A Unified and General Framework for Continual Learning

Zhenyi Wang, Yan Li, Li Shen, Heng Huang

A Unified Framework for Continual Learning

Table 1: A unified framework for CL. We define a generalized CL optimization objective as $\mathcal{L}^{CL} = \mathcal{L}_{CE}(\boldsymbol{x}, y) + \alpha D_{\boldsymbol{\Phi}}(h_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{z}) + \beta D_{\boldsymbol{\Psi}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{old})$. Where $\alpha \geq 0, \beta \geq 0$, $\mathcal{L}_{CE}(\boldsymbol{x}, y)$ is the loss function on new task, $D_{\boldsymbol{\Phi}}(h_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{z})$ is *output space* regularization represented as a Bregman divergence associated with function $\boldsymbol{\Phi}$, $D_{\boldsymbol{\Psi}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{old})$ is *weight space* regularization represented as a Bregman divergence associated with function $\boldsymbol{\Psi}$. Several existing representative CL methods can be recovered from this general optimization objective by setting different $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and Bregman divergence.

Category	Method	Ref	Recover Setting
Bayesian-based	VCL	Nguyen et al. (2018)	$\alpha = 0, \boldsymbol{\Psi}(p) = \int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}$
	NCL	Kao et al. (2021)	$\boldsymbol{\Phi}(\boldsymbol{p}) = \sum_{i=1}^{i=n} \boldsymbol{p}_i \log \boldsymbol{p}_i$. $\boldsymbol{\Psi} = \frac{1}{2}\|\|\boldsymbol{\theta}\|\|^2$
Regularization-based	EWC	Kirkpatrick et al. (2017)	$\alpha = 0, \boldsymbol{\Psi}(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^T F \boldsymbol{\theta}$
	CPR	Cha et al. (2021)	$\boldsymbol{\Phi}(\boldsymbol{p}) = \sum_{i=1}^{i=n} \boldsymbol{p}_i \log \boldsymbol{p}_i$
Memory-replay-based	ER	Chaudhry et al. (2019b)	$\beta = 0, \boldsymbol{\Phi}(\boldsymbol{p}) = \sum_{i=1}^{i=n} \boldsymbol{p}_i \log \boldsymbol{p}_i$
	DER	Buzzega et al. (2020)	$\beta = 0, \boldsymbol{\Phi}(\boldsymbol{x}) = \|\|\boldsymbol{x}\|\|^2$
Novel CL method	`Refresh Learning`	Ours	Unlearn-relearn plug-in

Refresh Learning for Continual Learning

- Forgetting can be beneficial for the human brain in various situations, as it helps in efficient information processing and decision-making.

- CL involves adapting to new tasks and acquiring new knowledge over time. If a model were to remember every detail from all previous tasks, it could quickly become impractical and resource-intensive.

- Forgetting less relevant information helps in managing memory resources efficiently, allowing the model to focus on the most pertinent knowledge.

Refresh Learning for Continual Learning

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\rho_{opt}} \mathcal{L}^{CL} \quad \text{(relearn)}$$

$$s.t. \;\; \rho_{opt} = \min_{\rho} [\mathcal{E}(\rho) = -\mathbb{E}_{\rho} \mathcal{L}^{CL} + \mathbb{E}_{\rho} \log \rho] \quad \text{(unlearn)}$$

Algorithm 1 Refresh Learning for General CL.

REQUIRE: model parameters $\boldsymbol{\theta}$, CL model learning rate η,
for $k = 1$ to K **do** (number of CL steps)
  for $j = 1$ to J **do** (unlearn steps)
    $\boldsymbol{\theta}_k^j = \boldsymbol{\theta}_k^{j-1} + \gamma[F^{-1} \nabla \mathcal{L}^{CL}(\boldsymbol{\theta}_k^{j-1})] + \mathcal{N}(0, 2\gamma F^{-1})$
  end for
  $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta \nabla \mathcal{L}^{CL}(\boldsymbol{\theta}_k^j)$ (relearn step)
end for

Experiments

Table 2: **Task-IL and class-IL** overall accuracy on CIFAR10, CIFAR-100 and Tiny-ImageNet, respectively with memory size 500. '—' indicates not applicable.

Algorithm Method	**CIFAR-10** Class-IL	Task-IL	**CIFAR-100** Class-IL	Task-IL	**Tiny-ImageNet** Class-IL	Task-IL
fine-tuning	19.62 ± 0.05	61.02 ± 3.33	9.29 ± 0.33	33.78 ± 0.42	7.92 ± 0.26	18.31 ± 0.68
Joint train	92.20 ± 0.15	98.31 ± 0.12	71.32 ± 0.21	91.31 ± 0.17	59.99 ± 0.19	82.04 ± 0.10
SI	19.48 ± 0.17	68.05 ± 5.91	9.41 ± 0.24	31.08 ± 1.65	6.58 ± 0.31	36.32 ± 0.13
LwF	19.61 ± 0.05	63.29 ± 2.35	9.70 ± 0.23	28.07 ± 1.96	8.46 ± 0.22	15.85 ± 0.58
NCL	19.53 ± 0.32	64.49 ± 4.06	8.12 ± 0.28	20.92 ± 2.32	7.56 ± 0.36	16.29 ± 0.87
GPM	—	90.68 ± 3.29	—	72.48 ± 0.40	—	—
UCB	—	79.28 ± 1.87	—	57.15 ± 1.67	—	—
HAT	—	92.56 ± 0.78	—	72.06 ± 0.50	—	—
A-GEM	22.67 ± 0.57	89.48 ± 1.45	9.30 ± 0.32	48.06 ± 0.57	8.06 ± 0.04	25.33 ± 0.49
GSS	49.73 ± 4.78	91.02 ± 1.57	13.60 ± 2.98	57.50 ± 1.93	—	—
HAL	41.79 ± 4.46	84.54 ± 2.36	9.05 ± 2.76	42.94 ± 1.80	—	—
oEWC	19.49 ± 0.12	64.31 ± 4.31	8.24 ± 0.21	21.2 ± 2.08	7.42 ± 0.31	15.19 ± 0.82
oEWC+refresh	$\mathbf{20.37 \pm 0.65}$	$\mathbf{66.89 \pm 2.57}$	$\mathbf{8.78 \pm 0.42}$	$\mathbf{23.31 \pm 1.87}$	$\mathbf{7.83 \pm 0.15}$	$\mathbf{17.32 \pm 0.85}$
CPR(EWC)	19.61 ± 3.67	65.23 ± 3.87	8.42 ± 0.37	21.43 ± 2.57	7.67 ± 0.23	15.58 ± 0.91
CPR(EWC)+refresh	$\mathbf{20.53 \pm 2.42}$	$\mathbf{67.36 \pm 3.68}$	$\mathbf{9.06 \pm 0.58}$	$\mathbf{22.90 \pm 1.71}$	$\mathbf{8.06 \pm 0.43}$	$\mathbf{17.90 \pm 0.77}$
ER	57.74 ± 0.27	93.61 ± 0.27	20.98 ± 0.35	73.37 ± 0.43	9.99 ± 0.29	48.64 ± 0.46
ER+refresh	$\mathbf{61.86 \pm 1.35}$	$\mathbf{94.15 \pm 0.46}$	$\mathbf{22.23 \pm 0.73}$	$\mathbf{75.45 \pm 0.67}$	$\mathbf{11.09 \pm 0.46}$	$\mathbf{50.85 \pm 0.53}$
DER++	72.70 ± 1.36	93.88 ± 0.50	36.37 ± 0.85	75.64 ± 0.60	19.38 ± 1.41	51.91 ± 0.68
DER+++refresh	$\mathbf{74.42 \pm 0.82}$	$\mathbf{94.64 \pm 0.38}$	$\mathbf{38.49 \pm 0.76}$	$\mathbf{77.71 \pm 0.85}$	$\mathbf{20.81 \pm 1.28}$	$\mathbf{54.06 \pm 0.79}$

Thank you