KFC: Knowledge Reconstruction and Feedback Consolidation Enable Efficient and Effective Generative Learning

ICLR-24 (tiny paper)

[1]Libo Huang [1]Zhulin An [2]Yan Zeng [1]Xiang Zhi [1]Yongjun Xu

[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]School of Mathematics and Statistics, Beijing Technology and Business University

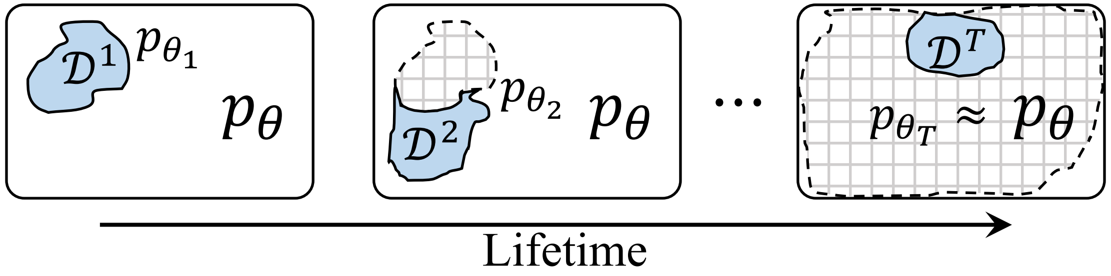
Outline

- Background
- Problem
- Method
- Experiment
- Conclusion

Background

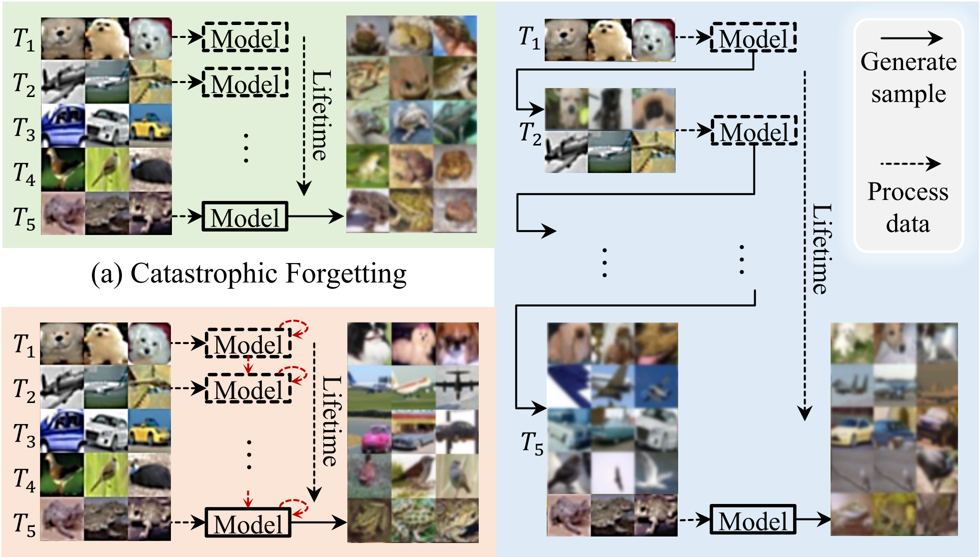
Continual Generative Learning

- Learning sequence tasks
- Catastrophic forgetting



Existing solutions

- Generative replay
  - High time complexity
- Train on current-task data
  - Unstable training
  - Inferior-quality sample



(a) Catastrophic Forgetting

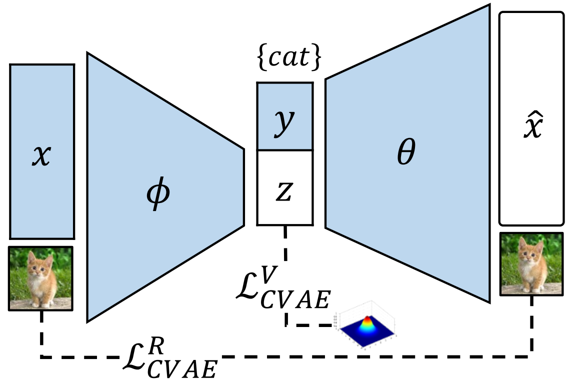
(c) KFC

(b) Generative Replay

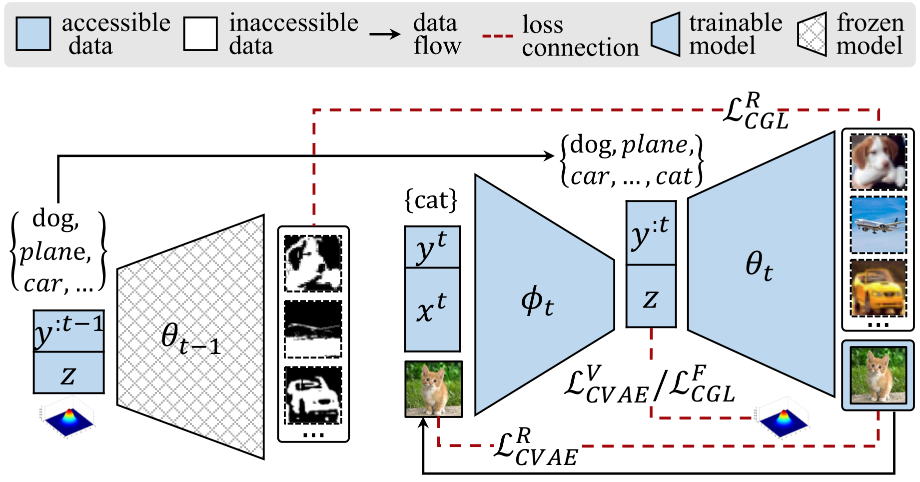
Problem

- Sequentially generating T tasks without forgetting
  - With only current-task data
  - With stable training
  - Without inferior-quality sample

- Motivation from VAE
  - Sample reconstruction character of VAE/Diffusion
  - Stable training of reconstruction training (not adversarial training)
  - Discriminator-driven generation of GAN



(a) CVAE framework



(b) KFC framework

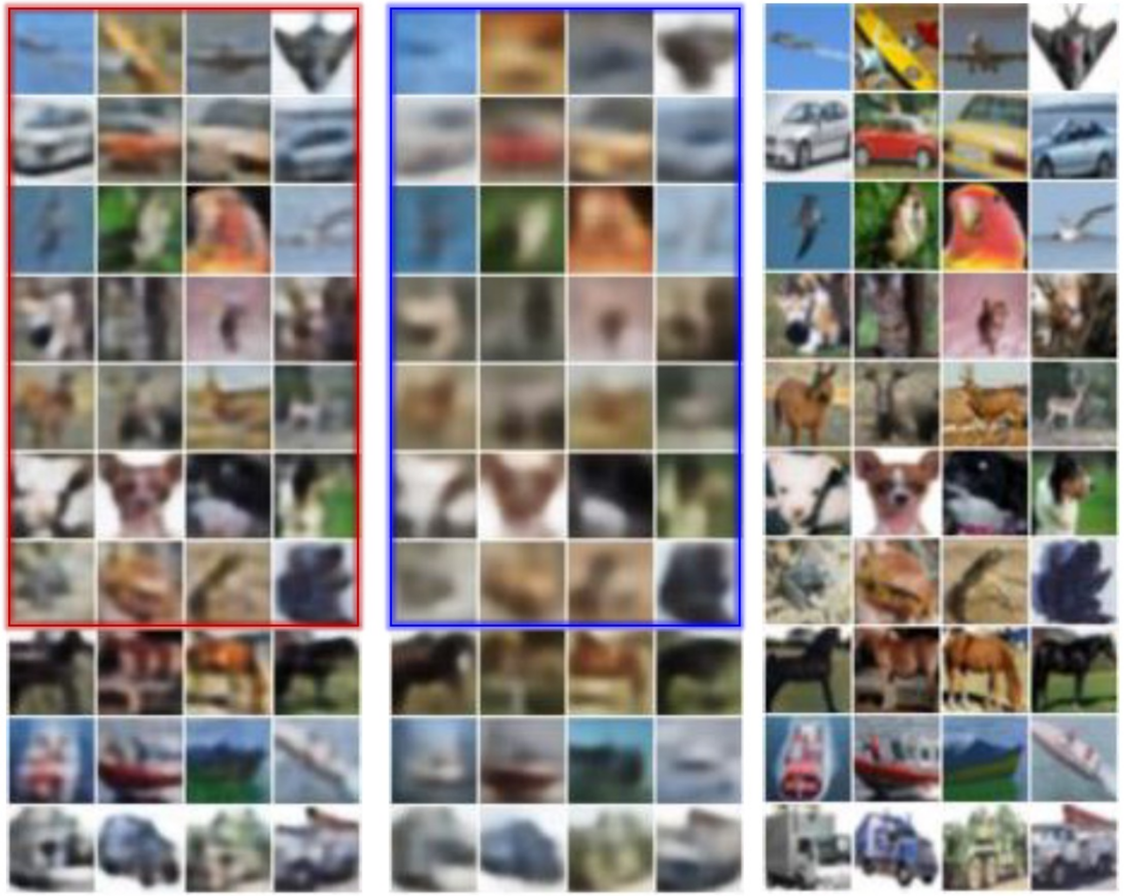
Method

- no VAE-based CGL without replay
  - VAE's stable training
  - $\min_{\phi_t,\theta_t}\{\mathcal{L}^{R}_{CVAE}(\phi_t,\theta_t)+\mathcal{L}^{V}_{CVAE}(\phi_t)+\lambda^{\mathrm{r}}_t\mathcal{L}^{R}_{\mathrm{CGL}}(\theta_t)+\lambda^{f}_t\mathcal{L}^{F}_{CGL}(\phi_t,\theta_t)\}$
- Knowledge reconstruction
  - Extend VAE's sample reconstruction
  - $\mathcal{L}^{R}_{\mathrm{CGL}}(\theta_t)=-\mathbb{E}_{p_{\theta_{t-1}}(\hat{\boldsymbol{x}}^{t-1}|y^{t-1},\boldsymbol{z}),y\sim U(1,|\boldsymbol{y}^{:t-1}|),\boldsymbol{z}\sim N(\boldsymbol{0},\boldsymbol{I})}\left[\log p_{\theta_t}(\hat{\boldsymbol{x}}^{t-1}|y,\boldsymbol{z})\right]$
- Feedback consolidation
  - Re-encode reconstructed data
  - $\mathcal{L}^{F}_{CGL}(\phi_t,\theta_t)=\mathbb{E}_{\log p_{\theta_t}(\hat{x}^t|y^t,\boldsymbol{z})}\left[\mathrm{KL}\left[q_{\varphi_t}(\boldsymbol{z}|y^t,\hat{x}^t)\|p(\boldsymbol{z})\right]\right]$

Experiments

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).



(a) KFC (b) GR (c) CIFAR10

Experiments: *limited* resource

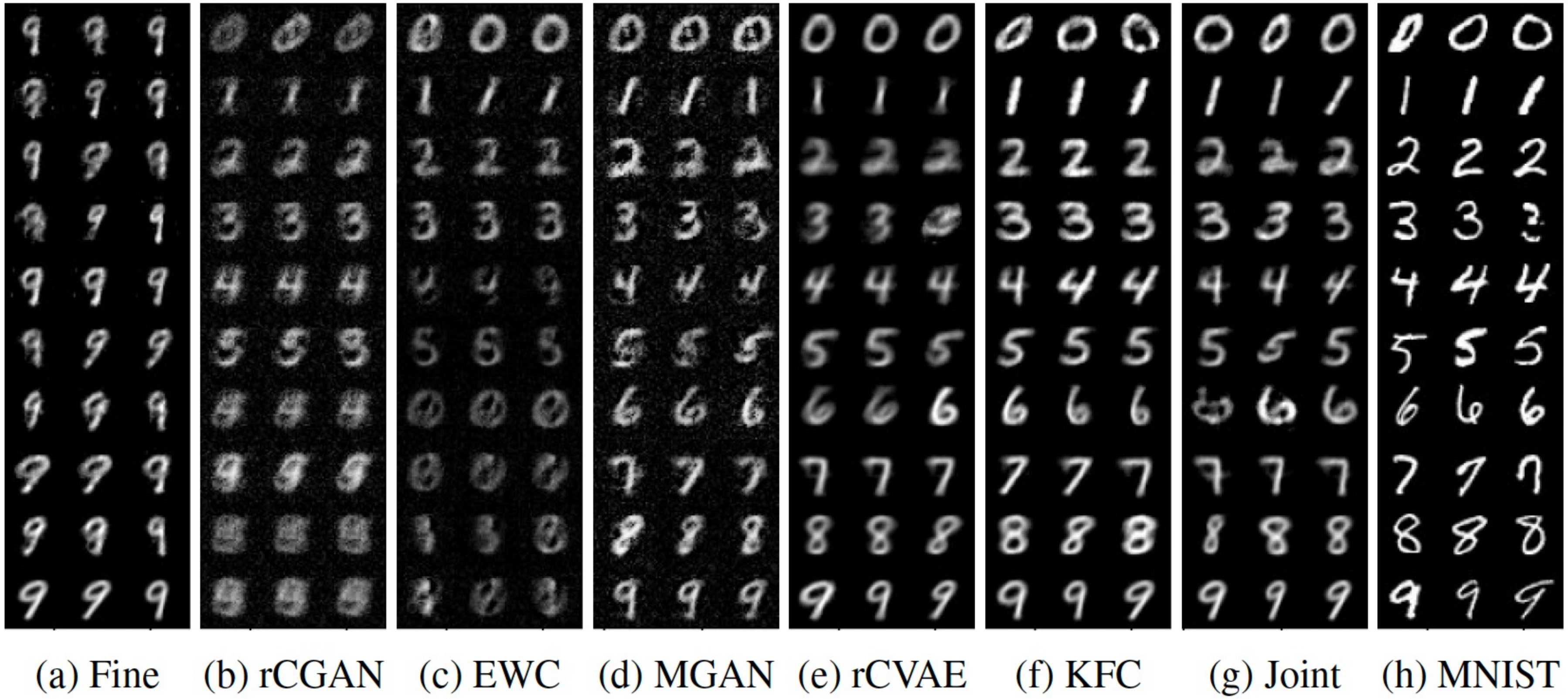
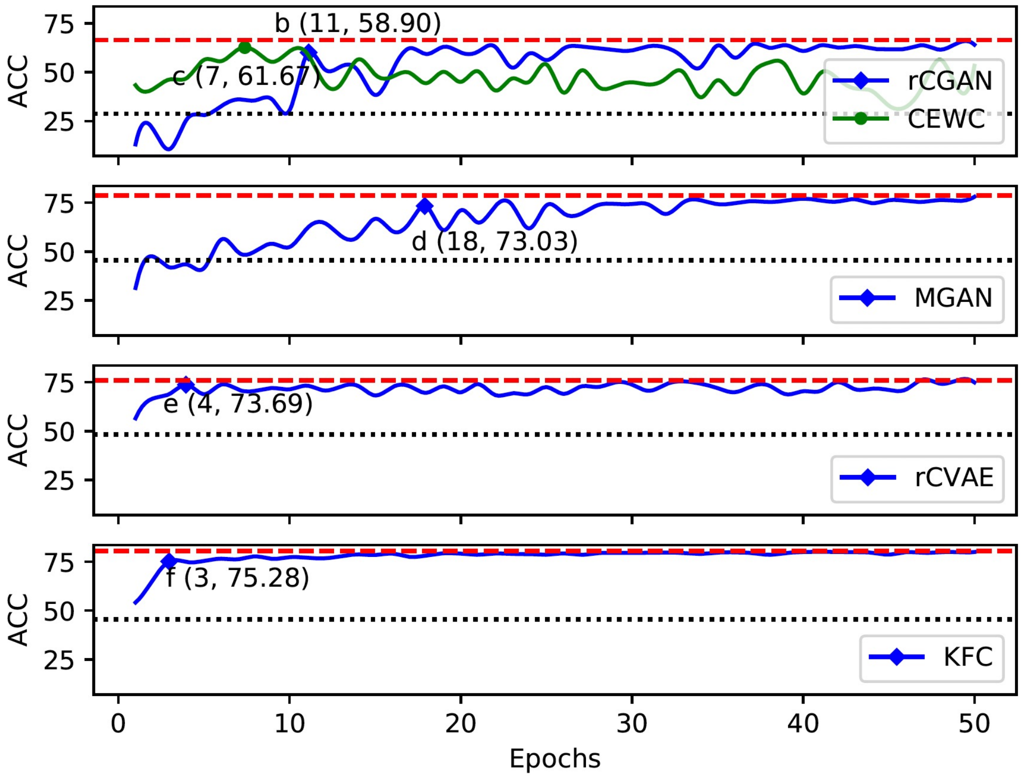
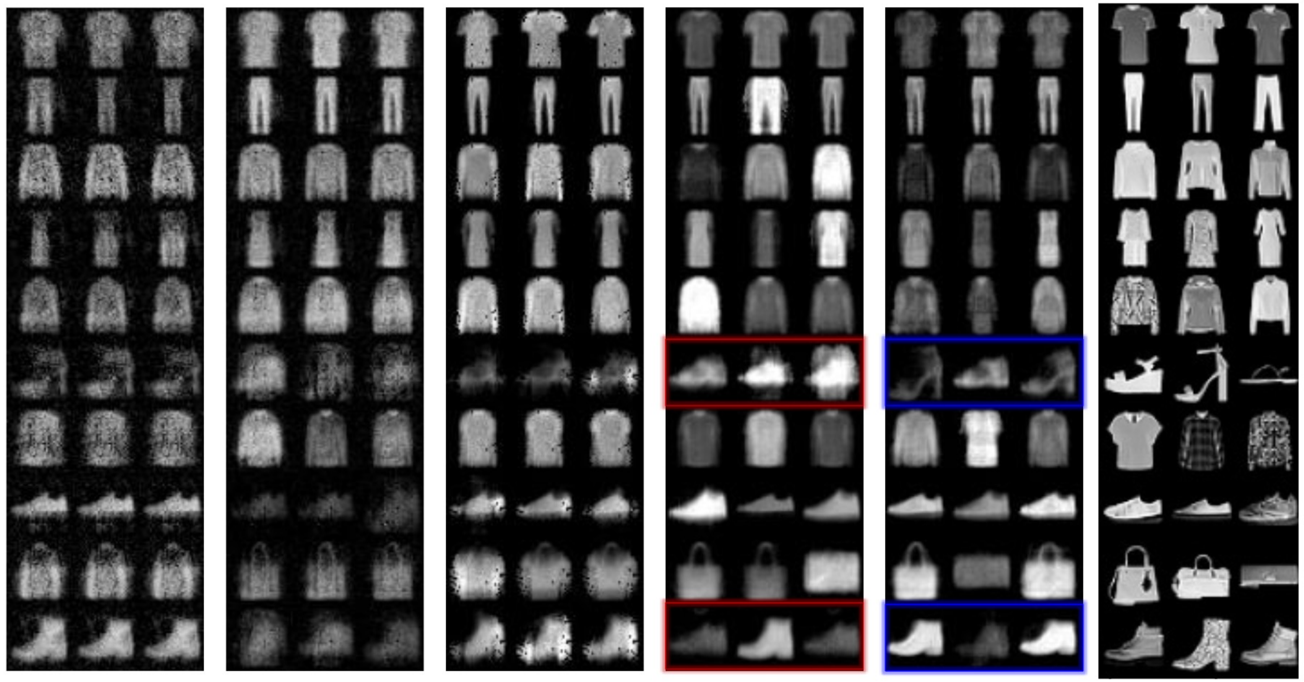


Figure A1: The generated digits after the each method of (a) Fine, (b) rCGAN, (c) CEWC, (d) MGAN, (e) rCVAE, (f) KFC, (g) Joint sequentially learn (h) 10 MNIST tasks.

Experiments: *unlimited* resource



(a) ACCs on FashionMNIST



(b) rCGAN (c) CEWC (d) MGAN (e) rCVAE (f) KFC (g) FashionMNIST

Experiments: *ablations*

Task	rCGAN		CEWC		MGAN		rCVAE		Fine-Tuning		Baseline1		Baseline2		KFC	
	ACC	FID	ACC	FID	ACC	FID	ACC	FID	ACC	FID	ACC	FID	ACC	FID	ACC	FID
1	87.95 (±2.92)	270.28	**88.45** (±9.01)	**202.65**	87.05 (±3.09)	233.43	87.35 (±0.72)	258.99	86.29 (±1.31)	271.00	86.69 (±1.33)	257.02	86.79 (±0.86)	246.32	87.95 (±0.57)	241.98
2	66.17 (±6.93)	194.66	50.05 (±2.80)	287.93	67.76 (±1.66)	204.61	66.65 (±0.59)	181.46	44.37 (±2.02)	215.40	46.96 (±2.21)	266.10	60.39 (±0.13)	195.10	**69.33** (±0.28)	**150.73**
3	55.75 (±4.49)	261.94	44.22 (±0.68)	275.06	59.92 (±0.60)	166.89	58.93 (±0.92)	176.80	38.46 (±2.24)	293.71	41.09 (±1.50)	232.41	54.42 (±0.88)	191.57	**64.45** (±0.70)	**141.74**
4	50.09 (±0.52)	237.22	38.47 (±0.91)	347.21	56.01 (±3.93)	174.95	55.11 (±0.30)	186.17	35.11 (±0.91)	214.60	36.83 (±0.96)	262.85	49.95 (±0.73)	187.46	**61.11** (±0.75)	**132.62**
T, P	$16.63h, 4.03m$		$10.42h, 4.03m$		$10.02hs, 4.01m$		$11.88h, 4.01m$		$5.98hs, 4.02m$		$6.76h, 4.02m$		$6.20h, 4.02m$		$6.89h, 4.02m$	

Table A1: ACC (%) (± std) and FID of various CGL methods evaluated on 4 sequential CIFAR-10 tasks with well trained models. T and P indicate the whole training time (in hours) and the number of network parameters (in mega), respectively.

Conclusion

- Existing CGL methods
  - *High time complexity* of generative replay (GAN & VAE)
  - *Inferior-quality sample* of training on current-task data (GAN only)
- We propose KFC
  - Use VAE's stable training
  - Propose knowledge reconstruction
  - Propose Feedback consolidation
- Improves VAE-based CGL method *efficiently* and *effectively*
  - MNIST, FashionMNIST, CIFAR10

Thanks for your attention!

Libo Huang
huanglibo@ict.ac.cn

openreview.net/pdf?id=pVTcR8ig3R

github.com/libo-huang/KFC

linkedin.com/in/lbhuang/

WeChat: Libo Huang