Elastic Weight Consolidation Done Right for Continual Learning

Supplementary Material

I. Pseudocode of Our Method

We present the pseudocode of our proposed method in a PyTorch-like style in Listing A to demonstrate its simplicity and effectiveness. After training the current network, we estimate the importance of each weight using Logits Reversal (LR) combined with cross-entropy loss. In contrast to EWC, which does not incorporate an LR step, our approach leverages this additional operation to achieve more accurate importance estimation.

Listing A: Pseudocode of Logits Reversal in PyTorch-like style.

```
"""
Input:
images: Tensor[bs, C, H, W]
targets: Tensor[bs,]

Output:
omega: dict{name: tensor, shape=param.shape}
"""
logits = network(images)["logits"]
logits_r = -logits
loss = torch.nn.functional.cross_entropy(
    logits_r, targets)
optimizer.zero_grad()
loss.backward()
for name, param in network.named_parameters():
    if param.grad is not None:
        omega[n]+= param.grad.pow(2).clone()
```

II. Comparisons with Continual Learning Methods from Other Categories

Table A. EFCIL performance on CIFAR-100 compared against strong baselines. Reported results represent means across three independent trials.

method	category	T=5	T=10	T=20
iCaRL-CNN†	Replay	51.07	48.66	44.43
iCaRL-NCM†	Replay	58.56	54.19	50.51
LWF-MC	Regularization	45.93	27.43	20.07
MUC	Regularization	49.42	30.19	21.27
GPM	Optimization	41.51	37.78	41.27
ADAM-NSCL	Optimization	22.67	12.72	8.79
EWC	Regularization	32.82	27.02	19.61
EWC-DR	Regularization	**63.75**	**60.94**	**53.45**

†Replay-based method which stores and reuses samples from previous tasks during training.

In the CIFAR-100 big-start exemplar-free CIL setting, we report the average incremental accuracy (A_{avg}). Besides ours, a replay-based† method (iCaRL [34] stores 20 samples per old class), two regularization-based methods (LWF-MC [34], MUC [27]), and two optimization-based ones (GPM [38], ADAM-NSCL [46] reimplemented within our task-free framework) are included in Tab. A. The EFCIL setting is significantly more challenging than replay-based ones since no previous task data can be stored or accessed during incremental learning.

Table B. EFCIL results on CIFAR-100 comparing EWC-DR with modern methods. Average incremental accuracy (A_{avg}) is reported using seed 1993.

method	Big start			Equally split		
	T=5	T=10	T=20	T=5	T=10	T=20
FeTrIL	66.3	65.2	61.5	60.4	52.1	43.2
PASS	63.5	61.8	58.1	63.4	52.2	41.8
SSRE	65.9	65.0	61.7	56.6	44.4	33.6
EWC	33.4	26.4	19.2	39.2	27.3	17.7
EWC-DR	63.9	62.1	53.5	61.5	49.6	35.3

Tab. B showcases the comparison of EWC-DR with other modern PyCIL methods on CIFAR-100 under seed 1993, our method substantially elevates the original EWC from a lower baseline to a level competitive with contemporary techniques. Notably, in the Equally-split setting, EWC-DR even surpasses the SSRE.

Our proposed EWC-DR significantly improves upon EWC, making it a competitive baseline in continual learning especially in exemplar-free setting, though it is not intended to be state-of-the-art.

III. Sensitivity Analysis of λ

As with all weight regularization methods, our EWC-DR introduces a hyperparameter λ, which penalizes changes to the model weights, as defined in Equation (1). We conducted a sensitivity analysis to assess the impact of λ on the average accuracy (A_{avg}). Specifically, for the CIFAR-100 10-task equally split EFCIL setting, λ was varied from 1,000 to 50,000. As depicted in Figure A, A_{avg} increases steadily as λ grows from 1,000 and reaches its peak around λ = 10,000–20,000. Beyond this point, further increasing λ leads to a gradual decline in accuracy. This trend suggests that an appropriate choice of λ is crucial for balancing the trade-off between learning new knowledge and mitigating

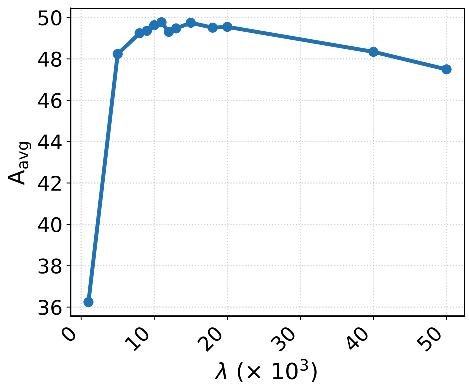

Figure A. Sensitivity analysis of the average accuracy (A_{avg}) with respect to the regularization parameter λ on CIFAR-100 under the 10-task equally split EFCIL setting.

forgetting. However, the proposed method is not sensitive to λ, as shown in Figure A.

IV. Statistics of FC Layer Weight Importance Matrices

We report statistics on the importance of the FC layer weights in tabular form, corresponding to the illustrations in Figure 5. Importance values per class are further summed for tabulation in Tab. C. EWC has low importance due to gradient vanishing. MAS has high importance and a relatively even distribution across classes. Our EWC-DR obtains high importance, with a notable peak in the GT class.

Table C. Statistics of the FC layer weight importance.

class	0	1	2 (GT)	3	4
EWC	0.003	0.013	0.219	0.136	0.018
MAS	14.44	8.53	25.48	10.26	18.08
EWC-DR	5.30	6.55	41.38	2.51	11.12