Temporal Imbalance of Positive and Negative Supervision in Class-Incremental Learning

Jinge Ma, Fengqing Zhu

Purdue University, West Lafayette, Indiana, U.S.A.

类增量学习 Class-Incremental Learning， CIL

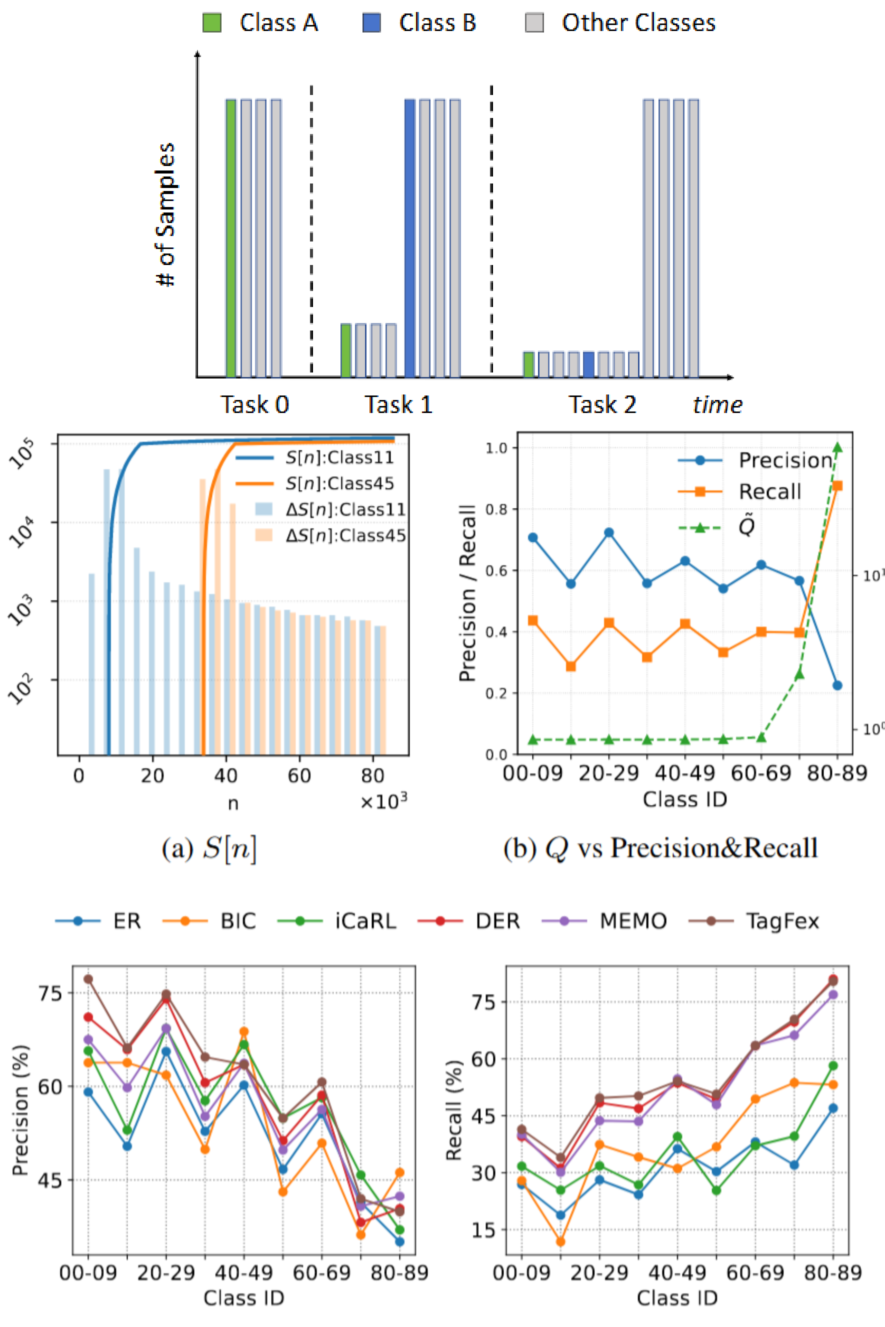
- 任务序列$D = D_1, \ldots, D_T$
- 标签空间$\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset, \mathrm{i} \neq \mathrm{j}$
- 时间步 t
  - 在D_t中训练
  - 在$\cup_{i=1}^{t}(D_i)$ 中测试

类别不平衡问题

- 生成旧类别样本，旧类别特征
  - http://arxiv.org/abs/2511.17973
  - http://arxiv.org/abs/2505.23744
- 修正分类头
  - http://arxiv.org/abs/2506.15720
- 特征原型相似度分类器
  - https://arxiv.org/abs/2303.07338v2

时间不平衡问题

- 在Replay-base的方法中，每个类别存储相同数量的样本
- 在类别平衡的情况下，模型偏向于靠后的类别
  - 观察不同类别的PR变化曲线
- 类似指数平滑、指数衰减



(c) Precision/Recall asymmetry in various baselines

时间不平衡问题

- 损失函数中的正、负监督

$$\ell_{\mathrm{CE}}(y,z) = -\log \frac{e^{z_y}}{e^{z_y} + \sum_{k\neq y} e^{z_k}} = \underbrace{\log\left(e^{z_y} + \sum_{k\neq y} e^{z_k}\right)}_{\text{negative supervision}} - \underbrace{z_y}_{\text{positive supervision}}$$

- 早期类别受到负监督的频次高于后期类别
- 降低负监督的强度

$$\ell_{\mathrm{TAL}}(y,z,Q[N]) = -\log\left(\frac{e^{z_y}}{e^{z_y} + \alpha \sum_{k\neq y} w(Q_k[N])\, e^{z_k}}\right)$$

$$\text{where } w(Q_k[N]) = \left(\frac{Q_k[N]}{Q_{\max}}\right)^r$$

方法 Temporal-Adjusted Loss ℓ_{TAL}

- 每个已见类别有一个监督极性序列a_k
- 对于第N次迭代的样本(x, y)
  - 记录监督极性序列$a_{\mathrm{k}}[\mathrm{N}-1] = \begin{cases} +1, \mathrm{k} = \mathrm{y} \\ -1, \mathrm{k} \neq \mathrm{y} \end{cases}$
  - 令正监督强度$Q_k[N] = \sum_{n=0}^{N-1}(f[N-1-n] \cdot a_k[n])$，$f$为衰减函数
- 取$f[n] = \lambda^{n+1}$为指数衰减函数
- 此时正监督强度更新是马尔可夫过程

$$Q_k[N+1] = \lambda(Q_k[N] + a_k[N])$$

方法 Temporal-Adjusted Loss $\boldsymbol{\ell_{TAL}}$

- $Q_k[N+1] = \lambda(Q_k[N] + a_k[N])$存在负数
- 修正为全正数形式进行更新

$$Q_k[N+1] = \lambda(Q_k[N] + \begin{cases} a_k[N], & a_k[N] = +1 \\ w(Q_k[N])a_k[N], & a_k[N] = -1 \end{cases}$$

$$w(Q_k[N]) = \left(\frac{Q_k[N]}{Q_{\max}}\right)^r$$

$$Q_{max} = \lambda/(1-\lambda)$$

- 利用Q_k[N]得到最终损失

$$\ell_{TAL} = -log\frac{e^{z_y}}{e^{z_y} + \alpha\sum_{k\neq y} w(Q_k[N])e^{z_k}}$$

- 其中α为修正常数，使时间、类别平衡时，损失退化为CE Loss

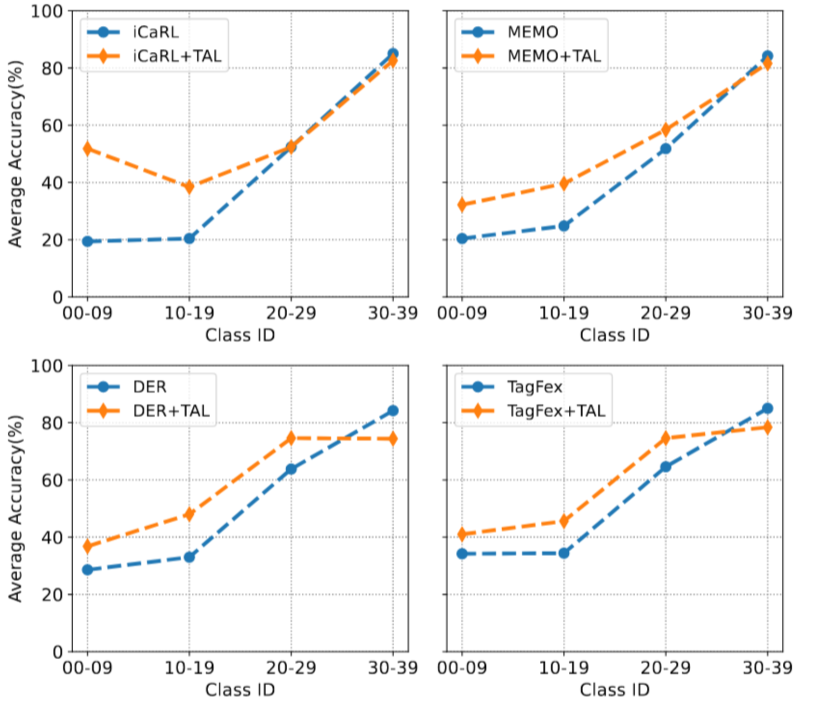
实验 ResNet+CIFAR100/IN100/Food101

- $A_{Mean} = \frac{1}{T}\sum_{t=1}^{T}(Acc_t), A_{Last} = Acc_T$
- 每个任务类别数量相等

Method	CIFAR-100				ImageNet-100				Food101	
	10-task		20-task		10-task		20-task		10-task	
	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}
iCaRL [29]	58.76±0.12	45.39±0.22	55.98±0.14	43.20±0.24	43.71±0.12	24.38±0.24	33.45±0.14	17.62±0.28	63.46±0.12	50.13±0.22
+ Ours	**60.82**±0.10	**47.36**±0.18	**58.68**±0.12	**44.79**±0.20	**52.19**±0.10	**32.78**±0.20	**42.54**±0.12	**23.90**±0.24	**68.57**±0.10	**56.03**±0.18
FOSTER [44]	62.04±0.42	49.30±0.64	56.99±0.46	47.34±0.66	48.39±0.46	20.88±0.70	35.67±0.48	15.28±0.72	63.88±0.46	52.60±0.66
+ Ours	**64.80**±0.40	**52.76**±0.60	**58.21**±0.42	**49.28**±0.64	**51.49**±0.42	**23.38**±0.66	**38.07**±0.46	**19.72**±0.70	**64.61**±0.42	**53.90**±0.60
MEMO [53]	61.68±0.16	51.70±0.30	56.98±0.18	48.93±0.34	47.01±0.16	34.98±0.34	34.73±0.18	22.78±0.36	62.25±0.16	54.75±0.30
+ Ours	**63.81**±0.12	**52.85**±0.28	**58.04**±0.16	**49.86**±0.30	**51.05**±0.12	**37.28**±0.30	**38.76**±0.16	**26.16**±0.34	**64.73**±0.12	**55.83**±0.28
DER [50]	63.53±0.19	50.75±0.35	57.47±0.17	50.34±0.31	52.25±0.25	40.28±0.49	43.95±0.19	25.82±0.35	69.20±0.23	61.42±0.47
+ Ours	**66.33**±0.17	**53.82**±0.31	**60.34**±0.13	**53.49**±0.29	**54.57**±0.23	**42.62**±0.43	**47.80**±0.17	**29.79**±0.31	**71.58**±0.19	**63.98**±0.41
TagFex [52]	65.97±0.20	55.99±0.41	59.55±0.20	51.48±0.41	54.73±0.35	41.70±0.59	47.32±0.23	31.36±0.35	71.55±0.31	66.67±0.53
+ Ours	**68.68**±0.18	**57.91**±0.37	**62.34**±0.18	**54.70**±0.37	**57.05**±0.31	**43.01**±0.55	**51.08**±0.19	**35.78**±0.31	**74.57**±0.29	**68.64**±0.49

实验

- IN100
- 早期类别准确率均上升
- 特征提取器对早期类别样本的特征也能更好的区分



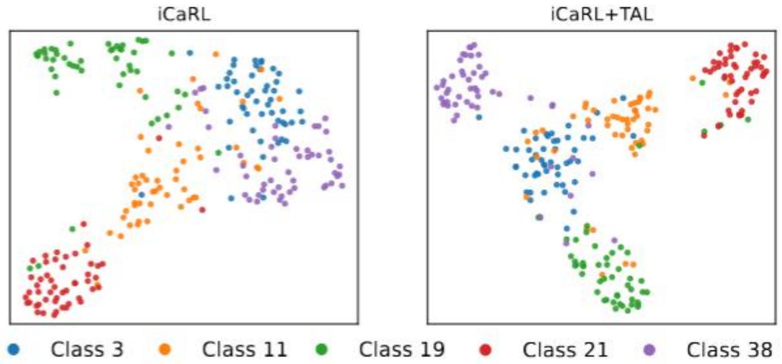


Figure 5. This figure illustrates the impact of applying TAL on the feature space of iCaRL. We visualize the distribution of test images from several new and old classes in the ResNet18 feature space on ImageNet100 at Task ID = 3.

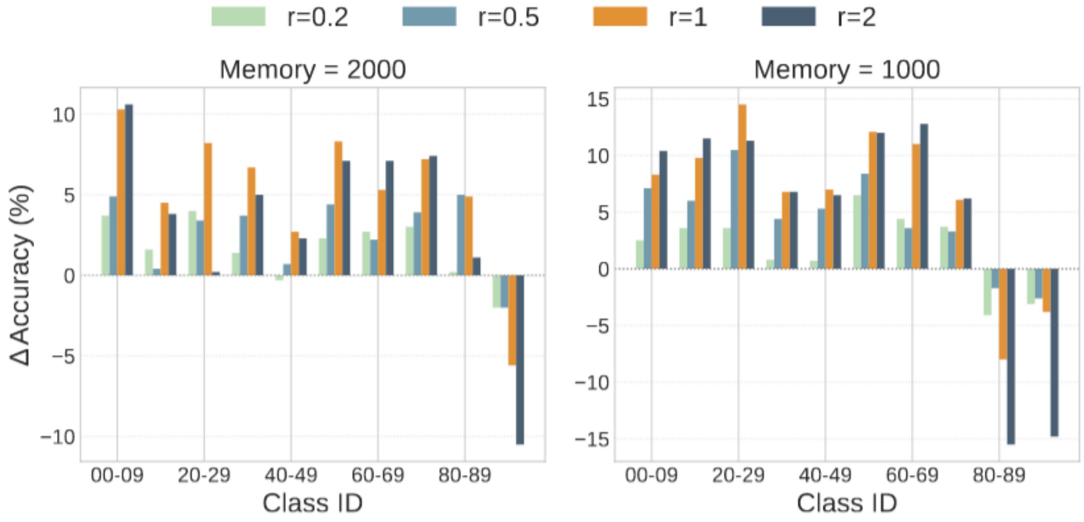
Ablation 超参分析

$$\ell_{TAL} = -log\frac{e^{z_y}}{e^{z_y} + \alpha\sum_{k\neq y} w(Q_k[N])e^{z_k}}$$

$$w(Q_k[N]) = \left(\frac{Q_k[N]}{Q_{max}}\right)^r$$

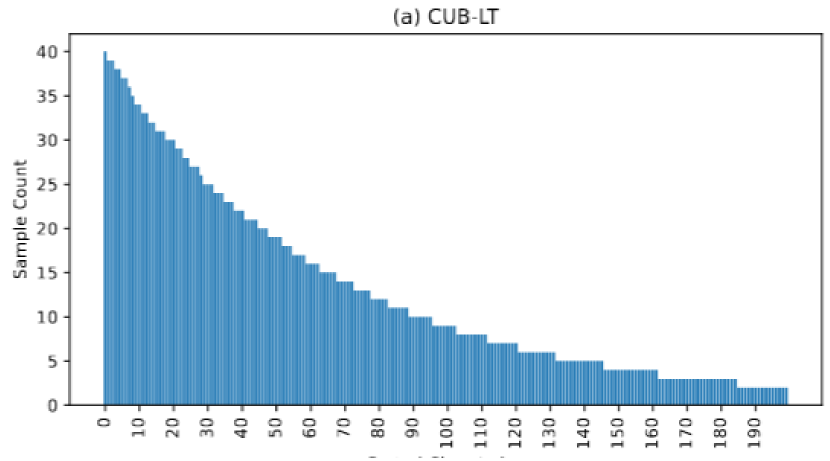
- r越大，对负监督的抑制越强，早期类别准确率更高，后期类别准确率越低。

r	λ							
	0.99		0.995		0.999		0.9995	
	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}	A_{Mean}	A_{Last}
0.2	61.30	44.39	61.06	44.05	60.69	43.54	60.26	43.18
0.5	62.12	45.36	61.27	45.05	61.72	44.50	61.36	44.28
1.0	**62.60**	**46.13**	<u>**63.36**</u>	<u>**47.64**</u>	62.46	**46.66**	61.79	45.44
2.0	62.51	45.38	60.24	45.80	**62.85**	46.16	**62.50**	**44.85**
5.0	56.70	36.90	57.15	37.97	58.62	38.83	58.08	38.94
CE	59.96	42.39	-	-	-	-	-	-



实验 预训练模型ViT+长尾数据集

- 每个任务有单独的可训练参数
- 早期类别在后续训练中不会加入负监督



Method	CUB-LT [42]				ImageNetR-LT [16]				Food101-LT [15]			
	5-task		10-task		5-task		10-task		5-task		10-task	
	A_{Mean}	A_{Tail}	A_{Mean}	A_{Tail}	A_{Mean}	A_{Tail}	A_{Mean}	A_{Tail}	A_{Mean}	A_{Tail}	A_{Mean}	A_{Tail}
L2P (CVPR'22)[46]	74.22	57.20	68.91	46.05	70.03	56.32	69.94	54.52	55.36	29.33	47.46	16.11
+ TAL	**75.91**	**61.19**	**70.71**	**49.80**	**72.84**	**63.50**	**71.63**	**60.47**	**58.53**	**40.93**	**50.84**	**23.84**
DualPrompt (ECCV'22)[45]	73.85	51.30	69.87	50.75	65.40	48.28	65.14	46.52	52.24	25.55	49.84	22.56
+ TAL	**75.06**	**54.65**	**70.68**	**52.50**	**66.45**	**52.04**	**66.00**	**50.06**	**54.28**	**30.01**	**50.77**	**24.89**
CODAPrompt (CVPR'23)[33]	74.14	53.26	69.10	49.54	72.92	57.16	70.61	53.60	62.64	34.71	56.11	28.33
+ TAL	**75.51**	**56.04**	**70.45**	**51.68**	**75.34**	**62.08**	**72.86**	**58.00**	**65.19**	**39.44**	**58.12**	**34.10**
LAE (ICCV'23) [13]	74.54	57.90	72.03	50.20	69.13	53.52	69.54	56.00	62.14	37.96	56.43	29.91
+ TAL	**76.97**	**61.40**	**74.83**	**57.10**	**70.54**	**58.48**	**71.00**	**58.64**	**64.34**	**42.35**	**59.38**	**35.89**
MOS (AAAI'25)[35]	73.95	51.40	71.52	48.93	69.15	53.80	68.73	55.44	60.63	29.30	55.46	28.08
+ TAL	**75.30**	**54.86**	**72.14**	**50.95**	**71.47**	**59.29**	**70.31**	**58.77**	**63.24**	**33.27**	**57.92**	**33.73**

非CIL设定下的一般图像分类任务

- ResNet+CIFAR/IN随机抽几个类别的作为子集进行实验
- 因为shuffle，batchsize，即使非增量学习条件下，仍然可以改善batch中类别不平衡的问题。

Dataset	Method	Batch=64	Batch=128	Batch=256
5 classes				
CIFAR	CE	88.56	88.68	85.88
	TAL	**89.36**	**89.24**	**89.36**
ImageNet	CE	96.83	94.84	96.43
	TAL	**97.62**	**95.24**	**97.22**
10 classes				
CIFAR	CE	85.64	84.56	84.28
	TAL	**87.42**	**85.76**	**84.88**
ImageNet	CE	89.00	86.60	88.24
	TAL	**89.80**	**88.40**	**88.68**