

ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ АВТОМОБИЛЬНЫХ НОМЕРОВ НА ОСНОВЕ ПОЛЬЗОВАТЕЛЬСКОЙ ВЫБОРКИ

Аннотация. В статье представлена разработка инструментария генерации больших наборов изображений автомобильных номеров на основе небольшой пользовательской выборки реальных изображений. Разработка выполнена на основе модели генеративно-состязательной нейронной сети. Созданный программный продукт предназначен для последующей интеграции с решениями по распознаванию автомобильных номеров.

Ключевые слова: генерация изображений, генеративно-состязательные нейронные сети, распознавание автомобильных номеров, модель pix2pix GAN.

Введение. Одной из важнейших и трудоемких задач в машинном обучении является формирование наборов обучающих данных. Компания Ivideon [1] занимается разработкой продуктов в сфере видеоаналитики. Среди решений, предлагаемых компанией, имеется программный продукт, использующий технологию распознавания автомобильных номеров. Данный продукт представляет интерес не только для отечественных, но и для зарубежных заказчиков.

Для обучения модели распознавания требуется не менее нескольких тысяч изображений автомобильных номеров той страны, представителем которой является заказчик. Во многих случаях данные, необходимые для развития проекта, найти в открытом доступе не представляется возможным. В таких случаях разработчики вынуждены либо формировать наборы данных самостоятельно, либо оплачивать услуги специалистов по сбору данных.

Процесс формирования выборок из тысяч изображений номеров, необходимых для многих проектов, является трудоемкой и дорогостоящей задачей. Разработка и реализация метода генерации «синтетического» набора данных на основе имеющейся сравнительно небольшой выборки может очень упростить подготовку входных наборов данных для обучения моделей.

Можно рассматривать несколько подходов к решению задачи генерации изображений с помощью искусственных нейронных сетей глубокого обучения. Одним из популярных современных направлений является применение генеративно-сопоставительных нейронных сетей — Generative Adversarial Networks (GAN). Такие сети используются в дизайне, программировании, написании текстов, сфере развлечений. Принцип работы моделей, использующих в своей основе генеративно-сопоставительную структуру, был впервые описан в работе [2]. Основной чертой GAN является нахождение отличительных особенностей обучающей выборки и последующее их воспроизведение [3].

Для решения задачи генерации данных, в зависимости от поставленных задач, могут использоваться различные структуры генеративно-сопоставительных моделей [3-5]. Так, для решения задач генерации случайных данных, не имеющих основы, может использоваться стандартная архитектура (GAN), где сеть-Генератор принимает на вход случайную последовательность чисел и имеет входной полносвязный слой. Для задач генерации данных с упором на классы генерируемых объектов может использоваться условная генеративно-сопоставительная модель (CGAN). Если основой для генерации изображения служит другое изображение, то используют модель, переводящую изображение в изображение (pix2pix GAN). Другие популярные архитектуры GAN, принципы использования и практики их разработки описаны в работе [6].

Выбор конкретной архитектуры генеративно-сопоставительной сети и подхода к ее реализации должен выполняться с учетом всех особенностей поставленной задачи.

Проблема исследования. В рамках данного исследования была поставлена цель разработать инструментарий генерации больших наборов искусственных изображений автомобильных номеров на основе небольшой пользовательской выборки реальных изображений. Следующим этапом должна стать интеграция реализованного генератора изображений с программными решениями по распознаванию номеров.

В качестве основы для разработки была выбрана генеративно-состязательная модель. Стандартные структуры генеративно-состязательных сетей не эффективны в задачах создания на изображениях читаемого текста. По этой причине было принято решение разрабатывать генеративно-состязательную сеть на основе модели `pix2pix GAN`. Эта модель позволяет использовать необработанное изображение автомобильного номера в качестве основы для генерируемого изображения. Такой подход дает возможность модели «запомнить» особенности пользовательской выборки и применить их к исходному изображению.

Для обеспечения требуемой функциональности необходимо было решить ряд задач:

- подготовить набор изображений для обучения модели;
- реализовать процедуры предварительной обработки изображений (препроцессинг);
- реализовать модель генеративно-состязательной сети для формирования набора изображений номеров:
 - реализовать структуру сети-Генератора изображений номеров;
 - реализовать структуру сети-Дискриминатора распознавания сгенерированных изображений;
 - сформировать алгоритм обучения весов Генератора и Дискриминатора;
- реализовать возможность получения различных типов выходных данных сети (в зависимости от входных данных, введенных пользователем);
- реализовать возможность выгрузки сгенерированного набора изображений в указанную пользователем директорию.

Материалы и методы. В качестве программной платформы для разработки модели использовалась библиотека TensorFlow с высокоуровневым API Keras [7, 8]. Данная платформа предоставляет возможность создания оптимизированной структуры модели нейронной сети, а также сохранения обученной модели — параметров сети в виде файла формата *.h5.

В процессе обучения сети алгоритм обучения будет стремиться минимизировать ошибку каждой составляющей GAN. Обучение модели-Дискриминатора заключается в минимизации ошибки этой модели при определении типа поступившего на вход изображения номера: реальное или сгенерированное изображение. Обучение модели-Генератора состоит в максимизации ошибки Дискриминатора (цель — генерация таких изображений, которые трудно отличить от реальных).

Таким образом, задача обучения GAN — это минимаксная задача: минимизация ошибки Генератора при максимизации вероятности найти ошибку со стороны Дискриминатора. Идеальным вариантом будет считаться ситуация, когда Дискриминатор попадает в состояние неопределенности и не может определить тип изображения (реальное или сгенерированное).

Для реализации качественной генерации необходимо, чтобы обучающие данные удовлетворяли ряду требований:

- 1) для обучения одной модели необходимо не менее 80 изображений;
- 2) изображения номеров должны быть хорошо читаемыми;
- 3) угол съемки при получении изображений для обучения одной модели должен быть примерно одинаковым.

При наличии в обучающей выборке значительной части некачественных, размытых изображений, эта особенность будет «запоминаться» и воспроизводиться моделью, следствием чего будет являться снижение качества генерации.

Для обучения модели была сформирована обучающая выборка, содержащая 110 реальных изображений автомобильных номеров размером 64x128 пикселей. Был произведен первичный отбор данных по углу поворота изображения автомобильного номера относительно камеры и качеству изображений. Пример полученного изображения представлен на рис. 1.



Рис. 1. Пример элемента обучающей выборки

Все собранные изображения должны пройти этап предобработки для приведения к единому виду, воспринимаемому моделью-Генератором. Предобработка была реализована с использованием сервиса PlatesMania [9] посредством организации запросов к API сервиса. С помощью PlatesMania создаются искусственные (нереалистичные) изображения номеров, которые могут использоваться моделью-Генератором в качестве основы. Пример работы конструктора сервиса представлен на рис. 2.

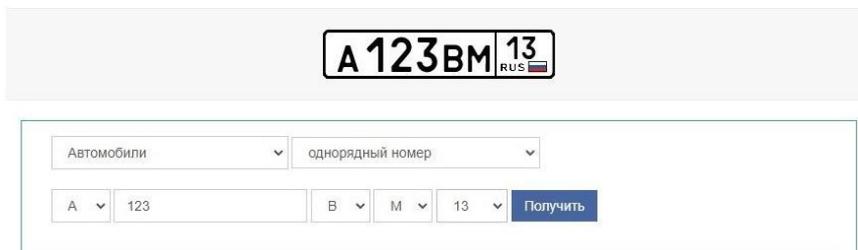


Рис. 2. Пример работы конструктора изображений автомобильных номеров сервиса PlatesMania

После конкатенации реального и искусственного изображений формируется единое изображение размером 64x256 пикселей, подаваемое на вход модели-Генератору в процессе обучения. Пример получаемого таким образом входного изображения представлен на рис. 3.



Рис. 3. Пример входного изображения

Для разработки генеративно-состязательной нейронной сети за основу была взята модель, представленная на рис. 4.

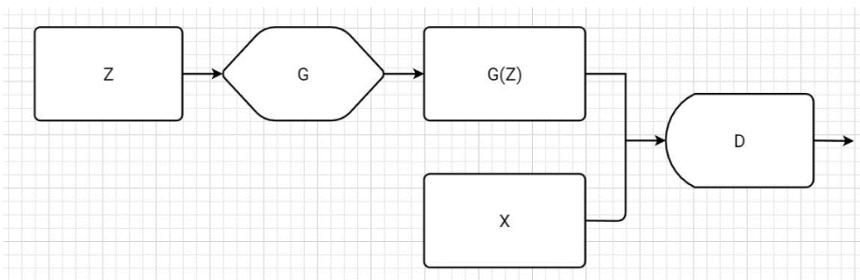


Рис. 4. Модель генеративно-сопоставительной нейронной сети:
 G — модель-Генератор; D — модель-Дискриминатор;
 Z — сформированное входное изображение; G(Z) — сгенерированное
 а основе Z изображение; X — пользовательская выборка данных

Выход Дискриминатора D — значение от 0 до 1, которое интерпретируется как степень уверенности в том, что переданное на вход дискриминатора изображение принадлежит к обучающей выборке реальных изображений.

Для реализации этой модели использовалась архитектура `pix2pix` GAN. На рис. 5 и 6 представлены, соответственно, структуры модели-Генератора и модели-Дискриминатора. В качестве функций активации на промежуточных слоях использовалась функция ReLu, а на выходном слое — гиперболический тангенс.

Обучение сети происходило в течение 60000 эпох и заняло 12 часов. Обучение производилось посредством вычислительных средств персонального компьютера. Технические характеристики устройства:

- Видеокарта: RTX 2060 SUPER 8гб;
- Процессор: Intel Core i5-11600KF;
- Объем оперативной памяти: 32гб.

Результаты. В ходе разработки реализована процедура предобработки первичных изображений автомобильных номеров, построена и обучена генеративно-сопоставительная модель для генерации новых изображений номеров на основе пользовательской выборки.

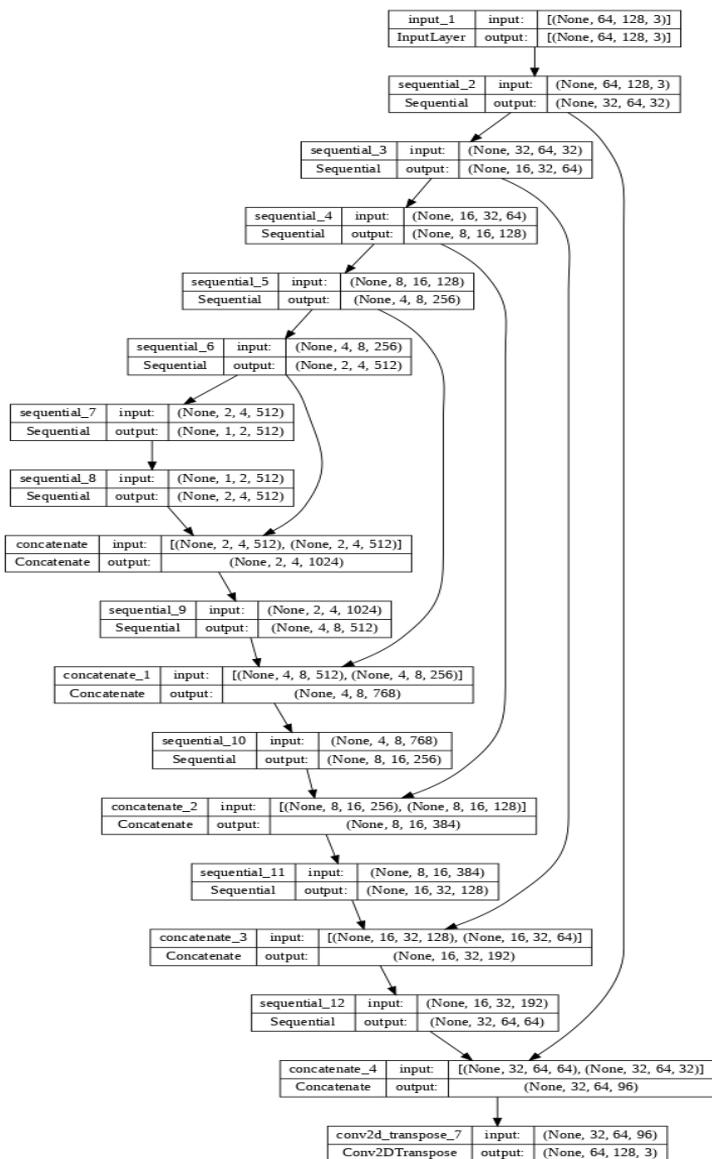


Рис. 5. Структура модели-Генератора

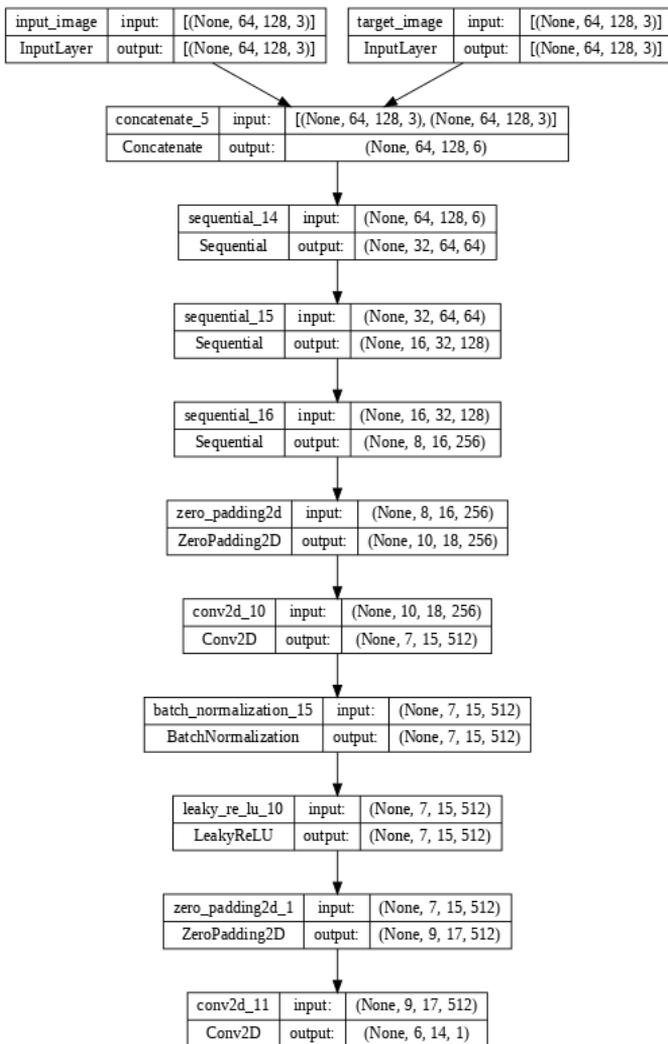


Рис. 6. Структура модели-Дискриминатора

Общий алгоритм взаимодействия пользователя с разработанным программным продуктом выглядит следующим образом.

1. Пользователь формирует набор изображений автомобильных номеров, помещает их в заданную директорию и дает каждому изображению наименование, содержащее параметры изображенного на нем номера.

2. Пользователь формирует соответствующий собранной обучающей выборке набор искусственных нереалистичных изображений (например, посредством конструктора автомобильных номеров сервиса PlatesMania) и дает каждому элементу наименование, соответствующее его реальному прототипу из обучающей выборки.

3. Формируется обучающая выборка: пользователь передает в метод предобработки директории наборов реальных и нереалистичных изображений автомобильных номеров. Метод производит итерацию по данным директориям и конкатенирует изображения по наименованию. Полученные посредством слияния изображения приводятся к размеру 64x256.

4. На вход модели (в метод `fit()`) передаются данные для обучения и желаемое количество эпох обучения. Рекомендуемым количеством итераций является 50000. Производится обучение модели.

5. По окончании обучения, в коренной директории пользователя формируется файл модели формата *.h5; имеется возможность подачи искусственного изображения размером 64x128 на вход обученной модели. На выход будет передано изображение, сгенерированное на основе входного. Сгенерированное моделью изображение «перенимает» особенности обучающей выборки. Пример работы обученной модели представлен на рис. 7.



Рис. 7. Пример входных и выходных данных обученной модели

6. Пользователь указывает директорию с набором искусственных изображений, на основе которых необходимо сформировать новый набор реалистичных данных.

7. Алгоритм производит генерацию и сохраняет сгенерированные изображения в указанной пользователем директории.

Заключение Разработан инструментарий генерации больших наборов изображений автомобильных номеров на основе небольшой пользовательской выборки реальных изображений. Разработка выполнена на основе модели генеративно-состязательной нейронной сети архитектуры pix2pix GAN. Обученная модель позволяет создавать реалистичные изображения автомобильных номеров на основе искусственно созданных нереалистичных изображений.

Созданный программный продукт предназначен для последующей интеграции с решениями по распознаванию автомобильных номеров.

СПИСОК ЛИТЕРАТУРЫ

1. Ivideon: Облачный сервис для видеонаблюдения: [сайт]. — URL: <https://ru.ivideon.com/> (дата обращения 28.05.2023). — Текст: электронный.
2. Goodfellow I. J. Generative Adversarial Networks / I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, Sh. Ozair, A. Courville, Y. Bengio. — URL: <https://arxiv.org/abs/1406.2661> (date of the application 28.05.2023). — Text: electronic.
3. Langr J. GANs in Action: Deep learning with Generative Adversarial Networks. / J. Langr, V. Bok. — New York: Manning, 2019. 276 p. — Direct text.
4. Фостер Д. Генеративное глубокое обучение. Творческий потенциал нейронных сетей / Д. Фостер. — СПб.: Питер, 2020. — 336 с. — Текст: непосредственный.
5. Траск Э. Грокаем глубокое обучение / Э. Трак. — СПб.: Питер, 2020. — 352 с. — Текст: непосредственный.
6. Iglesias G. A survey on GANs for computer vision: Recent research, analysis and taxonomy / G. Iglesias, E. Talavera, A. Diaz-Alvarez. — URL: <https://arxiv.org/pdf/2203.11242.pdf> (date of the application 28.05.2023). — Text: electronic.
7. TensorFlow: Сквозная платформа машинного обучения с открытым исходным кодом: [сайт]. — URL: <https://www.tensorflow.org/> (дата обращения: 28.05.2023). — Текст: электронный.
8. Keras: Высокоуровневый API библиотеки TensorFlow: [сайт]. — URL: https://www.tensorflow.org/api_docs/python/tf/keras (дата обращения: 28.05.2023). — Текст: электронный.
9. PlatesMania: галерея автомобильных номеров: [сайт]. — URL: <https://platesmania.com/> (дата обращения: 28.05.2023). — Текст: электронный.