

*В. Е. Карнаухов<sup>1</sup>, А. В. Хвостиков<sup>2</sup>, А. С. Крылов<sup>3</sup>*

## **МЕТОДЫ ГЕНЕРАТИВНОЙ АУГМЕНТАЦИИ ДЛЯ ЗАДАЧИ АНАЛИЗА ГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ В УСЛОВИЯХ ОГРАНИЧЕННОГО КОЛИЧЕСТВА ДАННЫХ\***

### Введение

На текущий момент методы глубокого обучения демонстрируют выдающиеся результаты в задачах классификации, сегментации и анализа биомедицинских изображений и, в частности, гистологических изображений [1], [2]. Несмотря на свою высокую эффективность, подходы, основанные на применении нейросетевых моделей, требуют использования большого количества размеченных изображений для корректного обучения, а с увеличением сложности модели растет и требуемый объем размеченных данных [3]. В случае с гистологическими изображениями получение новых данных стало относительно доступным с появлением сканеров нового поколения, способных получать качественные изображения, чего нельзя сказать о разметке данных. Ручная разметка гистологических изображений, в особенности полнослайдовых, может занимать до десятков часов на одно изображение в случае задачи сегментации и должна выполняться специалистами-гистологами.

Наиболее реалистичной ситуацией в области современной цифровой патологии является наличие небольшого количества размеченных изображений и, возможно, большого количества неразмеченных.

В данной работе в качестве базовой задачи рассматривается задача классификации фрагментов гистологических тканей. Предлагается метод генеративной аугментации гистологических изображений, основанный на использовании генеративно-сопоставительных моделей. По относительно

---

<sup>1</sup>Аспирант факультета ВМК МГУ имени М.В. Ломоносова, e-mail: valera-96@yandex.ru.

<sup>2</sup>К.ф.-м.н., м.н.с. факультета ВМК МГУ имени М.В. Ломоносова, e-mail: khvostikov@cs.msu.ru.

<sup>3</sup>Д.ф.-м.н., профессор факультета ВМК МГУ имени М.В. Ломоносова, e-mail: kryl@cs.msu.ru.

\*Работа выполнена при поддержке гранта РФФИ 22-21-00081

небольшому количеству размеченных данных генеративно-состязательные сети способны создавать новые синтетические изображения с соответствующей разметкой, которые уже, в свою очередь, могут быть использованы для обучения нейросетевых моделей анализа гистологических изображений. Показано, что при таком подходе расширяется обучающая выборка и, как следствие, повышается точность моделей классификации на тестовых данных.

### Генеративно-состязательные сети

Генеративно-состязательные сети являются одним из самых популярных подходов в машинном обучении для генерации синтетических данных [4]. Они состоят из двух нейронных сетей: генератора и дискриминатора. Генератор на вход принимает случайный вектор фиксированной длины и генерирует примеры, похожие на тренировочные, в то время как дискриминатор классифицирует входные данные как реальные (из тренировочного набора) или поддельные (сгенерированные). Обе сети обучаются одновременно. В результате такой конкуренции дискриминатор в процессе обучения становится лучше различать реальные и сгенерированные данные, а генератор — качественнее создавать реалистичные примеры. Схема описанной модели представлена на рис. 1.

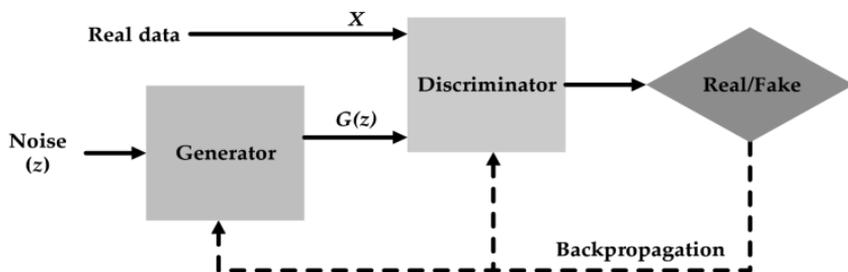


Рис. 1. Общая схема генеративно-состязательной сети

Цель дискриминатора — правильно классифицировать настоящие и сгенерированные данные. Наиболее стандартным способом обучения дискриминатора является максимизация следующей функции потерь:

$$D_{loss} = E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))],$$

где  $x$  — реальный пример,  $z$  — случайный вектор шума,  $D(\cdot)$  — оценка дискриминатором вероятности того, что пример является реальным,  $G(\cdot)$  — выход генератора при некотором шуме,  $E_x$  — математическое ожидание по всей выборке из реального набора данных,  $E_z$  — математическое ожидание по всем входным данным генератора.

Цель генератора — генерировать примеры, которые дискриминатор не может отличить от реальных данных, поэтому он пытается минимизировать следующую функцию потерь:

$$G_{loss} = E_z[\log(1 - D(G(z)))].$$

Объединяя два выражения в одно, получаем функцию оптимизации генеративно-сопоставительной модели:

$$\min_G \max_D V(G, D) = E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))].$$

### Перенос обучения

Перенос обучения — это распространенный подход в машинном обучении, при котором модель, обученная для решения одной задачи, используется для решения другой задачи [5]. Он применяется в различных задачах машинного обучения, в частности, в задачах классификации и генерации. Перенос обучения позволяет избежать необходимости в большом количестве новых данных, поскольку модель, которая была уже обучена на большом наборе тренировочных данных, может справиться с новой, но похожей задачей с весьма меньшим объемом данных. Также перенос обучения позволяет сократить время обучения новой модели, поскольку обучение начинается не со случайной инициализации весов.

Одним из эффективных способов переноса обучения является заморозка первых слоев предобученной модели и обучение оставшихся слоев [6]. На первых слоях сеть изучает базовые характеристики изображений и определяет наиболее общие признаки, в то время как на последних слоях сеть выявляет специфические признаки, которые зависят от выбранного набора данных и задачи.

Перенос обучения изначально использовался для простых моделей классификации, но он так же может быть использован и для дообучения генеративно-сопоставительных моделей. Так, в работе [7] был предложен подход, в котором замораживались первые слои дискриминатора, а в работе [8] было показано, что заморозка первых слоев генератора также повышает качество генерации изображений.

### Аугментация изображений для генеративно-сопоставительных сетей

Генеративно-сопоставительные сети требуют достаточно большого количества тренировочных данных, чтобы качественно генерировать новые изображения [9]. Обучение таких генеративно-сопоставительных сетей с использованием ограниченного набора данных обычно приводит к переобучению дискриминатора, что негативно сказывается на обучении генератора.

Практически во всех областях глубокого обучения аугментация данных является стандартным решением проблемы переобучения [10]. Например, обучение классификатора с добавлением к изображениям

случайных поворотов, шумов приводит к повышению инвариантности к данным преобразованиям, что является желательным качеством классификатора [11]. В то же время генеративно-состязательные сети, обучаемые с использованием аналогичной аугментации данных, учатся генерировать распределение, включая применяемые преобразования [12]. Например, добавление шума в обучающие данные приводит к зашумленным сгенерированным результатам, даже если их нет в исходном наборе данных.

В работе [13] был предложен подход, который позволяет использовать ряд классических методов аугментации (отражение по вертикали и горизонтали, повороты на произвольное число градусов, сдвиги, масштабирование, изменение интенсивности пикселей), чтобы предотвратить переобучение дискриминатора, при этом генератор будет создавать только чистые изображения. Все преобразования применяются с одинаковой вероятностью, которая изменяется автоматически во время обучения.

### Используемые наборы данных

В данной работе для задач классификации и генерации использовались четыре разных набора данных с гистологическими изображениями.

Первый набор данных — набор NCT-CRC-HE-100K [14]. Он состоит из 100000 неперекрывающихся фрагментов (патчей) изображений с оптическим увеличением 20X. Каждый патч имеет разрешение  $224 \times 224$  пикселей и соответствует одной из 9 меток классов в зависимости от конкретного типа ткани или фона: ADI, BACK, DEB, LYM, MUC, MUS, NORM, STR, TUM. Ко всем патчам была применена цветовая нормализация методом Масенко [15].

Второй набор данных — набор NCT-CRC-HE-100K-NONORM [14]. Он так же, как и первый набор, состоит из 100000 патчей с оптическим увеличением 20X, которые имеют размер  $224 \times 224$  пикселей и соответствуют тем же 9 классам. Однако, в отличие от первого набора, к данным изображениям не была применена цветовая нормализация.

Третий набор данных — объединение наборов WSS1 и WSS2 набора данных PATH-DT-MSU<sup>1</sup>, каждый из которых состоит из 10 полнослайдовых изображений (5 изображений для обучения и 5 изображений для валидации с оптическим увеличением 40X). Участки изображений соответствуют разным типам тканей, всего 6 классов: AT, BG, DYS, LP, MM, TUM. Каждое полнослайдовое изображение было разбито на патчи размера  $224 \times 224$  с перекрытием 112 пикселей. В итоге обучающий набор состоял из 66073, а валидационный — из 49460 изображений.

---

<sup>1</sup><https://imaging.cs.msu.ru/en/research/histology/path-dt-msu>

Четвертый набор данных — набор VACH [16], состоящий из 400 изображений с оптическим увеличением 20X и разрешением 2048×1536 пикселей. Набор разделен на 4 класса: Normal, Benign, In situ carcinoma, Invasive carcinoma, в каждом из которых по 100 изображений. Каждая группа изображений была разделена на 75 тренировочных и 25 валидационных изображений. Поскольку в данном наборе метка присваивается всему изображению, то нет гарантии, что небольшие области содержат информацию о классе исходного изображения. Поэтому, как и в работах [17], [18], каждое изображение было разбито на патчи размера 512×512 пикселей с перекрытием 256 пикселей, чтобы обеспечить более надежную метку для каждого участка изображения. Всего из данного набора получилось 10500 тренировочных и 3500 валидационных изображений.

На рис. 2 приведены примеры изображений из рассмотренных наборов данных.

### Общая схема экспериментов

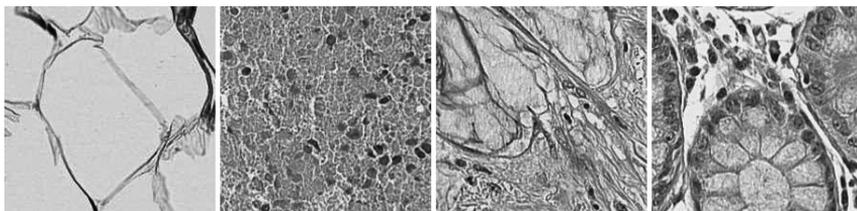
В данной работе предложена следующая схема экспериментов.

1. На относительно большом наборе данных обучить генеративно-состязательную сеть, которая будет использоваться в качестве предобученной модели на других наборах данных.
2. На более маленьких наборах данных обучить (или дообучить) генеративно-состязательные модели со случайными начальными весами, с предобученными весами без заморозки слоев, с предобученными весами с заморозкой первых слоев дискриминатора или генератора.
3. С помощью обученных генеративных моделей создать новые синтетические изображения, которые будут использоваться для расширения обучающей выборки исходных наборов данных.
4. Сравнить результаты работы моделей классификации разной архитектуры на новых (расширенных) обучающих выборках, которые являются более объемными по сравнению с исходными.

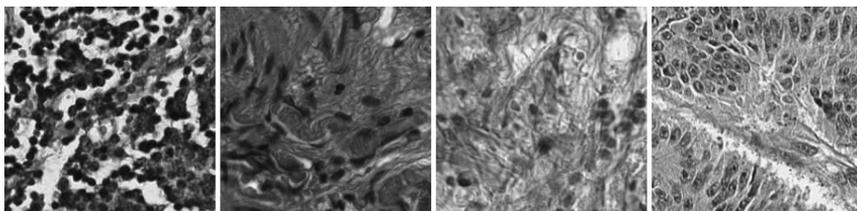
### Исходные данные для экспериментов

В настоящее время модель StyleGAN2 [19] является одной из самых эффективных и широко применяемых генеративно-состязательных сетей, поэтому она была выбрана в качестве модели генерации гистологических изображений.

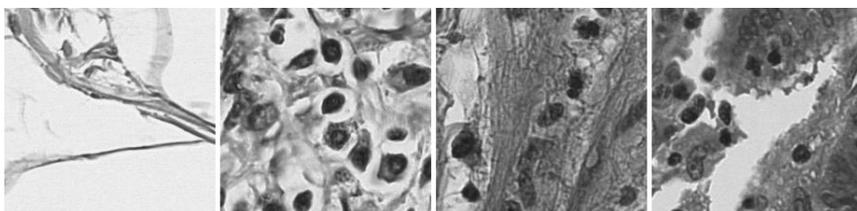
Большинство параметров обучения StyleGAN2 остались такими же, как и в оригинальной версии статьи [19]. Поскольку при генерации гистологических изображений необходимо учитывать классовую информацию, в качестве условной функции потерь была использована



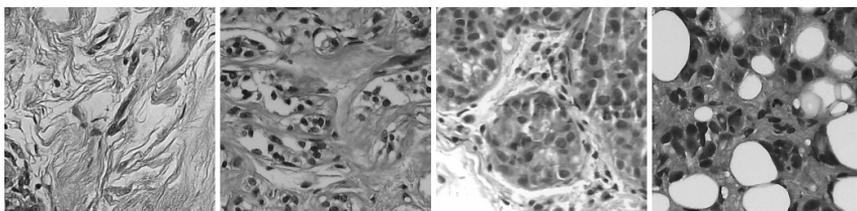
Набор NCT-CRC-HE-100K



Набор NCT-CRC-HE-100K-NONORM



Набор PATH-DT-MSU



Набор VACH

Рис. 2. Примеры изображений из разных наборов данных

Data-to-Data Cross-Entropy [20], при которой генеративно-состязательные сети достигают более высоких результатов по сравнению с другими условными функциями потерь. Для повышения качества генерации была добавлена адаптивная аугментация, указанная ранее. Для оценки качества генеративно-состязательных моделей использовались меры оценки качества FID [21], Improved Precision & Recall [22], Density & Coverage

[23]. Критерием остановки обучения моделей являлось то, что их дальнейшая работа не приводила к повышению точности по указанным меркам качества на обучающей выборке.

Для итоговой оценки результатов генеративно-сопоставительных моделей использовались три сверточных нейронных сети разной архитектуры для классификации гистологических изображений: ResNet34 [24], DenseNet169 [25], EfficientNet-B2 [26], которые были предобучены на наборе ImageNet [27]. Точность классификаторов оценивалась с помощью мер качества Accuracy и Balanced Accuracy [28]. В качестве методов аугментации использовались только базовые преобразования: повороты на  $90^\circ$ , отражение по вертикали и горизонтали, изменение яркости, контрастности и насыщенности пикселей в диапазоне [0.8, 1.2] от исходных значений.

Стоит отметить, что в наборе PATH-DT-MSU классы не сбалансированы (например, в тренировочном наборе количество изображений класса AT составляет 33966, а количество изображений класса DYS — 1002), поэтому при обучении как моделей классификации, так и генеративно-сопоставительных моделей необходимо выравнять соотношение классов изображений. Для этого был применен метод сэмплирования, который для загрузчика данных автоматически устанавливает равные вероятности появления каждого класса.

### Базовая генеративно-сопоставительная модель

Базовая модель — это модель, которая будет использоваться для переноса обучения на другие наборы данных. Для этого необходимо обучить ее на сравнительно большом наборе, чтобы дискриминатор мог извлекать из изображений большое количество разных признаков, а генератор мог генерировать как можно более разнообразные данные. Поэтому в качестве тренировочного набора были выбраны наборы NCT-CRC-HE-100K и NCT-CRC-HE-100K-NONORM, которые были объединены в одно множество. Таким образом, в качестве базовой модели выступал StyleGAN2, который обучался на наборе, состоящем из 200000 изображений.

### Результаты экспериментов

На наборах PATH-DT-MSU и BACH были обучены генеративно-сопоставительные модели:

- со случайной инициализацией весов;
- с предобученными весами без заморозки слоев;
- с предобученными весами с заморозкой:
  - первого, первых двух, первых трех блоков дискриминатора;
  - первого, первых двух, первых трех блоков генератора;

- первого, первых двух блоков дискриминатора и первых двух блоков генератора.

После обучения каждой моделью по отдельности генерировались синтетические изображения. Экспериментально было выявлено, что классификаторы ResNet34, DenseNet169, EfficientNet-B2 на валидационном наборе PATH-DT-MSU показывали точность выше, если обучались только на сгенерированных изображениях, в то время как на валидационном наборе VACH классификаторы достигали более высокой точности, если обучались на объединении реальных и сгенерированных изображений.

Генеративно-состязательные сети могут создавать сколь угодно большое количество данных, но было показано, что, начиная с некоторого места, с увеличением числа сгенерированных изображений точность классификаторов, обученных на этих данных, не повышалась. Поэтому для того, чтобы оценить работу генеративных моделей, обученных на наборе PATH-DT-MSU, каждый классификатор три раза обучался (для усреднения результатов) на 30000 изображениях (5000 изображений на каждый класс), т.е. каждой моделью генерировалось по 90000 изображений.

Для оценивания работы генеративных моделей, обученных на наборе VACH, каждый классификатор три раза обучался на наборе из 10500 реальных и 10500 синтетических изображений (5250 изображений на каждый класс), т.е. каждой моделью генерировалось по 31500 изображений.

В таблицах 1 – 2 представлены результаты классификаторов для наборов PATH-DT-MSU и VACH. Для удобства в таблицах приведены только ключевые результаты для сравнения. Первая строчка – использование только реальных изображений, вторая строчка – использование синтетических изображений, созданных генеративными моделями, обученными с нуля, третья строчка – использование синтетических изображений, созданных генеративными моделями, предобученными на 200000 гистологических изображениях, четвертая строчка – использование синтетических изображений, созданных предобученными генеративными моделями с заморозкой указанных блоков, при которых была достигнута самая высокая точность. Надпись freezing d1, g2 обозначает заморозку первого блока дискриминатора и первых двух блоков генератора, надпись freezing g2 – заморозку только первых двух блоков генератора. Во второй таблице не приведены значения меры качества Balanced Accuracy, поскольку классы в валидационном наборе VACH сбалансированы, поэтому значения Accuracy и Balanced Accuracy совпадают.

Табл. 1. Результаты работы моделей классификации для набора PATH-DT-MSU

Training data	ResNet34		DenseNet169		EfficientNet-B2	
	Acc	Bal Acc	Acc	Bal Acc	Acc	Bal Acc
66073 real	87.13	76.23	87.99	77.68	87.69	78.20
30000 generated with non-pretrained	85.89	79.55	86.68	80.72	86.51	81.67
30000 generated with pretrained	86.69	80.71	88.06	81.95	87.88	83.40
30000 generated with freezing d1, g2	<b>88.36</b>	<b>80.87</b>	<b>89.34</b>	<b>82.12</b>	<b>90.09</b>	<b>83.78</b>

Табл. 2. Результаты работы моделей классификации для набора VACH

Training data	ResNet34	DenseNet169	EfficientNet-B2
	Accuracy	Accuracy	Accuracy
10500 real	83.68	85.43	85.78
10500 real 10500 generated with non-pretrained	84.66	86.05	86.91
10500 real 10500 generated with pretrained	85.20	86.32	87.24
10500 real 10500 generated with freezing g2	<b>85.82</b>	<b>86.69</b>	<b>87.75</b>

Помимо результатов, представленных в таблице, были проведены эксперименты с заморозкой и других блоков, указанных в начале раздела. Было показано, что:

- использование синтетических изображений, созданных генеративно-состязательными моделями, обученными с нуля, уменьшает точность классификации (на 1.2–1.3% на наборе PATH-DT-MSU), но повышает сбалансированную точность (на 3.0–3.5% на наборе PATH-DT-MSU и на 0.6–1.1% на наборе VACH) по сравнению с тем, если использовать только исходные реальные изображения;
- использование синтетических изображений, созданных генеративно-состязательными моделями, предобученными на большом наборе данных, повышает точность классификации (на 0.8–1.4% на наборе PATH-DT-MSU) и сбалансированную точность (на

1.2–1.7% на наборе PATH-DT-MSU и на 0.3–0.5% на наборе BACH) по сравнению с тем, если использовать генеративные модели, обученные с нуля;

- использование синтетических изображений, созданных генеративно-сопоставительными моделями, предобученными на большом наборе данных с заморозкой первых слоев, повышает точность классификации (на 1.3–2.2% на наборе PATH-DT-MSU) и сбалансированную точность (на 0.2–0.4% на наборе PATH-DT-MSU и на 0.4–0.6% на наборе BACH) по сравнению с тем, если использовать предобученные генеративные модели без заморозки слоев.

Таким образом, наиболее эффективно можно использовать генеративно-сопоставительные модели в задаче классификации гистологических изображений, если начинать их обучение с модели, предварительно обученной на большом наборе данных, с заморозкой первых слоев, что позволяет повысить точность классификации (на 1.2–2.4% на наборе PATH-DT-MSU) и сбалансированную точность (на 4.4–5.6% на наборе PATH-DT-MSU и на 1.3–2.1% на наборе BACH) по сравнению с тем, если использовать только исходные реальные изображения. Особенно отмечается повышение качества генерации изображений при заморозке первых двух блоков генератора.

На рис. 3 приведены примеры сгенерированных изображений.

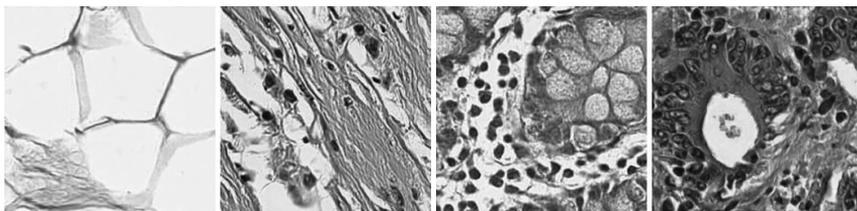
### Детали реализации

Обучение генеративно-сопоставительных моделей и нейросетевых моделей классификации выполнялось на языке Python 3.9. В качестве фреймворка машинного обучения использовался PyTorch 1.10 с поддержкой технологии CUDA 11.1. Вычисления производились на сервере с процессором Intel(R) Xeon(R) Gold 6226R CPU 2.90 GHz и видеокартой NVIDIA RTX A6000 48 GB.

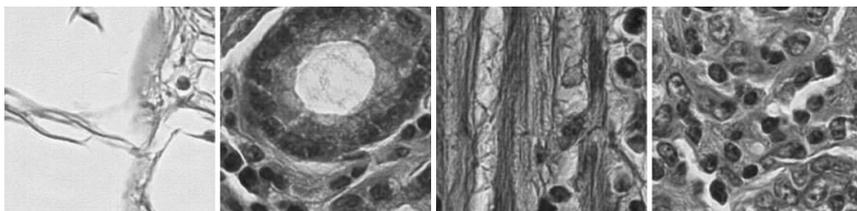
Обучение одной генеративно-сопоставительной сети на одной видеокарте на объединенном наборе из 200000 изображений занимало около 450ч, на наборе PATH-DT-MSU — 180ч, на наборе BACH — 80ч. Обучение одной модели классификации с тремя запусками на одной видеокарте на наборе PATH-DT-MSU занимало в среднем 7ч, на наборе BACH — 4ч. Суммарное время обучения всех моделей составило около 5000ч.

### Заключение

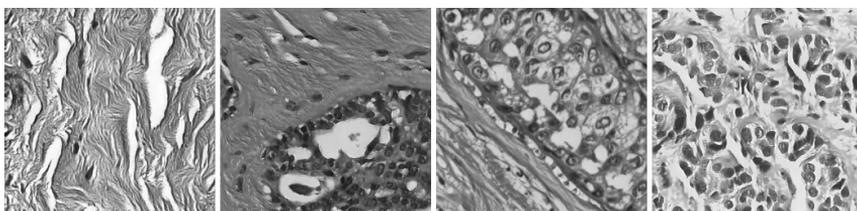
В данной работе предложен новый метод генеративной аугментации гистологических изображений, основанный на использовании генеративно-сопоставительных моделей. Показано, что при работе с наборами с ограниченным количеством данных генеративные модели повышают точность нейросетевых моделей классификации за счет



Изображения, сгенерированные базовой генеративной моделью



Изображения, сгенерированные для набора PATH-DT-MSU



Изображения, сгенерированные для набора VACH

Рис. 3. Примеры сгенерированных изображений

добавления в обучающую выборку новых синтетических изображений. Продемонстрировано, что перенос обучения генеративно-состязательных моделей повышает качество генерации изображений. Показано, что заморозка первых слоев генератора или дискриминатора повышает эффективность работы генеративно-состязательных моделей. Особенно это отмечается при заморозке первых двух блоков генератора.

#### Благодарности

Работа выполнена при поддержке гранта РФФ 22-21-00081.

#### Литература

1. *Bošnački, D., van Riel, N., and Veta, M.* Deep learning with convolutional neural networks for histopathology image analysis //Automated Reasoning for Systems Biology and Medicine. — Springer, Cham, 2019. — P. 453–469.
2. *Hägele, M., Seegerer, P., Lapuschkin, S., Bockmayr, M. et al.* Resolving

- challenges in deep learning-based analyses of histopathological images using explanation methods //Scientific Reports. — 2020. — V. 10. — №. 6423. — P. 1–12.
3. *Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F.* The computational limits of deep learning //arXiv preprint arXiv:2007.05558. — 2020.
  4. *Razavi-Far, R., Ruiz-Garcia, A., Palade, V., and Schmidhuber, J.* Generative adversarial learning: architectures and applications. — Springer, Cham, 2022. — P. 7–12.
  5. *Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.* How transferable are features in deep neural networks? //Advances in Neural Information Processing Systems. — 2014. — V. 27. — P. 3320–3328.
  6. *Kumar, J. S., Anuar, S., and Hassan, N. H.* Transfer learning based performance comparison of the pre-trained deep neural networks //International Journal of Advanced Computer Science and Applications. — 2022. — V. 13. — №. 1. — P. 797–805.
  7. *Mo, S., Cho, M., and Shin, J.* Freeze the discriminator: a simple baseline for fine-tuning GANs //arXiv preprint arXiv:2002.10964. — 2020.
  8. *Back, J.* Fine-tuning StyleGAN2 for cartoon face generation //arXiv preprint arXiv:2106.12445. — 2021.
  9. *Jiang, L., Dai, B., Wu, W., and Loy, C. C.* Deceive D: Adaptive pseudo augmentation for GAN training with limited data //Advances in Neural Information Processing Systems. — 2021. — V. 34. — P. 21655–21667.
  10. *Shorten C., and Khoshgoftaar T. M.* A survey on image data augmentation for deep learning //Journal of Big Data. — 2019. — V. 6. — №. 1. — P. 1–48.
  11. *Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V.* Randaugment: Practical automated data augmentation with a reduced search space //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. — 2020. — P. 3008–3017.
  12. *Zhao, Z., Singh, S., Lee, H., Zhang, Z. et al.* Improved consistency regularization for GANs //AAAI Conference on Artificial Intelligence. — 2021. — V. 35. — №. 12. — P. 11033–11041.
  13. *Karras, T., Aittala, M., Hellsten, J., Laine, S. et al.* Training generative adversarial networks with limited data //Advances in Neural Information Processing Systems. — 2020. — V. 33. — P. 12104–12114.
  14. *Kather, J. N., Halama, N., and Marx, A.* 100,000 histological images of human colorectal cancer and healthy tissue (v0.1) — 2018. — [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.1214456>

15. *Macenko, M., Niethammer, M., Marron, J. S., Borland, D. et al.* A method for normalizing histology slides for quantitative analysis //2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. — 2009. — P. 1107–1110.
16. *Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. et al.* BACH: Grand challenge on breast cancer histology images //Medical Image Analysis. — 2019. — V. 56. — P. 122–139.
17. *Araújo, T., Aresta, G., Castro, E., Rouco, J. et al.* Classification of breast cancer histology images using convolutional neural networks //PLoS ONE. — 2017. — V. 12. — №. 6. — P. e0177544.
18. *Nazeri, K., Aminpour, A., and Ebrahimi, M.* Two-stage convolutional neural network for breast cancer histology image classification //International Conference Image Analysis and Recognition. — Springer, Cham, 2018. — P. 717–726.
19. *Karras, T., Laine, S., Aittala, M., Hellsten, J. et al.* Analyzing and improving the image quality of StyleGAN //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2020. — P. 8110–8119.
20. *Kang, M., Shim, W., Cho, M., and Park, J.* Rebooting ACGAN: Auxiliary classifier GANs with stable training //Advances in Neural Information Processing Systems. — 2021. — V. 34. — P. 23505–23518.
21. *Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. et al.* GANs trained by a two time-scale update rule converge to a local Nash equilibrium //Advances in Neural Information Processing Systems. — 2017. — V. 30. — P. 6626–6637.
22. *Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J. et al.* Improved precision and recall metric for assessing generative models //Advances in Neural Information Processing Systems. — 2019. — V. 32. — P. 3929–3938.
23. *Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y. et al.* Reliable fidelity and diversity metrics for generative models //37th International Conference on Machine Learning. — PMLR, 2020. — V. 119. — №. 665. — P. 7176–7185.
24. *He, K., Zhang, X., Ren, S., and Sun, J.* Deep residual learning for image recognition //2016 IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 770–778.
25. *Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q.* Densely connected convolutional networks //2017 IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 2261–2269.
26. *Tan, M., and Le, Q.* EfficientNet: rethinking model scaling for convolutional neural networks //36th International Conference on Machine Learning. — PMLR, 2019. — P. 6105–6114.

27. *Deng, J., Dong, W., Socher, R., Li, L. et al.* ImageNet: A large-scale hierarchical image database //2009 IEEE Conference on Computer Vision and Pattern Recognition. — 2009. — P. 248–255.
28. *Grandini, M., Bagli, E., and Visani, G.* Metrics for multi-class classification: an overview //arXiv preprint arXiv:2008.05756. — 2020.