The 2nd Place Solution for CVPR 2022 Workshop on Continual Learning (CLVision, 3rd Edition) Challenge–Track 1: A Replay-based Continual Learning Approach

Tao Xu[1,2], Hangtao Feng[1,2,*], Suiwu Zheng[1,2,4], Xu Yang[1,2,3], and Zhiyong Liu [1,2,3 †]
1 School of Artificial Intelligence, University of Chinese Academy of Sciences
2 The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation Chinese Academy of Sciences
3 Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences
4 Huizhou Zhongke Advanced Manufacturing Limited Company, Huizhou, China
{xutao2020, fenghangtao2018, suiwu.zheng, xu.yang, zhiyong.liu}@ia.ac.cn

Abstract

Class incremental learning (CIL) means learning new classes step by step, and in each learning process, only the samples of the new class can be acquired, and the samples of the already learned classes are not available. The deep model tends to forget the learned classes when learning new classes, which is called catastrophic forgetting. Therefore, we propose several tricks to improve the anti-forgetting ability of the network based on classes balanced replay strategy. Specifically, we divide incremental class learning into two phases: the first phase is the learning ability enhancement phase, which is to improve the network's ability to learn classes by building an efficient classification framework; the second phase is the forgetting resistance enhancement phase, which is to reduce the forgetting of learned knowledge by using classes balanced replay strategy. By setting up these two phases, our method can significantly improve the performance and ranks 2nd place on $CVPR2022\ Continual\ Learning\ Challenge-Track1$.

1. Introduction

In recent years, deep networks have witnessed the development of the computer vision field [5, 8, 12, 14]. However, the current approaches are mainly tuned for static data sets, i.e., the number of classes of targets to be learned is fixed and known. However, real-life scenarios are not static, and the intelligences need to continuously update their knowledge and skills based on dynamic scenarios. Therefore, continuous learning is proposed to model this dynamic continuous learning process.

As one of the most fundamental tasks in computer vision, the continuous learning classification task has also received a great of attention from researchers. The continuous learning classification task requires that the model retains the ability to classify learned categories while learning new ones. However, the deep model tends to forget the learned classes when learning new classes, which is called catastrophic forgetting [4, 10] . Nowadays, the mainstream continuous learning methods are classified as follows: (1) Regularization methods [1, 7, 18]: i.e., regularization operations on the weights of the model so that the model limits the weight changes when learning new knowledge or in skills. (2) Knowledge distillation [6, 11]: i.e., retaining the old model and transferring the knowledge of the old model to the new model by distillation methods. (3) Replay methods [2, 3]: i.e., generating or retaining a small number of samples of learned categories and training them together with the new class data model. However, the first method, because it limits the weights of the model, leads to the capacity of the model is also limited, which affects the performance of the model; the second method requires high hardware requirements of the training equipment during the training process, which requires large-capacity memory, and, the lack of old class samples seriously affects the forgetting resistance of the model; the performance of the last method is very dependent on the quality and quantity of the retained samples closely related.

Therefore, in this paper, we propose a replay-based class incremental learning algorithm based on class balancing.

[*1] Equal contribution.
[†2] Corresponding author.

Specifically, we divide the continuous learning process into two parts, namely: the learning ability enhancement part and the incremental learning part. In the learning ability enhancement part, we improve the fast learning ability of the model by selecting and adjusting the backbone, and in the incremental learning part, we use a replay strategy based on category balancing to improve the forgetting resistance of the model, and finally rank 2nd place on $CVPR2022\ Continual\ Learning\ Challenge$.

2. Method

In this section, we first describe the setting of CIL and then introduce our methods.

2.1. Class incremental learning

The goal of class incremental learning is to obtain a classifier that performs well on all learned classes. For the t_{th} task, the training data for the classifier consists of x_i, the input image, and y_i, the corresponding label, where y_i belongs to C_t, denoting all the labels for the t_{th} task.

At the t_{th} task, the training data only includes samples from the current classes and lacks samples seen previously, the model's effectiveness on the previous task decreases significantly, which is also known as catastrophic forgetting. This phenomenon is a relatively common and challenging problem in continuous learning tasks. Therefore, we propose a classes-balanced replay strategy for class incremental learning to attenuate the forgetting of the network and improve the model's effectiveness in class incremental learning.

2.2. The learning ability enhancement part

Pre-trained models on large-scale datasets can quickly and well transfer to other datasets. Considering that the existing dataset is relatively small, it is easy to lead to overfitting if we choose to learn from scratch.

Therefore, we choose to fine-tune a series of competitive pre-trained classification models on our dataset. Resnet [5] introduces the residual module into CNN to deal with the Gradient explosion and extinction problems caused by network deepening. Efficientnet [17] aims at achieving the same effect with fewer parameters by carefully balancing network depth, width, and resolution. Regnet [14] designs network design spaces which provide simple and fast networks that work well across a wide range of flop regimes. The above pre-trained models have shown significant generalization ability.

2.3. The incremental learning part

The second part is the incremental learning component, which aims to reduce the network's forgetfulness of the categories already learned. We focus on reducing the forgetting of the old categories by the model by classes-balanced replay strategy. The classes-balanced replay strategy is shown in the Figure 1, which trains the network by retaining some of the learned knowledge and combining it with the training data of the current task.

Considering the category imbalance in the learned data, we select the learned samples according to the category to attenuate the effect of category imbalance on the training results.

3. Experiments

3.1. Challenge setting

The "Continual instance-level object classification" tack features a stream of 15 experiences. This track features a fully supervised (labels are given for the training set) learning scheme in which the stream of incremental experiences is modeled by following the Class-Incremental scenario. No task labels or other additional signals are provided at test time.

The main rules of the competition are as follows

- The benchmark creation procedure can't be changed. Instances will be loaded as 224x224 RGB images. A different input size can be used in the solution, but the image must first be loaded as 224x224.
- Model initialization can be done by randomly initializing weights or by pretraining using the ImageNet-1K (ImageNet 2012) [8] dataset (recommended). Apart from the pre-trained weights, the solution must not access/use data from the pretraining dataset.
- Solutions can exploit a replay buffer containing data coming from up to 3500 training samples (image+label).
- Test-time training or tuning is not allowed. The maximum allowed execution time for the whole solution (training+test) is 12 hours.

3.2. Dataset and Implementation Details

CLVISION Image Classification Challenge proposes the topic that how to incrementally learn image classification. The objective of Challenge is to increase the average TOP-1 accuracy on the EgoObject dataset by incremental learning. The dataset is divided into two splits, including a training set with labels and a testing set without labels. Each split includes 1,110 classes, and the training split is evenly divided into 15 experiences for incremental training.

In the challenge, we use the EgoObject dataset. 15 experiences are tested with totally 1,110 classes. At each incremental experience, a RegNet [14] is trained with 1 GPU and 4 instances per GPU by stochastic gradient descent with 3 epochs. The learning rate is adjusted according to a cosine decaying policy and the initial learning rate is 0.02,

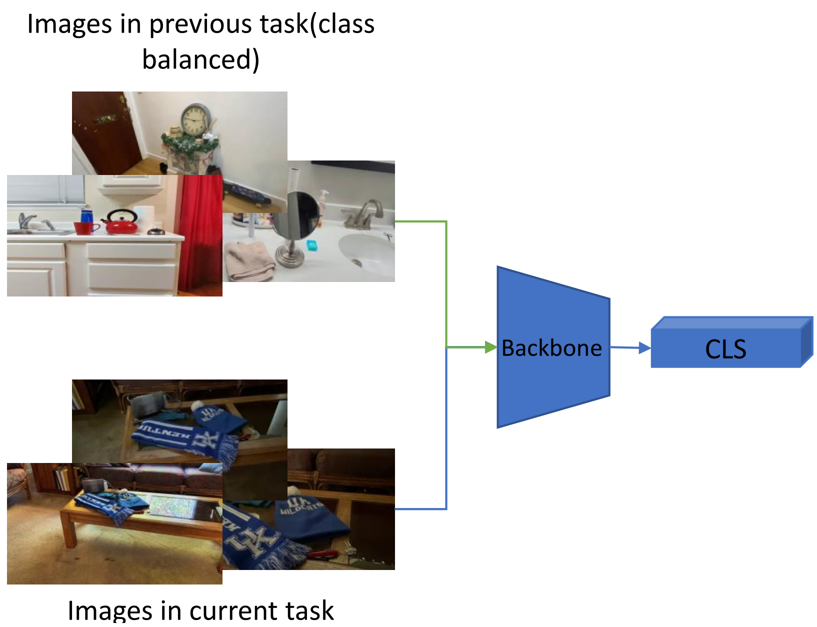

Figure 1. The classes-balanced replay strategy

weight decay is 1e-5 and momentum is 0.9. Model is evaluated by average TOP-1 accuracy on the test set consisting of all classes. The image resolution is 224x224 during the training and testing. And the model is pretrained on the ImageNet-1K.

3.3. Results

Stronger backbones. We evaluate the effectiveness of different pre-trained backbones in Table 1. As we can see in Table 1, the Regnet_x_16gf achieves the best performance with the optimal test hyper-parameters.

Table 1. Comparisons of different backbones.

Backbone	**Accuracy rate**
Resnet-18 [5]	0.5002
Resnet-50 [5]	0.5118
Resnet-101 [5]	0.5055
Efficientnet-b2 [17]	0.5123
Efficientnet-b3 [17]	0.5168
Efficientnet-b4 [17]	0.5147
Regnet_x_3.2gf [14]	0.5186
Regnet_x_8gf [14]	0.5193
Regnet_x_16gf [14]	**0.5205**

Batch Normalization. To accelerate the rate of convergence during model training and make the training process more stable, we introduce a Batch Normalization (BN) layer before the fully-connected layer. As we can see in Table 2, the backbones with BN layer outperform others.

Table 2. Comparisons of the backbones with/without BN layer.

Backbone	**BN layer**	**Accuracy rate**
Resnet-50 [5]	×	0.5099
Resnet-50 [5]	✓	0.5118
Efficientnet-b3 [17]	×	0.5142
Efficientnet-b3 [17]	✓	0.5168
Regnet_x_16gf [14]	×	0.5187
Regnet_x_16gf [14]	✓	**0.5205**

Label smoothing. To avoid overfitting and improve the generalization ability of model, we use label smoothing CE Loss (LSCE) [16] to train our model. As indicated in Table 3, the model trained with label smoothing CE Loss is more generalized.

Table 3. Comparisons of loss function.

Backbone	**LSCE**	**Accuracy rate**
Regnet_x_16gf [14]	×	0.5190
Regnet_x_16gf [14]	✓	**0.5205**

Continuous Learning strategy

We randomly selected some data from the training data for training and testing in order to compare the effectiveness of different continuous learning strategies. And the training data and the testing data are exactly the same, which is only

Table 4. Comparisons of continuous learning strategies.

Strategies	Accuracy rate
LWF [9]	0.1175
EWC [7]	0.0748
SI [18]	0.0748
Gdumb [13]	0.0748
MAS [1]	0.1515
ICARL [15]	0.4862
Replay	**0.4939**

Table 5. Comparisons of mixed continuous learning strategies.

Strategies	Accuracy rate
Replay	0.4939
+LWF [9]	0.4950
+MAS [1]	0.4939
+ICARL [15]	0.4950
+class_balance	**0.4961**

used for testing the continuous learning strategy. Table 4 shows the performances. We can see that the Replay strategy performs the best.

We also compared the effect of different strategies for combination, and the results are shown in Table 5. The results show that $Replay + class\_balance$ combination has the best accuracy.

4. Conclusion

Continuous learning is a challenge because the model easily forgets the old classes as it learns new classes of data. We use a classes balanced replay strategy to improve the learning ability and forgetting resistance of the network. The method consists of two parts: ability learning enhancement and forgetting resistance enhancement, and experiments are conducted to demonstrate the effectiveness of our method in continuous learning. Finally, we rank 2nd place on $CVPR2022\ Continual\ Learning\ Challenge$.

5. Acknowledgement

This work was supported in part by the National Key Research and Development Plan of China under Grant 2020AAA0108902 and 2020AAA0105900, partly by Youth Innovation Promotion Association CAS, and partly by Beijing Science and Technology Plan Project (grant Z201100008320029), part by the Strategic Priority Research Program of Chinese Academy of Science under Grant XDB32050100.

References

[1] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 139–154, 2018. 1, 4

[2] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*, pages 233–248, 2018. 1

[3] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420*, 2018. 1

[4] Ian J Goodfellow, Mehdi Mirza, Da Xiao, Aaron Courville, and Yoshua Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*, 2013. 1

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2, 3

[6] Khurram Javed and Faisal Shafait. Revisiting distillation and incremental classifier learning. In *Asian conference on computer vision*, pages 3–17. Springer, 2018. 1

[7] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 1, 4

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 1, 2

[9] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 4

[10] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989. 1

[11] Umberto Michieli and Pietro Zanuttigh. Knowledge distillation for incremental learning in semantic segmentation. *Computer Vision and Image Understanding*, 205:103167, 2021. 1

[12] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017. 1

[13] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *European conference on computer vision*, pages 524–540. Springer, 2020. 4

[14] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10428–10436, 2020. 1, 2, 3

[15] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017. 4

[16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 3

[17] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2, 3

[18] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International Conference on Machine Learning*, pages 3987–3995. PMLR, 2017. 1, 4