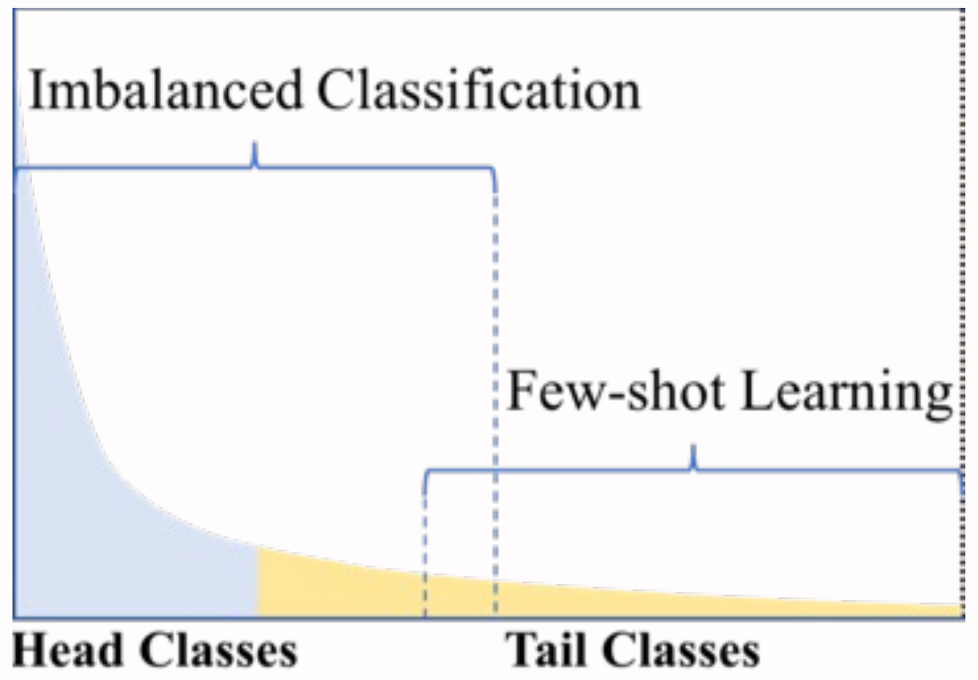
Emanuel Sanchez Aimar[1] Arvi Jonnarth[1,2] Michael Felsberg[1] Marco Kuhlmann[1]

[1]Linköping University, Sweden [2]Husqvarna Group, Sweden

{emanuel.sanchez.aimar, arvi.jonnarth, michael.felsberg, marco.kuhlmann}@liu.se

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Challenges of Long-Tailed Recognition

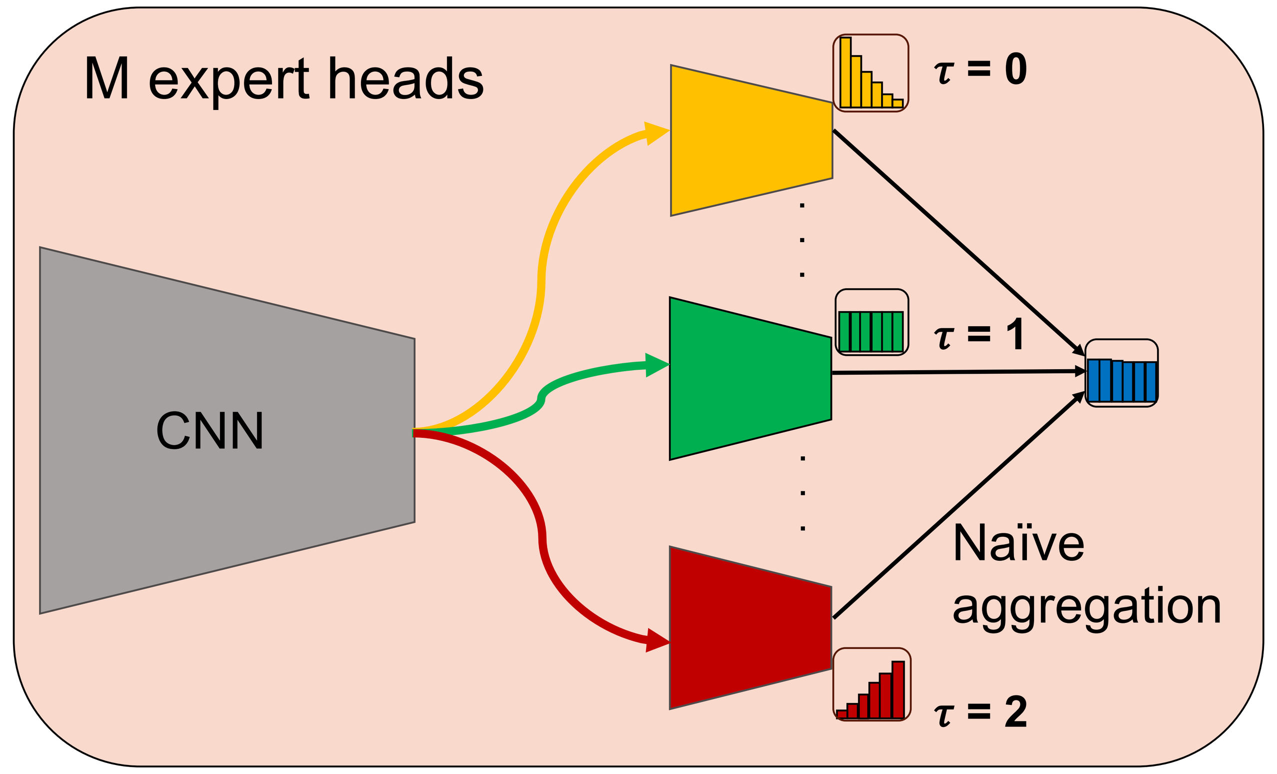


- **Class imbalance** and **data scarcity**
- **Label distribution shift**: training and test distributions tend to be different (*long-tailed* vs *uniform*)
- Lead to ***poorly-calibrated models***

Contributions

- We propose **BalPoE**, an ensemble-based framework composed of **skill-diverse calibrated experts**, to address long-tailed (LT) recognition
- We validate its properties:
  - **Unbiased**: loss formulation is **fisher-consistent** for minimizing the balanced error
  - **Well-calibrated**
  - **SOTA performance** on LT benchmarks

Balanced Product of Experts (BalPoE)



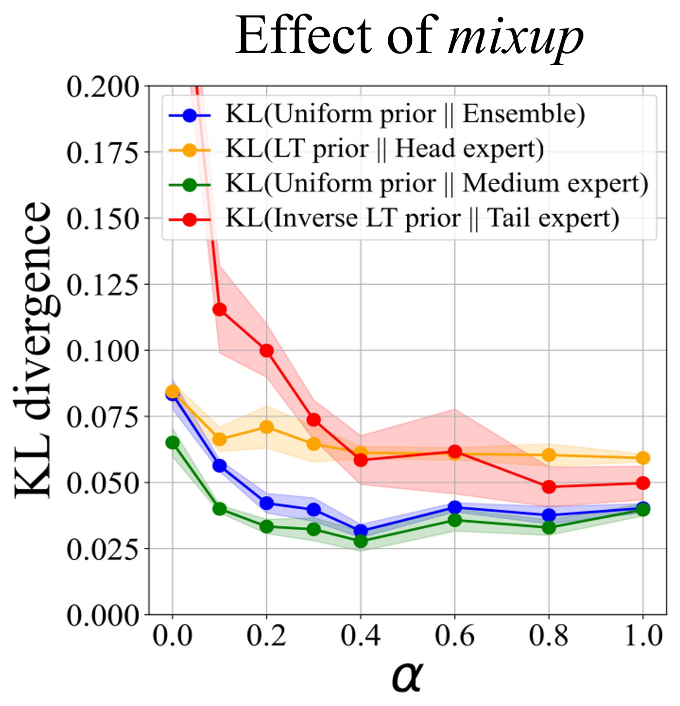
- **Logit-adjusted loss for (1 - τ)-expert**:

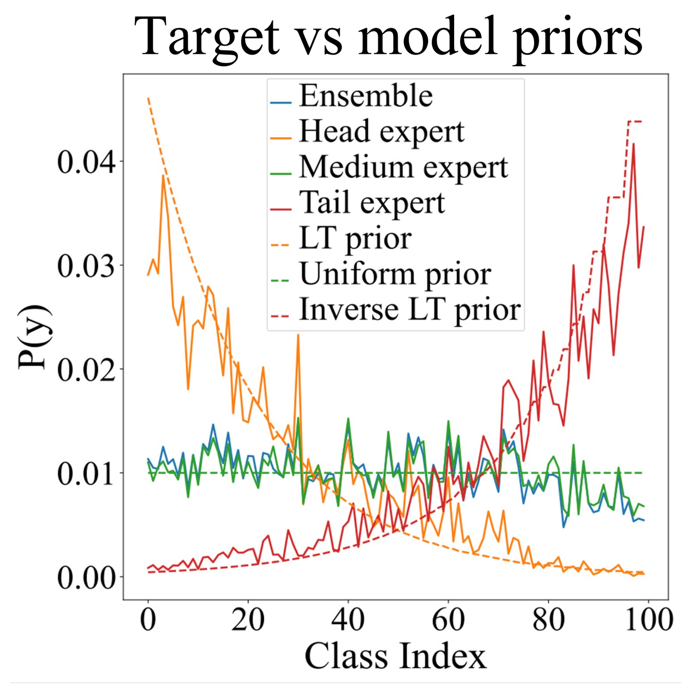
$$\ell_{\tau}(y, f(x)) = -\log \frac{e^{f_y(x)+\tau_y \log \mathbb{P}^{\text{train}}(y)}}{\sum_{j\in\mathcal{Y}} e^{f_j(x)+\tau_j \log \mathbb{P}^{\text{train}}(j)}}$$

- BalPoE attains the average bias of all its experts
- **Simple constraint**: average tau is one => unbiased ensemble
- Logit adjustment assumes calibrated model!

Meeting the calibration assumption

- **Mixup** promotes expert calibration for target distributions
  => improves **ensemble calibration**!



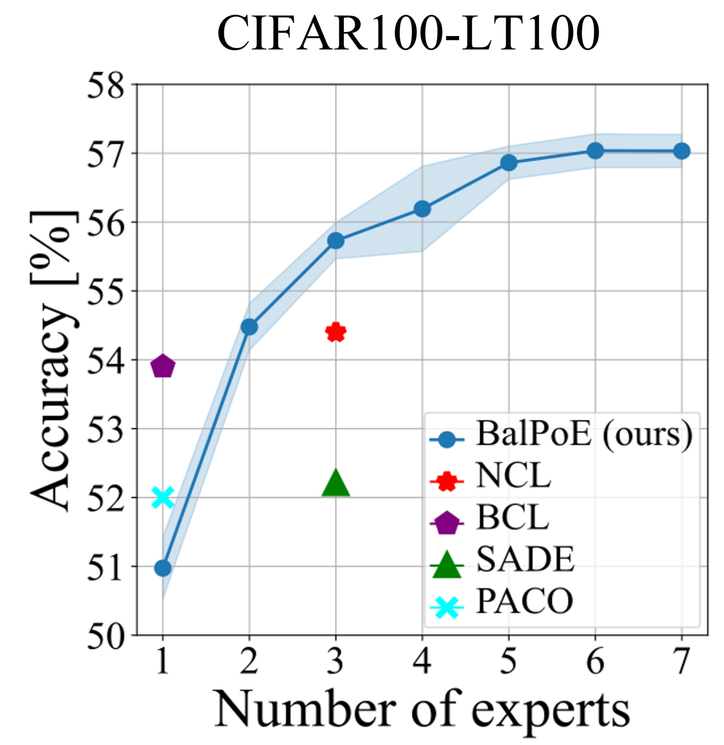


Experiments

Classification results

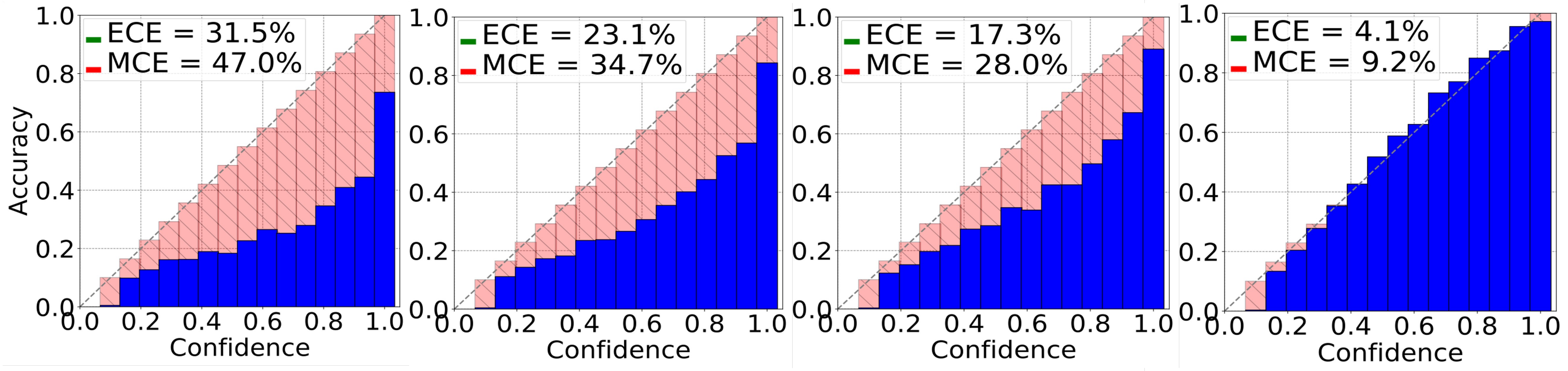
- Train on long-tailed data, test on balanced data, M=3 experts
- IR: Imbalance ratio, BB: backbone. e.g., R50 -> ResNet50

		CIFAR-100-LT			ImageNet-LT		iNat
	IR →	10	50	100	256	256	500
Method ↓	BB →	R32	R32	R32	R50	RX50	R50
PaCo [12]		64.2	56.0	52.0	57.0	58.2	73.2
CMO+BS [49]		65.3	56.7	51.7	58.0	-	74.0
BCL [72]		64.9	56.6	53.9	56.0	57.1	71.8
NCL [38]		-	58.2	54.2	59.5	60.5	74.9
SADE [70]		65.3	57.3	52.2	-	61.2	74.5
BalPoE (ours)		**68.1**	**60.1**	**55.9**	**60.8**	**62.0**	**76.9**



Calibration results

- Expected Calibration Error (ECE) and Maximum Calibration Error (MCE) over CIFAR100-LT100 test set



(a) CE (b) BS (c) SADE (d) **BalPoE** (ours)

References

[Mixup] Zhang et al. Mixup: Beyond empirical risk minimization. ICLR, 2018.
[BS] Ren et al. Balanced meta-softmax for long-tailed visual recognition. NeurIPS, 2020.
[PaCo] Cui et al. Parametric contrastive learning. ICCV, 2021.
[NCL] Li et al. Nested collaborative learning for long-tailed visual recognition. CVPR, 2022.
[BCL] Zhu et al. Balanced contrastive learning for long-tailed visual recognition. CVPR, 2022.
[CMO] Park et al. The majority can help the minority: Context-rich minority oversampling for long-tailed classification. CVPR, 2022.
[SADE] Zhang et al. Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition. NeurIPS, 2022.

SCAN ME