GopGAN: Gradients Orthogonal Projection Generative Adversarial Network With Continual Learning

Xiaobin Li, *Graduate Student Member, IEEE*, and Weiqiang Wang, *Member, IEEE*

***Abstract*— The generative adversarial networks (GANs) in continual learning suffer from catastrophic forgetting. In continual learning, GANs tend to forget about previous generation tasks and only remember the tasks they just learned. In this article, we present a novel conditional GAN, called the gradients orthogonal projection GAN (GopGAN), which updates the weights in the orthogonal subspace of the space spanned by the representations of training examples, and we also mathematically demonstrate its ability to retain the old knowledge about learned tasks in learning a new task. Furthermore, the orthogonal projection matrix for modulating gradients is mathematically derived and its iterative calculation algorithm for continual learning is given so that training examples for learned tasks do not need to be stored when learning a new task. In addition, a task-dependent latent vector construction is presented and the constructed conditional latent vectors are used as the inputs of generator in GopGAN to avoid the disappearance of orthogonal subspace of learned tasks. Extensive experiments on MNIST, EMNIST, SVHN, CIFAR10, and ImageNet-200 generation tasks show that the proposed GopGAN can effectively cope with the issue of catastrophic forgetting and stably retain learned knowledge.**

***Index Terms*— Catastrophic forgetting, continual learning, generative adversarial networks (GANs), orthogonal projection matrix.**

I. Introduction

GENERATIVE adversarial networks (GANs) [6] are the game of two players, i.e., generator and discriminator, which are widely applied in synthesizing high-quality data. For image synthesis, the generator takes latent vectors as input and is expected to generate high-quality synthesized images, whereas the discriminator is expected to effectively discriminate them from real images. With more effective losses [1], [7], [22] and new architectures [16], [27], [28], GANs have shown the remarkable generative ability.

Conditional GANs [24] are able to generate semantic images by embedding the conditions into the input of generator and discriminator. The various conditions in CGANs include label [27], text [41] and images [14], [42], and labels are the most common one [27].

It is unrealistic for agents to obtain all training data and learn once in practice, and we expect agents to acquire knowledge such as human beings, i.e., constantly learn in their lives and effectively learn new tasks quickly without forgetting the knowledge learned in the past. To achieve general artificial intelligence, it is necessary to endow agents with continual learning ability. However, the experiments show that it is still challenging now for deep neural networks to reserve the knowledge obtained in previously learned tasks when learning a new task, known as catastrophic forgetting [4], [23]. GANs also seriously suffer from catastrophic forgetting. For example, in sequential tasks training, a generator can only generate satisfactory images for the last task and forgets the ability ever owned for other previous tasks.

Recent years have witnessed some efforts of researchers to mitigate catastrophic forgetting. For example, the conceptor-aided backprop (CAB) method [10] and the orthogonal weight modification (OWM) method [39] have been proposed to cope with the issue in a novel way for continual learning. Concretely, an orthogonal projection matrix in each layer of networks is constructed to preserve the learned knowledge, and during learning a new task, the gradients in neural networks are projected to the orthogonal direction of all previously learned features by the orthogonal projection matrix. The updated weights can not only fit the new task but also retain the validity of encoding the old knowledge. However, for CGANs, the inputs of a generator are random noises independently sampled from the same distribution except label condition, and it brings about the practical disappearance of free memory space for unlearned tasks (we will explain the point in Section III-C).

Inspired by the OWM algorithms [10], [39], we present a new conditional GAN, called gradients orthogonal projection GAN (GopGAN), to generate synthetic data in the way of continual learning, and apply it for the sequential learning task of image synthesis. Fig. 1 shows the schematic of GopGAN. Here, we summarize the main contributions of this article as follows. First, we give a rigorous mathematical proof to show updating weights in the orthogonal subspace of feature space can guarantee that the learning of a new task does not result in forgetting the knowledge acquired in old tasks. Then, by defining and solving an optimization problem, we derive a new optimal solution of projection matrix. Furthermore,

Manuscript received 14 July 2020; revised 3 January 2021 and 31 May 2021; accepted 25 June 2021. Date of publication 16 July 2021; date of current version 5 January 2023. This work was supported in part by the National Key Research and Development Program of China under Contract 2017YFB1002203, in part by NSFC under Grant 61976201, in part by the NSFC Key Projects of International (Regional) Cooperation and Exchanges under Grant 61860206004, and in part by the Ningbo 2025 Key Project of Science and Technology Innovation under Grant 2018B10071. *(Corresponding author: Weiqiang Wang.)*

The authors are with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: lixiaobin161@mails.ucas.ac.cn; wqwang@ucas.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3093319.

Digital Object Identifier 10.1109/TNNLS.2021.3093319