Class-Incremental Learning with Strong Pre-trained Models Supplemental Material

Tz-Ying Wu[1,2] Gurumurthy Swaminathan[1] Zhizhong Li[1]
Avinash Ravichandran[1] Nuno Vasconcelos[2] Rahul Bhotika[1] Stefano Soatto[1]
[1]AWS AI Labs [2]UC San Diego
{gurumurs,lzhizhon,ravinash,bhotikar,soattos}@amazon.com {tzw001,nuno}@ucsd.edu

A. Additional experiments

A.1. Disjoint-CIL

Robustness. To test the robustness of the proposed method for different backbones, we study score fusion with ResNet18 and ResNet50. The results are shown in Tables 7, and 8. As shown for the ResNet10 backbone in the main paper, the score fusion approach results in better performance even for deeper backbones, but the relative trends remain the same. In addition, We also study score fusion with a larger number of novel classes. In the main paper, we showed results for 800 base and 40 novel classes. Here, we show the results for 800 base and 200 novel classes in Table 9. The results for 200 classes are similar to the ones for 40 classes (refer Table 3). This shows that our score fusion method can be generalized to larger number of novel classes.

Table 7. Disjoint-CIL analysis with deeper backbone ResNet18 (ResNet50 in Table 8), 40 novel classes using random class split. Our observations generalize to deeper architectures.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{avg}
confidence-based routing	49.91	47.98	88.50	68.24
learning-based routing w/ L_{rt-bal}	65.59	65.02	77.00	71.01
oracle routing w/ L_{rt-bal}	65.85	65.02	82.50	73.76
FA (ours) + BiC [1]	69.12	69.70	57.70	63.70
FA (ours) + WA [2]	68.88	69.02	66.10	67.56
score fusion (ours) best-Acc_{all}	**69.45**	70.01	58.13	64.07
score fusion (ours) best-balanced	67.36	66.61	82.37	74.49
score fusion (ours) best-Acc_{avg}	65.83	64.85	85.50	**75.17**
joint learning (oracle)	70.32	70.43	68.20	69.32

Table 8. Disjoint-CIL analysis with deeper backbone ResNet50 (see Table 7 for ResNet18), 40 novel classes using random class split. Our observations generalize to these architectures.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{avg}
confidence-based routing	60.15	58.65	90.10	74.38
learning-based routing w/ L_{rt-bal}	73.21	72.83	80.97	76.90
score fusion (ours) best-Acc_{all}	**75.40**	75.94	64.57	70.25
score fusion (ours) best-balanced	74.44	74.00	83.13	78.57
score fusion (ours) best-Acc_{avg}	72.02	71.24	87.43	**79.34**
joint learning (oracle)	76.59	76.71	74.10	75.41

A.2. Overlapping-CIL

As a recap of Section 5.1, in the overlapping-CIL scenario where a subset of base and novel classes overlap, we analyze three different ways to split base and novel classes and two different ways to split the samples within each class. In particular, we showed the results for splits based on style in Table 5. In this section, we show the results for the rest of the splits.

We show in Table 10 results for random split and within that splitting the samples either randomly or using clustering. In Table 11, we show the result for the domain change (inanimate and animate) based split, again splitting the samples randomly or using clustering. In both the cases, our score fusion method performs better than other baselines thereby showing the generalizability of the approach to various practical scenrios.

B. Implementation Details

In all the experiments, we adopt batch size 256 with learning rate starts at 0.1, and normalize the features and weights. The pretrained model is trained for 90 epochs with the SGD optimizer, where the learning rate decays every 30 epochs. In the first stage, we finetune the network for 30 epochs with learning rate decays every 10 epochs. As mentioned in section 4 in the main paper, The second-stage training adopts class-balanced sampling, which is trained for 10 epochs. Experiments are repeated with different random seeds.

Table 9. Disjoint-CIL analysis with ResNet10, larger novel set with 200 classes using random class splits. Our observations hold.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{avg}
confidence-based routing	51.25	44.84	76.88	60.86
learning-based routing w/ L_{rt-bal}	56.91	54.39	67.01	60.70
oracle routing w/ L_{rt-bal}	57.80	54.95	69.18	62.07
FA (ours) + BiC [1]	61.09	61.62	58.98	60.30
FA (ours) + WA [2]	54.42	64.31	14.88	39.60
score fusion (ours) best-Acc_{all}	**61.56**	61.39	62.21	61.80
score fusion (ours) best-balanced	61.29	59.63	67.93	63.78
score fusion (ours) best-Acc_{avg}	59.57	56.31	72.60	**64.45**
joint learning (oracle)	61.68	61.42	62.76	62.09

Table 10. Overlapping-CIL results of random class splits.

(a) Overlapping classes samples split randomly.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{ovlp}	Acc_{avg}
confidence-based routing	39.91	37.67	84.80	83.20	68.56
learning-based routing w/ L_{rt-bal}	55.32	54.72	68.34	59.20	60.75
oracle routing w/ L_{rt-bal}	55.54	54.71	74.06	58.40	62.39
logit concatenation (avg pool)	59.74	59.95	56.00	51.20	55.72
logit concatenation (max pool)	59.65	59.80	55.89	60.80	58.83
score fusion (ours) best-Acc_{all}	**59.88**	60.40	48.61	55.47	54.83
score fusion (ours) best-balanced	57.99	57.24	74.02	65.33	65.53
score fusion (ours) best-Acc_{avg}	53.09	51.70	81.64	72.27	**68.54**
joint learning (oracle)	60.25	60.35	56.91	68.00	61.75

(b) Overlapping classes samples split by clustering.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{ovlp}	Acc_{avg}
confidence-based routing	37.37	35.26	79.43	80.00	64.90
learning-based routing w/ L_{rt-bal}	52.97	52.49	65.30	44.27	54.02
oracle routing w/ L_{rt-bal}	52.74	51.95	71.66	45.60	56.40
logit concatenation (avg pool)	56.41	56.57	54.86	42.40	51.28
logit concatenation (max pool)	56.34	56.35	54.86	64.80	58.67
score fusion (ours) best-Acc_{all}	**57.00**	57.41	48.11	54.13	53.22
score fusion (ours) best-balanced	55.33	54.54	71.24	69.60	65.13
score fusion (ours) best-Acc_{avg}	49.51	48.13	76.76	77.33	**67.41**
joint learning (oracle)	57.29	57.51	51.77	60.00	56.43

Table 11. Overlapping-CIL results of domain-changing (inanimate-animate) splits.

(a) Overlapping classes samples split randomly.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{ovlp}	Acc_{avg}
confidence-based routing	40.82	37.75	83.43	91.20	70.79
learning-based routing w/ L_{rt-bal}	56.18	54.58	81.71	58.67	64.99
oracle routing w/ L_{rt-bal}	56.70	55.01	83.77	58.40	65.73
logit concatenation (avg pool)	57.34	56.25	74.17	64.00	64.81
logit concatenation (max pool)	55.84	54.55	74.06	75.20	67.94
score fusion (ours) best-Acc_{all}	**57.44**	56.27	75.81	61.87	64.65
score fusion (ours) best-balanced	55.66	53.88	81.75	75.20	70.28
score fusion (ours) best-Acc_{avg}	52.55	50.44	82.40	84.27	**72.37**
joint learning (oracle)	57.95	56.52	80.11	65.60	67.41

(b) Overlapping classes samples split by clustering.

Method	Acc_{all}	Acc_{base}	Acc_{novel}	Acc_{ovlp}	Acc_{avg}
confidence-based routing	38.28	35.30	80.23	82.40	65.98
learning-based routing w/ L_{rt-bal}	52.89	51.37	78.90	44.00	58.09
oracle routing w/ L_{rt-bal}	53.30	51.66	81.03	45.60	59.43
logit concatenation (avg pool)	53.57	52.61	70.17	46.40	56.39
logit concatenation (max pool)	52.34	51.06	70.17	72.00	64.41
score fusion (ours) best-Acc_{all}	**54.11**	52.84	74.40	56.80	61.35
score fusion (ours) best-balanced	52.84	51.13	78.06	69.87	66.35
score fusion (ours) best-Acc_{avg}	51.55	49.73	77.94	73.33	**67.00**
joint learning (oracle)	54.28	52.70	78.06	67.20	65.99

References

[1] Yue Wu and et al. Large scale incremental learning. In *CVPR 2019*. 1, 2

[2] Bowen Zhao, Xi Xiao, Guojun Gan, Bin Zhang, and Shu-Tao Xia. Maintaining discrimination and fairness in class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13208–13217, 2020. 1, 2