IEEE ICIP 2023

Rethinking Long-tailed Visual Recognition with Dynamic Probability Smoothing and Frequency Weighted Focusing

Wan Jun Nah[1], Chun Chet Ng[1], Che-Tsung Lin[2], Yeong Khang Lee[3], Jie Long Kew[1], Zhi Qin Tan[1], Chee Seng Chan[1], Christopher Zach[2], Shang-Hong Lai[4]

[1]Universiti Malaya. [2]Chalmers University of Technology.
[3]ViTrox Corporation Berhad. [4]National Tsing Hua University.

Long-Tailed Recognition

- Real-world data often exhibit **long-tailed distribution.**
- Models trained with such data often **favor head classes, affecting fairness.**

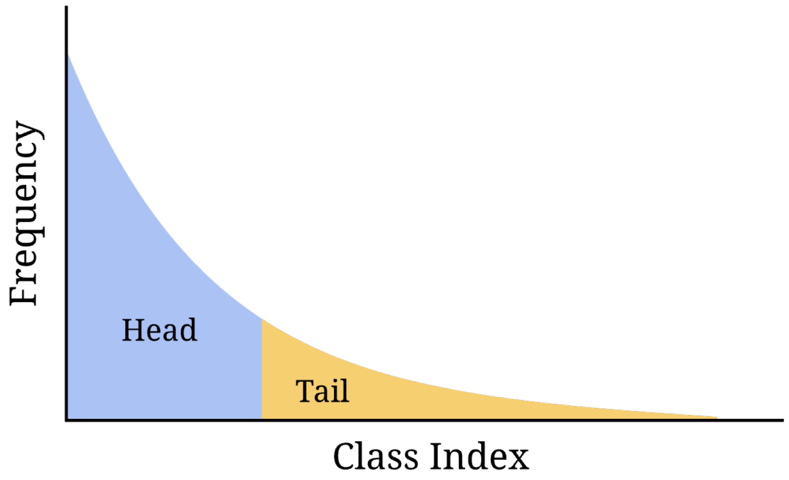

Figure 1: Long-tailed distribution.

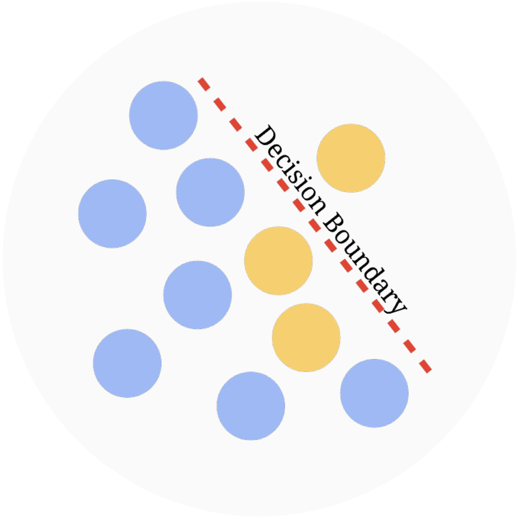

Figure 2: Biased decision boundary.

Common Approaches

1. **Re-sampling**
    - Under-sampling / Over-sampling to change data distribution (Chawla et al., Buda et al.).
    - **Information loss / overfitting.**
2. **Re-weighting**
    - Assign higher penalties to under-represented samples to alter their impact on loss.
    - Class-wise (CB Loss [Cui et al.], LDAM [Cao et al.])
        - **Assume every instance has same difficulty.**
    - Instance-wise (FL Loss [Lin et al.], IB Loss [Park et al.])
        - **Memorization of hard examples, unable to generalize.**
3. **Training scheme**
    - Decoupling representation learning and classifier learning (Kang et al.).
    - Deferred re-weighting / re-sampling (DRW/DRS) (LDAM [Cao et al.]).

Motivations

- Naively combining instance- and class-level re-weighting (CB-FL Loss [Cui et al.]) **causes overfitting.**
- Overfitting is **exacerbated** when the number of **samples decreases** as high class-level weights amplify the loss of outliers.
- Most popular LT datasets only cover natural objects captured under ideal lighting conditions, **less representative of real-world scenarios.**

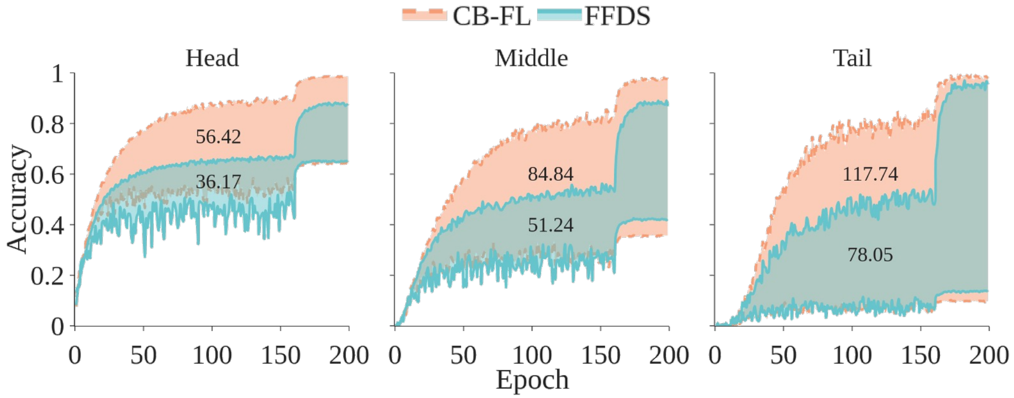

Figure 3: Overfitting of CB-FL on CIFAR-100-LT, IF = 100.
Colored areas: Training Acc. - Testing Acc.

Contributions

1. **Novel loss function: Frequency weighted Focusing with Dynamic Smoothing (FFDS) & Deferred-FFDS (D-FFDS)**
   - **Address intra- and inter-class imbalance** without over-focusing outliers.
2. **New dataset: ICText-LT**
   - **Diverse qualities** of integrated circuits (IC).
   - Collected under real-world industrial conditions.
3. **Promising results**
   - **Outperforming state-of-the-art alternatives.**
   - LT Benchmarks: CIFAR-10/100-LT, Tiny ImageNet-LT, and ICText-LT.

FFDS

Let $z = [z_1, z_2, \cdots, z_C]^T$ predicted logits of a model for all C classes,
$p_i = e^{z_i} / \sum_{j=1}^{C} e^{z_j}, \forall i \in \{1,2,\cdots,C\}$ predicted probabilities.

- **Overall loss function:**

$$L_{FFDS}(z,y) = -w_y (1 - \hat{p}_y)^{\gamma_y} \sum_j^C Q(j) \log(p_j), \quad (1)$$

$$Q(j) = (1-\epsilon)\mathbb{1}_{\{j=y\}} + \big(\epsilon/(C-1)\big)\mathbb{1}_{\{j\neq y\}}, \quad (2)$$ *Label smoothing*

$$w_y = (1-\beta)/\big(1-\beta^{N_y}\big). \quad (3)$$ *Class-balanced weight*

- **Modules:**
  - **Frequency-weighted focusing (FreqFocus):** To address intra-class imbalance.
  - **Dynamic probability smoothing (DynaSmooth):** To alleviate overfitting on outliers.

FFDS

$$L_{FFDS}(z,y) = -w_y(1-\hat{p}_y)^{\gamma_y}\sum_j^C Q(j)\log(p_j) \quad (1)$$

1. FreqFocus

- **Observations:**
  - **Head** classes: Presence of **hard examples.**
  - **Tail** classes: **Limited data** availability.
- **Idea:** Encourage the model to **attend to hard examples** (low p) **in head classes** while **treating tail examples equally.**
- **Three forms:**
  - Linear
    $$\gamma_y = \gamma_{\min} + (\gamma_{\max} - \gamma_{\min})\left(\frac{N_y - N_{\min}}{N_{\max} - N_{\min}}\right) \quad (4)$$
  - Convex
    $$\gamma_y = \gamma_{\min} + (\gamma_{\max} - \gamma_{\min})\left(\frac{N_y - N_{\min}}{N_{\max} - N_{\min}}\right)^3 \quad (5)$$
  - Concave
    $$\gamma_y = \gamma_{\min} + (\gamma_{\max} - \gamma_{\min})\tanh\left(4 \cdot \frac{N_y - N_{\min}}{N_{\max} - N_{\min}}\right) \quad (6)$$

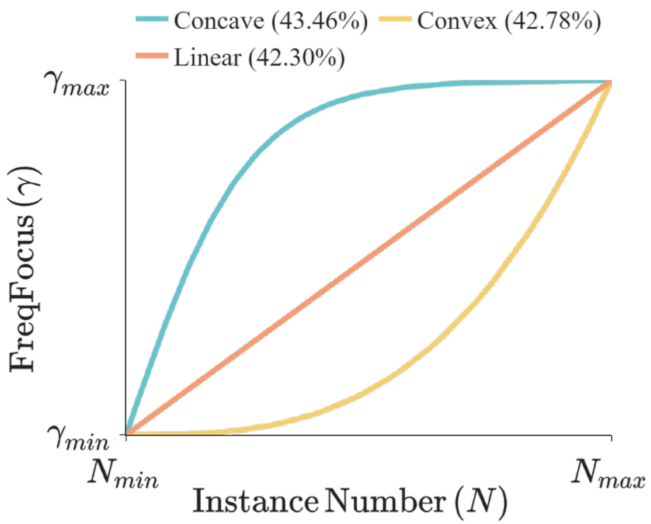

Figure 4: Relationship between class frequency and gamma, along with their respective accuracy on CIFAR-100-LT, IF = 100.

FFDS

$$L_{FFDS}(z,y) = -w_y(1-\hat{p}_y)^{\gamma_y}\sum_j^C Q(j)\log(p_j) \quad (1)$$

2. DynaSmooth

- **Observation: Over-focusing on outliers** (with extreme p) **exacerbate overfitting.**
- **Idea:** Dynamically **reduce instance-level weights** in proportional to their **(eq. 7) likelihood of being outliers.**
- To maintain training stability, the likelihood of being outliers is computed within groups partitioned according to class frequency.

$$f(d) = k\left(\sqrt{p_y} - \sqrt{\bar{p}_{m_y}}\right)^2 \quad (7)$$

$$\hat{p}_y = p_y - f(d)\left(p_y - \bar{p}_{m_y}\right) \quad (8)$$

$\bar{p}_{m_y}$: mean predicted probability of the group containing y from T-1 epoch
k: hyperparameter

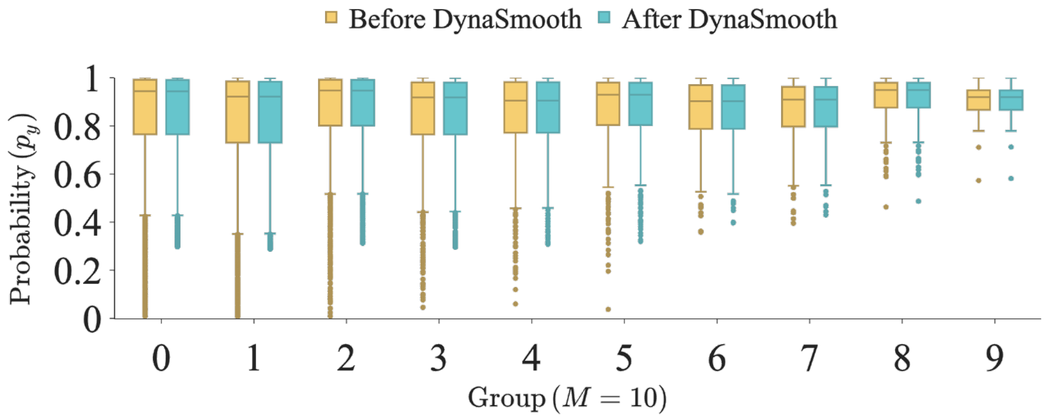

Figure 5: Boxplot of probability before and after applying DynaSmooth.

Deferred-FFDS (D-FFDS)

- **Idea:** To address convergence difficulties and training instability (Kang et al.), D-FFDS fine-tunes models only after learning meaningful representation with Cross Entropy.
- Training Scheme:

Algorithm 1 D-FFDS Training Scheme

Require: $B = (x_i, y_i)_{i=1}^{N}$: dataset, f_θ: θ parameterized model, δ: learning rate

Initialize the model with random parameters θ
$t \leftarrow 0$
while $t < T_{\text{phase\_1}}$ **do**
  $B_m \sim B$ ▷ Sample mini-batch with m samples
  $L(f_\theta) \leftarrow \frac{1}{m} \sum_{(x,y)\in B_m} L_{\text{CE}}(f(\theta), y)$
  $f_\theta \leftarrow f_\theta - \delta \nabla L(f_\theta)$
  $t \leftarrow t + 1$
end while
while $t < T_{\text{phase\_2}}$ **do**
  $B_m \sim B$ ▷ Sample mini-batch with m samples
  $L(f_\theta) \leftarrow \frac{1}{m} \sum_{(x,y)\in B_m} L_{\text{FFDS}}(f(\theta), y)$
  $f_\theta \leftarrow f_\theta - \delta \nabla L_{\text{FFDS}}(f_\theta)$
  $t \leftarrow t + 1$
end while

ICText-LT Dataset

- **Statistics:**
  - **36 classes** (A-Z, 0-9)
  - Training images: **68,307** (Imbalance)
    Testing images: **6,300** (Balance)
  - Imbalance factor: **18** (natural), **100** (re-sampled)

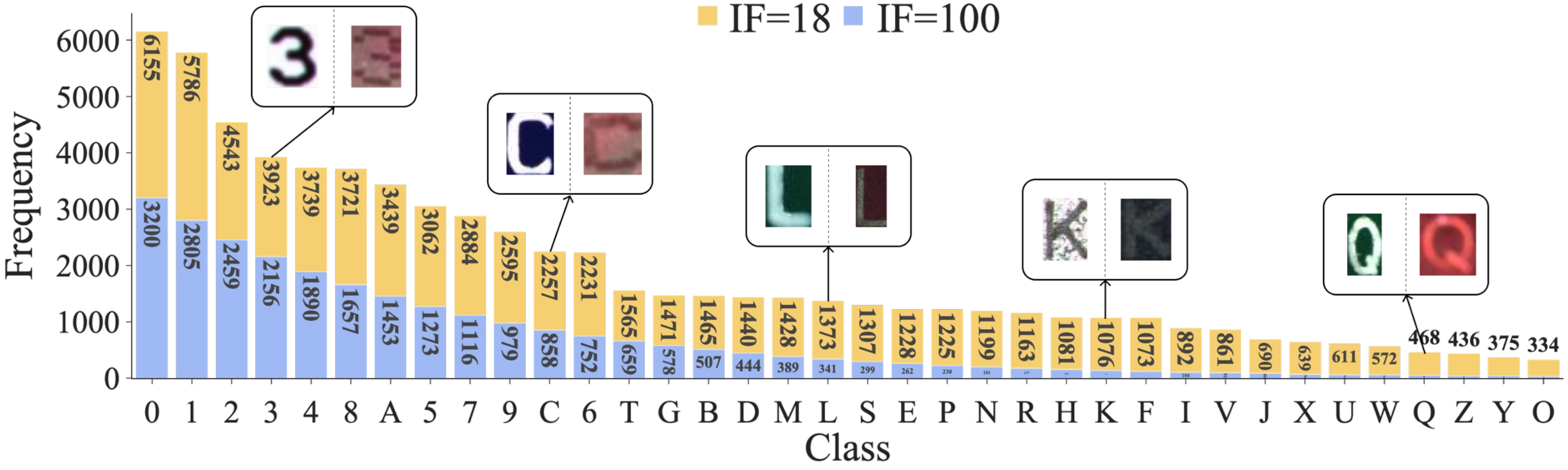

Figure 6: Distribution of ICText-LT's training set with IF ∈ {18, 100}.

Experimental Results

1. Datasets

- CIFAR-10/100-LT (10/100 classes), Tiny ImageNet-LT (200 classes), and ICText-LT (36 classes).
- Long-tailed imbalance factor (IF): 10 – 100.

2. Baselines

- **Cross Entropy Loss (CE):** Standard cross entropy loss.
- **Focal Loss (FL):** Re-weight samples proportional to their difficulties.
- **Class-Balanced Loss (CB):** Re-weight classes inversely proportional to the effective number of samples.
- **Label-Distribution Aware Loss (LDAM):** Assign larger class-dependent margins to tail classes.
- **Influence-Balanced Loss (IB):** Down-weight high influential samples.

Results Comparison

Our methods (FFDS & D-FFDS) **excel over state-of-the-art alternatives** on LT benchmarks with varying imbalance factors.

	Method	CIFAR-10-LT			CIFAR-100-LT			Tiny ImageNet-LT		ICText-LT	
		100	50	10	100	50	10	100	10	100	18
One-Phase	CE	70.47	77.21	86.40	38.85	43.24	56.09	38.19	53.22	74.22	83.51
	FL	70.38	76.71	86.66	38.41	44.32	55.78	38.95	54.02	74.21	83.70
	CB	70.36	74.81	87.03	38.32	43.85	55.71	41.37	54.82	75.29	84.86
	CB-FL	74.57	79.27	85.73	39.60	41.66	57.99	38.71	54.92	75.35	83.70
	LDAM	73.35	78.74	86.96	39.60	44.19	56.91	39.40	54.58	74.24	83.38
	FFDS	**75.60**	**79.82**	**87.46**	**40.74**	**45.67**	**58.66**	**42.34**	**56.11**	**76.89**	**85.40**
Two-Phase	LDAM-DRW	77.03	81.53	88.16	42.04	47.71	58.71	42.78	57.06	77.87	84.52
	IB	78.26	81.70	88.25	42.14	46.22	57.13	42.65	57.22	76.59	85.41
	IB-CB	78.04	81.54	88.09	41.31	46.16	56.78	40.15	55.79	75.86	85.22
	IB-FL	79.76	81.51	88.04	42.06	47.49	58.20	41.04	57.06	77.59	85.05
	D-FFDS	**79.93**	**82.94**	**88.48**	**43.46**	**48.48**	**58.82**	**43.86**	**58.31**	**79.56**	**85.98**

Table 1: Comparison of testing accuracy.

D-FFDS outperforms FFDS by starting with a better representation.

Ablation Studies

Grouping classes with similar characteristics when computing outlier likelihood **aids model performance** along with the proposed modules.

Number of Groups (M)	CIFAR-100-LT	
	100	10
1	42.66	58.24
5	42.07	**58.82**
10	**43.46**	58.29
50	43.17	58.41

Table 2: Effect of Number of Groups, M.

DynaSmooth	FreqFocus	CIFAR-100-LT	
		100	10
-	-	41.80	57.10
-	✓	42.43	57.68
✓	-	42.71	58.21
✓	✓	**43.46**	**58.82**

Table 3: Effectiveness of Proposed Modules.

Conclusion

- Our **novel loss functions (FFDS & D-FFDS)** address both intra- and inter-class imbalance without over-focusing on outliers.
- We introduce **ICText-LT dataset**, which presents challenging character representations collected in real-world industrial settings.
- Experimental results show that our method **outperforms existing approaches** on CIFAR-10/100-LT, Tiny ImageNet-LT, and ICText-LT.

Paper

Project Page