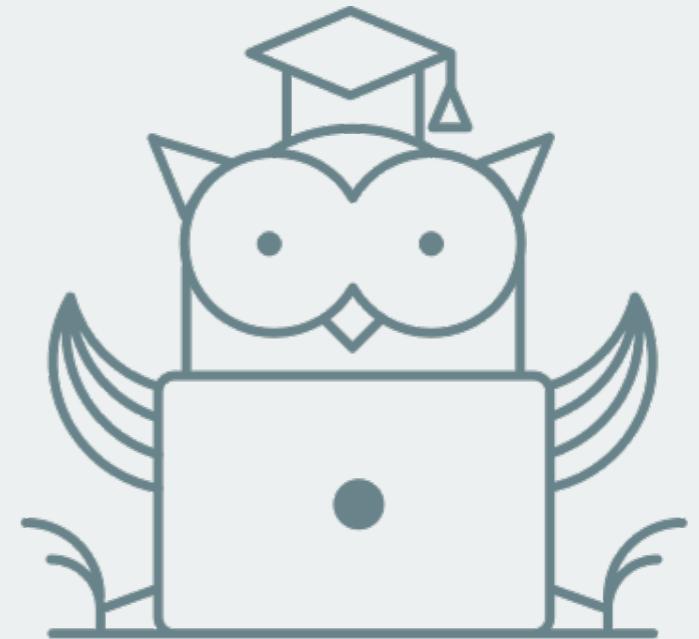




ОНЛАЙН-ОБРАЗОВАНИЕ

Катастрофическое забывание

Артур Кадурин
CEO Insilico Taiwan



- 1. Проблема катастрофического забывания**
2. Elastic Weight Consolidation
3. Deep Generative replay
4. Memory replay GAN



Transfer learning or fine-tuning



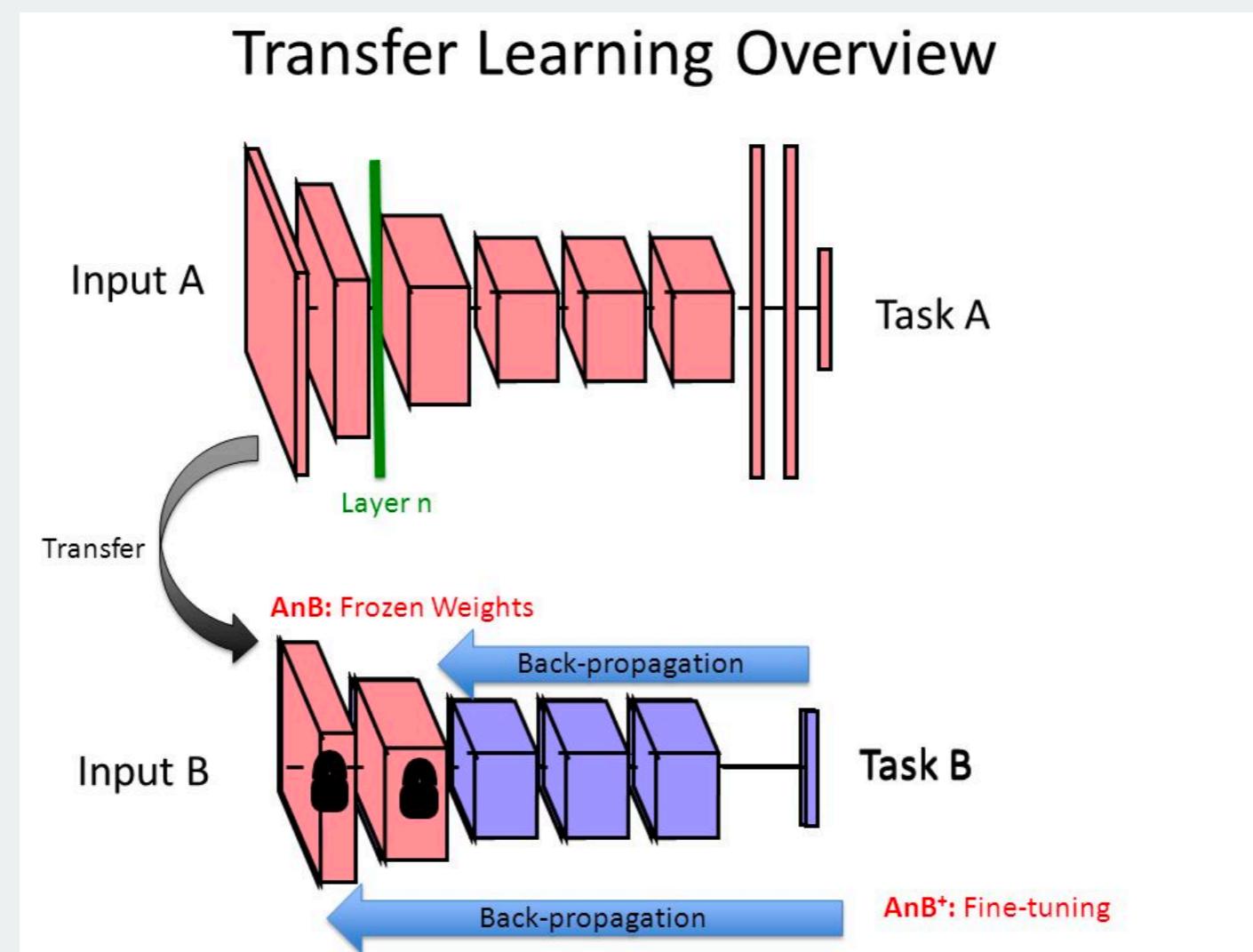
В реальной жизни очень часто возникают задачи когда на относительно небольшом датасете надо быстро поставить эксперимент по классификации. Если мы попробуем обучить какую-нибудь глубокую модель с нуля, то, скорее всего, она недообучится или переобучится. В этом случае часто применяется подход с переиспользованием уже обученных сеток.



Transfer learning or fine-tuning



В реальной жизни очень часто возникают задачи когда на относительно небольшом датасете надо быстро поставить эксперимент по классификации. Если мы попробуем обучить какую-нибудь глубокую модель с нуля, то, скорее всего, она недообучится или переобучится. В этом случае часто применяется подход с переиспользованием уже обученных сеток.

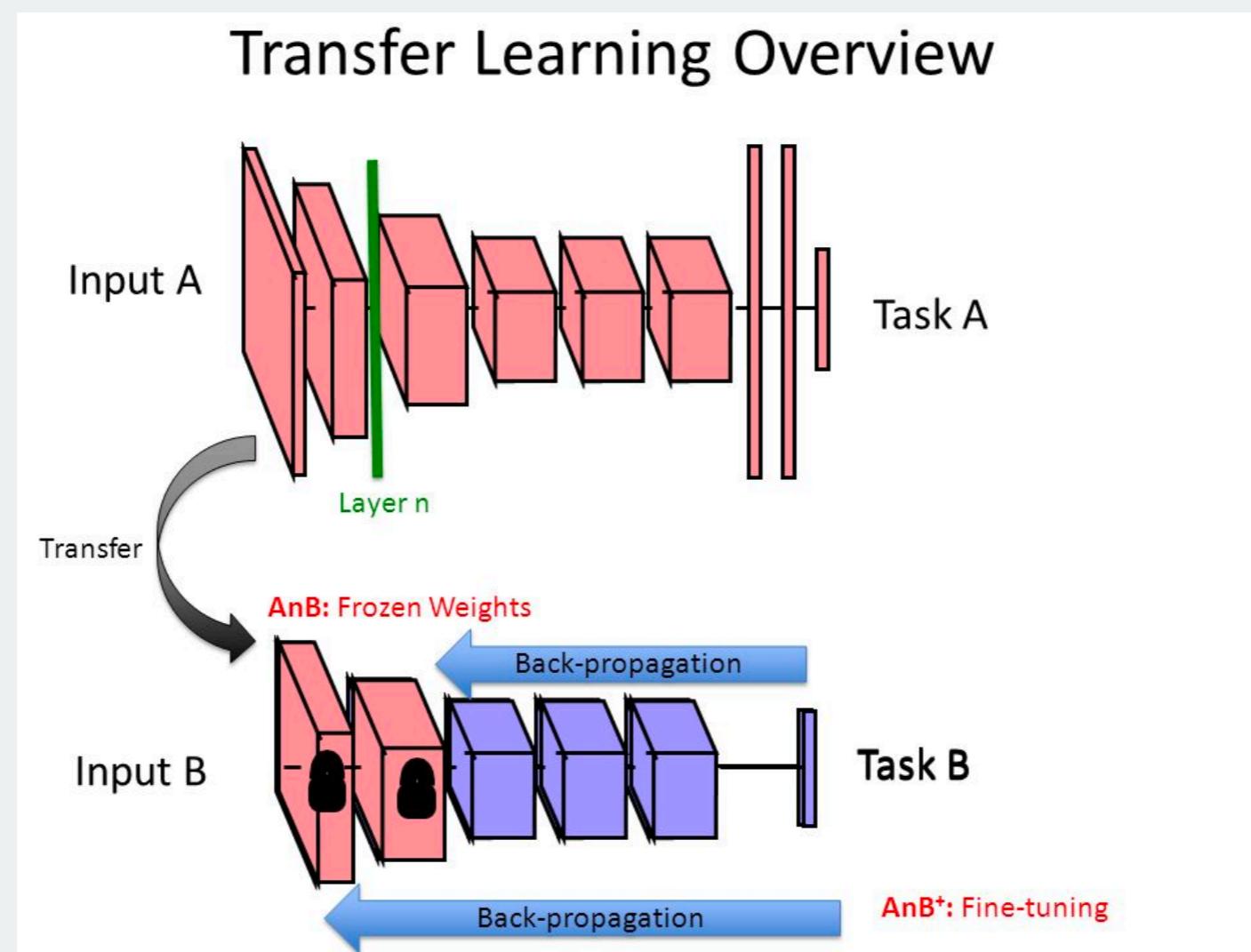


Catastrophic forgetting



Одна из больших проблем с таким подходом заключается в том, что нейронная сеть «забывает» как решать исходную задачу.

Какие могут быть способы решения этой проблемы?



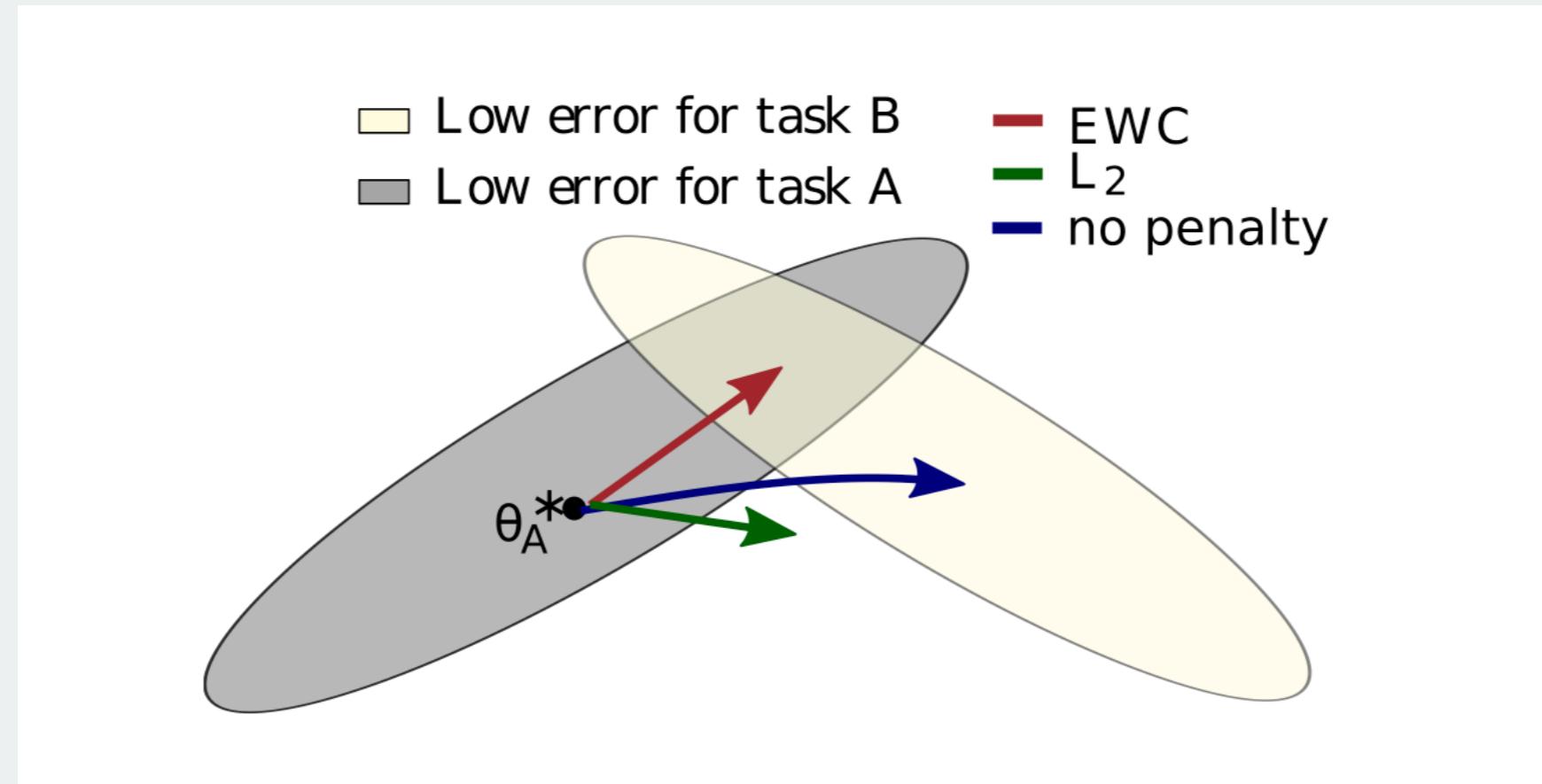
1. Проблема катастрофического забывания
2. **Elastic Weight Consolidation**
3. Deep Generative replay
4. Memory replay GAN



Elastic Weight Consolidation



Ключевая идея EWC - «правильно» взвесить параметры модели, обученной на предыдущей задаче, при обучении новой.



$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i (\theta_i - \theta_{A,i}^*)^2$$

James Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks



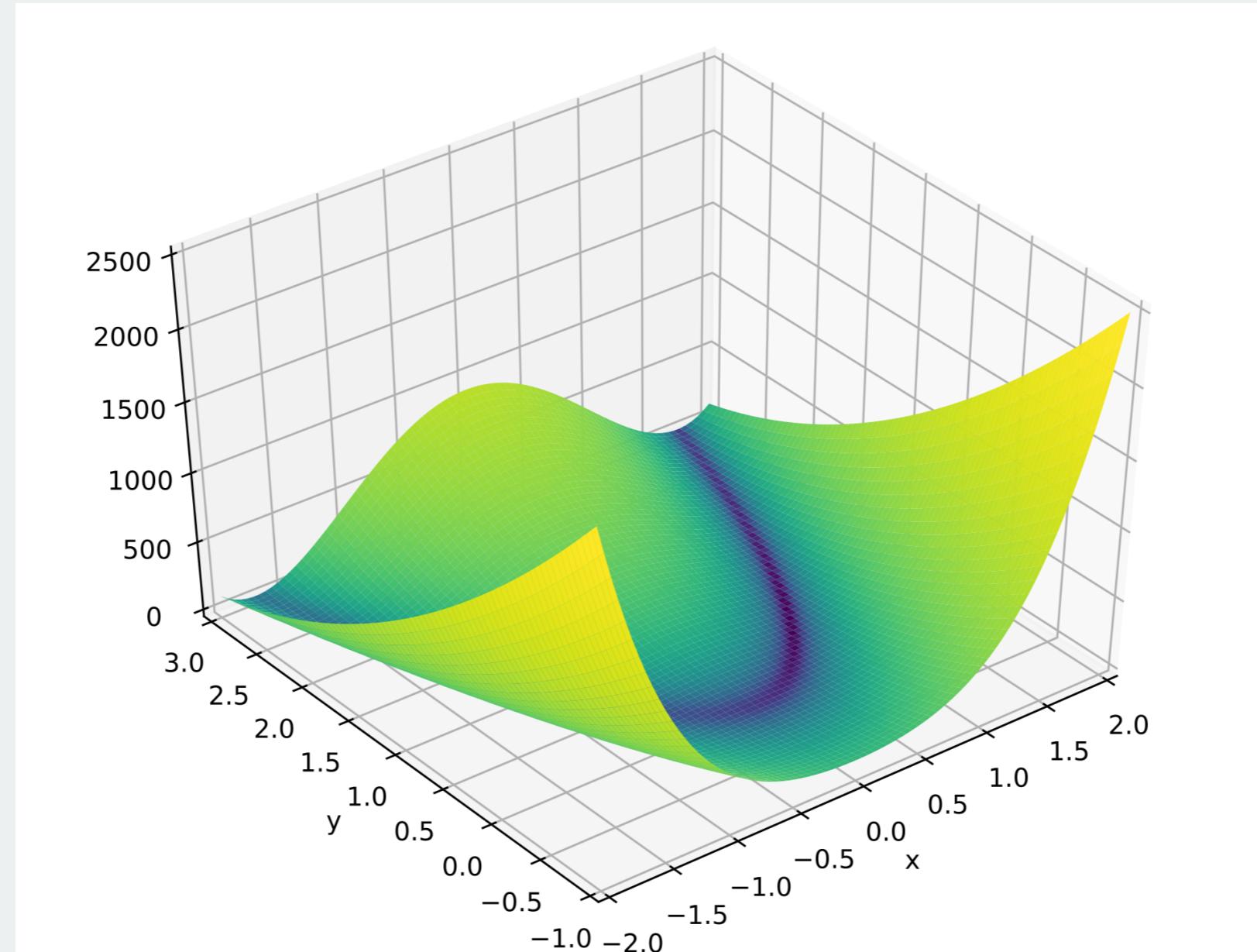
$$[\mathcal{I}(\theta)]_{i,j} = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \log f(X; \theta) \right) \left(\frac{\partial}{\partial \theta_j} \log f(X; \theta) \right) \middle| \theta \right].$$

1. Эквивалентна второй производной (гессиану) функции ошибки рядом с минимумом
2. Для вычисления достаточно знать первые производные
3. Гарантированно положительная полуопределенная функция

James Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks



Elastic Weight Consolidation



(Это функция Розенброка. Ее назначение - тестировать алгоритмы оптимизации. Но мы используем ее как визуализацию для EWC)

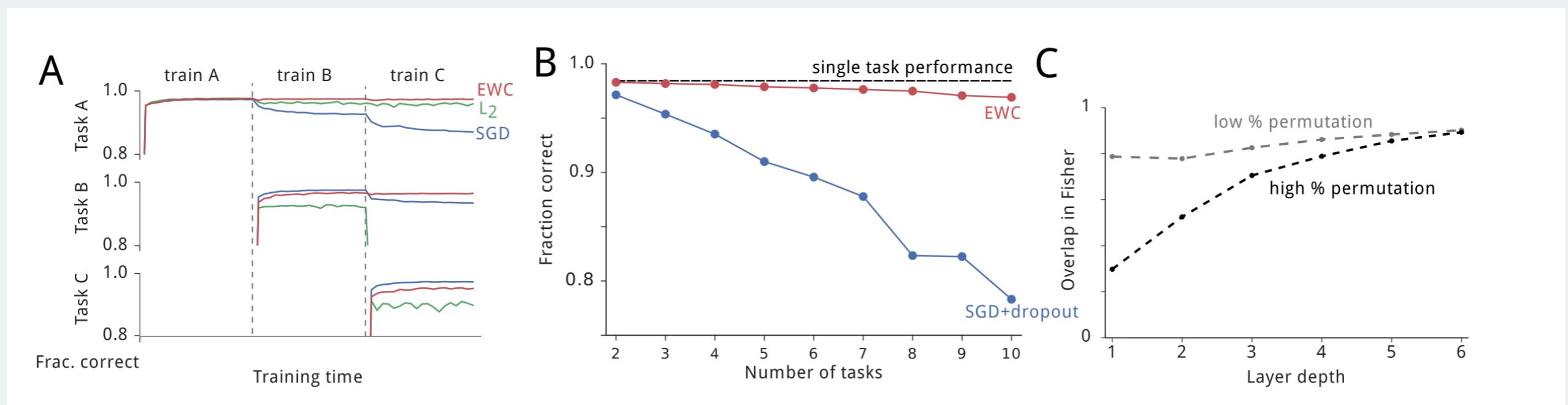
James Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks



Elastic Weight Consolidation



Новая задача создавалась как фиксированная перестановка части пикселей в изображениях MNIST



James Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks



1. Проблема катастрофического забывания
2. Elastic Weight Consolidation
3. **Deep Generative replay**
4. Memory replay GAN



Generative replay



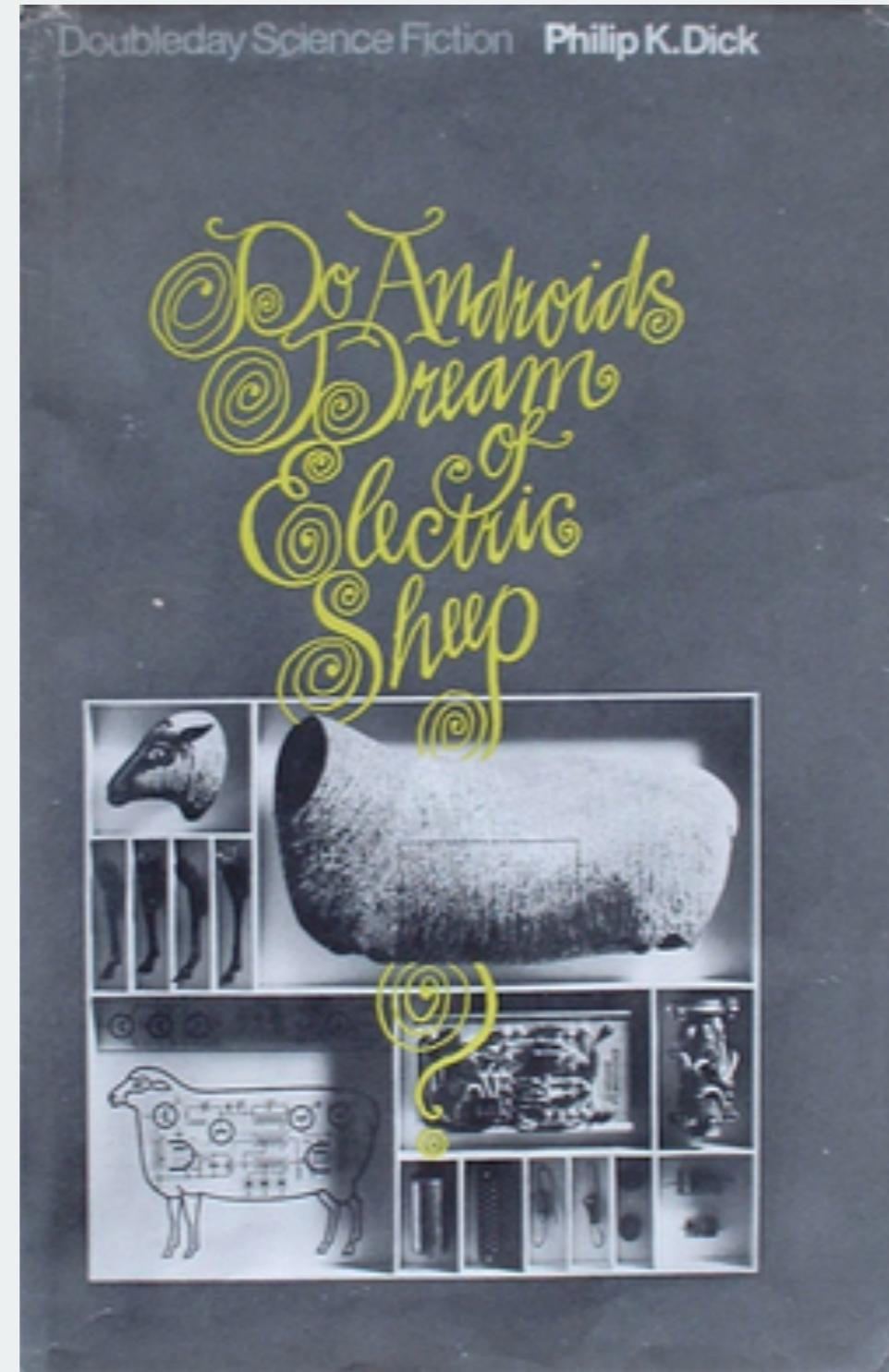
К счастью (или к сожалению), люди не сохраняют тренировочные примеры из прошлого для того чтобы не разучиться ходить в процессе обучения вождению. Современная теория в качестве возможного объяснения этой волшебной способности предлагает такой механизм: гиппокамп сохраняет новый опыт в течение дня, а во время сна, с помощью многократных повторов, эта информация объединяется с той что уже есть в коре.

Фактически, у нас есть две памяти и время от времени они синхронизируются.



Generative replay

O T U S

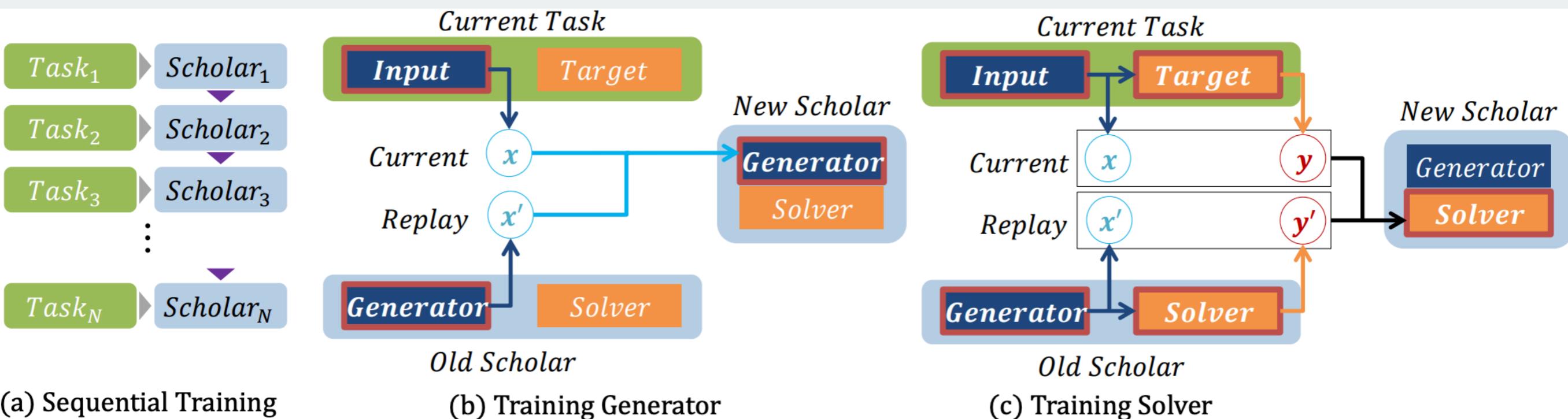


Generative replay



К счастью (или к сожалению), люди не сохраняют тренировочные примеры из прошлого для того чтобы не разучиться ходить в процессе обучения вождению. Современная теория в качестве возможного объяснения этой волшебной способности предлагает такой механизм: гиппокамп сохраняет новый опыт в течение дня, а во время сна, с помощью многократных повторов, эта информация объединяется с той что уже сохранена в коре.

Фактически, у нас есть две памяти и время от времени они синхронизируются.



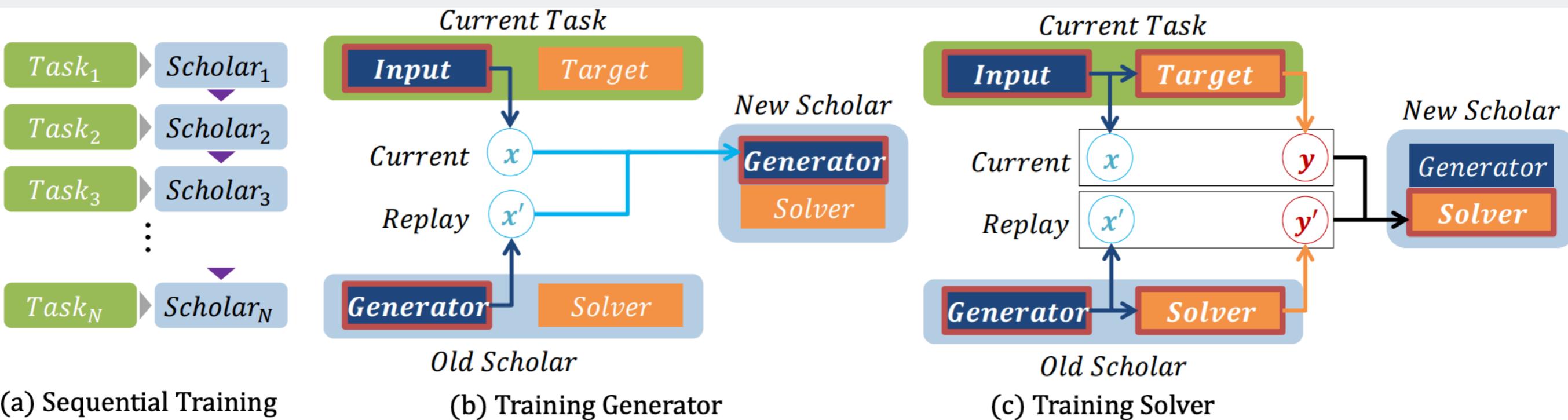
Hanul Shin et al. Continual Learning with Deep Generative Replay



Generative replay



- Обучение n+1 генератора: он учится генерировать распределение, воспроизводящее распределение n-го генератора и новых данных в заданной пропорции
- Генератор - не обязательно GAN
- Обучение n+1 солвера: он учится решать задачу на объединенных сгенерированных и реальных данных; метки для сгенерированных данных он получает от n-го генератора и солвера

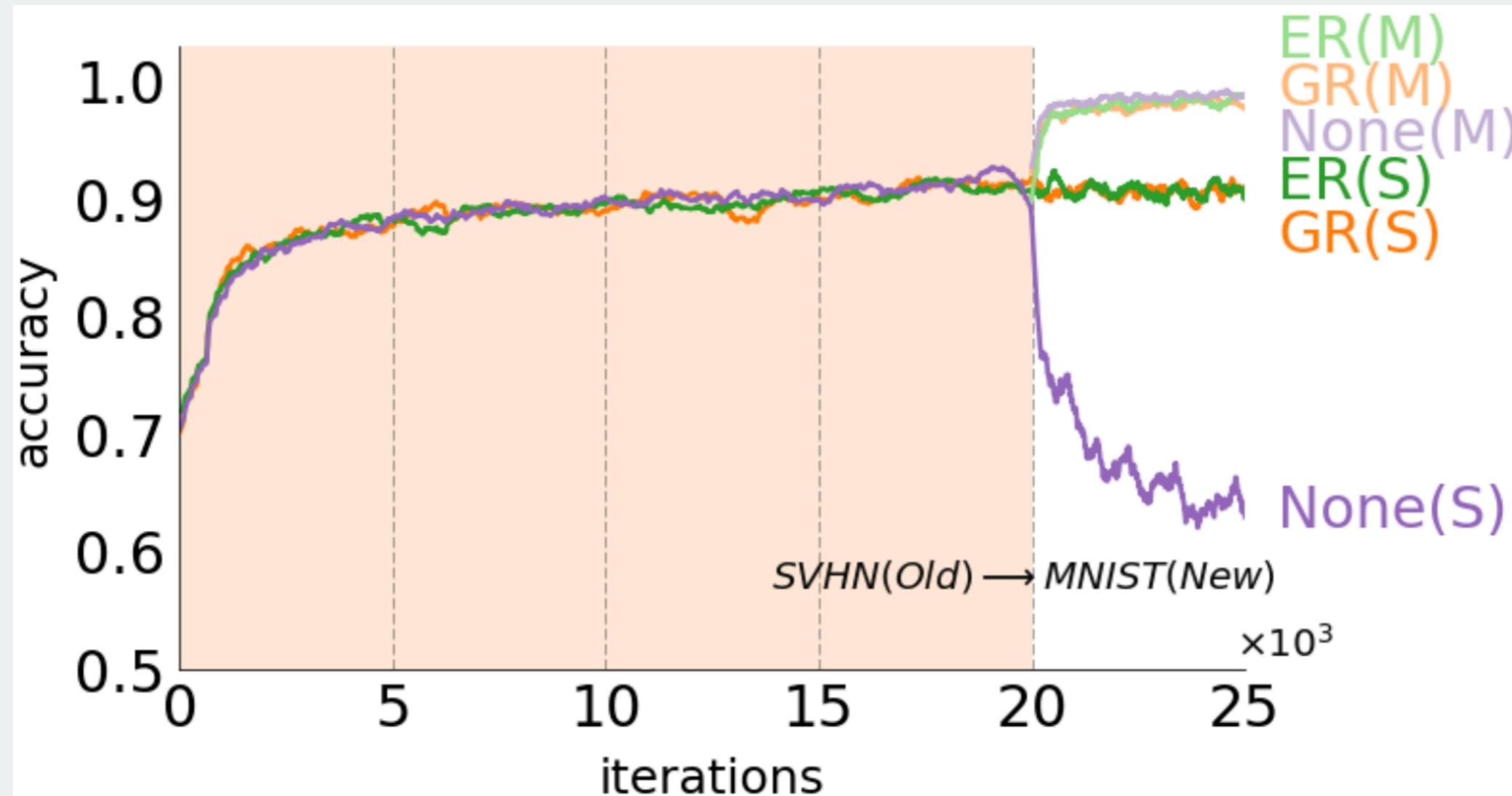


Hanul Shin et al. Continual Learning with Deep Generative Replay



Generative replay

O T U S



Hanul Shin et al. Continual Learning with Deep Generative Replay



Generative replay



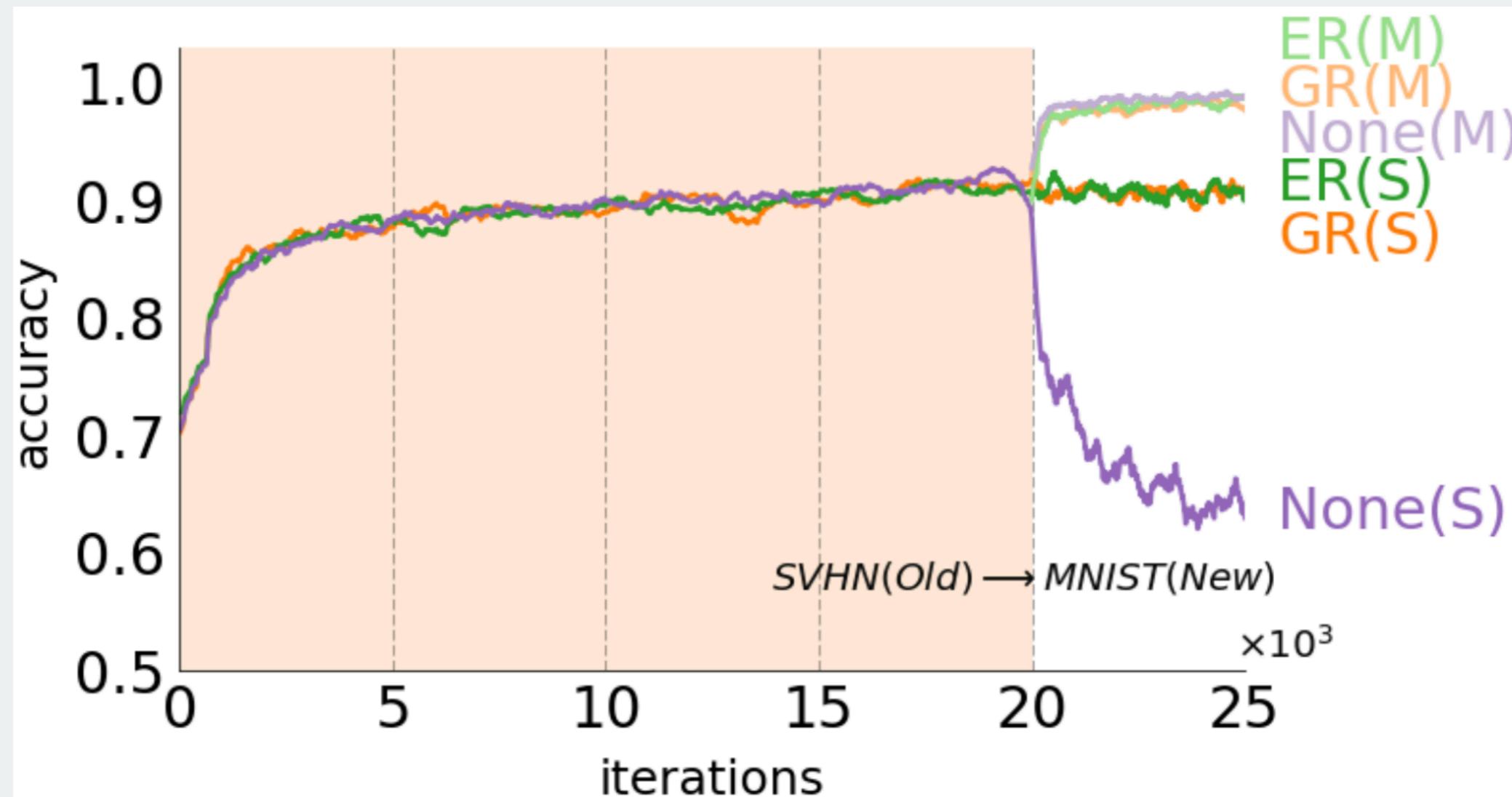
Генератор после каждой задачи создает примеры из кумулятивного распределения

11111010	00103131	58413555	20672562	07255220
11110111	31110212	32203055	00721086	10174356
10010111	31801102	64400454	64942054	74842091
11111111	32233311	42540145	35726160	19438419
01110010	18103202	93321554	21093514	48489082
01111110	32220831	40253410	65732342	64917195
00110000	00301011	01122511	57304633	93926378
01000010	07300303	42113204	66744210	14837538

Hanul Shin et al. Continual Learning with Deep Generative Replay



Generative replay



Идеи для проекта?

Hanul Shin et al. Continual Learning with Deep Generative Replay



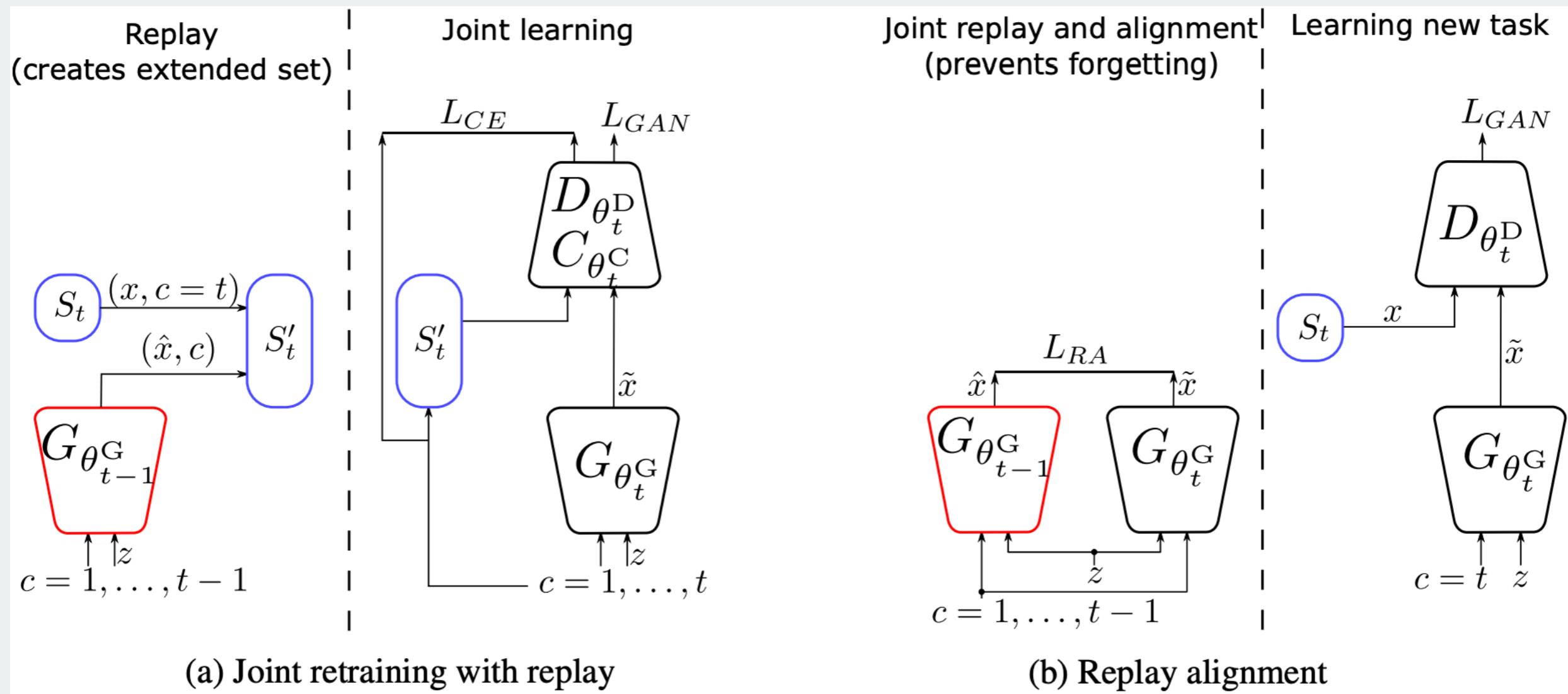
1. Проблема катастрофического забывания
2. Elastic Weight Consolidation
3. Deep Generative replay
- 4. Memory replay GAN**



Memory Replay GANs



Аналогичная проблема возникает, когда мы хотим научить генератор генерировать что-нибудь новое, но при этом не забыв и о старом.



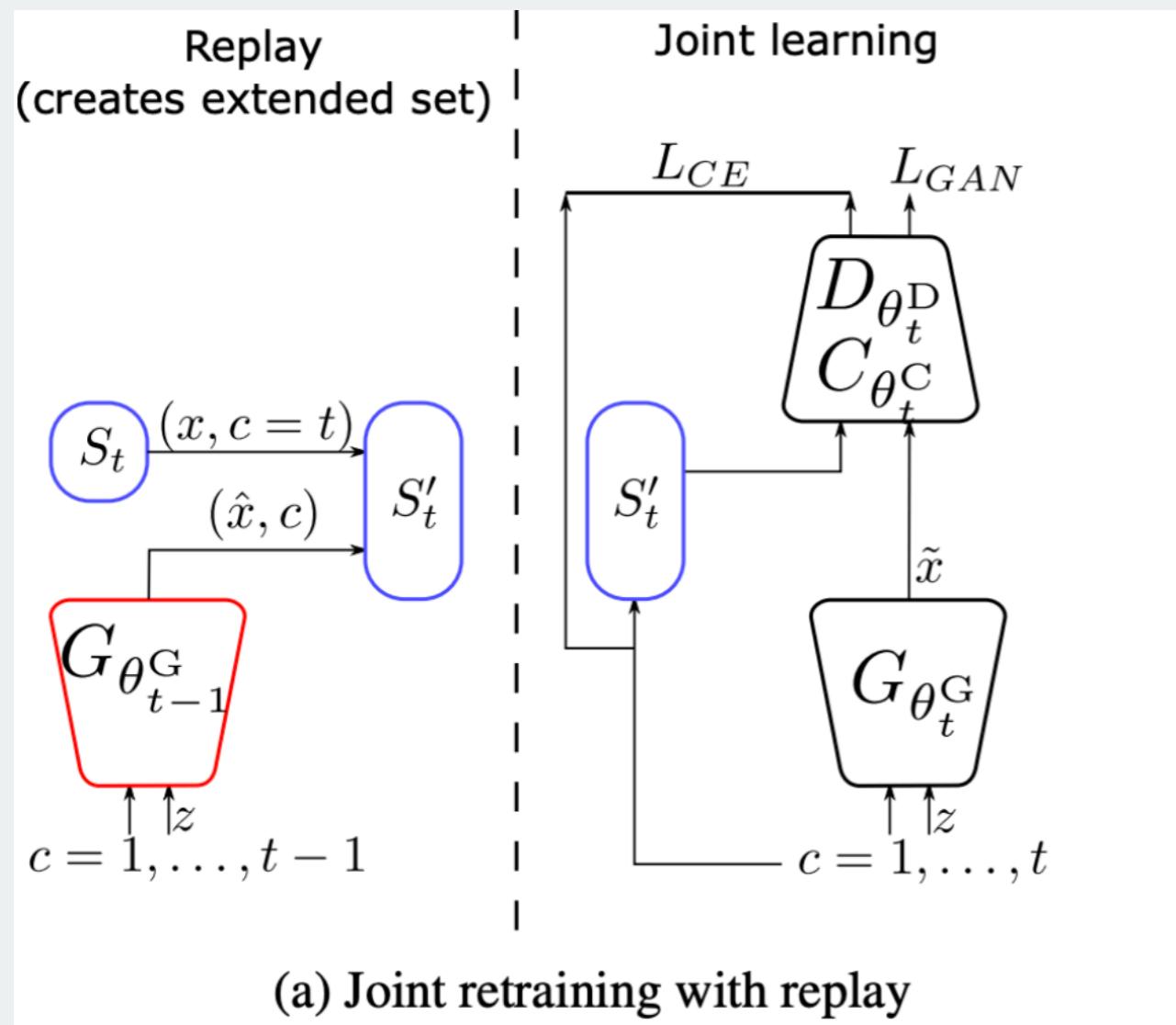
Chenshen Wu et al. Memory Replay GANs: learning to generate images from new categories without forgetting



Memory Replay GANs



Аналогичная проблема возникает когда мы хотим научить генератор генерировать что-нибудь новое, но при этом не забыв и о старом.



- Для обучения новой задаче создается расширенное множество примеров $S't$
- Для этого используются новые данные и данные, сгенерированные «старым» дискриминатором
- Новая пара генератор-дискриминатор обучается при помощи расширенного множества
- При обучении новой пары дополнительный сигнал - ошибка классификации Lce

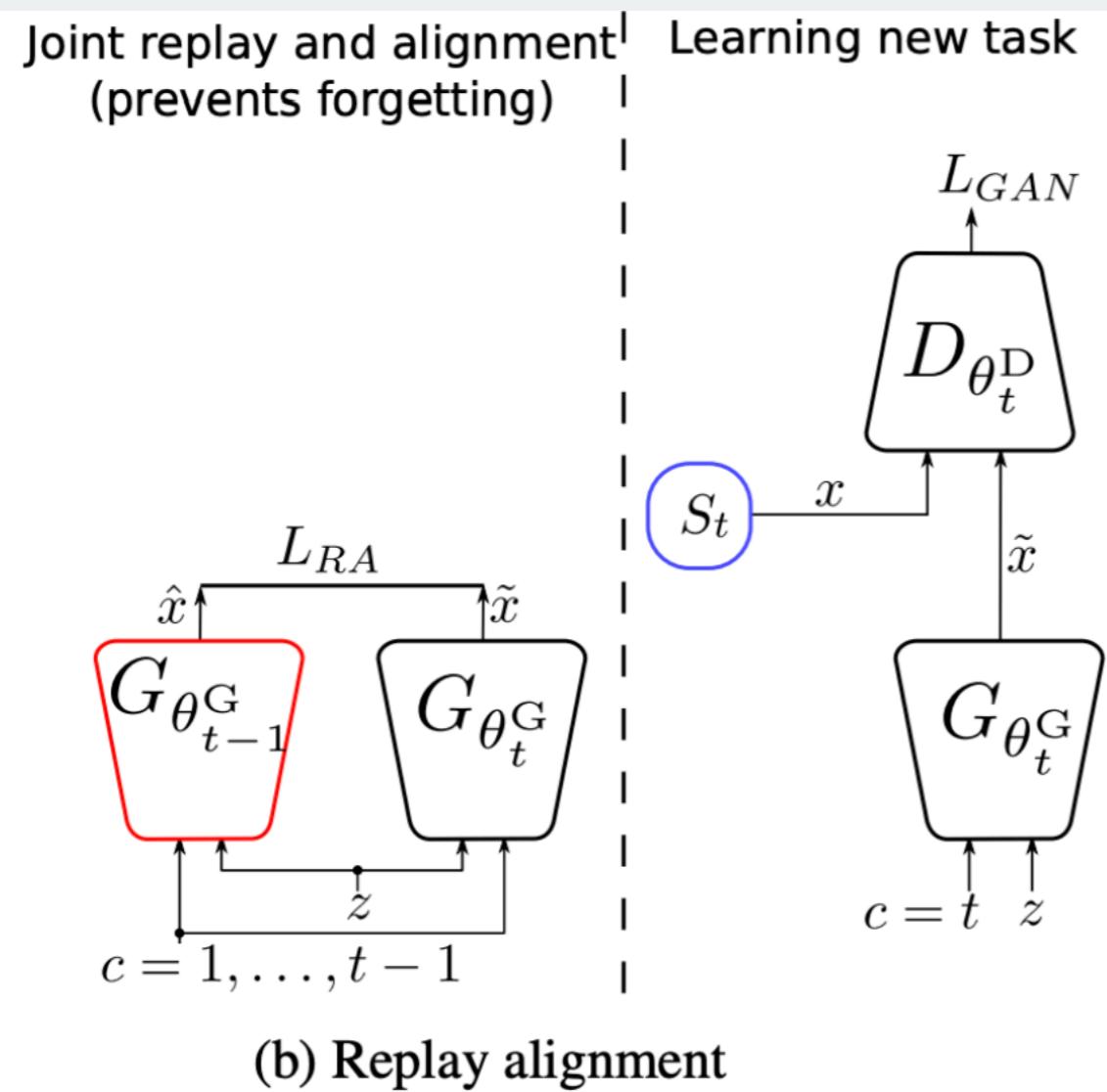
Chenshen Wu et al. Memory Replay GANs: learning to generate images from new categories without forgetting



Memory Replay GANs



Аналогичная проблема возникает когда мы хотим научить генератор генерировать что-нибудь новое, но при этом не забыв и о старом.



- Новый генератор обучается воспроизводить изображения, похожие на то, что создает «старый» генератор для той же метки и кода z .
- Точнее, он учится не «забывать», как создаются старые примеры
- S_t - датасет, содержащий только «новую» задачу.
- Пара нового генератора и дискриминатора учится только генерировать изображения из нового класса

Chenshen Wu et al. Memory Replay GANs: learning to generate images from new categories without forgetting



Memory Replay GANs

O T U S

Joint replay & alignment

$$\min_{\theta_t^G} L_{\text{GAN}}^G(\theta_t, S_t) + \lambda_{\text{RA}} L_{\text{RA}}(\theta_t, S_t)$$

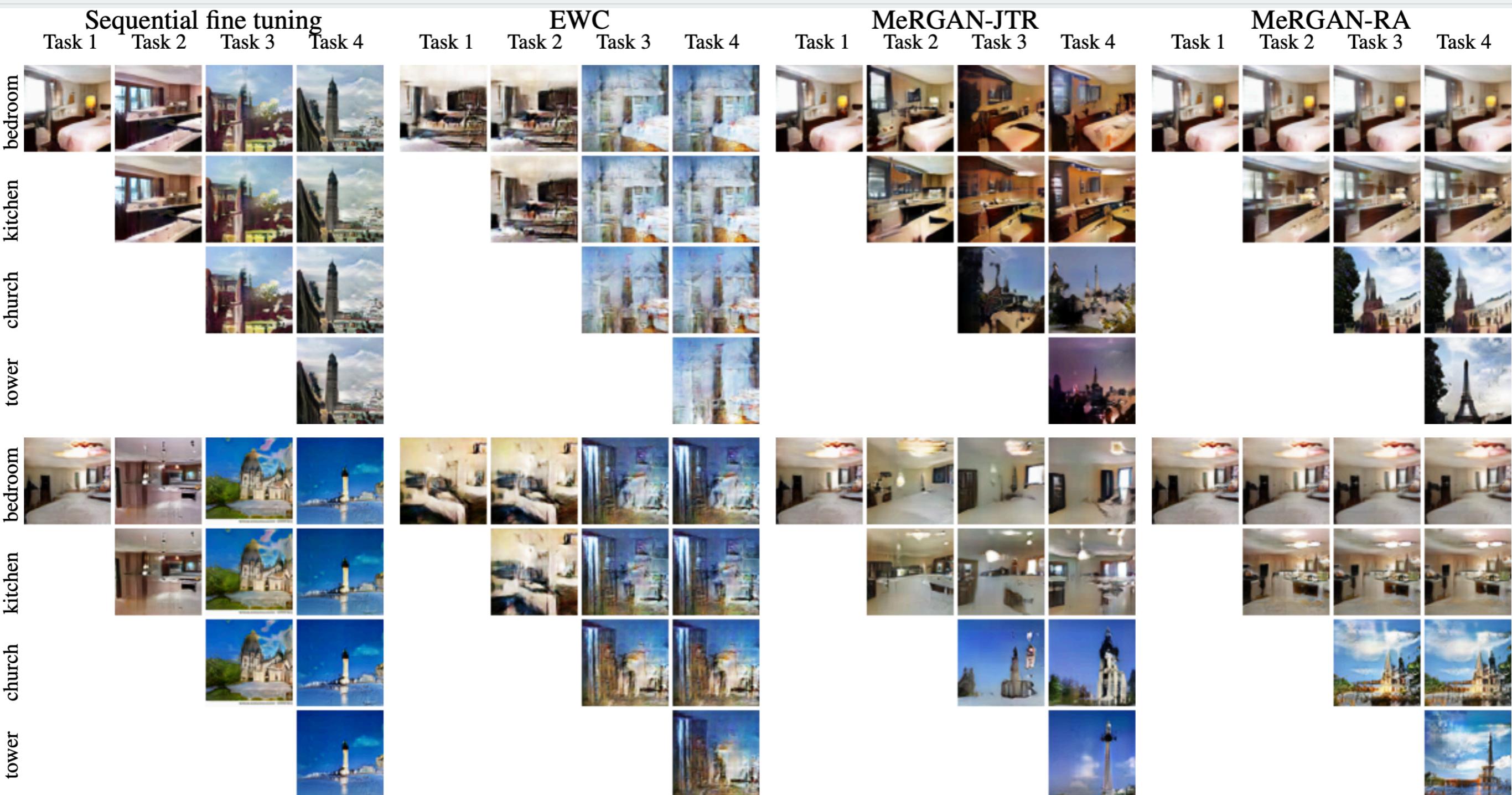
$$L_{\text{RA}}(\theta_t, S_t) = \mathbb{E}_{x \sim S, z \sim p_z, c \sim \mathcal{U}\{1, t-1\}} \left[\left\| G_{\theta_t^G}(z, c) - G_{\theta_{t-1}^G}(z, c) \right\|^2 \right]$$

Chenshen Wu et al. Memory Replay GANs: learning to generate images from new categories without forgetting



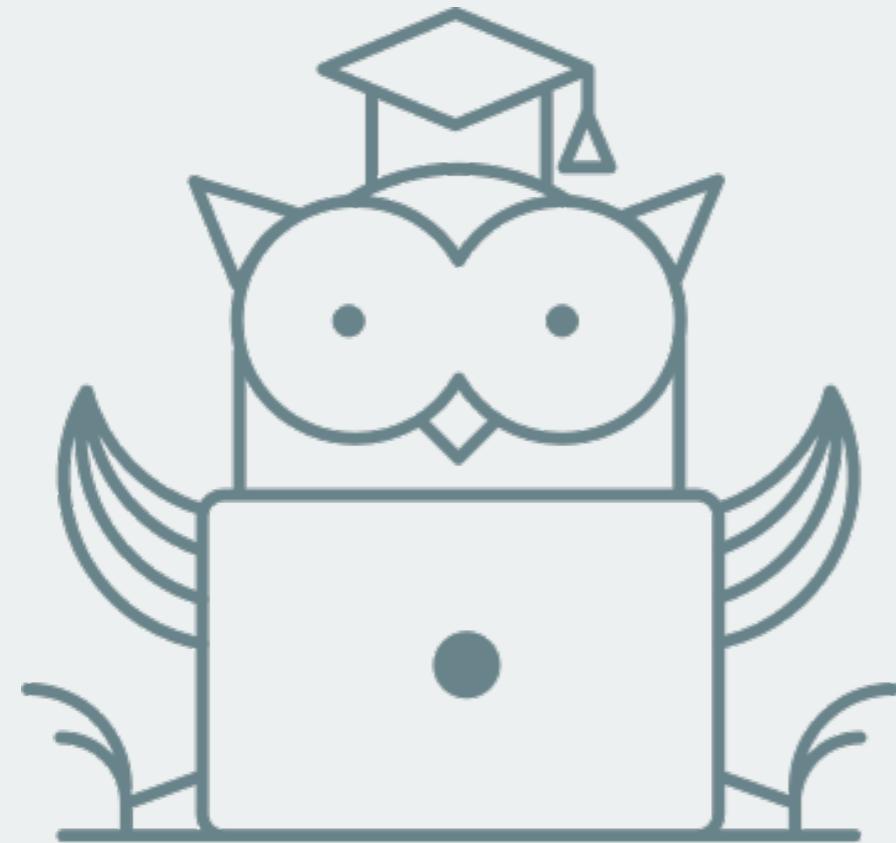
Memory Replay GANs

O T U S



Chenshen Wu et al. Memory Replay GANs: learning to generate images from new categories without forgetting





Спасибо
за внимание!