Generative Feature Replay For Class-Incremental Learning
(Supplementary Material)

Table 3. Ablation study of different regularization methods on CIFAR-100 for the 4-task scenario.

	T1	T2	T3	T4
EWC + GAN		40.8	26.8	21.2
MAS + GAN	81.9	40.2	26.0	20.9
Feature Distillation + GAN		58.4	48.8	42.2

A. Comparative analysis on ImageNet-1000

The average accuracy and forgetting on ImageNet-1000 are shown in Figure 6. We can see that our proposed method outperforms iCaRL by a large margin in 5, 10 and 25 tasks. Compared to the state-of-the-art method Rebalance, we obtain slightly better accuracy in 5 tasks, and the gap is enlarged in both 10 and 25 tasks. In terms of the average forgetting, our method surpasses all methods by more than 10%. It is important to note that for both iCaRL and Rebalance, they need to store 20000 exemplars in order to train in a continual setting. It takes about 3.8 Gb memory for these exemplar-based methods, while for our proposed method, we only need to store a generator and a discriminator with 4.5 Mb memory.

B. Ablation study on different regularization

For our ablation study we use CIFAR-100 with 4 tasks of equal number of classes. In Table 3 we compare different regularization methods in feature extractor, where feature distillation clearly outperforms MAS and EWC. This shows that adding constraints on features is superior to constraining in parameter space. This guarantees that the generated features are closer to the real ones.

C. T-SNE on generated features

Here we show the T-SNE visualization of generated features using GANs and real features extracted from images (see Figure 7). We can see that the distributions of generated features and real features are very close, which allows our proposed method to train the classifier jointly with current data. There are clusters in the figures, which represents the distributions of different classes.

D. Architecture details

Generator and Discriminator consist of two hidden layer of 512 neurons followed by LeakyReLU with parameter 0.2. We concatenate Gaussian noise $\mathbf{z}$ of 200 dimensions and one-hot vectors as input of Generator. More details can be seen in the available code.

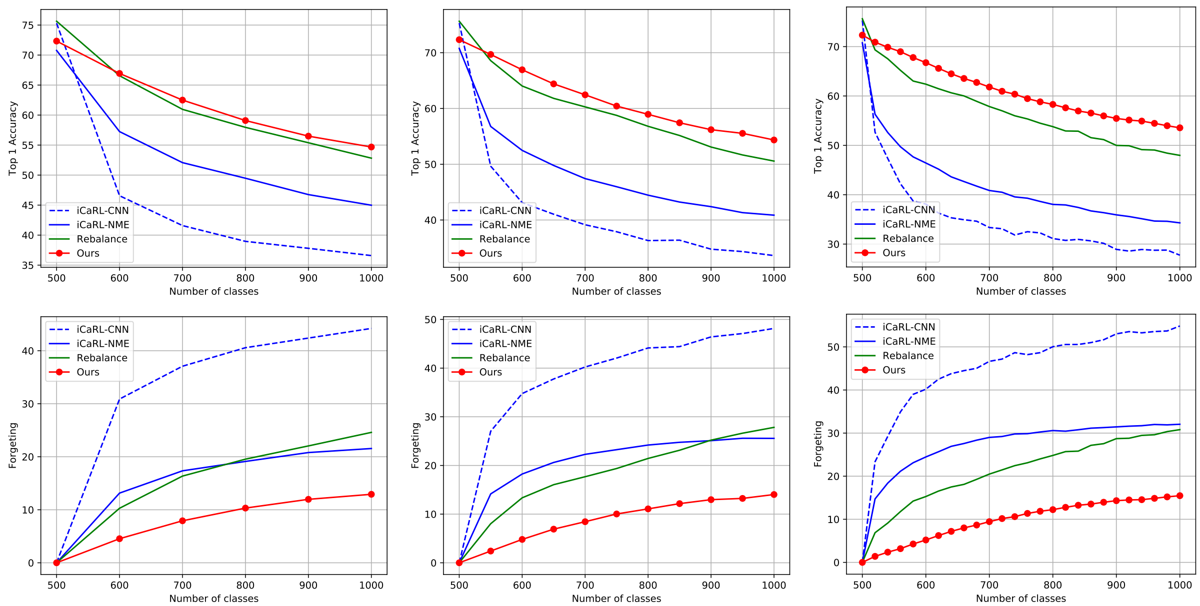

Figure 6. Comparison in the average accuracy (Top) and the average forgetting (Bottom) with various methods on ImageNet-1000. The first task has the half number of classes, and the remaining classes are divided into 5, 10 and 25 respectively. The lines with symbols are methods without using any exemplars, and without symbols are methods with 20000 exemplars.

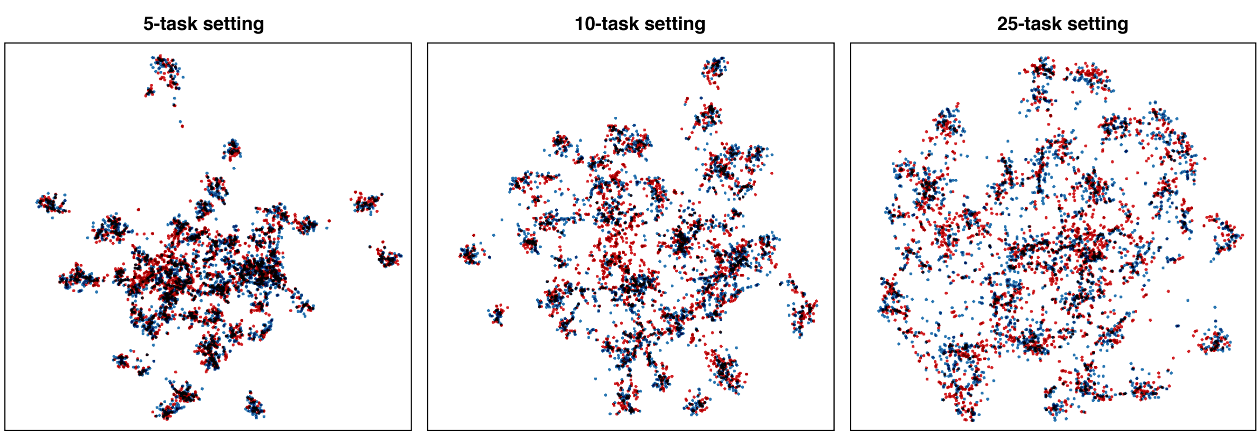

Figure 7. Real features (Red) and Generated features (Blue) on ImageNet-Subset of first task after training all tasks in 5, 10 and 25 tasks setting, respectively.