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.

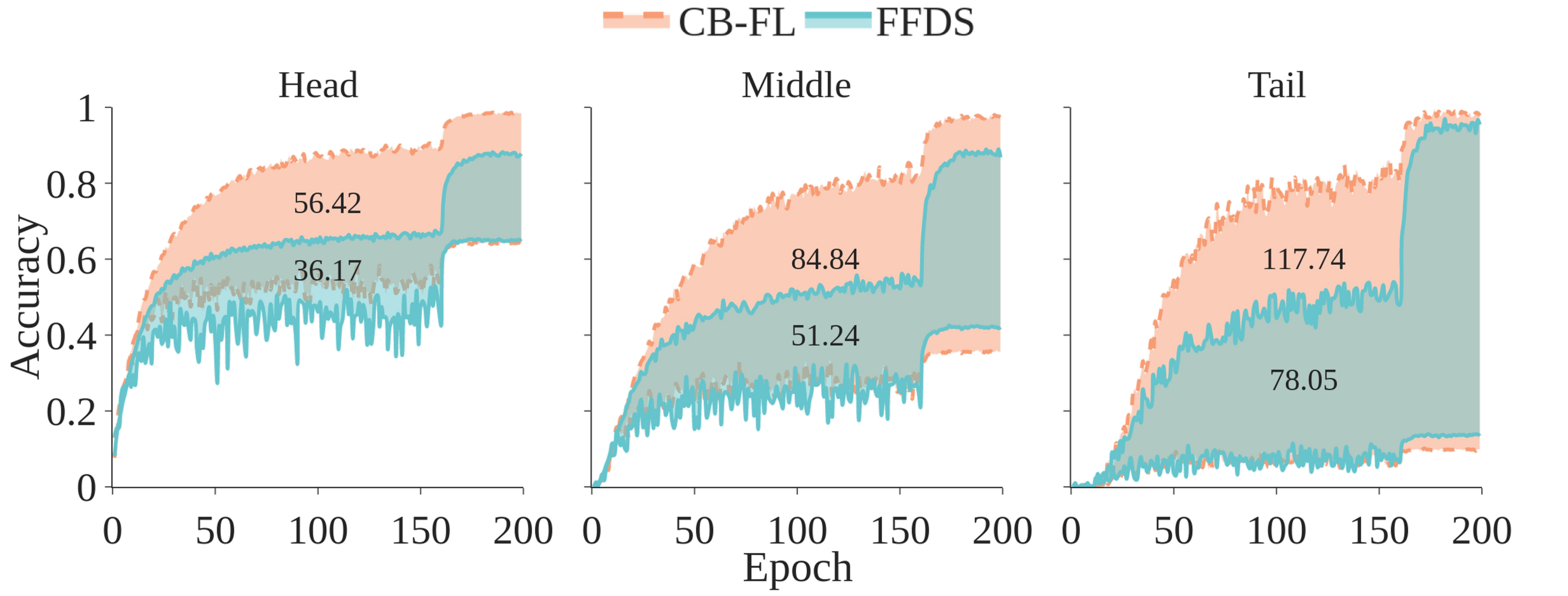
Project Page

Long-Tailed Recognition

- Real-world data often in **long-tailed (LT) distribution.**
- Models trained with such data often **favor head classes, impairing fairness.**

Motivations

- Naively combining instance- and class-level re-weighting **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, **less representative of real-world scenarios.**



CB-FL: Class-balanced Loss + Focal Loss FFDS: Proposed Loss

Colored areas: Training Acc. - Testing Acc.

Contributions

Loss Function FFDS & D-FFDS	Dataset ICText-LT	Promising Results

Proposed Method – FFDS & D-FFDS

Stage 1: CE
Stage 2: FFDS

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

$Q(j)$: label smoothing with parameter 0.1

w_y: class-balanced weight

a) FreqFocus

- **Head** Classes: Presence of **hard examples.**
  Tail Classes: **Limited data** availability.
- Encourage the model to **attend to hard examples** *(low p_j)* **in head classes** while **treating tail examples equally.**
- Three forms: **(eq. 2) Linear**, Convex, Concave.

$$\gamma_y = \gamma_{min} + (\gamma_{max} - \gamma_{min})\left(\frac{N_y - N_{min}}{N_{max} - N_{min}}\right) \tag{2}$$

N: Class frequency

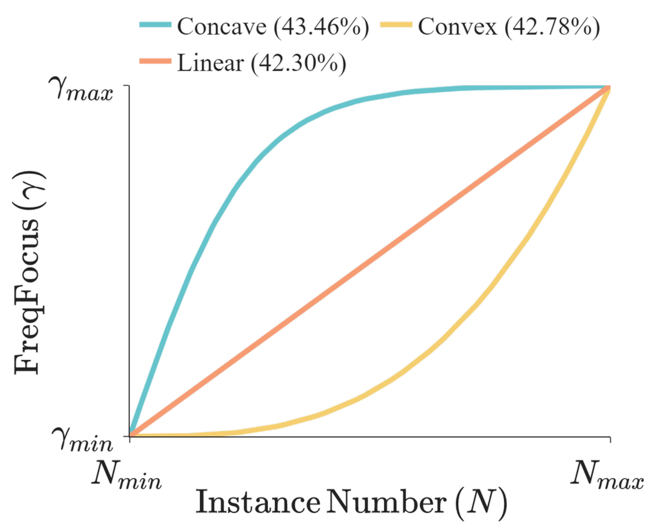
b) DynaSmooth

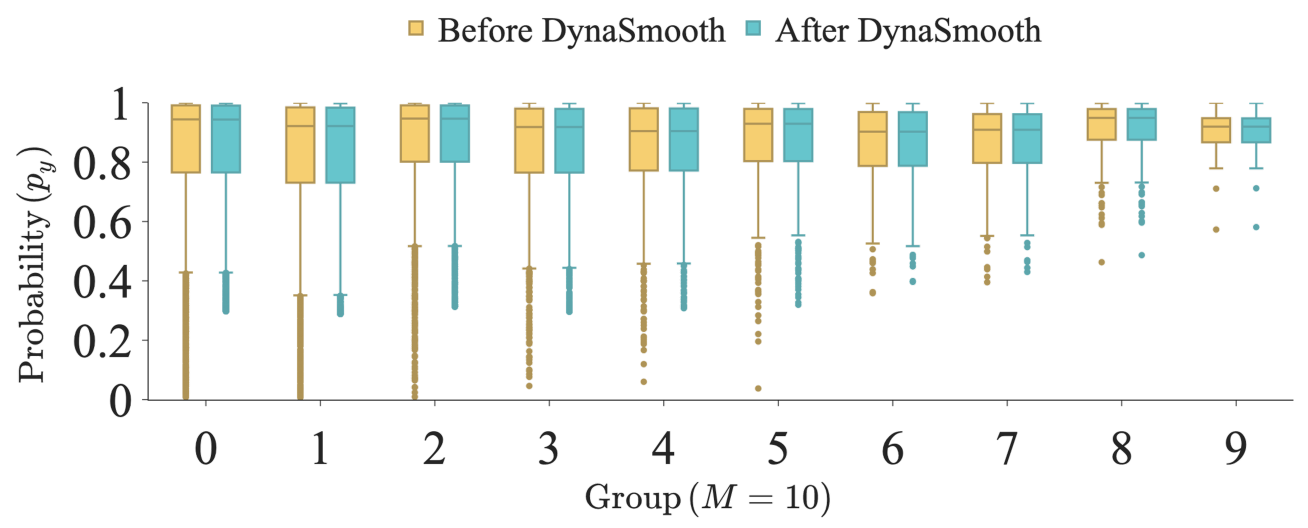
- **Over-focusing outliers** (with extreme p_j) could **exacerbate overfitting.**
- Dynamically **reduce instance-level weights in proportional** to their ***(eq. 3)* likelihood of being outliers.**

$$f(d) = k\left(\sqrt{p_y} - \sqrt{\bar{p}_{m_y}}\right)^2 \tag{3}$$

$$\hat{p}_y = p_y - f(d)\left(p_y - \bar{p}_{m_y}\right) \tag{4}$$

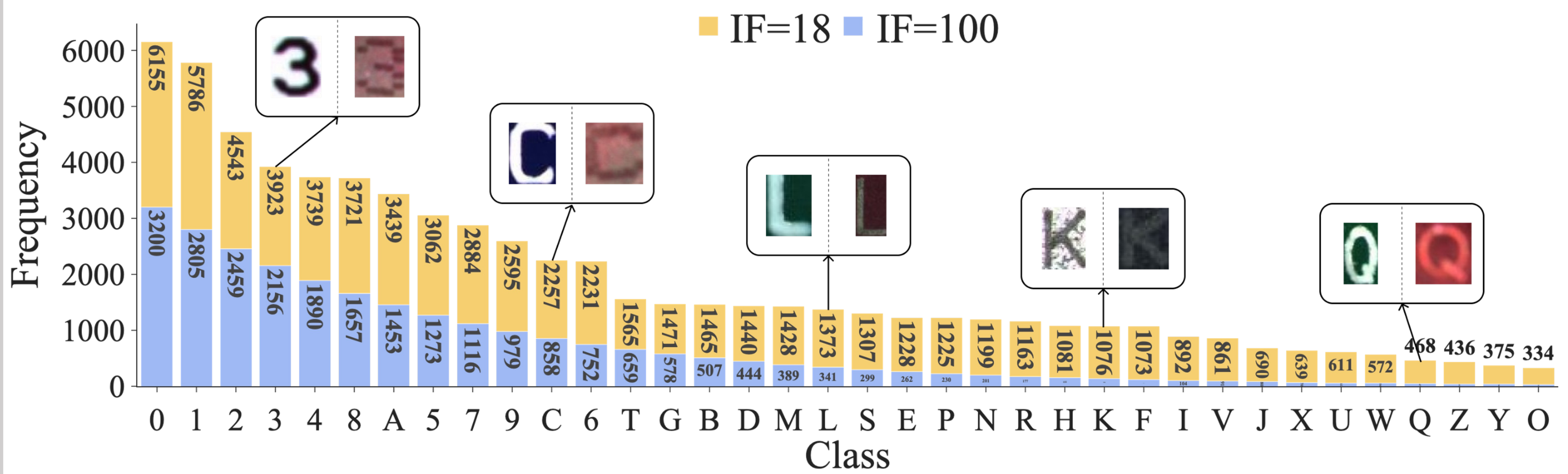
$\bar{p}_{m_y}$: mean predicted probability of the group containing y from T-1 epoch





Proposed Dataset – ICText-LT

- Industrial-based dataset focused on **detecting printed characters on chip components.**
- **36 classes** (A-Z, 0-9)
- Training images: **68,307** (Balance)
  Testing images: **6,300** (Imbalance)
- Imbalance factor: **18** (natural), **100** (re-sampled)



Experimental Results

	Method	CIFAR-100-LT 100	CIFAR-100-LT 10	Tiny ImageNet-LT 100	Tiny ImageNet-LT 10	ICText-LT 100	ICText-LT 18
One-Phase	CE	38.85	56.09	38.19	53.22	74.22	83.51
	FL	38.41	55.78	38.95	54.02	74.21	83.70
	CB	38.32	55.71	41.37	54.82	75.29	84.86
	CB-FL	39.60	57.99	38.71	54.92	75.35	83.70
	LDAM	39.60	56.91	39.40	54.58	74.24	83.38
	FFDS	**40.74**	**58.66**	**42.34**	**56.11**	**76.89**	**85.40**
Two-Phase	LDAM-DRW	42.04	58.71	42.78	57.06	77.87	84.52
	IB	42.14	57.13	42.65	57.22	76.59	85.41
	IB-CB	41.31	56.78	40.15	55.79	75.86	85.22
	IB-FL	42.06	58.20	41.04	57.06	77.59	85.05
	D-FFDS	**43.46**	**58.82**	**43.86**	**58.31**	**79.56**	**85.98**