Invariant Feature Learning for Generalized Long-Tailed Classification

Kaihua Tang[1], Mingyuan Tao[2], Jiaxin Qi[1], Zhenguang Liu[3], Hanwang Zhang[1]
[1]Nanyang Technological University, [2]Damo Academy, Alibaba Group, [3]Zhejiang University

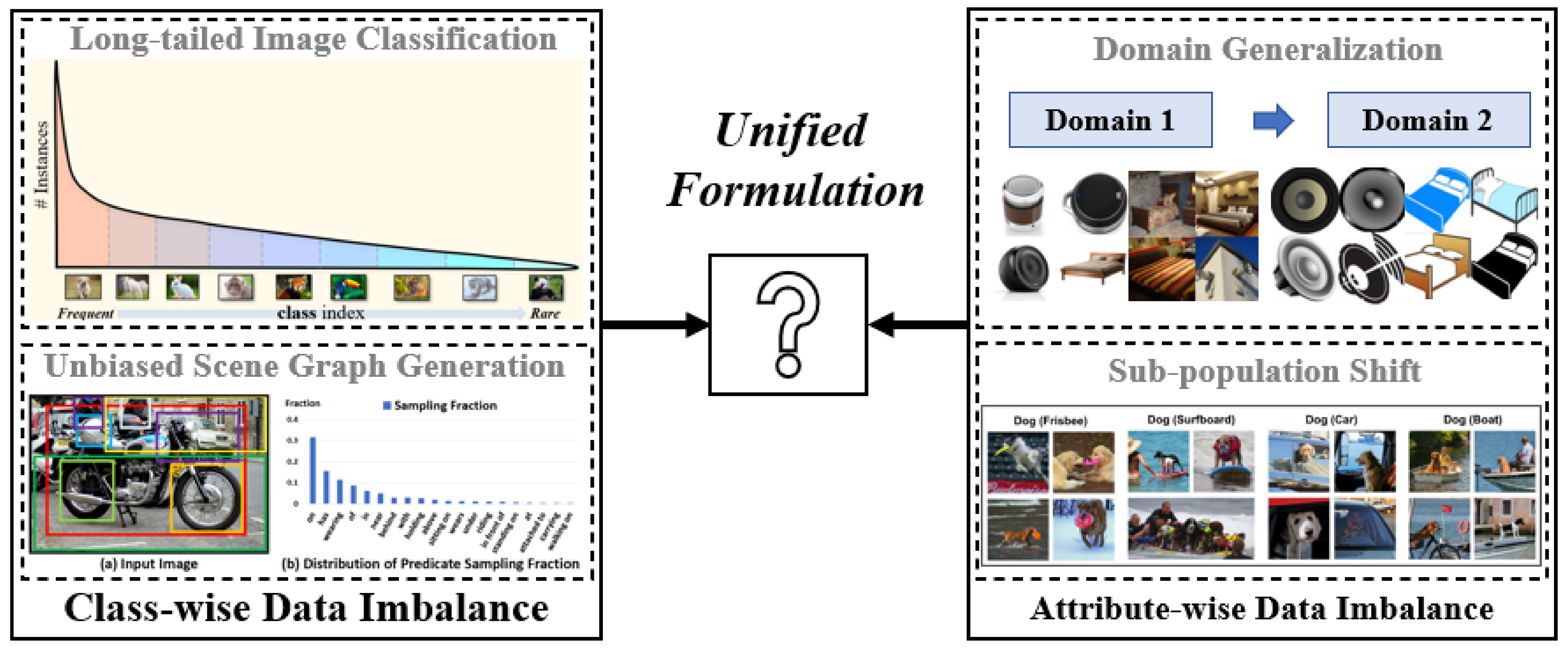
ECCV TEL AVIV 2022

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
DAMO ALIBABA DAMO ACADEMY

Motivation:

➢ Generalized Long-tailed Problem:

a) Class-wise Data Imbalance
b) Attribute-wise Data Imbalance



➢ Reviewing the Conventional Long Tail:

The **previous problem formulation** [1,2]:

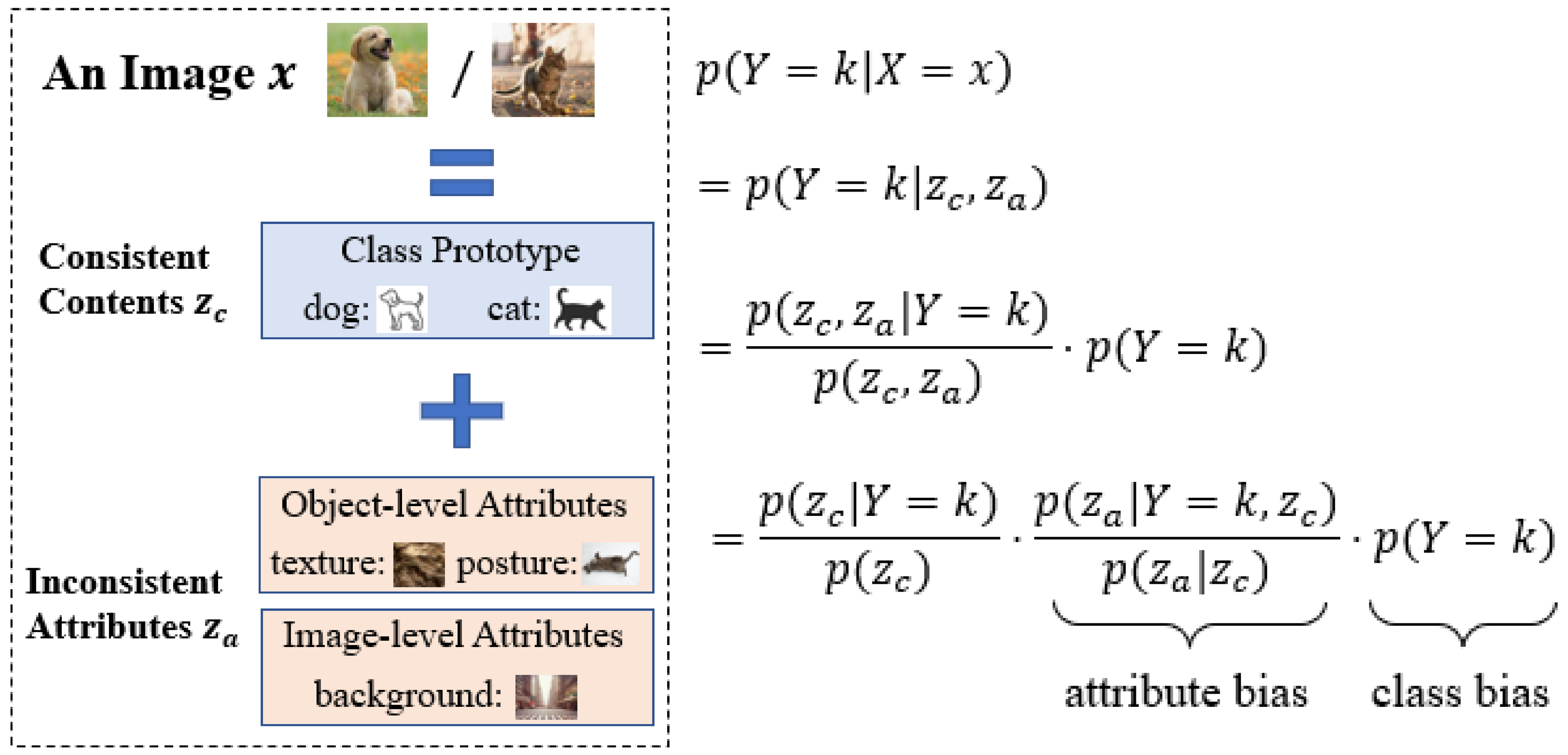
$$P(y|x) = \frac{P(x|y)}{P(x)} P(y) \propto P(x|y)\overbrace{P(y)}^{bias}$$

Unbiased Prediction:

$$\log(P(y|x)) - \log(P(y)) = \log\underbrace{\left(\frac{P(x,y)}{P(x)\cdot P(y)}\right)}_{\text{Point-wise mutual information}}$$

Underlying assumption **(Too Strong!)**:
Intra-class distribution is unchanged, $P_{tr}(x|y) = P_{te}(x|y)$

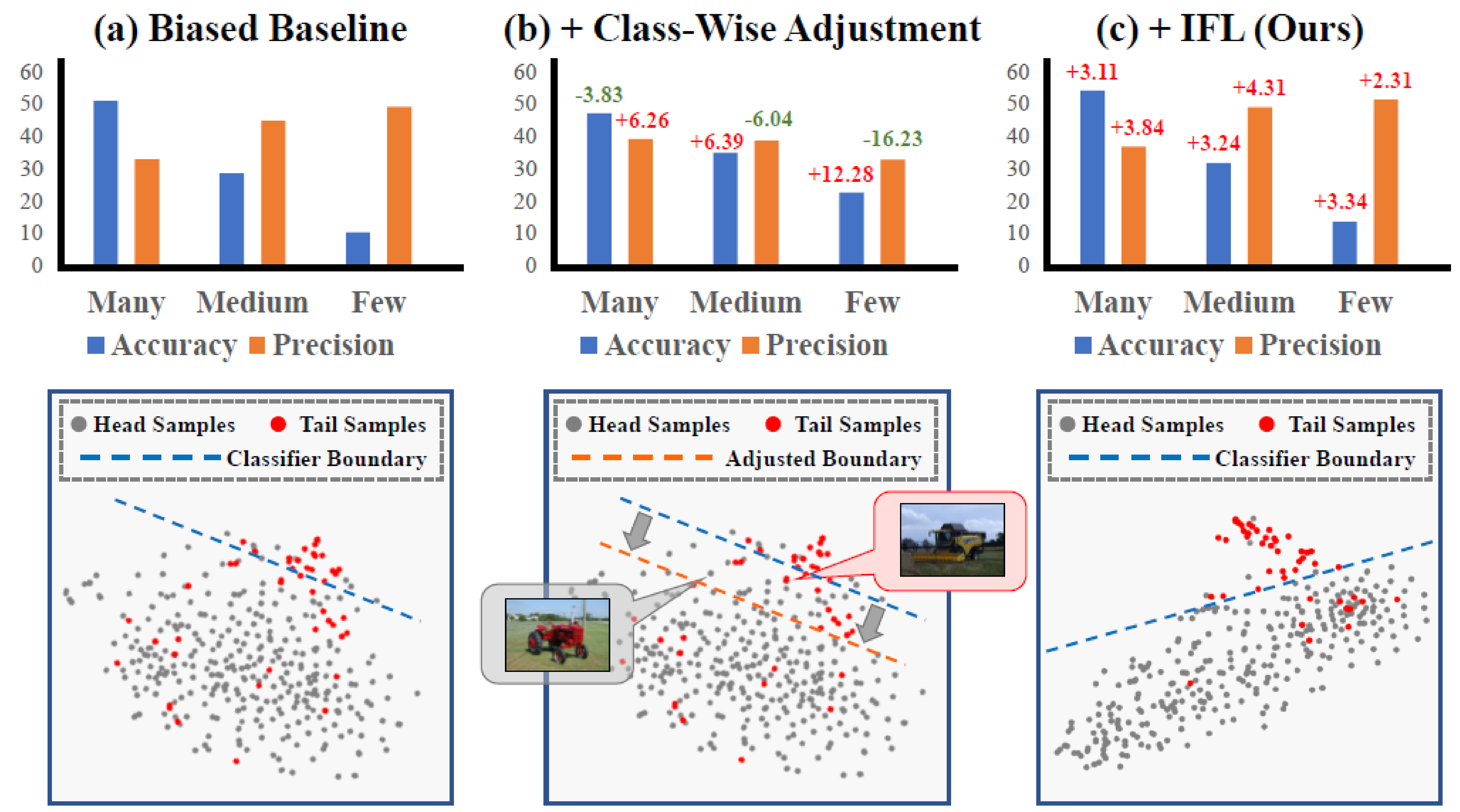
➢ The Generalized Problem Formulation:



Approach:

➢ The Limitation of Conventional LT:

a) Class-wise adjustment has precision-accuracy trade-off.
b) The confusing feature space is not well addressed.

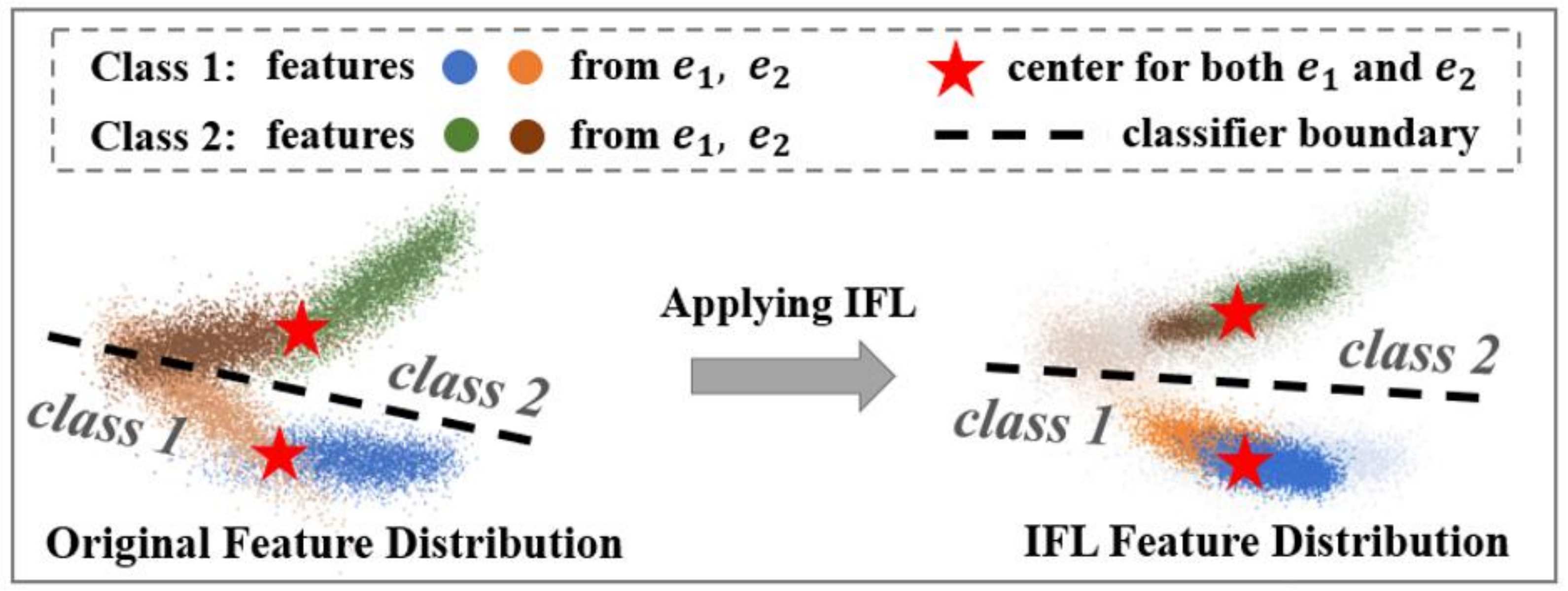


➢ The Proposed Invariant Feature Learning (IFL):

Invariant Risk Minimization [3] version of Center Loss [4]:

$$\min_{\theta, w} \sum_{e \in E} \sum_{i \in e} L_{cls}(f(x_i^e; \theta), y_i^e; w),$$

$$\text{subject to } \theta \in \mathrm{argmin}_\theta \sum_{e \in E} \sum_{i \in e} \left\| f(x_i^e; \theta) - C_{y_i^e} \right\|_2$$



➢ The Purposes of Applying IFL:

1. Estimating a set of more balanced and fairer class centers.
2. Increasing worst case performance to ensure the consistency.
3. Solving the attribute-wise data imbalance.

Experiment:

➢ Evaluation Protocols and Metrics:

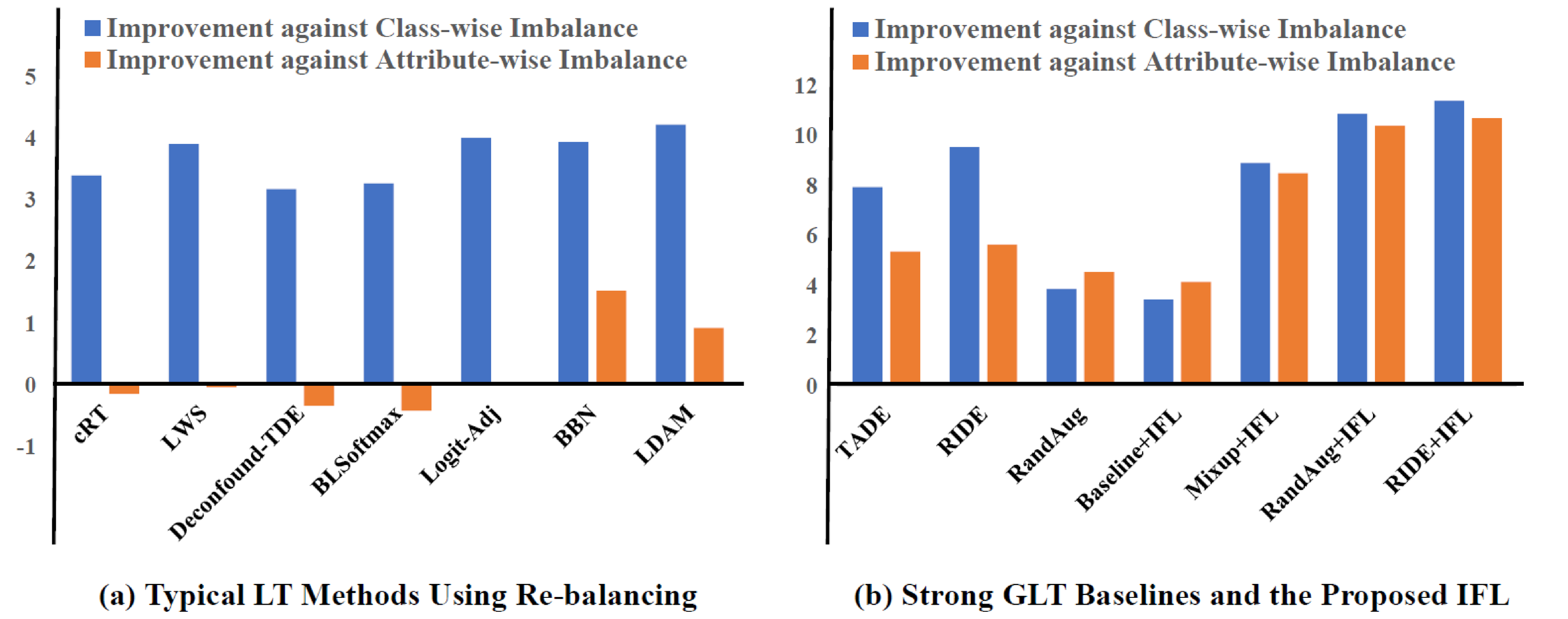
a) Protocol 1: Class-wise Long Tail (**CLT**)
b) Protocol 2: Attribute-wise Long Tail (**ALT**)
c) Protocol 3: Generalized Long Tail (**GLT**=CLT+ALT)
d) Metrics: Top-1 **Accuracy** and Top-1 **Precision**

➢ Plugging IFL into different types of LT algorithms:

	Methods	Class-Wise Long Tail (CLT) Protocol								Generalized Long Tail (GLT) Protocol							
	< Accuracy | Precision >	$Many_C$		$Medium_C$		Few_C		Overall		$Many_C$		$Medium_C$		Few_C		Overall	
Re-balance	Baseline	59.34	39.08	36.95	52.87	14.39	56.65	42.52	47.92	50.98	32.90	28.49	44.72	10.28	49.11	34.75	40.65
	cRT [19]	56.55	45.79	42.89	46.23	26.67	41.47	45.92	45.34	48.02	38.40	34.16	38.07	19.92	33.50	37.57	37.51
	LWS [19]	55.38	46.67	43.91	46.87	30.11	40.92	46.43	45.90	47.15	39.16	34.88	38.68	22.56	32.88	37.94	38.01
	Deconfound-TDE [50]	54.94	49.27	43.18	43.91	28.64	33.40	45.70	44.48	46.87	42.39	34.43	35.77	22.11	26.30	37.56	37.00
	BLSoftmax [41]	55.60	48.19	42.74	47.27	28.79	38.14	45.79	46.27	47.15	40.89	33.48	39.11	21.10	27.50	37.09	38.08
	Logit-Adj [33]	54.55	49.70	44.40	45.05	31.53	36.04	46.53	45.56	45.94	41.97	35.15	36.63	24.07	28.59	37.80	37.56
	BBN [71]	61.64	42.74	43.80	54.44	13.94	55.12	46.46	49.86	52.41	35.58	34.31	46.38	10.06	44.43	37.91	41.77
	LDAM [6]	59.05	45.39	43.23	48.80	24.44	44.99	46.74	46.86	51.02	38.78	34.13	40.39	18.46	35.91	38.54	39.08
	(ours) Baseline + IFL	**62.71**	42.98	40.10	**56.83**	18.92	**61.92**	45.97	**52.06**	**54.09**	36.74	31.73	**49.03**	13.62	**51.42**	37.96	**44.47**
	(ours) cRT + IFL	61.27	45.84	43.96	51.67	24.32	53.64	47.94	49.63	52.75	39.11	35.14	43.36	17.92	43.35	39.60	41.65
	(ours) LWS + IFL	61.50	45.43	43.79	52.85	23.86	55.58	47.89	50.29	53.21	38.92	34.99	44.44	17.42	45.90	39.64	42.45
	(ours) BLSoftmax + IFL	58.00	53.70	44.70	51.73	33.49	37.58	48.34	50.39	49.92	46.86	36.11	44.31	25.71	32.01	40.08	43.48
	(ours) Logit-Adj + IFL	56.96	**56.22**	**46.54**	50.10	**36.88**	33.29	**49.26**	50.02	48.25	**49.17**	**37.50**	41.65	**29.00**	25.77	**40.52**	42.28
Augment	Mixup [66]	59.68	37.96	30.83	55.74	7.09	34.33	38.81	45.41	51.04	31.85	23.10	47.25	4.94	22.88	31.55	37.44
	RandAug [10]	64.96	42.63	40.30	59.10	15.20	56.60	46.40	52.13	56.36	35.97	31.43	51.13	10.36	48.92	38.24	44.74
	(ours) Mixup + IFL	67.71	47.77	45.87	62.58	24.71	**67.77**	51.43	57.44	59.36	40.95	36.77	54.67	18.06	55.10	43.00	49.25
	(ours) RandAug + IFL	**69.35**	**49.42**	**48.05**	**63.19**	**26.92**	66.04	**53.40**	**58.11**	**60.79**	**42.41**	**39.07**	**55.15**	**20.04**	**57.90**	**44.90**	**50.47**
Ensemble	TADE [67]	58.44	**56.38**	48.01	51.41	**36.60**	41.08	50.47	51.85	50.29	**49.25**	38.74	43.74	**27.99**	31.75	41.75	44.15
	RIDE [56]	64.04	51.91	48.66	53.21	30.44	46.25	52.08	51.65	55.47	44.55	38.65	44.26	22.80	37.26	43.00	43.32
	(ours) TADE + IFL	61.71	55.59	48.87	53.42	34.02	40.93	51.78	52.41	53.75	48.73	39.90	45.28	26.77	35.34	43.47	45.17
	(ours) RIDE + IFL	**65.68**	54.13	**50.82**	**56.22**	31.91	**52.10**	**53.93**	**54.76**	**57.84**	47.00	**41.80**	**48.65**	24.63	**42.96**	**45.64**	**47.14**

➢ Why GLT is a "Generalized" Version of LT?

A: GLT algorithms can automatically solve the previous (class-wise) LT problem, but not vise versa.



(a) Typical LT Methods Using Re-balancing (b) Strong GLT Baselines and the Proposed IFL

Reference:

[1] Menon, Aditya Krishna, et al. "Long-tail learning via logit adjustment." ICLR 2021
[2] Ren, Jiawei, et al. "Balanced meta-softmax for long-tailed visual recognition." NeurIPS 2020
[3] Arjovsky, Martin, et al. "Invariant risk minimization." arXiv preprint arXiv:1907.02893 (2019).
[4] Wen, Yandong, et al. "A discriminative feature learning approach for deep face recognition." ECCV 2016.

QR Code to Github Link:

QR Code to Paper Link: