Supplementary - General Incremental Learning with Domain-aware Categorical Representations

1. More Implementation Details

We use Nvidia Titan XP as the computation platforms with CUDA 10.1. Our python is 3.7 and PyTorch is 1.71. We use seed 1993 for all the experiments.

Data Augmentation For iDigits benchmark, we only resize every image to 32x32. For iCIFAR-20 benchmark, we use RandomCrop with shape 32x32 and padding 4. We also use RandomHorizontalFlip, ColorJitter with brightness as 63/255 and Normalization with mean (0.5071,0.4867,0.4408) and standard deviation (0.2675,0.2565,0.2761). For iDomainNet, We apply RandomResizedCrop with size as 112 , RandomHorizontalFlip and Normalization with mean (0.5,0.5,0.5) and standard deviation (0.5,0.5,0.5) for all the channels.

2. Details of Splits

We introduce the NCD splits for the three benchmarks. For iDigits NCD split, numbers of new classes at each session are [4, 3, 2, 1, 0, 0, 0, 0, 0, 0]. For iDomainNet NCD split, numbers of new classes at each session are [60, 10, 10, 10, 10, 0, 0, 0, 0, 0]. For iCIFAR-20 NCD split, numbers of new classes at each session are [10, 5, 3, 1, 1, 0, 0, 0, 0, 0]. The number of new classes decreases as the number of session increases because unseen categories for models in real world should become fewer and fewer with the accumulation of knowledge. Each class at every session contains data from one domain. We will release the data of the three benchmarks with all the splits used in the experiments later.

3. More Curves

We include more curves of performance w.r.t. sessions on iDigits and iDomainNet benchmarks with three splits, which are shown in Fig. 1 and Fig. 2. We can see our method consistently perform better than other methods.

Table 1. Forgetting metric on iCIFAR-20.

Method	NC (%)	ND (%)	NCD (%)
UCIR	-12.22	-12.36	-6.48
UCIR w/ ours	-10.66	-10.27	-4.48
GeoDL	-11.85	-11.33	-4.06
GeoDL w/ ours	-9.12	-8.57	-3.73
DER	-5.92	-14.76	-5.79
DER w/ ours	-5.36	-3.44	-1.79

4. Forgetting Metric

We also compute the forgetting metric of different methods on the three splits of iCIFAR-20 benchmark as follows:

$$F = \frac{1}{N-1}\sum_{i=2}^{N} A_i^{i-1} - A_{i-1}^{i-1} \tag{1}$$

where N is the number of total sessions. A_i^{i-1} is the accuracy of model at session i evaluated on the test set of session $i-1$. $A_i^{i-1} - A_{i-1}^{i-1}$ measures model's forgetting on previously observed classes and domains.Results are presented in Tab. 1. This shows our method has less forgetting than previous methods, which is consistent with our conclusion.

5. More Sensitive Studies

We examine the influence of the number of new components m added at every new session, which conducted on the iCIFAR-20 ND split with 3 sessions. Session 1 has one domain for each class and sessions 2, 3 have two domains for each class. The results are shown in Tab. 4. We can see that model with small m cannot perform well on this split since classes in sessions 2 and 3 have more than one domain. However, model with larger m can perform closely to the oracle and is also robust to the change of number m. Moreover, in Table 2, we show the average number of components for all the classes in the final session of the iCIFAR-20 NCD split with varying δ. As each expansion adds 30 components, the result indicates that our reduction process can eliminate many redundant components. The av-

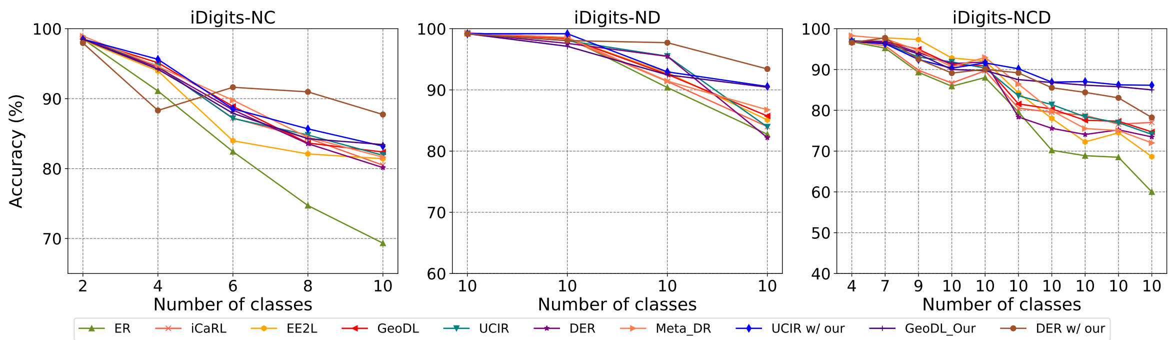

Figure 1. **Performances w.r.t sessions** on iDigits benchmark with three splits

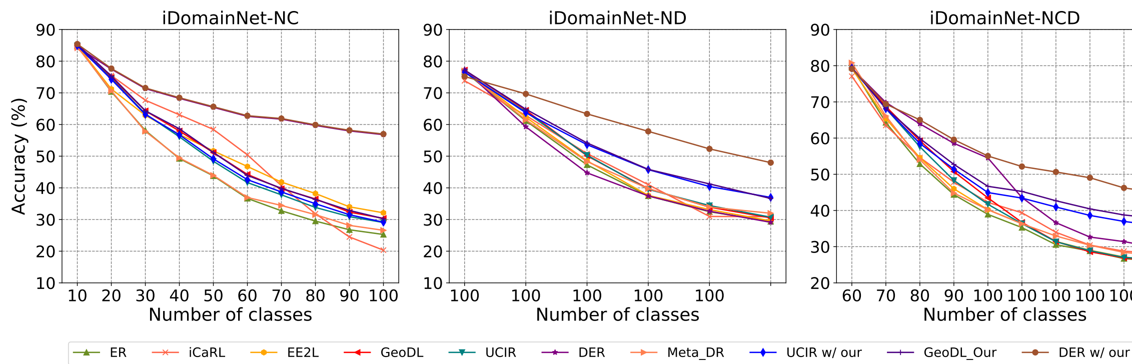

Figure 2. **Performances w.r.t sessions** on iDomainNet benchmark with three splits

Table 2. **Sensitive Study:** Influence of δ on the number of components.

Threshold δ	0.6	0.65	0.7	0.75	0.8	0.85	0.9
DER w/ ours	6.6	6.2	5.4	5.1	4.7	4.35	3.45

erage number of components decreases as the threshold δ increases, as expected. To measure the consistency of components and domain labels, we compute the purity of the components within each class and then average them on all the classes and sessions. The resulting average purity is 0.962 for the iCIFAR-20 NCD split, demonstrating the efficacy of our mixture model.

We also conduct sensitive study on memory size. As shown in Tab. 5, the performance of the model continuously improves as the memory size increases. Furthermore, we also evaluate the effect of randomness of memory selection. Concretely, as shown in Table 3, we provide results of DER w/ ours for iCIFAR-20 using five different random seeds, which shows that our random selection strategy for each class is relatively robust. That been said, we believe a better selection strategy is an interesting direction for future work.

Table 3. The average performance on five random seeds.

	NC	ND	NCD
Avg Acc(%)	82.58±0.47	84.19±0.50	82.00±0.22

Table 4. **Sensitive Study:** Influence of m for our method on iCIFAR-20.

Method	Final (%)	Avg (%)
DER	80.94	85.59
DER w/ ours ($m = 1$)	82.45	86.34
DER w/ ours ($m = 5$)	84.39	87.32
DER w/ ours ($m = 10$)	84.54	87.40
DER w/ ours ($m = 15$)	84.08	87.21

6. Analysis of Intra-class Imbalance Issue

Tab. 6 shows the accuracy from domain 1 to domain 5 within classes at the last session of iCIFAR-20 NCD split. As the results show, previous methods suffer from intra-class imbalance problem, which have high performance on new domains and low performance on old domains. By contrast, our method can obtain a more balanced results.

Table 5. **Sensitive Study:** Influence of memory size on our method for iCIFAR-20 NCD split.

Memory size	$M = 500$		$M = 1000$		$M = 1500$		$M = 2000$		$M = 2500$		$M = 3000$	
	Final	Avg	Final	Avg	Final	Avg	Final	Avg	Final	Avg	Final	Avg
DER w/ ours	66.45	75.04	72.47	78.02	75.13	80.63	76.40	82.17	77.56	82.61	78.80	82.80

Table 6. **Analysis:** Accuracies of each domains within classes in the last session of iCIFAR-20 NCD split.

Category	Method	Domain 1 (%)	Domain 2 (%)	Domain 3 (%)	Domain 4 (%)	Domain 5 (%)
reptiles	DER	30.23	52.01	25.24	48.27	76.28
	DER w/ ours	48.17	57.79	45.48	61.91	66.28
medium-sized mammals	DER	32.56	17.31	37.36	37.63	88.47
	DER w/ ours	84.27	46,87	46.14	79.16	85.25
large carnivores	DER	33.19	22.00	71.43	14.18	92.38
	DER w/ ours	69.33	68.59	85.35	42.74	85.98
large omnivores and herbivores	DER	42.41	66.74	17.65	45.22	86.11
	DER w/ ours	70.23	64.27	56.48	84.26	76.32

7. More t-SNE visualization

We provide more t-SNE visualization results of different digits in the iDigits ND split across different sessions. We can see that for most digits, shown in Fig. 3,4,5,6, our method can inference the correct domain labels.

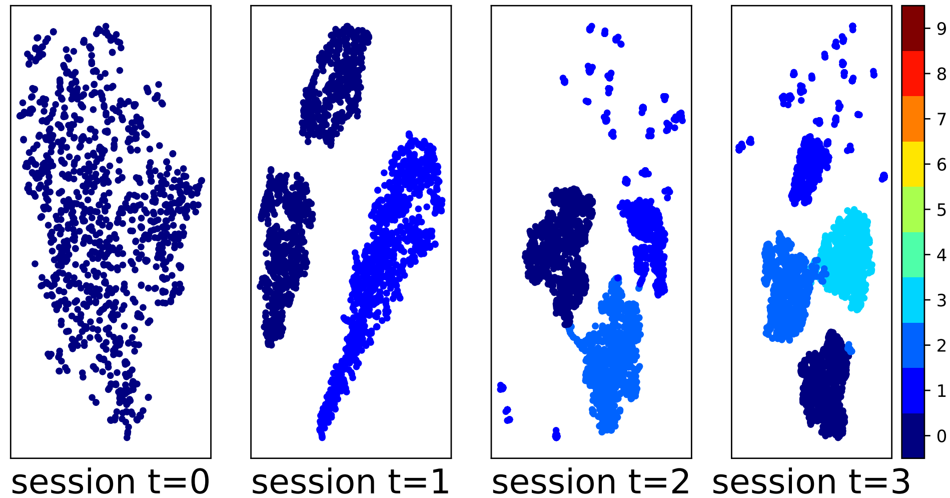

Figure 3. t-SNE visualization of digits 1.

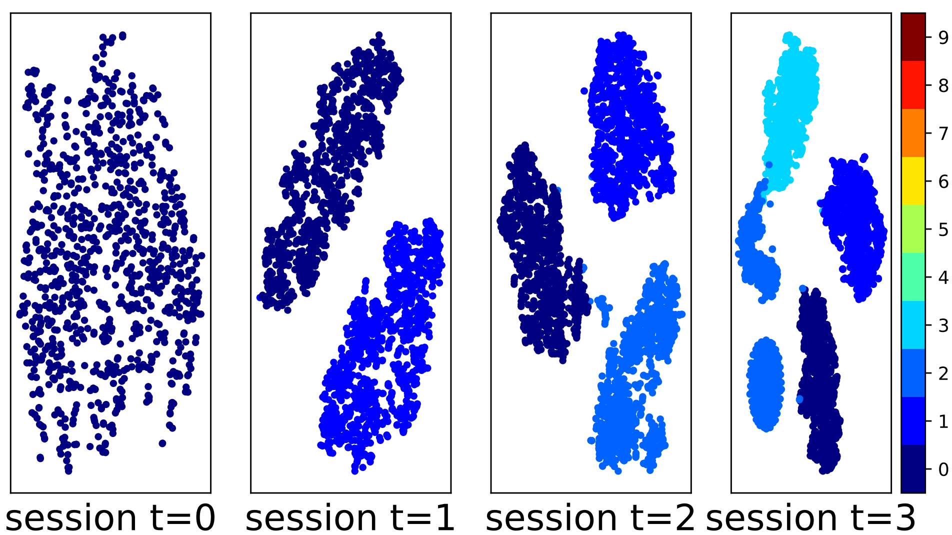

Figure 4. t-SNE visualization of digits 3.

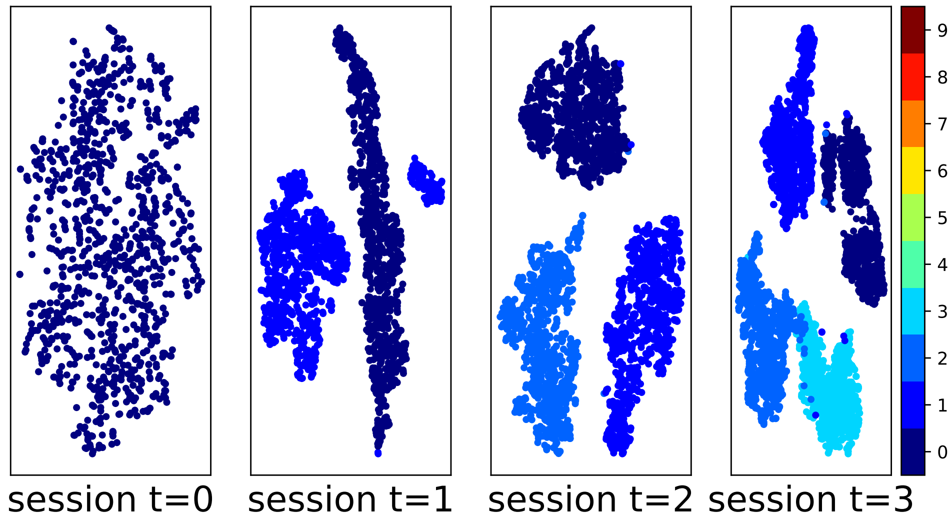

Figure 5. t-SNE visualization of digits 5.

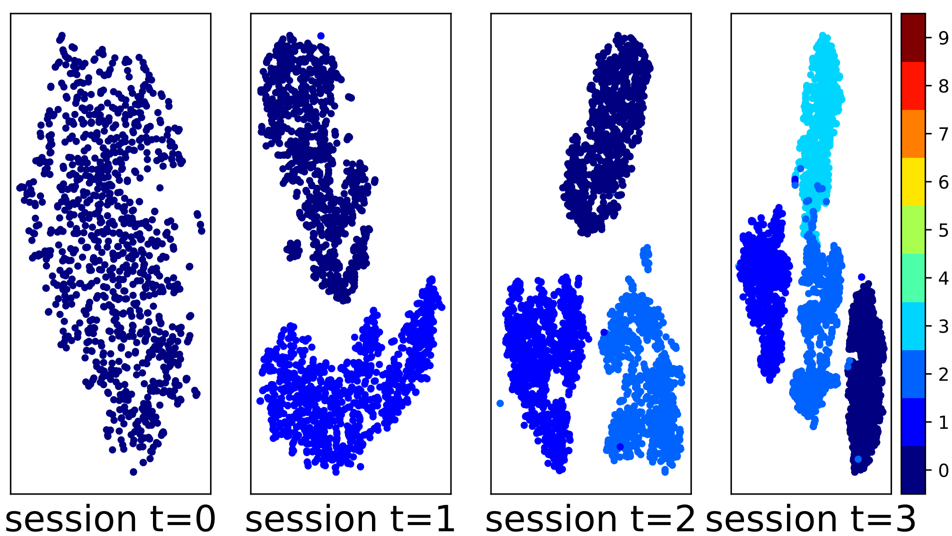

Figure 6. t-SNE visualization of digits 7.