ICoRe-SSL: Information- and Confusion-Regularized Semi-Supervised Learning for Long-Tailed Recognition

Xinyu Li[1], Xiaoyu Guo[2], Xiang Wei[1*]
[1]School of Software Engineering, Beijing Jiaotong University, Beijing, China
[2]Beijing Institute of Astronautical Systems Engineering, Beijing, China

***Abstract*—Class-imbalanced semi-supervised learning (CISSL) aims to exploit abundant unlabeled data under long-tailed class distributions, but is often hindered by biased training dynamics and unreliable pseudo-labels, where head classes dominate optimization and errors on tail classes accumulate over training. Most existing methods address this issue through reweighting schemes or adaptive pseudo-label selection, while largely overlooking how redundant sample participation and opaque class-wise prediction dynamics jointly distort the optimization process under long-tailed distributions. We propose ICoRe-SSL, an Information- and Confusion-Regularized Semi-Supervised Learning framework that tackles these limitations by jointly reshaping optimization dynamics and enhancing class-wise interpretability. ICoRe-SSL introduces an information-aware sample participation mechanism that suppresses low-information samples with expectation-preserving rescaling, reducing redundant head-class updates, and incorporates a confusion-aware analysis module to explicitly track pseudo-label statistics and confusion evolution during training. As a result, ICoRe-SSL maintains more balanced pseudo-label distributions and enables stable tail-class learning. Extensive experiments on standard long-tailed benchmarks demonstrate that ICoRe-SSL consistently outperforms state-of-the-art semi-supervised methods, achieving significant gains in balanced accuracy and geometric mean with negligible computational overhead.**

***Index Terms*—Semi-Supervised Learning; Class-Imbalanced Learning; Long-Tailed Recognition; Training Dynamics; Pseudo-Labeling**

I. Introduction

Semi-supervised learning (SSL) leverages abundant unlabeled data to improve performance when labeled data are limited [1]. However, real-world datasets typically exhibit long-tailed distributions, where a few head classes dominate while many tail classes have very limited samples [9], [10]. In such class-imbalanced semi-supervised learning (CISSL) settings, existing SSL methods often suffer from biased training dynamics, unreliable pseudo-labels, and inefficient data utilization, making it difficult to learn robust tail-class representations.

A key limitation of existing CISSL methods is the lack of explicit regulation of training dynamics under long-tailed distributions. Most methods treat all samples equally, leading to redundant optimization dominated by head-class samples. Pseudo-labeling further amplifies this issue, as head classes tend to generate more high-confidence pseudo-labels, reinforcing distribution bias during training [11], [18]. Meanwhile, the class-wise prediction behavior and confusion evolution remain largely unexplored, making bias accumulation difficult to diagnose and control.

To address these issues, we propose ICoRe-SSL, an Information- and Confusion-Regularized Semi-Supervised Learning framework for CISSL. Built upon adaptive balanced consistency learning [4], the proposed method regulates training from both sample and class perspectives. Specifically, an information-aware sample participation mechanism suppresses low-information samples with expectation-preserving rescaling to reduce redundant head-class updates. In addition, a confusion-aware analysis module models pseudo-label statistics and class-wise confusion to stabilize training dynamics and improve interpretability. Together, these components encourage more balanced learning under long-tailed distributions.

The main contributions are summarized as follows:

- **Information-regularized optimization for CISSL.** We introduce an expectation-preserving, information-aware sample participation mechanism that reshapes optimization dynamics under class imbalance without altering the original objective.
- **Confusion-aware and interpretable training dynamics.** We propose a unified framework that explicitly models class-wise confusion patterns and pseudo-label statistics, enabling fine-grained analysis of bias accumulation.
- **Strong performance under long-tailed settings.** Extensive experiments demonstrate that ICoRe-SSL consistently outperforms strong baselines, achieving superior tail-class performance with improved balanced metrics.

II. Related Work

A. Semi-Supervised Learning

Semi-supervised learning (SSL) aims to leverage large-scale unlabeled data to improve model performance when labeled data are scarce. Representative SSL approaches include consistency regularization methods such as Π-Model [12], Mean Teacher [14], and FixMatch [2], which enforce prediction consistency under perturbations, as well as entropy minimization and pseudo-labeling strategies [13]. More recent methods

*Corresponding author: xiangwei@bjtu.edu.cn

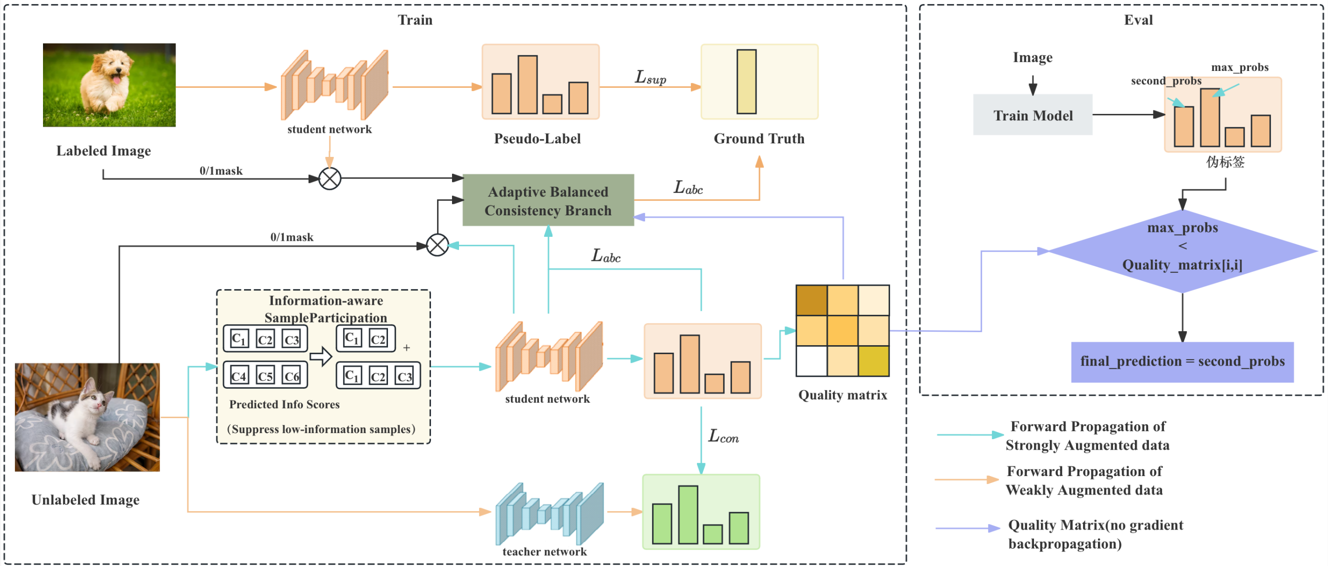


Fig. 1: Overview of the proposed ICoRe-SSL framework for class-imbalanced semi-supervised learning.

further integrate distribution alignment [1] and confidence-based filtering [2] to enhance pseudo-label reliability. Despite their effectiveness, most SSL methods implicitly assume that labeled and unlabeled data share a similar class distribution [3], which is often violated in real-world scenarios.

B. Class-Imbalanced Semi-Supervised Learning

Class-imbalanced semi-supervised learning (CISSL) addresses the challenge of distribution mismatch between labeled and unlabeled data. Early works focus on balancing pseudo-label generation or classifier training [6], [7]. CReST [15] progressively re-samples unlabeled data by over-emphasizing minority-class predictions, while ABC [4] and CoSSL [16] introduce auxiliary balanced classifiers to generate debiased pseudo-labels. Several methods explicitly refine pseudo-label distributions, including DARP [21], which solves a convex optimization problem for distribution alignment, and DASO [8], which blends semantic and linear pseudo-labels. Other approaches mitigate bias through reweighting or logit correction, such as SAW [17], UDAL [10], and logit-adjusted losses [9].

Recently, more principled strategies have been explored. CDMAD [18] proposes to reduce pseudo-label bias by explicitly modeling class distribution mismatch via dual distribution alignment, demonstrating improved robustness under severe imbalance. LCGC [19] further investigates class-wise generalization calibration by leveraging long-tailed priors and adaptive confidence modeling, highlighting the importance of class-dependent uncertainty estimation in CISSL.

Despite these advances, most methods rely on heuristic rebalancing or distribution assumptions, and the training dynamics under long-tailed SSL remain insufficiently explored, motivating frameworks that model both sample information and confusion.

III. Method

A. Problem Definition

We focus on CISSL, where a labeled dataset

$$\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l} \tag{1}$$

follows a long-tailed class distribution, and a large unlabeled dataset

$$\mathcal{D}_u = \{x_j\}_{j=1}^{N_u} \tag{2}$$

is drawn from the same underlying data distribution but lacks class labels. Here, $y_i \in \{1, \ldots, C\}$ represents class labels.

Our goal is to train a classifier f_θ that achieves **balanced performance across all classes**, particularly under the conditions of label scarcity and pseudo-label uncertainty.

B. Overview of ICoRe-SSL

We propose ICoRe-SSL, an Information- and Confusion-Regularized Semi-Supervised Learning framework, which extends adaptive balanced consistency learning by introducing two complementary components:

1) **Information-aware sample participation**, which dynamically adjusts the contribution of each sample during training based on its information content, effectively suppressing redundant updates from head-class samples;
2) **Confusion-aware class-wise analysis**, which tracks pseudo-label distributions and class-wise confusion patterns to identify and address class-wise prediction bias.

As shown in Figure 1, these two components are seamlessly integrated into a standard semi-supervised learning pipeline without introducing additional networks or supervision signals. Instead, they act by reshaping optimization dynamics, enabling more balanced and stable learning under class-imbalanced conditions. The complete training procedure of ICoRe-SSL is summarized in Algorithm 1.

Algorithm 1 Pseudo-code of ICoRe-SSL

Require: Labeled batch $\mathcal{B}_l$, unlabeled batch $\mathcal{B}_u$, model f_θ, pruning ratio r
Ensure: Trained classifier f_θ
while not converged **do**
  Apply standard augmentation to $\mathcal{B}_l$ and weak/strong augmentation to $\mathcal{B}_u$
  Compute supervised loss $\mathcal{L}_{sup}$ on $\mathcal{B}_l$
  Generate pseudo-labels and compute consistency loss $\mathcal{L}_{con}$ on $\mathcal{B}_u$
  Compute adaptive balanced consistency loss $\mathcal{L}_{abc}$ on $\mathcal{B}_u$
  Collect per-sample losses $\ell_i \in \{\mathcal{L}_{sup}, \mathcal{L}_{con}\}$
  Compute batch mean loss $\bar{\ell}$
  for each sample i **do**
    if $\ell_i < \bar{\ell}$ **then**
      Sample $z_i \sim \text{Bernoulli}(1-r)$
      $\tilde{\ell}_i \leftarrow z_i \cdot \ell_i/(1-r)$ {soft pruning}
    end if
  end for
  Combine losses:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda_{con}\mathcal{L}_{con} + \lambda_{abc}\mathcal{L}_{abc}$$

  Update model parameters: $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}$
  Update pseudo-label distribution and confusion matrix (no gradient)
end while
return f_θ

C. Information-aware Sample Participation

In long-tailed settings, head-class samples tend to dominate the optimization process, despite their limited contribution to the model's improvement [5]. To address this issue, we introduce an **information-aware sample participation mechanism** that modulates the contribution of each sample to the loss function, ensuring that the model is not dominated by redundant head-class samples.

For each mini-batch, the loss of sample x_i is denoted as ℓ_i. We identify low-information samples as those whose loss is smaller than the mean loss $\bar{\ell}$ of the mini-batch, i.e., $\ell_i < \bar{\ell}$.

Instead of discarding these low-information samples, we adopt a **probabilistic soft pruning** mechanism:

$$z_i \sim \text{Bernoulli}(1-r), \tag{3}$$

where r is the pruning ratio, and z_i is a binary random variable determining whether the sample participates in the loss calculation. This probabilistic approach ensures that all samples, including low-information ones, retain a non-zero probability of being involved in the training process.

To avoid bias in gradient estimation, we rescale the loss of retained low-information samples as:

$$\tilde{\ell}_i = \begin{cases} \frac{1}{1-r}\ell_i, & \text{if } \ell_i < \bar{\ell} \text{ and } z_i = 1, \\ \ell_i, & \text{otherwise.} \end{cases} \tag{4}$$

This rescaling guarantees that the expected gradient remains unbiased:

$$\mathbb{E}[\nabla_\theta \tilde{\ell}_i] = \nabla_\theta \ell_i. \tag{5}$$

By dynamically modulating the participation of each sample, we reduce redundant updates and mitigate head-class dominance, while preserving the original optimization objective.

D. Confusion-aware Class-wise Analysis

Beyond sample-level regulation, ICoRe-SSL further incorporates a **confusion-aware class-wise analysis** module to explicitly monitor and stabilize class-wise prediction dynamics under long-tailed distributions. This module tracks the evolution of pseudo-label distributions and class-wise confusion statistics throughout training, providing interpretable signals about how errors propagate across classes [8].

Confusion-aware analysis provides a class-level perspective that tracks and stabilizes prediction dynamics under long-tailed distributions. It functions as a class-level diagnostic and regularization mechanism, exposing systematic misclassification patterns, particularly the prevalent tail-to-head confusion observed in CISSL settings [18]. By making such dynamics explicit, the model can identify when pseudo-label errors are repeatedly reinforced and to react accordingly.

The collected confusion statistics are further used to guide training behavior, complementing information-aware sample participation. In this way, confusion-aware analysis prevents biased pseudo-label feedback loops from dominating the learning process, especially for underrepresented tail classes. Overall, this module provides a necessary class-level perspective that complements sample-level information regulation, enabling a more transparent and stable training process.

E. Training Objective

We emphasize that ICoRe-SSL does *not* formulate learning as a collection of independent objectives. Instead, it reshapes training dynamics through coordinated regulation at both the sample and class levels [20].

Specifically, information-aware sample participation operates at the data level rather than the loss level, dynamically prioritizing informative samples during training. Meanwhile,stabilizes prediction behavior under class imbalance. These components are integrated into a unified optimization framework rather than acting as competing objectives.

The overall training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda_{con}\mathcal{L}_{con} + \lambda_{abc}\mathcal{L}_{abc}, \tag{6}$$

where each term plays a complementary role in regulating learning dynamics.

a) Supervised learning: The supervised classification loss on labeled data is given by:

$$\mathcal{L}_{sup} = \frac{1}{|\mathcal{B}_l|} \sum_{(x_i,y_i)\in\mathcal{B}_l} \ell_{\text{ce}}(f_\theta(x_i), y_i). \tag{7}$$

b) Consistency learning: For unlabeled data, we adopt a standard consistency-based objective using weakly augmented pseudo-labels and strongly augmented predictions:

$$\mathcal{L}_{con} = \frac{1}{|\mathcal{B}_u|} \sum_{x_j \in \mathcal{B}_u} \mathbb{I}(\max p_\theta(x_j^w) \geq \tau) \cdot \ell_{\text{ce}}(f_\theta(x_j^s), \hat{y}_j). \quad (8)$$

c) Adaptive balanced consistency: To mitigate class imbalance in pseudo-label supervision, we introduce an adaptive class-aware reweighting:

$$\mathcal{L}_{abc} = \frac{1}{|\mathcal{B}_u|} \sum_{x_j \in \mathcal{B}_u} \alpha_{c(\hat{y}_j)} \cdot \ell_{\text{ce}}(f_\theta(x_j^s), \hat{y}_j), \quad (9)$$

where $\alpha_{c(\hat{y}_j)}$ dynamically adjusts the influence of head and tail classes. This term discourages systematic confusion from being repeatedly amplified through pseudo-label feedback.

d) Optimization: ICoRe-SSL is optimized using standard stochastic gradient descent (SGD) with momentum. It is important to note that information-aware sample participation does not directly participate in loss computation. Instead, it operates at the data sampling stage by dynamically selecting high-information samples, thereby improving the effectiveness of gradient updates.

Through the joint optimization of the above components, ICoRe-SSL establishes a multi-level regulation mechanism: confusion-aware class-wise analysis captures class confusion structure, adaptive balanced consistency balances class-wise gradient contributions, and information-aware sample participation optimizes the training sample distribution. This coordinated design enables more stable and robust semantic segmentation under long-tailed distributions.

IV. Experiments

A. Experimental Setup

Datasets. We evaluate ICoRe-SSL on widely used class-imbalanced semi-supervised learning benchmarks following standard long-tailed protocols. Long-tailed distributions are constructed by controlling the imbalance ratio ρ between the most frequent and least frequent classes:

$$\rho = \frac{N_{\max}}{N_{\min}}, \quad (10)$$

while keeping the unlabeled data distribution unchanged. This setting reflects realistic scenarios where class imbalance and label scarcity coexist.

Evaluation Metrics. In addition to standard top-1 accuracy, we report balanced accuracy (bACC) and geometric mean (GM) to emphasize class-wise and tail-class performance [18]:

$$\text{bACC} = \frac{1}{C} \sum_{c=1}^{C} \text{Acc}_c \quad (11)$$

$$\text{GM} = \left(\prod_{c=1}^{C} \text{Acc}_c \right)^{\frac{1}{C}} \quad (12)$$

Implementation Details. All methods adopt the same backbone architecture and data augmentation strategy as ABC for fair comparison. Models are trained using SGD with momentum, and all hyperparameters follow official implementations unless otherwise specified. Each experiment is repeated three times with different random seeds, and average results are reported.

TABLE I: Comparison of bACC on CIFAR-100-LT

Algorithm	$\gamma = 20$	$\gamma = 50$	$\gamma = 100$
FixMatch	49.6±0.8	42.1±0.3	37.6±0.5
FixMatch+DARP	50.8±0.8	43.1±0.5	38.3±0.5
FixMatch+DARP+cRT	51.4±0.7	44.9±0.5	40.4±0.8
FixMatch+CRest	51.8±0.7	44.9±0.5	40.1±0.7
FixMatch+CRest+LA	52.9±0.1	47.3±0.2	42.7±0.7
FixMatch+ABC	53.3±0.8	46.7±0.3	41.2±0.7
FixMatch+CoSSL	53.9±0.8	47.6±0.6	43.0±0.6
FixMatch+UDAL	–	48.0±0.6	43.7±0.4
FixMatch+CDMAD	54.3±0.4	48.8±0.8	44.1±0.3
FixMatch+ICoRe-SSL	**55.6±0.6**	**49.1±0.5**	**44.3±0.6**
ReMixMatch	51.6±0.4	44.2±0.6	39.3±0.4
ReMixMatch+DARP	51.9±0.4	44.7±0.7	39.8±0.5
ReMixMatch+DARP+cRT	54.5±0.4	48.5±0.9	43.7±0.8
ReMixMatch+CRest	51.3±0.3	45.5±0.8	41.0±0.8
ReMixMatch+CRest+LA	51.9±0.6	46.6±1.1	41.7±0.7
ReMixMatch+ABC	55.6±0.4	47.9±0.1	42.2±0.1
ReMixMatch+CoSSL	55.8±0.6	48.9±0.6	44.1±0.6
ReMixMatch+CDMAD	57.0±0.3	51.1±0.5	44.9±0.4
ReMixMatch+ICoRe-SSL	**57.3±0.7**	**51.2±0.7**	**45.1±0.6**

B. Comparison with State-of-the-Art

Overall Performance. Tables I and II reports the quantitative comparison with representative semi-supervised and class-imbalanced methods. ICoRe-SSL consistently outperforms all baselines across different imbalance ratios and label budgets. The performance gains become more pronounced under severe imbalance ($\rho > 100$), highlighting the robustness of our approach in challenging long-tailed settings.

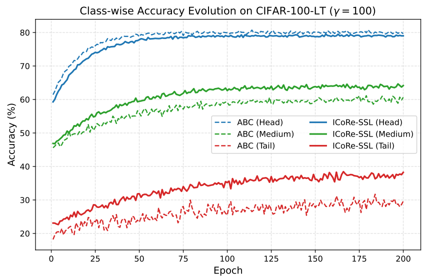
Tail-Class Performance. Compared with ABC, ICoRe-SSL achieves substantial improvements in bACC and GM, indicating significantly enhanced recognition ability for tail classes while maintaining competitive overall accuracy. This demonstrates that reshaping training dynamics effectively mitigates head-class dominance.

C. Training Dynamics and Class-wise Analysis

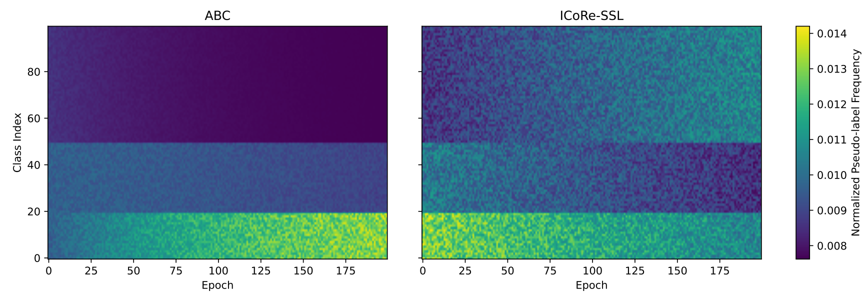
Figure 2 provides a dynamic view of how training imbalance emerges and is progressively corrected by ICoRe-SSL, offering evidence beyond final accuracy comparisons. Unless otherwise specified, all experiments follow the same semi-supervised protocol as the main results.

Specifically, Fig. 2(a) and (b) are conducted on CIFAR-100-LT under an imbalance ratio of $\gamma = 100$, where class-wise accuracy evolution and pseudo-label distribution dynamics are analyzed. Fig. 2(c) reports confusion matrix evolution on CIFAR-10-LT, which is adopted for clearer visualization and interpretability of class-wise confusion patterns.

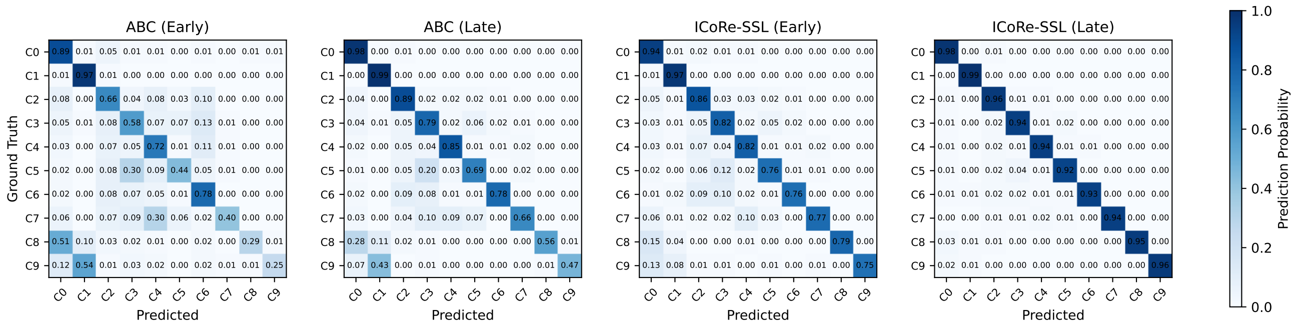
Class-wise Accuracy Evolution. Figure 2(a) illustrates the evolution of accuracy for head, medium, and tail classes. While ABC rapidly converges on head classes and exhibits unstable tail-class learning, ICoRe-SSL achieves steady and consistent



(a) Class-wise pseudo-label accuracy over training epochs.



(b) Pseudo-label confidence distribution heatmaps.



(c) Evolution of class-wise confusion matrices comparing ABC and ICoRe-SSL.

Fig. 2: Training dynamics analysis under class-imbalanced semi-supervised learning.

TABLE II: Comparison of bACC/GM on CIFAR-10-LT

Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
Vanilla	65.2±0.1 / 61.1±0.1	58.8±0.1 / 58.2±0.1	55.6±0.4 / 44.0±1.0
Re-sampling	64.3±0.5 / 60.6±0.7	55.8±0.5 / 45.1±0.3	52.2±0.1 / 38.2±1.5
LDAM-DRW	68.9±0.1 / 67.0±0.1	62.8±0.2 / 58.9±0.6	57.9±0.2 / 50.4±0.3
cRT	67.8±0.1 / 66.3±0.2	63.2±0.5 / 59.9±0.4	59.3±0.1 / 54.6±0.7
FixMatch	79.2±0.3 / 77.8±0.4	71.5±0.7 / 66.8±1.5	68.4±0.2 / 59.9±0.4
/+DARP+cRT	85.8±0.4 / 85.6±0.6	82.4±0.3 / 81.8±0.2	79.6±0.4 / 78.9±0.4
/+CREST+LA	85.6±0.4 / 81.9±0.5	81.2±0.7 / 74.5±1.0	71.9±2.2 / 64.4±1.8
/+ABC	85.6±0.3 / 85.2±0.3	81.1±1.1 / 80.3±1.3	77.3±1.3 / 75.6±1.7
/+CoSSL	86.8±0.3 / 86.6±0.3	83.2±0.5 / 82.7±0.6	80.3±0.6 / 79.6±0.6
/+SAW+LA	86.2±0.2 / 83.9±0.4	80.7±0.2 / 77.5±0.2	73.7±0.1 / 71.2±0.2
/+CDMAD	**87.3±0.1 / 87.0±0.2**	83.6±0.5 / 83.1±0.6	**80.8±0.9 / 79.9±1.1**
/+ICoRe-SSL	86.5±0.2 / 86.2±0.3	**83.8±0.5 / 83.3±0.8**	80.2±0.7 / 79.8±1.0
ReMixMatch	81.5±0.3 / 80.2±0.3	73.8±0.4 / 69.5±0.8	69.9±0.5 / 62.5±0.4
/+ABC	87.9±0.5 / 87.6±0.5	84.5±0.3 / 84.1±0.4	80.5±1.2 / 79.5±1.4
/+CDMAD	**88.3±0.4 / 88.1±0.4**	85.5±0.5 / 85.3±0.4	**82.5±0.2 / 82.0±0.3**
/+ICoRe-SSL	88.1±0.3 / 87.8±0.5	**85.8±0.6 / 85.5±0.8**	82.1±0.7 / 81.6±1.1

improvements for tail classes throughout training, significantly narrowing the head–tail performance gap.

Pseudo-label Distribution Evolution. Figure 2(b) shows the evolution of pseudo-label distributions. ABC suffers from severe pseudo-label collapse toward head classes, resulting in insufficient supervision for tail categories. In contrast, ICoRe-SSL maintains a substantially more balanced pseudo-label distribution over training epochs, enabling continuous and reliable tail-class learning.

Confusion Evolution. Figure 2(c) presents confusion matrices at early and late training stages. While both methods initially exhibit notable tail-to-head misclassification, ABC retains strong off-diagonal responses at later stages. ICoRe-SSL, however, progressively enhances diagonal dominance and suppresses systematic confusion, indicating more balanced class-wise decision boundaries.

D. Ablation Study

We conduct ablation studies to systematically evaluate the contribution of each component in ICoRe-SSL. All results are reported on CIFAR-100-LT with imbalance ratio $\gamma = 100$, as summarized in Table III.

Information-aware Sample Participation. Removing the information-aware module leads to noticeable degradation, particularly on tail classes, confirming its role in suppressing redundant head-class updates.

Confusion-aware Statistics. Disabling confusion-aware statistics results in reduced training stability and tail-class

TABLE III: Ablation study of ICoRe-SSL on CIFAR-100-LT with imbalance ratio $\gamma = 100$.

Method	bACC	Method	bACC
FixMatch w/ ICoRe-SSL	**44.3**	ReMixMatch w/ ICoRe-SSL	**45.1**
w/o information-aware module	43.2	w/o information-aware module	44.5
w/o confusion-aware statistics	42.6	w/o confusion-aware statistics	43.8
w/o information-aware & confusion-aware	41.2	w/o information-aware & confusion-aware	42.2
w/ uniform sample participation	43.8	w/ uniform sample participation	43.5
w/o confusion matrix tracking	41.9	w/o confusion matrix tracking	42.8

performance, highlighting the importance of explicit class-wise diagnostics.

Component Complementarity. Combining both components yields the best performance, indicating that information-aware optimization and confusion-aware analysis are complementary.

E. Computational Efficiency

ICoRe-SSL introduces negligible computational overhead. By suppressing redundant low-information updates, the proposed information-aware sample participation slightly accelerates convergence in practice. Across our experiments, ICoRe-SSL achieves comparable or even lower training time than ABC, with up to approximately 10–15% reduction in the number of effective optimization steps required to reach convergence.

V. Conclusion

We presented ICoRe-SSL, an information- and confusion-regularized framework for class-imbalanced semi-supervised learning. By integrating expectation-preserving information-aware sample participation with explicit class-wise confusion analysis, ICoRe-SSL reshapes training dynamics toward more balanced and stable learning under severe class imbalance. Extensive experiments demonstrate consistent improvements over strong baselines, particularly on tail classes, with negligible computational overhead. Our analysis highlights the importance of modeling information redundancy and class-wise prediction behavior, suggesting promising directions for future research in robust and interpretable semi-supervised learning. Beyond CISSL, our findings suggest that explicitly modeling information redundancy and prediction dynamics is a promising direction for efficient and interpretable learning under data imbalance.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFB3312202 and the Fundamental Research Funds for the Central Universities, China under Grant 2023JBMC001 and the National Natural Science Foundation of China (NSFC) under Grants 61906014.

References

[1] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "ReMixMatch: Semi-supervised learning with distribution alignment and augmentation anchoring," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020.

[2] K. Sohn, D. Berthelot, C. Li, Z. Zhang, N. Carlini, and A. Kurakin, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020.

[3] C. Wang, Z. Liu, J. Zhou, H. Zhang, and N. Yu, "FreeMatch: Self-adaptive thresholding for semi-supervised learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.

[4] H. Lee, S. Shin, and H. Kim, "ABC: Auxiliary balanced classifier for class-imbalanced semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2021.

[5] Z. Qin, K. Wang, Z. Zheng, J. Gu, X. Peng, and Z. Xu, "InfoBatch: Lossless training speed up by unbiased dynamic data pruning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024.

[6] I. Katharopoulos and F. Fleuret, "Not all samples are created equal: Deep learning with importance sampling," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018.

[7] G. He and C. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[8] Y. Oh, D.-J. Kim, and I. S. Kweon, "DASO: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[9] A. K. Menon, S. Jayasumana, and A. S. Rawat, "Long-tail learning via logit adjustment," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.

[10] J. Lazarow, K. Sohn, C.-Y. Lee, and C.-L. Li, "Unifying distribution alignment as a loss for imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023.

[11] X. Wang, Z. Wu, L. Lian, and S. X. Yu, "Debiased learning from naturally imbalanced pseudo-labels," *arXiv preprint arXiv:2202.09462*, 2022.

[12] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017.

[13] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML) Workshop*, 2013.

[14] A. Tarvainen and H. Valpola, "Mean teachers are better role models," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.

[15] C. Wei, K. Sohn, S. M. Lee, J. Li, and T. Pfister, "CReST: A class-rebalancing self-training framework for imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021.

[16] Y. Fan, D. Dai, A. Kukleva, and B. Schiele, "CoSSL: Co-learning of representation and classifier for imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[17] Z. Lai, C. Wang, H. Gunawan, S. C. S. Cheung, and C. N. Chuah, "Smoothed adaptive weighting for imbalanced semi-supervised learning: Improving reliability against unknown distribution data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[18] H. Lee and H. Kim, "CDMAD: Class-distribution-mismatch-aware debiasing for class-imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024.

[19] W. Xing, Y. Cheng, H. Yi, X. Gao, X. Wei, and X. Guo, "LCGC: Learning from consistency gradient conflicting for class-imbalanced semi-supervised debiasing," in *Proc. AAAI Conf. Artif. Intell.*, 2025.

[20] L.-Z. Guo and Y.-F. Li, "Class-imbalanced semi-supervised learning with adaptive thresholding," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 162, pp. 8082–8094, 2022.

[21] J. Kim, Y. Hur, S. Park, E. Yang, S. J. Hwang, and J. Shin, "Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, pp. 14567–14579, 2020.