УДК 004.052.42

ИССЛЕДОВАНИЕ, РАЗРАБОТКА И ВАЛИДАЦИЯ МЕТОДОВ КОНТРОЛИРУЕМОЙ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ ГЕНЕРАТИВНЫМИ НЕЙРОСЕТЕВЫМИ МОДЕЛЯМИ

А. А. Солодухо[1)], В. А. Ковалёв[2)]

[1)] *Белорусский государственный университет,*
Минск, Беларусь, art.saladukha@gmail.com
[2)] *Объединённый институт проблем информатики НАН Беларуси,*
Минск, Беларусь, vassili.kovalev@gmail.com

В данной работе исследуются методы контролируемой генерации изображений при помощи генеративных нейросетевых моделей. Описываются результаты экспериментов по генерации изображений различными вариациями архитектуры GAN, проводится сравнение качества генерации при помощи различных метрик, в том числе анализируется метрика, использующая нейронную сеть CLIP для получения признакового представления изображений, а также проектируется и определяется программный комплекс для контролируемой генерации изображений на примере датасетов MNIST и гистологических изображений.

Ключевые слова: нейросетевые модели; генерация изображений; методы генерации изображений; метрика; валидация; генеративно-состязательная сеть; методы понижения размерности.

RESEARCH, DEVELOPMENT AND VALIDATION OF METHODS FOR CONTROLLED IMAGE GENERATION WITH GENERATIVE ADVERSARIAL NETWORKS

A. A. Saladukha[a)], V. A. Kovalev[b)]

[a)] *Belarusian State University,*
Minsk, Belarus, art.saladukha@gmail.com
[b)] *The United Institute of Informatics Problems,*
Minsk, Belarus, vassili.kovalev@gmail.com

This work explores methods for controlled image generation using generative neural network models. The results of experiments on image generation using different variations of the GAN architecture are described, and the quality of generation is compared using different metrics. This includes analyzing a metric that uses the CLIP neural network to obtain a feature representation of images, and implementing the pipeline for controlled image generation using the MNIST datasets and histological images.

Keywords: Neural network model; image generation; image generation methods; metric; validation; generative adversarial networks; dimensionality reduction methods.

1. Введение

Генеративные нейронные сети GAN (генеративно-состязательная нейронная сеть) произвели революцию в задаче генерации изображений, продемонстрировав впечатляющую способность генерировать разнообразный и правдоподобный контент. Однако в силу особенностей внутреннего представления данных этими моделями затруднён контроль над конкретными характеристиками генерируемых изображений. В рамках этой работы будут исследованы и проведены эксперименты по генерации изображений основными вариациями архитектур генеративно-состязательных нейронных сетей (главным образом WGAN), кроме того будет рассмотрен новый подход к определению функции потерь для GAN (R3-GAN), проанализированы различные метрики оценки качества сгенерированных изображений, описаны их преимущества и недостатки, а также спроектирован и реализован программный комплекс для контролируемой генерации изображений, не требующий ручной разметки изображений [1].

2. Выбор архитектуры GAN

В качестве архитектуры GAN для контролируемой генерации выбрана та, что обладает наилучшим качеством сгенерированных изображений и наиболее стабильным процессом обучения. Для оценки используются метрики Inception Score (IS), Frechet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), Perceptual Path Length (PPL), а также новая метрика CLIP-based Maximum Mean Discrepancy (CMMD). В метриках IS и FID для получения векторов-признаков используется предобученная на ImageNet сеть Inception-V3, в LPIPS и PPL – AlexNet или VGG. Авторы метрики CMMD предложили новый подход, получая признаковое представление при помощи архитектуры нейронной сети Contrastive Language-Image Pre-training (CLIP) и применяя для оценки степени различий изображений метрику максимального среднего расхождения с гауссовским ядром RBF [2]. Авторами было продемонстрированы основные недостатки самой распространённой метрики (FID): признаковое представление, полученное при помощи Inception-V3, уступает по качеству и репрезентативности в сравнении с CLIP; неверные предположения о нормальности распределения данных (в FID считается расстояние между двумя многомерными распределениями по метрике Вассерштайна в предположении, что они нормально распределены, из-за чего при определенной организации данных два совершенно разных распределения по значению метрик оказываются эквивалентными). Кроме того, эмпирически продемонстрировано, что FID часто плохо согласуется с данными, размеченными вручную.

3. Результаты обучения. Сравнение архитектур

Для исследования качества генерации используется датасет из 6000 гистологических снимков (2000 здоровых и 4000 с раковыми клетками).

Ниже приведены результаты экспериментов с различными моделями.

1. WGAN с ограничением весов: сходимость к локальному оптимуму отсутствует (ограничение весов зависит от датасета и не является качественным ограничением для обобщённой модели WGAN в контексте ограничения значения нормы градиента функции критика) [3].

2. WGAN с ограничением нормы градиента критика (WGAN-GP): сходимость к локальному оптимуму присутствует, т.к. используется прямое ограничение на значение нормы градиента критика и, как следствие, выполнение условия 1-Липшецевости функции критика [4].

3. WGAN со спектральной нормализацией на весах критика (дополнительно, на весах генератора) (WGAN-SN): сходимость к локальному оптимуму отсутствует (происходит прямое ограничение весов, что не всегда одинаково влияет на норму градиента, а, следовательно, и на выполнение условия 1-Липшецевости функции критика). Несмотря на более строгое ограничение, данный подход нередко уступает обыкновенному ограничению нормы градиента критика и не приводит к сходимости в общем случае [5].

4. WGAN со спектральной нормализацией и с ограничением нормы градиента критика (WGAN-SNGP): сходимость к локальному оптимуму присутствует.

5. R3-GAN: сходимость к локальному оптимуму присутствует. Примечательность подхода R3-GAN заключается в использовании функции потерь, для которой, в отличие от большинства аналогичных для других нейронных сетей семейства GAN, гарантируется сходимость к локальному оптимуму при оптимизации с помощью стохастического градиентного спуска. Кроме того, присутствие двух регуляризационных слагаемых (одно для сгенерированных данных, второе – для реальных) приводит к лучшей стабильности процесса обучения.

На рис. 1 пример сгенерированных с помощью R3-GAN изображений.

В табл. 1 представлены значения метрик оценки качества генерации изображений. Они объясняются специфичным доменом медицинских изображений (в первую очередь – характер текстур) и особенностями метрик.

Например, IS опирается только на знания предобученной сети, а по той причине, что в датасете ImageNet гистологических изображений нет, сеть может быть не в состоянии отнести к какому-то конкретному классу изображения и как следствие значения метрики низкие.

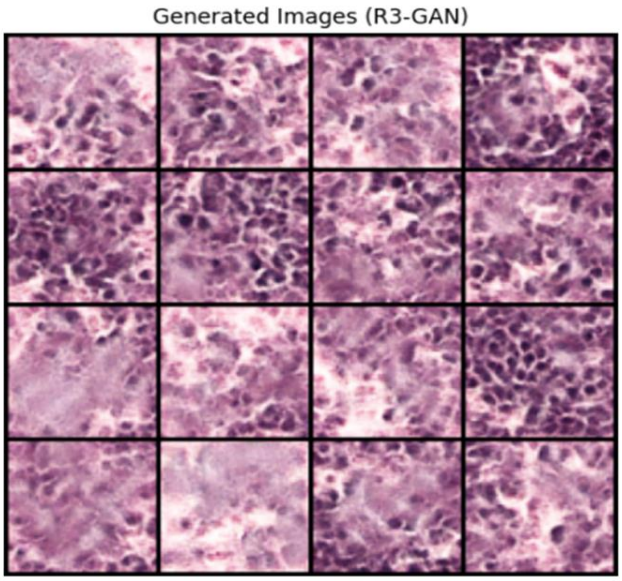

Рис. 1. Пример сгенерированных изображений (R3-GAN)

Таблица 1

Метрики качества генерации

Метрика	Архитектура				
	WGAN	WGAN-GP	WGAN-SN	WGAN-GP-SN	R3-GAN
IS	-	**2,29**	-	1,18	2,10
FID	-	273,53	-	**263,94**	267,90
LPIPS	-	0,69	-	0,70	**0,67**
PPL	-	1,5	-	1,51	**1,45**

Аналогичная ситуация касается FID: нет гарантий, что компоненты векторов признаков для гистологических изображений, полученных с помощью Inception-V3, будут соответствовать проявлению конкретных признаков, свойственных именно рассматриваемым изображениям. Для метрики LPIPS свойственна проблема требования выравнивания по классам для более получения корректного значения метрики. По той причине, что результатом вычисления метрики является взвешенная сумма попарных разностей компонент векторов признаков на разных глубинах нейронной сети между сгенерированными и реальными изображениями, соответственно, если классы изображений семантически сильно отличаются, то даже при высоком качестве генерации, значения метрики будут соответствовать менее качественным изображениям. В этом смысле признаковое представление, полученное с помощью CLIP, во-первых, более репрезентативно, во-вторых, сама нейронная сеть состоит из двух блоков, каждый из которых параллельно обучался для реализации задачи сопоставления признакового представления текстового описания изображений и соответствующих изображений.

4. Программный комплекс для контролируемой генерации изображений

Описание подхода контролируемой генерации:

1) генератор обучается на безусловную генерацию изображений;
2) генерируются векторы-латенты и соответствующие им изображения;
3) изображения подаются на вход предобученной нейронной сети для получения векторов признаков;
4) формируется датасет, в котором предсказываемые значения – векторы-латенты, являющиеся результатом преобразования векторов-признаков.
5) в результате обучения нейронная сеть представляет собой отображение из пространства признаков в латентное пространство генератора.

На рис. 2 представлен данный программный комплекс. Главное преимущество такого подхода по сравнению с другими заключается в понятной интерпретации (происходят манипуляции над признаковым представлением), а также в отсутствии необходимости в ручной разметке данных (при условии достаточно высокого уровня разнообразия сгенерированных изображений) [6].

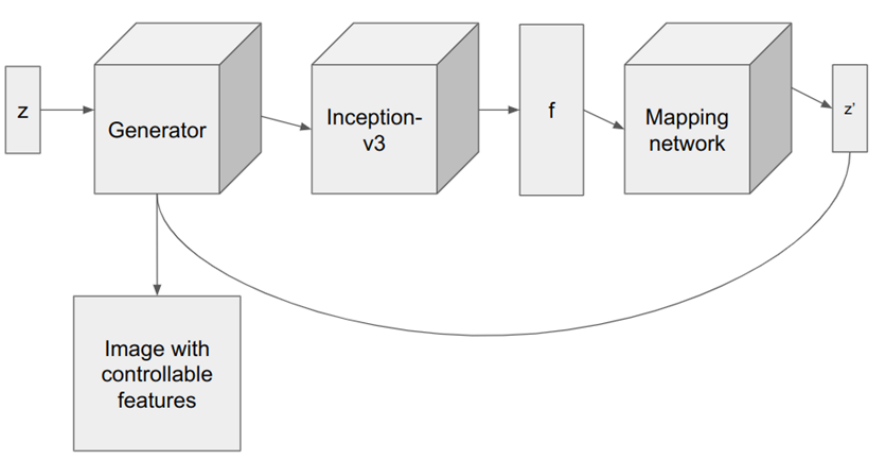

Рис. 2. Направление потоков данных

На рис. 3 представлено сравнение результатов операций над векторами в признаковом и латентном пространствах и результирующие изображения (на примере датасета MNIST). В случае операций над признаковыми векторами происходит отображение результирующего вектора в латентное пространство, и затем генерация изображения.

5. Направления дальнейшего исследования

В качестве перспективы для дальнейших исследований можно обозначить изучение влияния подхода self-supervised representation learning на получение признакового представления изображений, благодаря которому

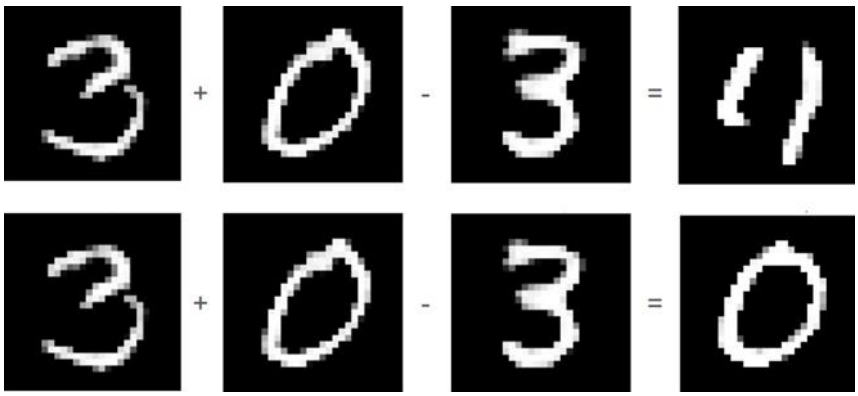

Рис. 3. Сравнение результатов операций над векторами в признаковом (нижний ряд) и латентном (верхний ряд) пространствах

возможно обучение сети, отображающей векторы из признакового в латентное пространство. Кроме того, использование метрики CMMD позволит получить более корректное представление о качестве сгенерированных изображений.

6. Заключение

В рамках работы были проведены эксперименты по генерации изображений при помощи вариаций GAN для выявления модели программного комплекса для контролируемой генерации, а также спроектирован и реализован программный комплекс для контролируемой генерации изображений, не требующий ручной разметки изображений.

Библиографические ссылки

1. The GAN is dead; long live the GAN! A Modern GAN Baseline / Y. Huang [et al.] // Advances in Neural Information Processing Systems (NeurIPS). 2024. Vol. 37, iss. 1. P. 44177–44215

2. Rethinking FID: Towards a Better Evaluation Metric for Image Generation / S. Jayasumana [et. at.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024, Vol. 33, iss. 1. P. 9307–9315.

3. *Arjovsky M., Chintala S., Bottou L.* Wasserstein Generative Adversarial Networks // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. Vol. 70, iss. 1. P. 214–223.

4. Improved Training of Wasserstein GANs / I. Gulrajani [et al.] // Advances in Neural Information Processing Systems (NeurIPS). 2017. Vol. 30, iss. 1. P. 5769–5779.

5. Analyzing and Improving the Image Quality of StyleGAN / T. Karras [et. al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. Vol. 33. P. 8107–8116.

6. *Shen Y., Zhou B.* Closed-Form Factorization of Latent Semantics in GANs // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. Vol. 30, iss. 1. P. 1532–1540.