Global and Local Mixture Consistency Cumulative Learning for Long-tailed Visual Recognitions

Fei Du[1,2,3], Peng Yang[2,3], Qi Jia[1,3], Fengtao Nan[1,2,3], Xiaoting Chen[1,3], Yun Yang[1,3*]

[1]National Pilot School of Software, Yunnan University, Kunming, China

[2]School of Information Science and Engineering, Yunnan University, Kunming, China

[3]Yunnan Key Laboratory of Software Engineering

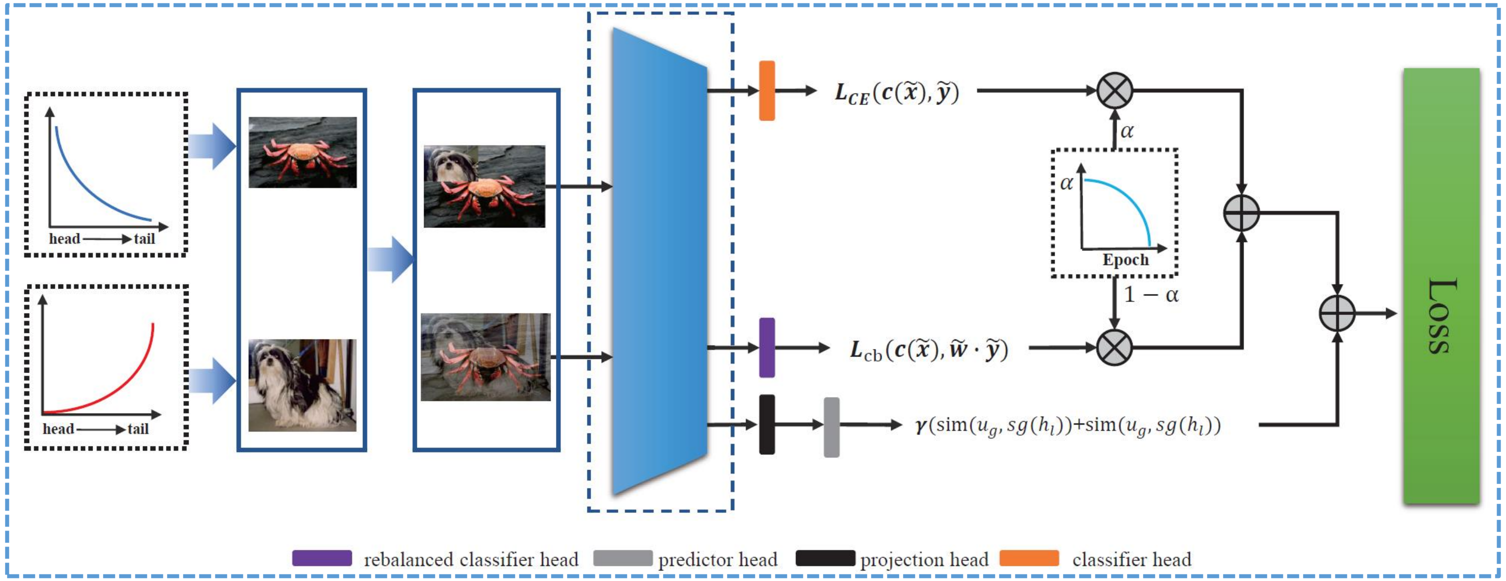
Paper tag : WED-PM-330

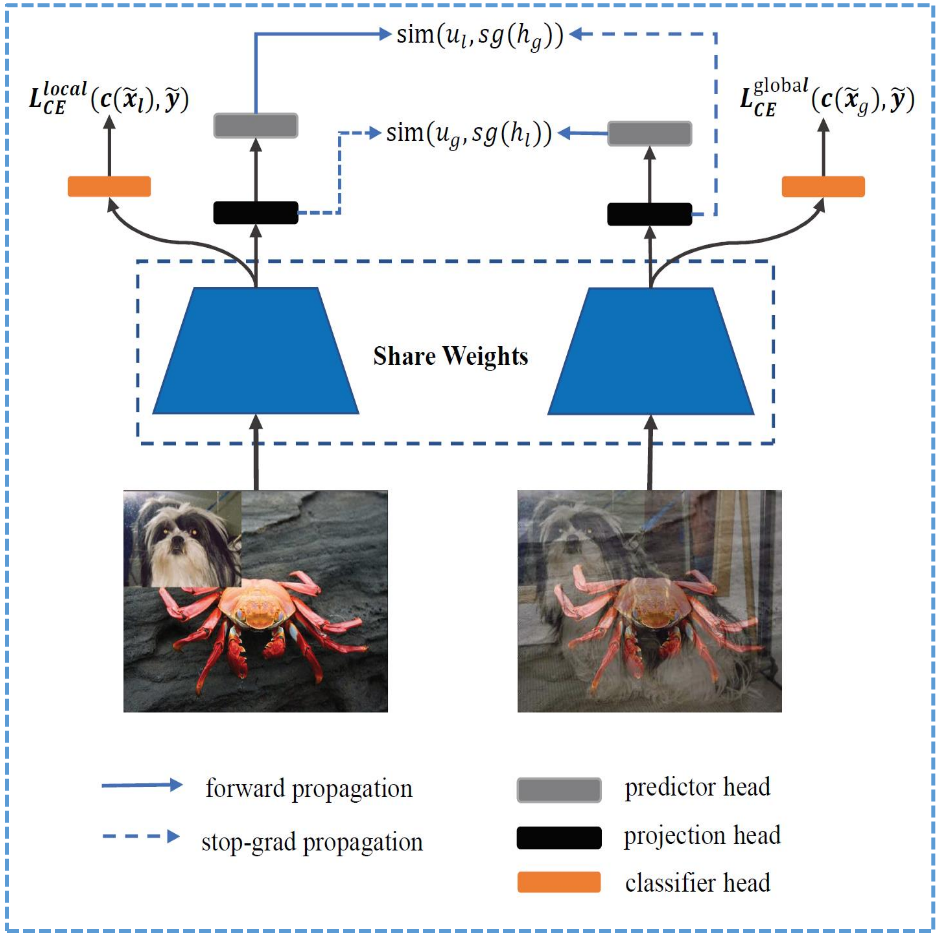
[Code Link: https://github.com/ynu-yangpeng/GLMC](https://github.com/ynu-yangpeng/GLMC)

Overview

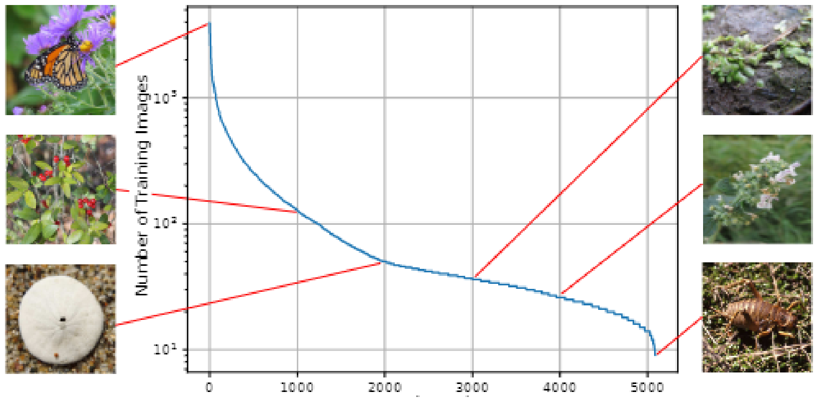
GLMC Core Idea:

- A Global and Local Mixture Consistency Loss improves the robustness of the feature extractor.
- A Cumulative Head-tail Soft Label Reweighted Loss mitigates the head class bias problem.

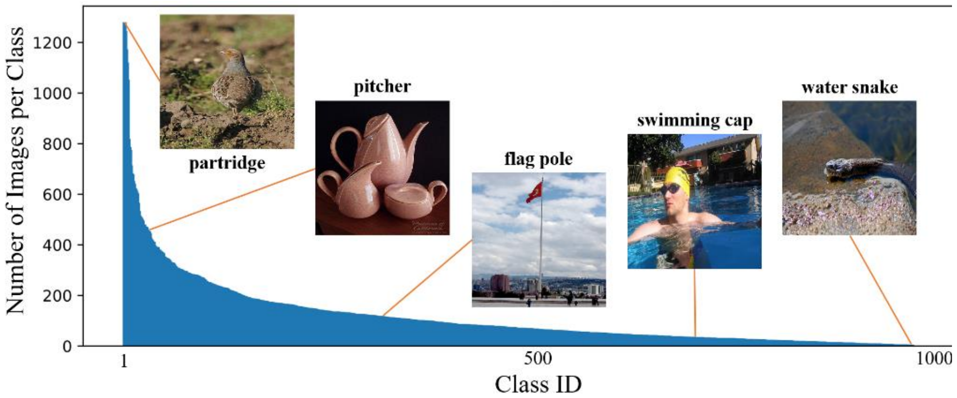




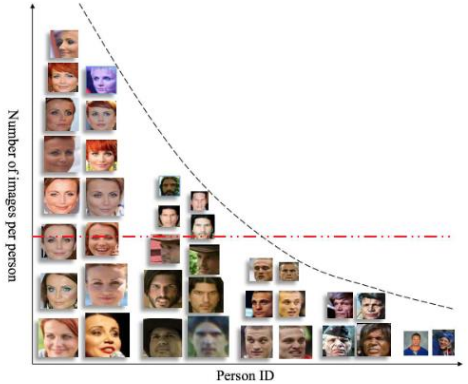
Real World Category Task Data Collection



iNaturalist2018 [1]



Objects [2]



Faces [3]

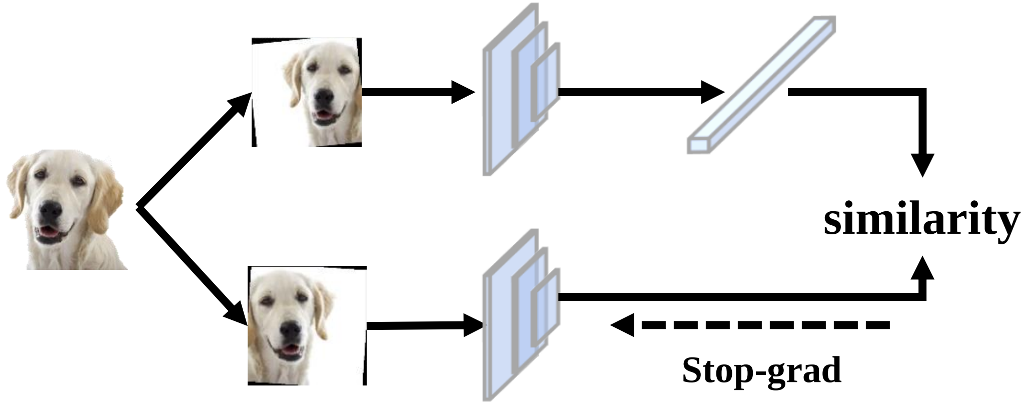
[1] Van Horn, G.; Mac Aodha, O.; Song, Y.; Cui, Y.; Sun, C.;Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. 2018.The inaturalist species classification and detection dataset.In Proceedings of the IEEE conference on computer vision and pattern recognition, 8769–8778.

[2] Liu, Z.; Miao, Z.; Zhan, X.; Wang, J.; Gong, B.; and Yu,S. X. 2019. Large-scale long-tailed recognition in an open world. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2537–2546.

[3] Zhang, X.; Fang, Z.; Wen, Y.; Li, Z.; and Qiao, Y. 2017.Range loss for deep face recognition with long-tailed training data. In Proceedings of the IEEE International Conference on Computer Vision, 5409–5418.

Related Work

➢ **Contrastive Representation Learning for long-tail recognition**



Target:

- To obtain a balanced representation space

Drawback:

- A multi-stage pipeline
- Large batches of negative examples for training
- Extensive training skills and memory overhead

➢ **Class Rebalance learning**

Target:

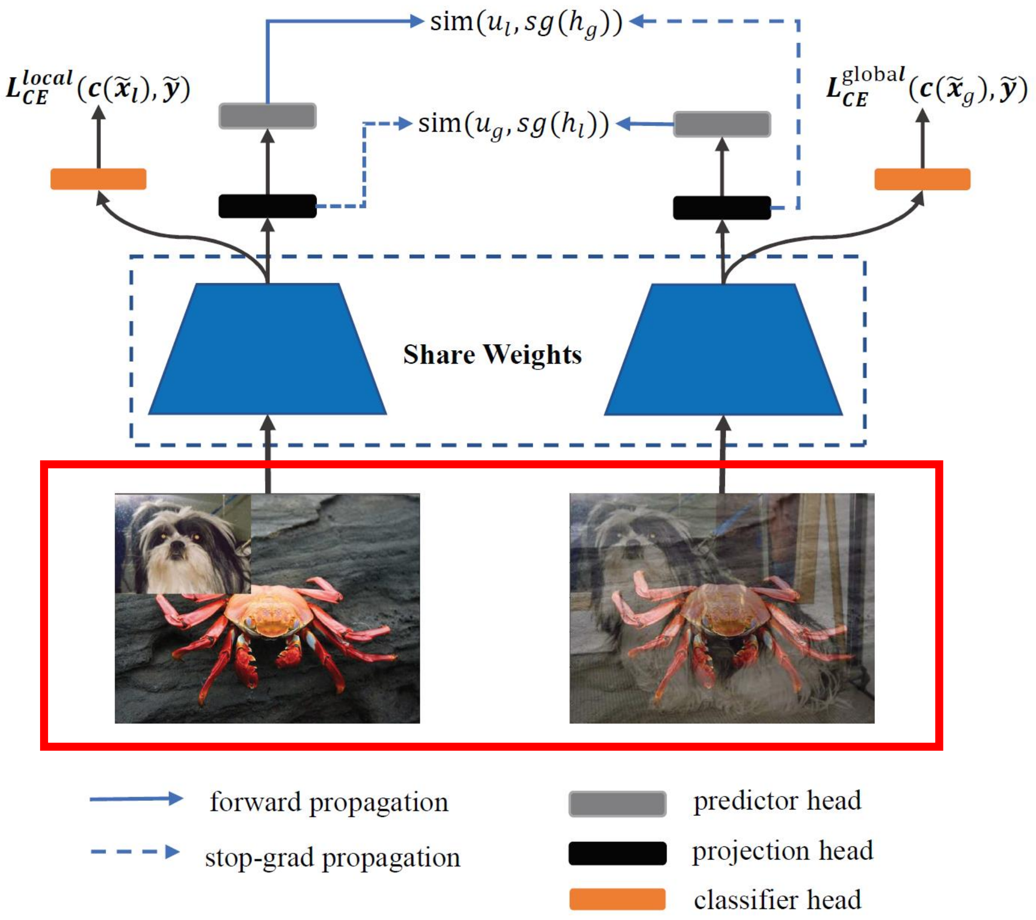
- To strengthen the tail class by oversampling or increasing weight.

Drawback:

- **over-learning** the tail class may increase the risk of **overfitting**
- **under-sampling** or reducing weight in the head class inevitably **sacrifice the performance of head classes**.

Global and Local Mixture Consistency Learning

- A stochastic mixed-label data augmentation module $\mathrm{Au}g(x,y)$. For each input batch samples, $\mathrm{Au}g(x,y)$ transforms x and their labels y in global and local augmentations pairs, respectively.



Global Mixture:

$\lambda \sim Beta(\beta, \beta)$

$\tilde{x}_g = \lambda x_i + (1-\lambda)x_j$

$\tilde{p}_g = \lambda p_i + (1-\lambda)p_j$

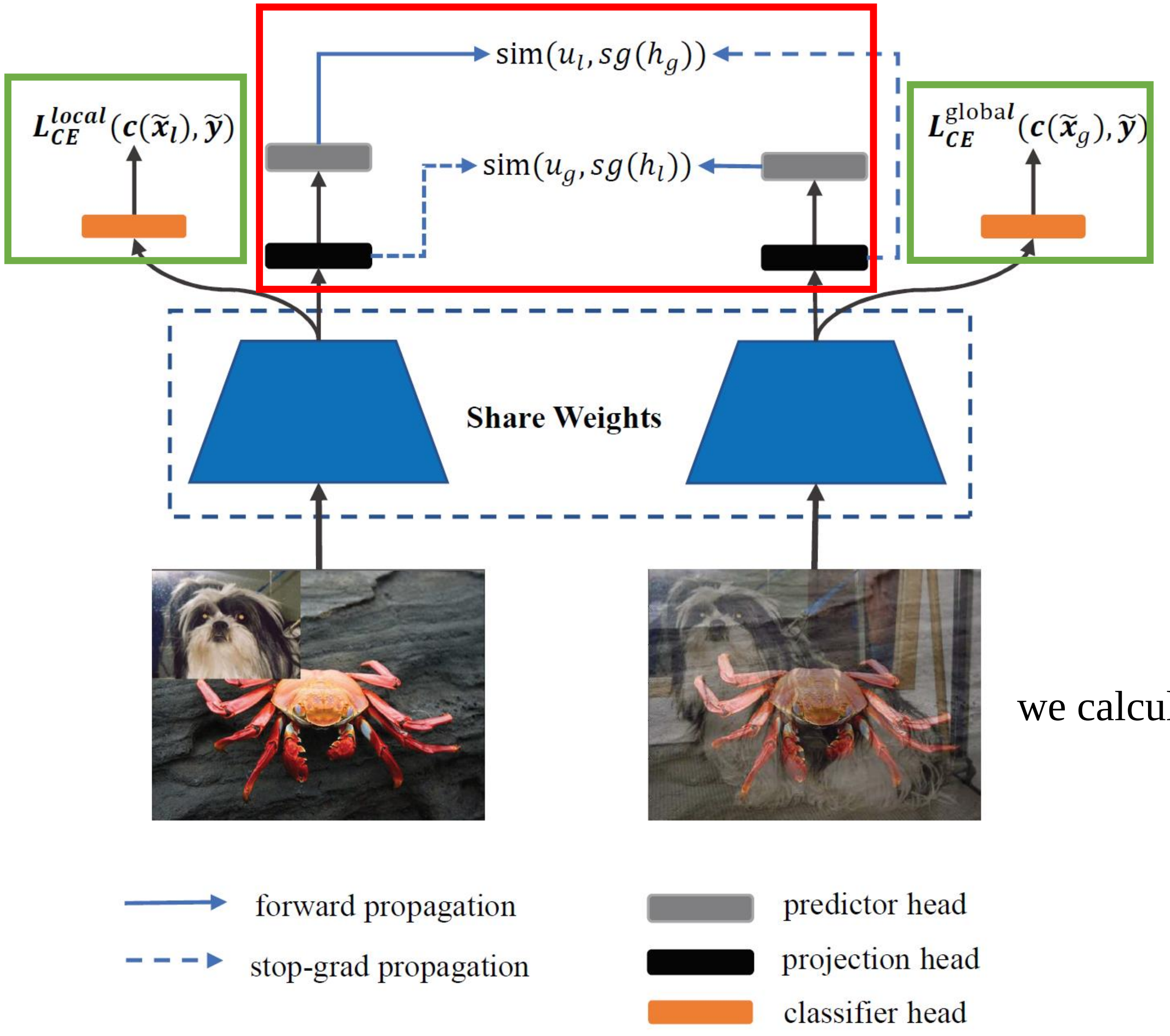
Local Mixture:

$\tilde{x}_l = M \odot x_i + (1-M) \odot x_j$

$r_x \sim Uniform(0, W), r_w = W\sqrt{1-\lambda}$

$r_y \sim Uniform(0, H), r_h = H\sqrt{1-\lambda}$

Global and Local Mixture Consistency Learning



$$sim(u_g, h_l) = -\frac{u_g}{\|u_g\|} \bullet \frac{h_l}{\|h_l\|}$$

$$sim(u_l, h_g) = -\frac{u_l}{\|u_l\|} \bullet \frac{h_g}{\|h_g\|}$$

$$\mathcal{L}_{sim} = sim(u_g, sg(h_l)) + sim(u_l, sg(h_g))$$

we calculate the mixed-label cross-entropy loss:

$$\mathcal{L}_c = -\frac{1}{2N}\sum_{i=1}^{N} \left(\tilde{p}_g^i \left(log f(\tilde{x}_g^i)\right) + \tilde{p}_l^i \left(log f(\tilde{x}_l^i)\right)\right)$$

Experiments

◆ Full ImageNet and CIFAR Recognition

Top-1 accuracy (%) on full CIFAR-10 and CIFAR-100 dataset with ResNet-50 backbone.

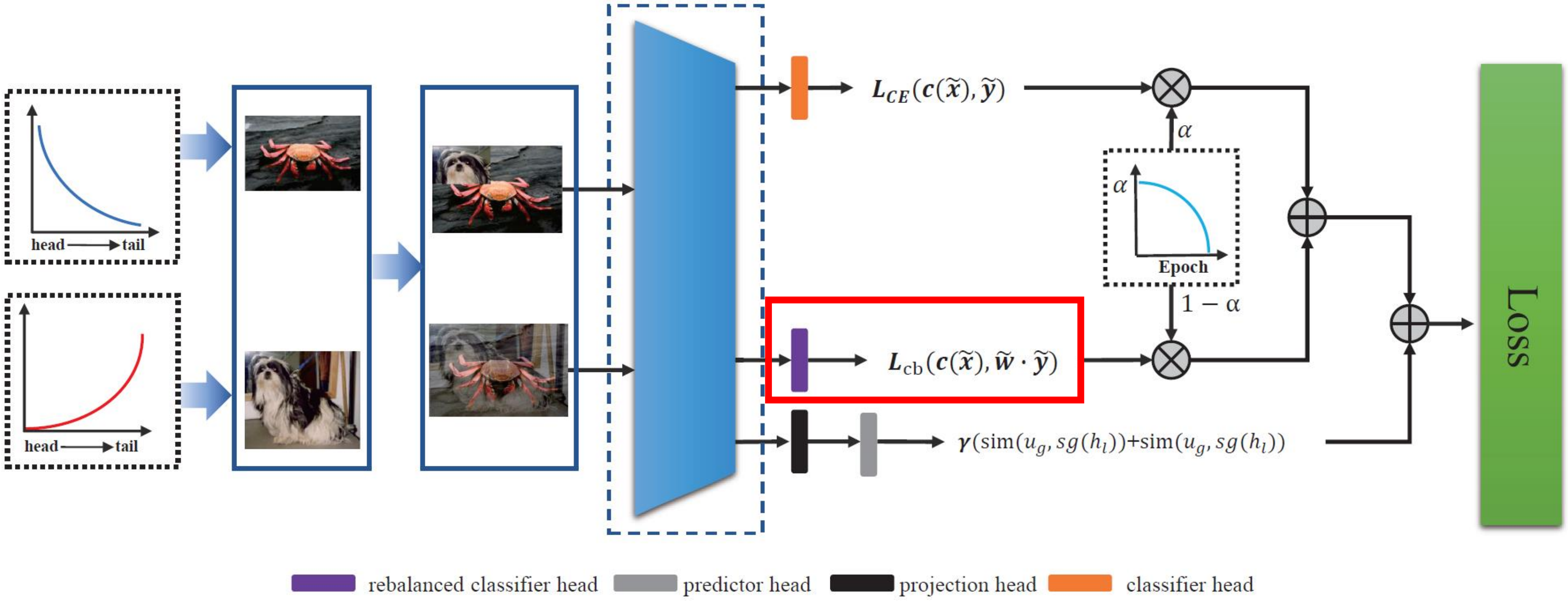
Method	CIFAR-10	CIFAR-100
vanilla	94.85	75.28
MixUp [41]	95.95	77.99
CutMix [40]	95.41	78.03
SupCon [21]	96	76.5
PaCo [8]	-	79.1
ours	**97.23**	**83.05**

Top-1 accuracy (%) on full ImageNet dataset with ResNet-50 backbone.

Method	Augmentation	Top-1 acc
vanilla	Simple Augment	76.4
vanilla	MixUp [41]	77.9
vanilla	CutMix [40]	78.6
Supcon [21]	RandAugment	78.4
PaCo [8]	Simple Augment	78.7
PaCo [8]	RandAugment	79.3
ours	MixUp + CutMix	**80.2**

Cumulative Class-Balanced Learning

- A linear rebalanced classifier head $cb(x)$ that maps vectors r to rebalanced category space. The rebalanced classifier calculates mixed cross entropy loss with the reweighted data distribution.

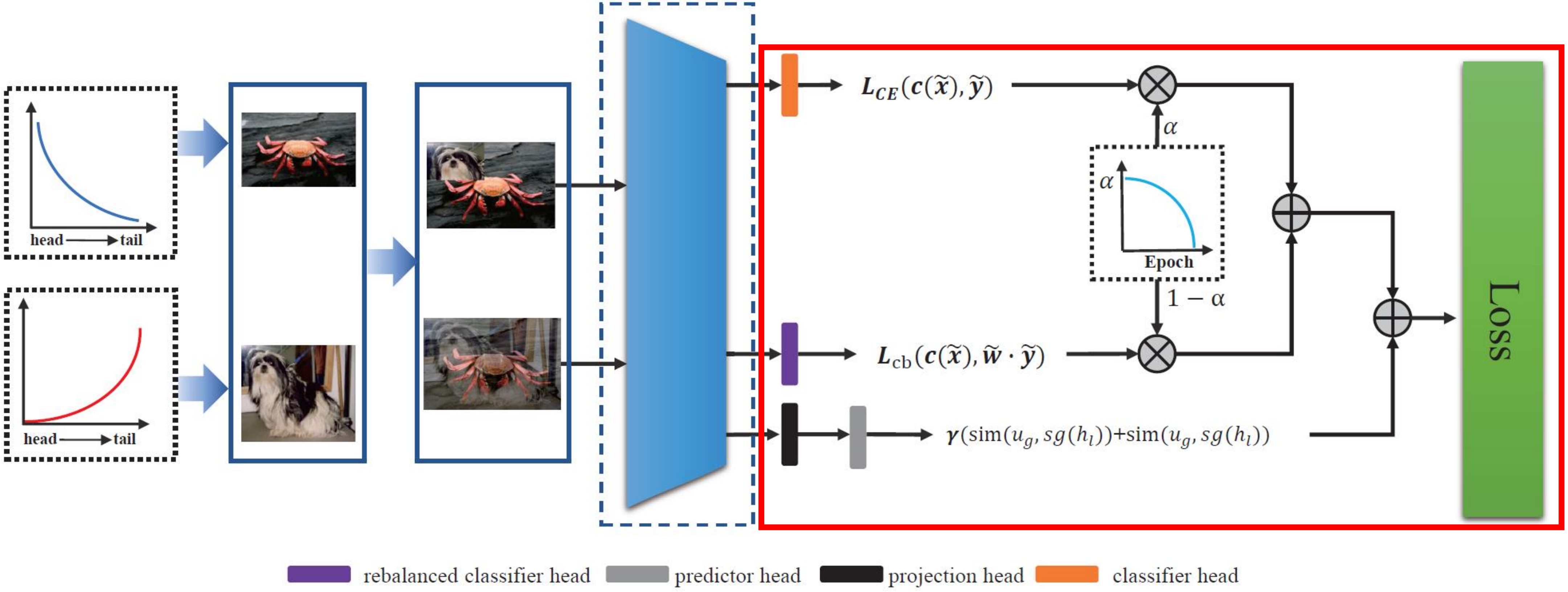


Weighting factor: $w_i = \dfrac{C \cdot (1/r_i)^k}{\sum_{i=1}^{C}(1/r_i)^k}$

Mixing weight vectors of the two images: $\tilde{w} = \lambda w_i + (1-\lambda) w_j$

$$\mathcal{L}_{\text{cb}} = -\frac{1}{2N}\sum_{i=1}^{N} \tilde{w}^i (\tilde{p}_g^i \left(logf\left(\tilde{x}_g^i\right)\right) + \tilde{p}_l^i \left(logf\left(\tilde{x}_l^i\right)\right))$$

Cumulative Class-Balanced Learning



$$\mathcal{L}_{total} = \alpha\mathcal{L}_c + (1-\alpha)\mathcal{L}_{cb} + \gamma\mathcal{L}_{sim}$$

$$\alpha = 1 - \left(\frac{T}{T_{max}}\right)^2$$

Experiments

Top-1 accuracy (%) of ResNet-32 on CIFAR-10-LT and CIFAR-100-LT with different imbalance factors [100, 50, 10]. GLMC consistently outperformed the previous best method only in the one-stage.

	Method	CIFAR-10-LT			CIFAR-100-LT		
		IF=100	50	10	100	50	10
	CE	70.4	74.8	86.4	38.3	43.9	55.7
rebalance classifier	BBN [45]	79.82	82.18	88.32	42.56	47.02	59.12
	CB-Focal [9]	74.6	79.3	87.1	39.6	45.2	58
	LogitAjust [29]	80.92	-	-	42.01	47.03	57.74
	weight balancing [1]	-	-	-	53.35	57.71	68.67
augmentation	Mixup [42]	73.06	77.82	87.1	39.54	54.99	58.02
	RISDA [6]	79.89	79.89	79.89	50.16	53.84	62.38
	CMO [32]	-	-	-	47.2	51.7	58.4
self-supervised pretraining	KCL [18]	77.6	81.7	88	42.8	46.3	57.6
	TSC [25]	79.7	82.9	88.7	42.8	46.3	57.6
	BCL [47]	84.32	87.24	91.12	51.93	56.59	64.87
	PaCo [8]	-	-	-	52	56	64.2
	SSD [26]	-	-	-	46	50.5	62.3
ensemble classifier	RIDE (3 experts) + CMO [32]	-	-	-	50	53	60.2
	RIDE (3 experts) [37]	-	-	-	48.6	51.4	59.8
one-stage training	**ours**	**87.75**	**90.18**	**94.04**	**55.88**	**61.08**	**70.74**
finetune classifier	**ours + MaxNorm [1]**	**87.57**	**90.22**	**94.03**	**57.11**	**62.32**	**72.33**

Experiments

Top-1 accuracy (%) on ImageNet-LT dataset. Comparison to the state-of-the-art methods with different backbone.
† denotes results reproduced by BCL with 180 epochs.

Method	Backbone	ImageNet-LT			
		Many	Med	Few	All
CE	ResNet-50	64	33.8	5.8	41.6
CB-Focal [9]	ResNet-50	39.6	32.7	16.8	33.2
LDAM [3]	ResNet-50	60.4	46.9	30.7	49.8
KCL [18]	ResNet-50	61.8	49.4	30.9	51.5
TSC [25]	ResNet-50	63.5	49.7	30.4	52.4
RISDA [6]	ResNet-50	-	-	-	49.3
BCL (90 epochs) [46]	ResNeXt-50	67.2	53.9	36.5	56.7
BCL (180 epochs) [46]	ResNeXt-50	67.9	54.2	36.6	57.1
PaCo† (180 epochs) [8]	ResNeXt-50	64.4	55.7	33.7	56.0
Balanced Softmax† (180 epochs) [34]	ResNeXt-50	65.8	53.2	34.1	55.4
SSD [26]	ResNeXt-50	66.8	53.1	35.4	56
RIDE (3 experts) + CMO [32]	ResNet-50	66.4	53.9	35.6	56.2
RIDE (3 experts) [37]	Swin-S	66.9	52.8	37.4	56
weight balancing + MaxNorm [1]	ResNeXt-50	62.5	50.4	41.5	53.9
ours	ResNeXt-50	**70.1**	52.4	30.4	56.3
ours + MaxNorm [1]	ResNeXt-50	60.8	**55.9**	**45.5**	56.7
ours + BS [34]	ResNeXt-50	64.76	55.67	42.19	**57.21**

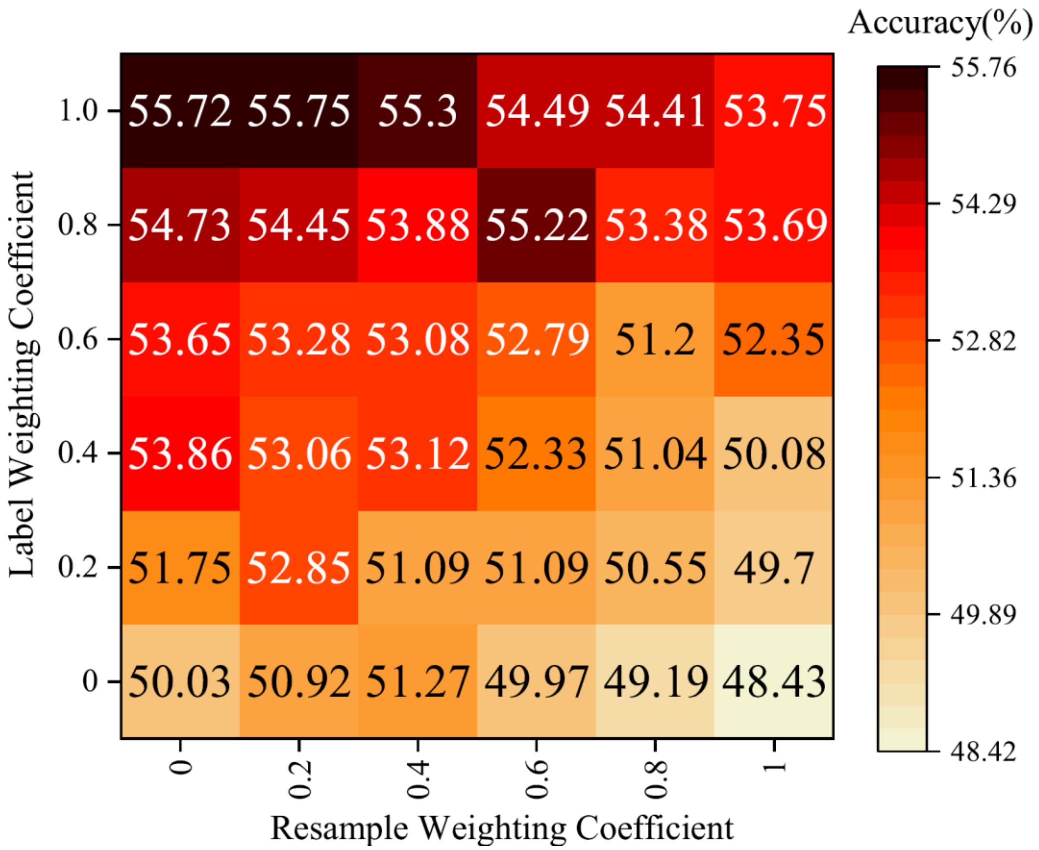
Ablation studies

Ablations of the different key components of GLMC architecture. We report the accuracies (%) on CIFAR100-LT (IF=100) with ResNet-32 backbone. Note that all model use one stage training.

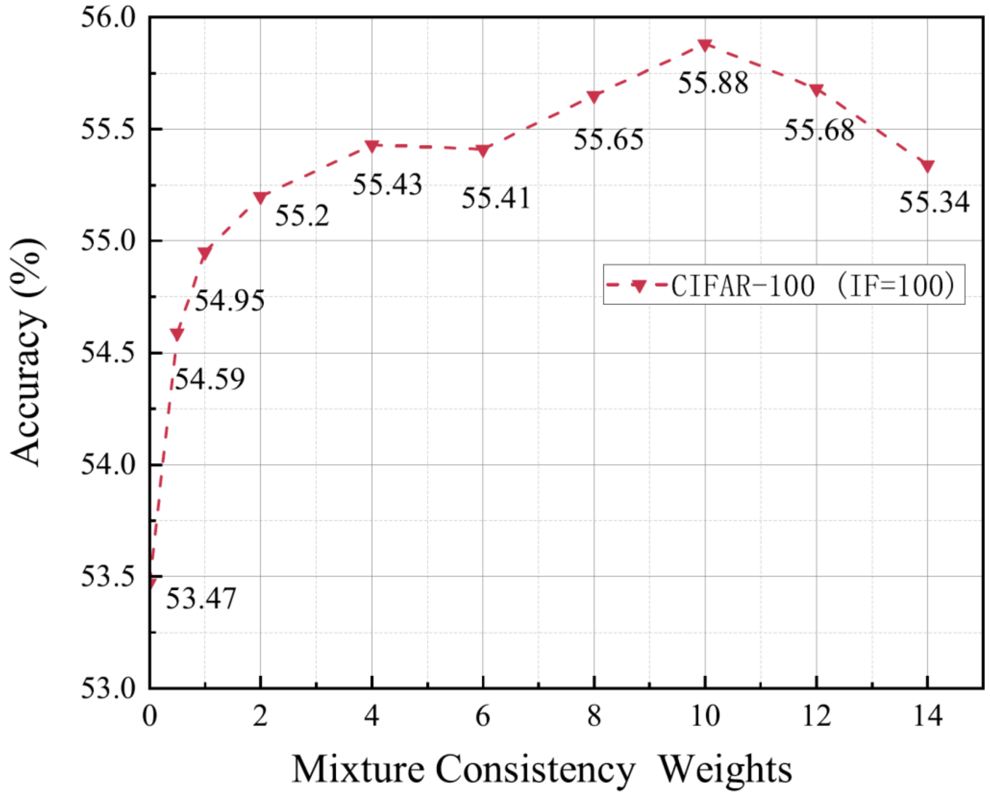
Global and Local Mixture Consistency	Cumulative Class-Balanced	Accuracies(%)
×	×	38.3
×	✓	44.63
✓	×	50.11
✓	✓	55.88

Ablation studies

Confusion matrices of different label reweighting and resample coefficient k on CIFAR-100-LT with an imbalance ratio of 100.



Different global and local mixture consistency weights on CIFAR-100-LT (IF = 100).



Thanks for listening!

https://github.com/ynu-yangpeng/GLMC

云南省软件工程重点实验室
Yunnan Key Laboratory of Software Engineering