Invariant Feature Learning for Generalized Long-Tailed Classification

Kaihua Tang[1], **Mingyuan Tao**[2], **Jiaxin Qi**[1], **Zhenguang Liu**[3], **Hanwang Zhang**[1]

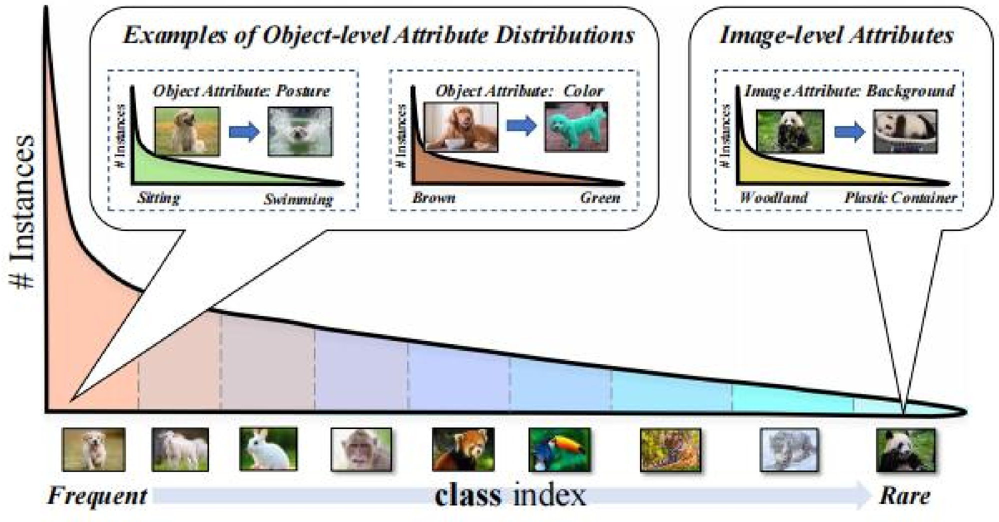
[1]Nanyang Technological University, [2]Damo Academy, Alibaba Group, [3]Zhejiang University

kaihua.tang@ntu.edu.sg, juchen.tmy@alibaba-inc.com, jiaxin003@e.ntu.edu.sg

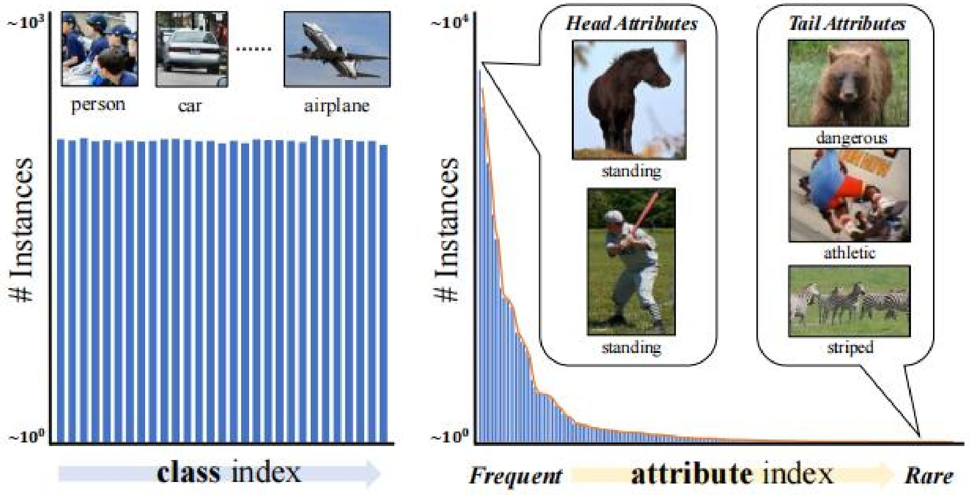
liuzhenguang2008@gmail.com, hanwangzhang@ntu.edu.sg

eccv2022

Introduce



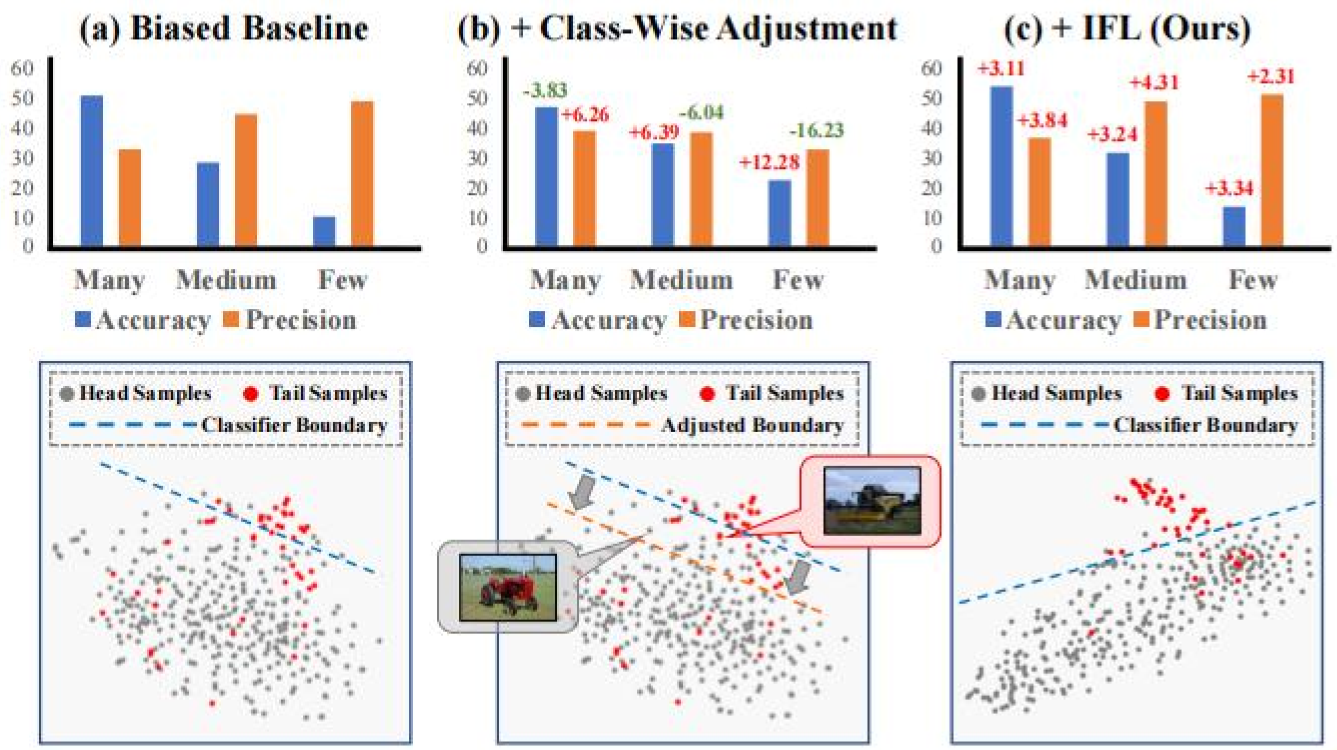
(a) Long-Tailed Distribution in Real-world Images



(b) Class-wise Balanced Data and its Imbalanced Attribute Distribution

Generalized Long-Tailed classification(GLT): class-wise imbalance + attribute-wise imbalance

Introduce



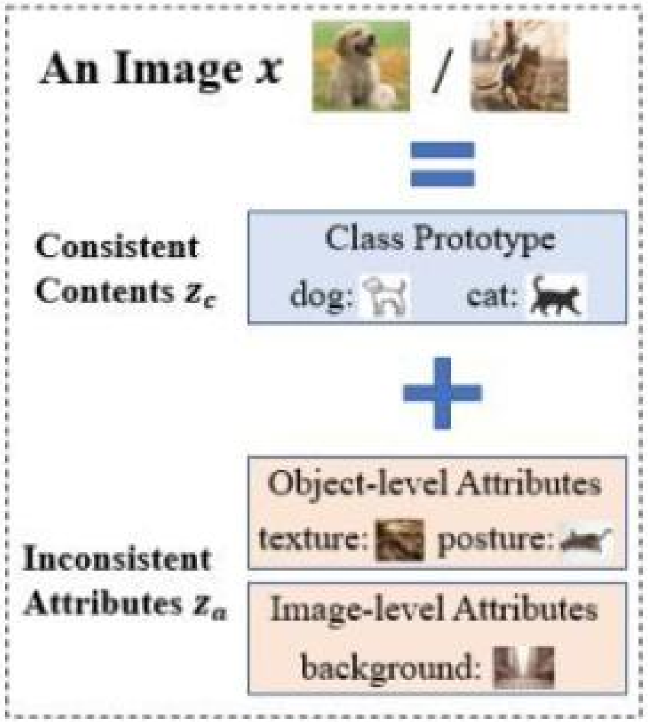
existing LT methods fail to tackle the attribute-wise imbalance:

- They rely on class-wise adjustment, while the attribute-wise traits are hidden in GLT;
- They are based on lifting the tail class boundary to welcome more samples to increase the tail accuracy, leaving the confused region of similar attributes unchanged in the feature space.

Introduce

Problem Formulation

Previous Assumption: $P(y \mid x) = \frac{P(x|y)}{P(x)} P(y) \propto P(x \mid y) P(y)$ $\quad p_{\text{train}}(X|Y) = p_{\text{test}}(X|Y)$



$$X = (z_c, z_a)$$

New Assumption:

$$p_{\text{train}}(z_c|Y) = p_{\text{test}}(z_c|Y)$$

Problem Formulation:

$$\begin{aligned} p(Y = k|X = x) &= p(Y = k|z_c, z_a) \\ &= \frac{p(z_c, z_a|Y = k)}{p(z_c, z_a)} \cdot p(Y = k) \\ &= \frac{p(z_c|Y = k)}{p(z_c)} \cdot \underbrace{\frac{p(z_a|Y = k, z_c)}{p(z_a|z_c)}}_{attribute\ bias} \cdot \underbrace{p(Y = k)}_{class\ bias}, \end{aligned}$$

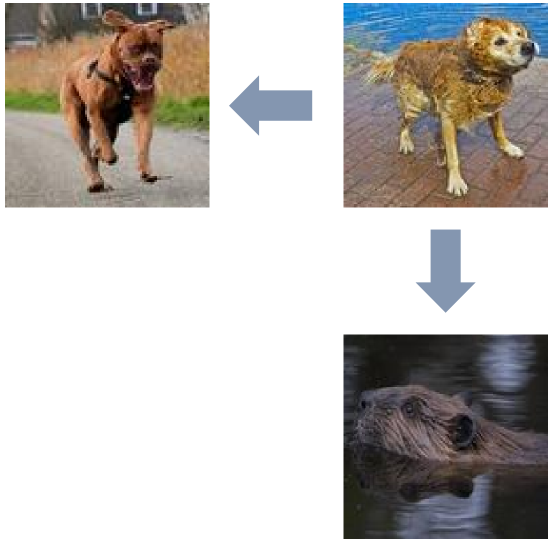
Introduce

Attribute Bias

- inconsistent performances within each class

$$\frac{p(z_a = wet \mid Y = dog, z_c = fur)}{p(z_a = wet \mid z_c = fur)} < \frac{p(z_a = fluffy \mid Y = dog, z_c = fur)}{p(z_a = fluffy \mid z_c = fur)}$$

$$\Rightarrow \quad p(Y = dog \mid z_c = fur, z_a = wet) < p(Y = dog \mid z_c = fur, z_a = fluffy)$$



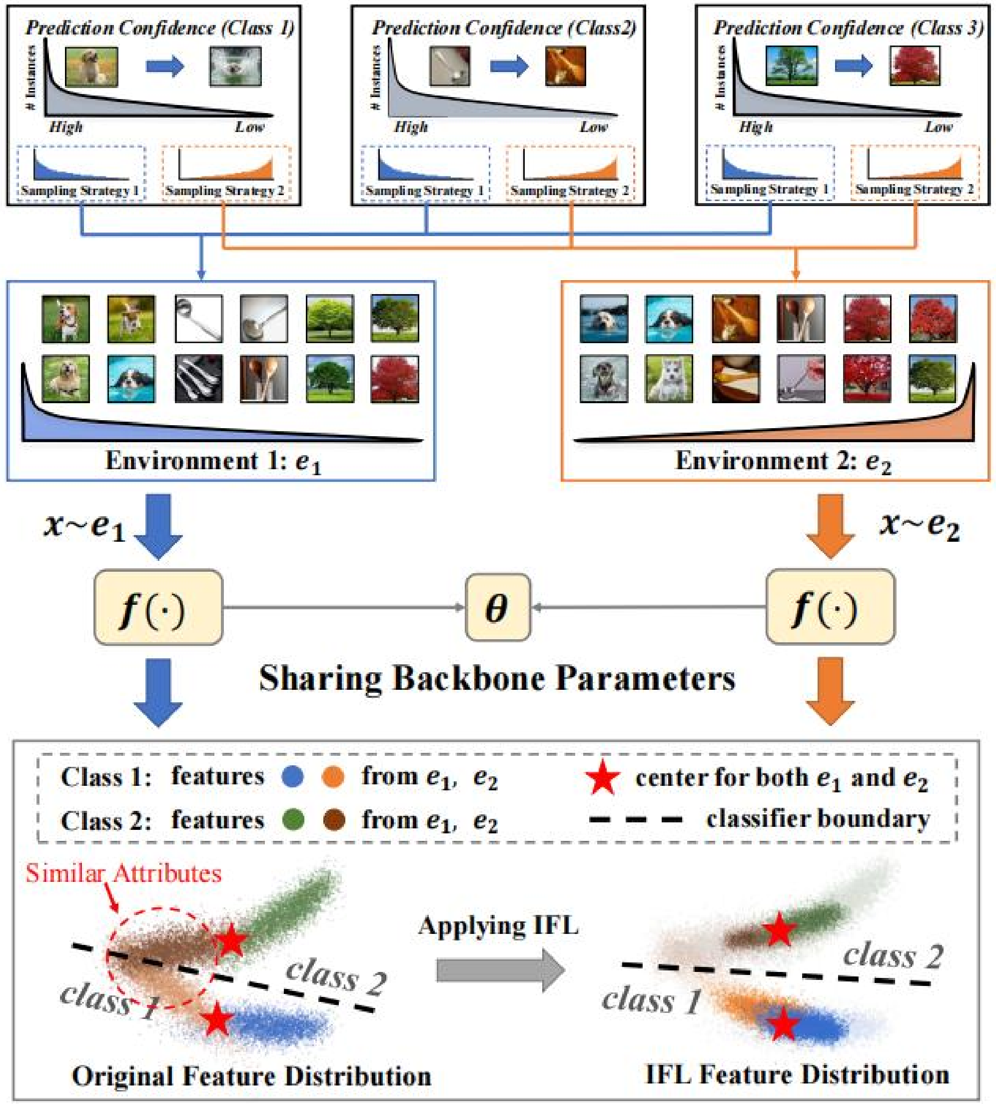
- spurious correlations

$$\frac{p(z_a = wet \mid Y = beave, z_c = fur)}{p(z_a = wet \mid z_c = fur)} >> 1$$

$$\Rightarrow \quad p(Y = beaver \mid z_c = fur, z_a = wet) \uparrow$$

Class Bias
Attribute Bias
$\}$ Generalized Long-Tailed classification(GLT)

Method



IRM(Invariant Risk Minimization)

Definition 3. *We say that a data representation $\Phi : \mathcal{X} \to \mathcal{H}$ elicits an invariant predictor $w \circ \Phi$ across environments $\mathcal{E}$ if there is a classifier $w : \mathcal{H} \to \mathcal{Y}$ simultaneously optimal for all environments, that is, $w \in \arg\min_{\bar{w}:\mathcal{H}\to\mathcal{Y}} R^e(\bar{w} \circ \Phi)$ for all $e \in \mathcal{E}$.*

- environment construction

➡ we use the current classification confidence of each training sample as an imbalance indicator of attributes inside the class

$$(1 - p(Y = k|z_c, z_a))^\beta$$

- optimization problem

➡

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

$$\text{subject to } \theta \in \arg\min_{\theta} \sum_{e\in\mathcal{E}} \sum_{i\in e} ||f(x_i^e;\theta) - C_{y_i^e}||_2$$

$$L = L_{cls} + \alpha \cdot L_{IFL}, \text{ where } L_{IFL} = ||f(x_i^e;\theta) - C_{y_i^e}||_2$$

Experiments

- **Class-wise Long Tail (CLT) Protocol**

  Train-GLT: class-wise LT and attribute-wise LT

  Test-CBL: class-wise balanced and attribute-wise LT

- **Attribute-wise Long Tail (ALT) Protocol**

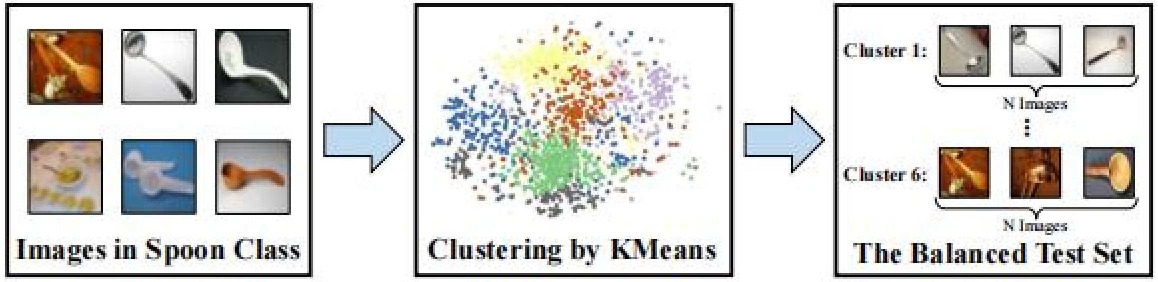
  Train-CBL: class-wise balanced and attribute-wise LT

  Test-GBL: class-wise balanced and attribute-wise balanced

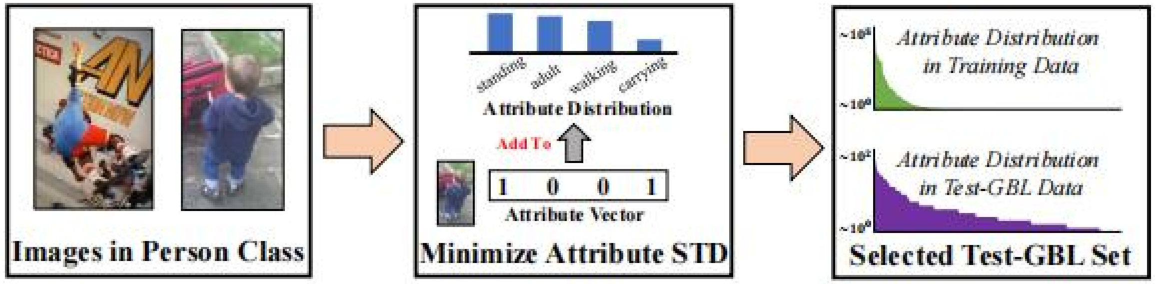
- **Generalized Long Tail (GLT) Protocol**

  Train-GLT: class-wise LT and attribute-wise LT

  Test-GBL: class-wise balanced and attribute-wise balanced



(a) Collecting an "Attribute-Wise Balanced" Test Set for ImageNet



(b) Balancing Attribute Distribution for MSCOCO-Attribute

Experiments

Table 1: **Evaluation of CLT and GLT Protocols on ImageNet-GLT**: Accuracy (*left in each cell*) and Precision (*right in each cell*) are reported. All methods are re-implemented under the same codebase with ResNext-50 backbone

	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 [22]	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 [22]	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 [55]	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 [46]	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 [37]	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 [78]	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 [7]	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 [73]	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 [11]	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 [74]	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 [61]	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**

Experiments

Table 2: **Evaluation of ALT Protocol on ImageNet-GLT**

	Methods < **Accuracy** | **Precision** >	Attribute-Wise Long Tail (ALT) Protocol Many$_A$	Medium$_A$	Few$_A$	Overall
Re-balance	Baseline	56.95 | 55.83	40.11 | 39.17	28.12 | 28.16	41.73 | 41.74
	cRT [22]	57.45 | 56.28	39.72 | 38.65	27.58 | 27.35	41.59 | 41.43
	LWS [22]	56.95 | 55.85	40.11 | 39.30	28.03 | 27.98	41.70 | 41.71
	Deconfound-TDE [55]	57.10 | 56.58	39.80 | 40.08	27.29 | 27.96	41.40 | 42.36
	BLSoftmax [46]	56.48 | 55.56	39.81 | 38.96	27.64 | 27.60	41.32 | 41.37
	BBN [78]	60.90 | 60.17	41.08 | 40.81	27.79 | 28.26	43.26 | 43.86
	LDAM [7]	59.04 | 56.51	40.96 | 39.21	27.96 | 27.22	42.66 | 41.80
	(ours) Baseline + IFL	**61.38** | **60.78**	**44.79** | **44.21**	**31.49** | **31.98**	**45.89** | **46.42**
	(ours) cRT + IFL	61.12 | 60.25	44.26 | 43.65	31.02 | 31.31	45.47 | 45.81
	(ours) LWS + IFL	61.19 | 60.45	44.66 | 44.07	31.43 | 31.91	45.76 | 46.25
	(ours) BLSoftmax + IFL	60.19 | 59.46	43.54 | 43.14	30.85 | 31.46	44.86 | 45.43
Augment	Mixup [73]	58.71 | 58.04	40.09 | 38.99	27.52 | 27.54	42.11 | 42.42
	RandAug [11]	62.35 | 61.25	45.04 | 44.27	31.47 | 31.26	46.29 | 46.32
	(ours) Mixup + IFL	65.90 | 65.88	49.43 | 49.43	35.40 | 35.89	50.24 | 51.04
	(ours) RandAug + IFL	**67.39** | **66.81**	**51.55** | **51.28**	**37.47** | **37.97**	**52.14** | **52.74**
Ensemble	TADE [74]	62.63 | 61.91	45.84 | 45.21	32.82 | 32.82	47.10 | 47.32
	RIDE [61]	63.48 | 61.42	45.62 | 44.16	32.59 | 32.26	47.24 | 46.67
	(ours) TADE + IFL	63.50 | 62.67	48.03 | 47.32	34.69 | 34.52	48.74 | 48.78
	(ours) RIDE + IFL	**67.54** | **67.13**	**51.92** | **51.72**	**37.84** | **38.46**	**52.44** | **53.17**

Table 3: **Evaluation on MSCOCO-GLT:** overall performances are reported

	Protocols < **Accuracy** | **Precision** >	CLT Overall	GLT Overall	ALT Overall
Re-balance	Baseline	72.34 | 76.61	63.79 | 70.52	50.17 | 50.94
	cRT [22]	73.64 | 75.84	64.69 | 68.33	49.97 | 50.37
	LWS [22]	72.60 | 75.66	63.60 | 68.81	50.14 | 50.61
	Deconfound-TDE [55]	73.79 | 74.90	66.07 | 68.20	50.76 | 51.68
	BLSoftmax [46]	72.64 | 75.25	64.07 | 68.59	49.72 | 50.65
	Logit-Adj [37]	75.50 | 76.88	66.17 | 68.35	50.17 | 50.94
	BBN [78]	73.69 | 77.35	64.48 | 70.20	51.83 | 51.77
	LDAM [7]	75.57 | 77.70	67.26 | 70.70	**55.52** | **56.21**
	(ours) Baseline + IFL	74.31 | 78.90	65.31 | **72.24**	52.86 | 53.49
	(ours) cRT + IFL	76.21 | 79.11	66.90 | 71.34	52.07 | 52.85
	(ours) LWS + IFL	75.98 | **79.18**	66.55 | 71.49	52.07 | 52.90
	(ours) BLSoftmax + IFL	73.72 | 77.08	64.76 | 70.00	52.97 | 53.52
	(ours) Logit-Adj + IFL	**77.16** | 79.09	**67.53** | 70.18	52.86 | 53.49
Augment	Mixup [73]	74.22 | 78.61	64.45 | 71.13	48.90 | 49.53
	RandAug [11]	76.81 | 79.88	67.71 | 72.73	53.69 | 54.71
	(ours) Mixup + IFL	77.55 | **81.78**	**68.83** | **74.84**	53.79 | 54.60
	(ours) RandAug + IFL	**77.71** | 81.10	68.16 | 73.97	**56.62** | **57.12**
Ensemble	TADE [74]	76.22 | 78.84	66.98 | 71.22	54.93 | 55.48
	RIDE [61]	78.29 | 80.33	68.59 | 72.20	58.90 | 59.43
	(ours) TADE + IFL	76.53 | 79.15	67.38 | 72.42	56.76 | 57.43
	(ours) RIDE + IFL	**78.86** | **80.70**	**69.09** | **72.57**	**58.93** | **59.84**

Experiments

Table 4: Ablation Studies on ImageNet-GLT, where overall results are reported; BLS, Focal, and IFF are balanced softmax loss [46], focal loss [31], and learning from failure [39], respectively

Ablation Settings					Evaluation Protocols		
#Env	Loss	IFL	Augment	Backbone	CLT Protocol	GLT Protocol	ALT Protocol
1	CE	-	-	ResNext-50	42.52 | 47.92	34.75 | 40.65	41.73 | 41.74
1	Focal	-	-	ResNext-50	39.93 | 46.99	32.52 | 39.12	39.58 | 39.85
1	LFF	-	-	ResNext-50	41.07 | 45.79	33.84 | 38.46	40.14 | 40.58
1	CE	✔	-	ResNext-50	39.74 | 47.06	32.82 | 40.86	39.99 | 41.38
2	IRM	-	-	ResNext-50	43.70 | 48.06	36.03 | 40.61	44.47 | 44.60
2	CE	✔	-	ResNext-50	45.97 | 52.06	37.96 | 44.47	45.89 | 46.42
3	CE	✔	-	ResNext-50	46.06 | 52.81	38.32 | 45.55	45.95 | 46.43
2	BLS	✔	-	ResNext-50	48.34 | 50.39	40.08 | 43.48	44.86 | 45.43
2	CE	✔	Mixup	ResNext-50	51.43 | 57.44	43.00 | 49.25	50.24 | 51.04
2	CE	✔	RandAug	ResNext-50	**53.40 | 58.11**	**44.90 | 50.47**	**52.14 | 52.74**
1	CE	-	-	RIDE-50	46.14 | 52.98	38.25 | 45.80	46.32 | 46.56
2	CE	✔	-	RIDE-50	49.20 | 54.64	41.35 | 47.67	48.62 | 48.62
2	TADE	✔	-	RIDE-50	51.78 | 52.41	43.47 | 45.17	48.74 | 48.78
2	LDAM	✔	-	RIDE-50	53.93 | 54.76	45.64 | 47.14	52.44 | 53.17
2	LDAM	✔	Mixup	RIDE-50	56.48 | 57.67	47.54 | 49.86	53.25 | 54.27
2	LDAM	✔	RandAug	RIDE-50	**58.70 | 59.61**	**49.80 | 51.62**	**55.65 | 55.81**

Experiments

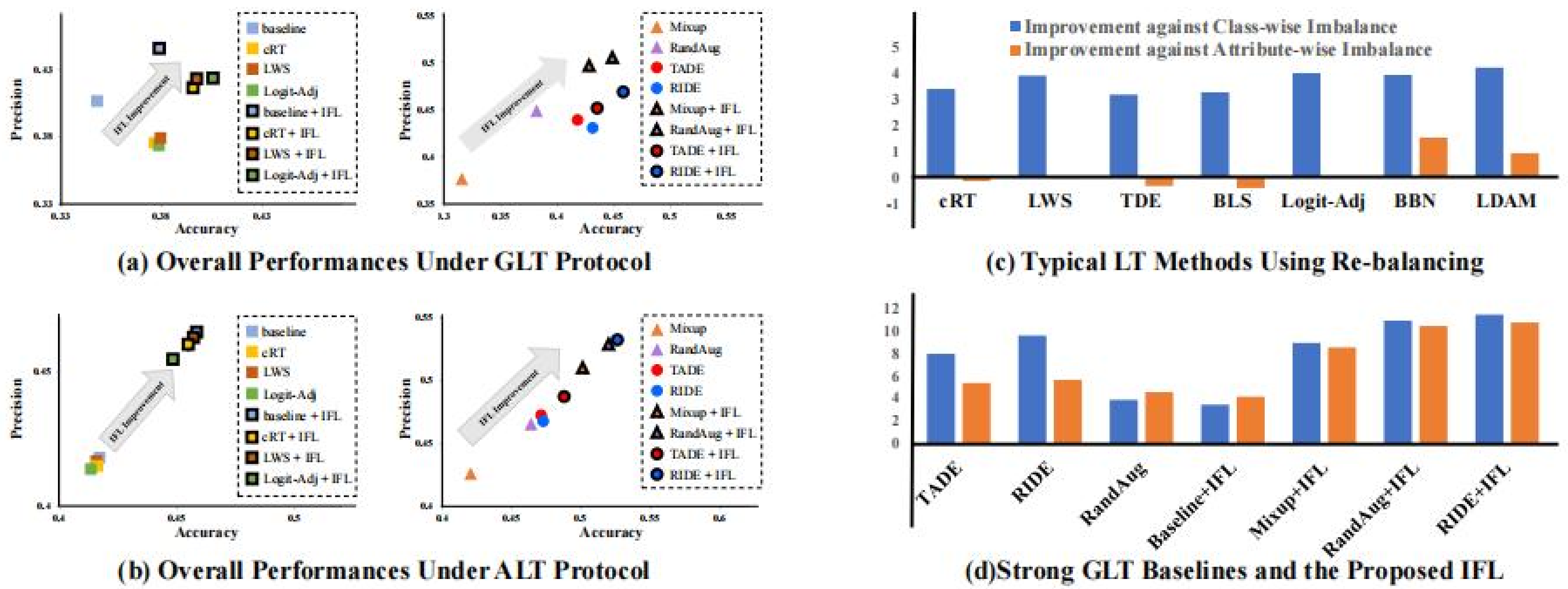


Figure 5: (a-b) The trending of precision and accuracy after applying the IFL; (c-d) GLT baselines will automatically improve class-wise LT, while conventional LT re-balancing algorithms won't improve the attribute-wise imbalance in GLT

Thank you