KFC: KNOWLEDGE RECONSTRUCTION AND FEEDBACK CONSOLIDATION ENABLE EFFICIENT AND EFFECTIVE CONTINUAL GENERATIVE LEARNING

Libo Huang[1], **Zhulin An**[1*], **Yan Zeng**[2], **Xiang Zhi**[1], **Yongjun Xu**[1]
[1] Institute of Computing Technology, Chinese Academy of Sciences
[2] School of Mathematics and Statistics, Beijing Technology and Business University
{www.huanglibo, yanazeng013}@gmail.com, {anzhulin, zhixiang20g, xyj}@ict.ac.cn

ABSTRACT

To address the issues of catastrophic forgetting in Continual Generative Learning (CGL), dominant methods leverage the generative replay strategy. However, they often suffer from high time complexity and inferior generative sample quality. In this work, we develop an efficient and effective CGL method via *K*nowledge reconstruction and *F*eedback *C*onsolidation (*KFC*). KFC extends the inherent data reconstruction properties of the variational autoencoder framework to historical knowledge reconstruction and re-encodes the current task's reconstructed data to the same posterior distribution as the original data. Experiments showcase that KFC achieves state-of-the-art performances in time complexity, sample quality, and accuracy on various CGL tasks. Code is in github.com/libo-huang/KFC.

1 INTRODUCTION

Adapting generative models to continual learning (a.k.a. CGL) has triggered considerable interest in computer vision recently (Huang et al., 2024; Belouadah et al., 2021). The notorious problem of CGL is *catastrophic forgetting*, which reflects the fact that when a generator learns a new task, it forgets its previously learned tasks (Parisi et al., 2019). The dominant CGL method is *generative replay* (GR) (Shin et al., 2017; van de Ven et al., 2020), which retrains a new generator on a mixture dataset that combines the pseudo samples generated from the previous generator and the real samples of the current task. Some extended CGL methods train the generator only on the current-task data, e.g., CEWC (Seff et al., 2017) and MGAN (Wu et al., 2018; Liu et al., 2020), etc. However, these methods are majorly investigated on Conditional Generative Adversarial Networks (CGAN), and they are feasible for a single incremental task, whereas for multiple sequential tasks, CGAN could induce unstable training, resulting in inferior-quality samples (Cong et al., 2020).

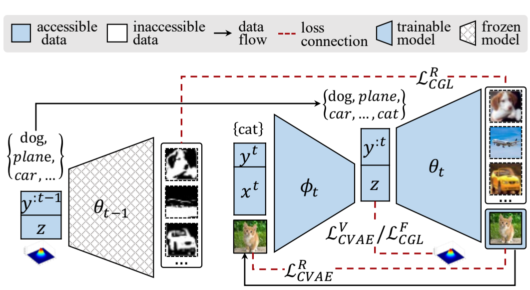

Figure 1: KFC enables CVAE (with $\mathcal{L}^V_{\mathrm{CVAE}}$ and $\mathcal{L}^R_{\mathrm{CVAE}}$) continual generative learning by reconstructing the historical knowledge retained in the frozen decoder with $\mathcal{L}^R_{\mathrm{CGL}}$ and feedback consolidating the reconstructed distribution with $\mathcal{L}^F_{\mathrm{CGL}}$.

FashionMNIST	Epoch	Time	ACC ↑
rCGAN	11	700s	58.90
CEWC	7	388s	61.67
MGAN	18	1230s	73.03
rCVAE	4	417s	73.69
KFC	**3**	**157**s	**75.28**

CIFAR10	Time	FID ↓
rCVAE/GR	12h	186.17
KFC	**7**h	**132.62**

Table 1: CGL models first reached optimal accuracy (ACC) on learning 10 FashionMNIST tasks with the consumed training epochs and times (in seconds). Well-trained CGL models get the Fréchet Inception Distance (FID) on learning 4 CIFAR10 tasks with the training times (in hours).

*Corresponding author: Zhulin An. This work was supported by the Beijing Natural Science Foundation under grant 4244098.

Motivated by the fact that Conditional Variational Autoencoder (CVAE) not only provides a stable training mechanism but keeps satisfactory sample diversity (Ramapuram et al., 2020), in this paper, we develop an efficient and effective CGL method based on CVAE via *K*nowledge reconstruction and *F*eedback *C*onsolidation (*KFC*). As shown in Fig.1, KFC continually trains the generator on the current-task data without pseudo samples from the previous generator as inputs. Inspired by the intrinsic reconstruction character of CVAE, KFC reconstructs the learned knowledge by training the current decoder guided by the old decoder's outputs with the same noise and label inputs. To improve the sample quality, KFC re-encodes the current reconstructed samples, consolidating these samples to hold the same posterior distribution as real ones. Experiments on the FashionMNIST and CIFAR10 CGL tasks verify that KFC enjoys high training efficiency and superior sample quality.

2 KFC: EFFICIENT AND EFFECTIVE CGL METHOD

KFC Framework. As shown in Fig.1, the whole training loss of KFC is,

$$\min_{\phi_t,\theta_t}\left\{\mathcal{L}^{R}_{CVAE}(\phi_t,\theta_t)+\mathcal{L}^{V}_{CVAE}(\phi_t)+\lambda^{r}_{t}\mathcal{L}^{R}_{CGL}\left(\theta_t\right)+\lambda^{f}_{t}\mathcal{L}^{F}_{CGL}\left(\phi_t,\theta_t\right)\right\}, \tag{1}$$

By minimizing conventional CVAE losses, $\mathcal{L}^{R}_{CVAE}(\phi_t,\theta_t)+\mathcal{L}^{V}_{CVAE}(\phi_t)$, the conditional encoder, ϕ_t, promotes $\boldsymbol{z}$ to follow the Gaussian distribution from the sample, $\boldsymbol{x}^t$, and the label, y^t, while the conditional decoder, θ_t, reconstructs the surrogate sample of the original $\boldsymbol{x}^t$ from y^t and $\boldsymbol{z}$. By minimizing the proposed knowledge reconstruction loss, $\mathcal{L}^{R}_{CGL}(\theta_t)$, θ_t could reconstruct the knowledge embedded in the historical decoder, θ_{t-1}, and by minimizing the proposed feedback consolidation loss, $\mathcal{L}^{F}_{CGL}\left(\phi_t,\theta_t\right)$, θ_t could improve the samples' quality as ϕ_t promotes the reconstructed data obtained from θ_t to follow the same posterior distribution as the current task data.

Knowledge Reconstruction. Inspired by the CVAEs' intrinsic reconstruction character (Kingma & Welling, 2014; Paisley et al., 2012), we propose a knowledge reconstruction loss, $\mathcal{L}^{R}_{\mathrm{CGL}}(\cdot)$, to track the forgetting problem.

$$\mathcal{L}^{R}_{\mathrm{CGL}}(\theta_t)=-\mathbb{E}_{p_{\theta_{t-1}}(\hat{\boldsymbol{x}}^{t-1}|y^{t-1},\boldsymbol{z}),y\sim U(1,|\boldsymbol{y}^{:t-1}|),\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})}\left[\log p_{\theta_t}(\hat{\boldsymbol{x}}^{t-1}|y,\boldsymbol{z})\right], \tag{2}$$

where $\mathbb{E}$ is the expectation operation, $|\cdot|$ is the cardinality operator. $U(\cdot)$ and $\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$ are discrete uniform and Gaussian distribution, respectively. $\boldsymbol{y}^{:t-1}=\{y^1,...,y^{t-1}\}$ indicates the set of task labels learned so far. By minimizing $\mathcal{L}^{R}_{\mathrm{CGL}}(\cdot)$, the current encoder, p_{θ_t}, could *reconstruct* the historical knowledge with the stored $p_{\theta_{t-1}}$ and the number of previous class labels, $\left|\boldsymbol{y}^{:t-1}\right|$.

Feedback Consolidation. Inspired by the CGANs' discriminator-driven generation (Che et al., 2020; Verma et al., 2018), we propose a feedback consolidation loss, $\mathcal{L}^{F}_{\mathrm{CGL}}(\cdot)$, to improve the sample quality.

$$\mathcal{L}^{F}_{\mathrm{CGL}}(\phi_t,\theta_t)=\mathbb{E}_{p_{\theta_t}(\hat{\boldsymbol{x}}^t|y^t,\boldsymbol{z})}\left[\mathrm{KL}\left[q_{\phi_t}\left(\boldsymbol{z}\mid y^t,\hat{\boldsymbol{x}}^t\right)\parallel p(\boldsymbol{z})\right]\right]. \tag{3}$$

It decodes the current task's reconstructed data, $\hat{\boldsymbol{x}}^t$, by using $p_{\theta_t}(\hat{\boldsymbol{x}}^t|y^t,\boldsymbol{z})$ at first, then *re-encode* $\hat{\boldsymbol{x}}^t$ to a latent variable $\boldsymbol{z}$ by using $q_{\phi_t}\left(\boldsymbol{z}\mid y^t,\hat{\boldsymbol{x}}^t\right)$. The Kullback–Leibler (KL) divergence ensures the real data of the current task and those obtained from the decoder follow the same posterior distribution, improving generative samples' quality (Che et al., 2020)[1].

3 EXPERIMENTS

We compare KFC with several state-of-the-art baselines, rCGAN (Ye & Bors, 2021), CEWC (Seff et al., 2017), MGAN (Liu et al., 2020), and rCVAE (van de Ven et al., 2020) on FashionMNIST and CIFAR10. As depicted in Tab. 1, KFC surpasses others evidently in terms of efficiency (lower training epochs and time to reach its better optimal ACC for the first time on 10 FashionMNIST tasks) and effectiveness (lower training time to reach its better FID on 4 CIFAR10 tasks)[2].

4 CONCLUSION

We developed KFC, an efficient and effective continual generative learning method based on the CVAE framework. It bridges the gap that no VAE-based CGL methods exist taking the current-task data as inputs to train the generative model continually. Experiments on FashionMNIST and CIFAR10 CGL tasks verify that KFC achieves continual generation more effectively and efficiently than existing CGL methods.

[1] Detailed method is provided in Appendix A.

[2] Extensive experiments are provided in Appendix B

URM STATEMENT

The authors acknowledge that key authors of this work meet the URM criteria of the ICLR 2024 Tiny Papers Track.

REFERENCES

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning (ICML)*, pp. 214–223. PMLR, 2017.

Benedikt Bagus, Alexander Gepperth, and Timothée Lesort. Beyond supervised continual learning: a review. *arXiv preprint arXiv:2208.14307*, 2022.

Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks*, 135:38–54, 2021.

Tong Che, Ruixiang Zhang, Jascha Sohl-Dickstein, Hugo Larochelle, Liam Paull, Yuan Cao, and Yoshua Bengio. Your gan is secretly an energy-based model and you should use discriminator driven latent sampling. *Advances in Neural Information Processing Systems (NeurIPS)*, 33: 12275–12287, 2020.

Yulai Cong, Miaoyun Zhao, Jianqiao Li, Sijia Wang, and Lawrence Carin. Gan memory with no forgetting. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:16481–16494, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision (ECCV)*, pp. 630–645. Springer, 2016.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems (NIPS)*, 30, 2017.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Libo Huang, Yan Zeng, Chuanguang Yang, Zhulin An, Boyu Diao, and Yongjun Xu. etag: Class-incremental learning via embedding distillation and task-oriented generation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 12591–12599, 2024.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

Xialei Liu, Chenshen Wu, Mikel Menta, Luis Herranz, Bogdan Raducanu, Andrew D Bagdanov, Shangling Jui, and Joost van de Weijer. Generative feature replay for class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 226–227, 2020.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. *Workshop on Deep Learning and Unsupervised Feature Learning, Neural Information Processing System (NIPSW)*, 2011.

John Paisley, David M Blei, and Michael I Jordan. Variational bayesian inference with stochastic search. In *In International Conference on Machine Learning (ICML)*, pp. 1367–1374, 2012.

German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: a review. *Neural Networks*, 113:54–71, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Jason Ramapuram, Magda Gregorova, and Alexandros Kalousis. Lifelong generative modeling. *Neurocomputing*, 404:381–400, 2020.

Ari Seff, Alex Beatson, Daniel Suo, and Han Liu. Continual learning in generative adversarial nets. *arXiv preprint arXiv:1705.08395*, 2017.

Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In *International Conference on Neural Information Processing Systems (NIPS)*, pp. 2994–3003, 2017.

Teik Toe Teoh and Zheng Rong. Deep convolutional generative adversarial network. In *Artificial Intelligence with Python*, pp. 289–301. Springer, 2022.

Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. In *International Conference on Learning Representations (ICLR)*, 2018.

Gido M van de Ven, Hava T Siegelmann, and Andreas S Tolias. Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, 11(1):1–14, 2020.

Vinay Kumar Verma, Gundeep Arora, Ashish Mishra, and Piyush Rai. Generalized zero-shot learning via synthesized examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4281–4289, 2018.

Chenshen Wu, Luis Herranz, Xialei Liu, Joost Van De Weijer, Bogdan Raducanu, et al. Memory replay gans: Learning to generate new categories without forgetting. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Fei Ye and Adrian Bors. Lifelong teacher-student network learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Hongxu Yin, Pavlo Molchanov, Jose M Alvarez, Zhizhong Li, Arun Mallya, Derek Hoiem, Niraj K Jha, and Jan Kautz. Dreaming to distill: Data-free knowledge transfer via deepinversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8715–8724, 2020.

Mengyao Zhai, Lei Chen, Frederick Tung, Jiawei He, Megha Nawhal, and Greg Mori. Lifelong gan: continual learning for conditional image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2759–2768, 2019.