

УДК 004.932.4, 004.81, 004.052.42

## Исследование устойчивости условной генеративно-состязательной сети Pix2Pix к искажению входных данных разметки изображений

© 2021 г. В. О. Ячная\*, \*\*, магистр; В. Р. Луцив\*, доктор техн. наук

\*Государственный университет аэрокосмического приборостроения, Санкт-Петербург

\*\*Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург

E-mail: tamimio.yvo@hotmail.com, vluciv@mail.ru

Поступила в редакцию 13.09.2021

DOI:10.17586/1023-5086-2021-88-11-46-55

Исследуются особенности генерации изображений предварительно обученной Условной генеративно-состязательной сетью на основе семантической разметки сцены. Семантическая разметка может быть составлена неточно и содержать дефекты, возникшие, например, вследствие преобразований графических форматов, в которых она хранилась или передавалась. Рассматриваются случаи генерации изображений на основе таких некорректных данных — при модификации цветового тона, насыщенности и яркости цветов в цветовых метках различных классов объектов. Определено, что особенно сильное влияние оказывает изменение цветового тона метки, часто приводящее к смене класса, которому она соответствует, поэтому при выборе цветового набора для аннотации пикселей следует учитывать равномерность распределения цветовых параметров по цветовой шкале и следить за точностью представления цветовых семантических меток.

**Ключевые слова:** искусственный интеллект, компьютерное зрение, искусственная нейронная сеть, синтетические данные, условная генеративно-состязательная нейронная сеть, семантическая разметка.

**Коды OCIS:** 110.0110, 110.2960, 100.0100, 100.2000, 100.4994

### ВВЕДЕНИЕ

Синтез искусственных данных, в частности изображений, открывает новую нишу в области обработки видеoinформации и становится востребованным во многих областях деятельности — от коррекции и восстановления фотографий [1] до подготовки баз данных для обучения искусственных нейронных сетей [2]. Автоматизация генерации изображений позволяет значительно сократить как временные, так и финансовые затраты процессов, выполняемых вручную.

Ввиду сложности структуры изображений на сегодняшний момент не создано нейрон-

ных сетей для практического использования, генерирующих изображения без предварительных данных, описывающих их общее построение [3]. Поэтому в качестве основы для генерации изображения следует использовать контурную или семантическую разметку [3], текстовое [4] или иное описание сцены и содержащихся в ней объектов.

Семантическая разметка (или карта) может, в частности, представлять собой изображение, каждый пиксел которого, а именно его цвет, соответствует определенному классу объектов, присутствующих в сцене [5]. Генерация

изображения на основе такой семантической карты представляет собой задачу трансляции изображения, т.е. требуется построить соответствия между входным и выходным изображениями [6].

Для создания синтетических изображений на основе рассматриваемого типа семантической разметки с помощью генеративных моделей искусственных нейронных сетей применяются специальные тренировочные базы данных. Каждый элемент такой базы данных должен состоять из двух изображений: целевого изображения объекта или сцены (например фотография) и изображения, представляющего собой попиксельную аннотацию целевого изображения — семантическую разметку.

Следует отметить, что не имеет значения, какой именно цвет выбран для описания каждого класса семантической разметки, но важно, чтобы один и тот же класс на каждом примере из набора обозначался одним и тем же цветом, иначе возникают ошибки генерации изображения, которым и посвящена эта статья.

В таком наборе данных размеченные объекты переднего плана не должны иметь промежутков, т.е. если какой-то фон виден «сквозь» какой-либо объект переднего плана, он считается частью переднего плана. Например, в случае использования в качестве целевых изображений фотографий городских пейзажей небольшие промежутки между листьями деревьев на фоне дома или неба принимаются за листву. Это относится и к тем областям, где два или более класса сильно «смешаны» друг с другом: они будут помечены классом переднего плана.

Соответствующие наборы данных могут быть получены, например, путем ручной разметки целевых изображений или в результате работы какой-либо сегментирующей системы. Ручная разметка данных — это трудоемкий процесс, требующий больших временных затрат, например, при подготовке баз данных для такой распространенной задачи, как сегментация, требуется набор изображений с соответствующими метками классов, на которых необходимо присвоить соответствующую метку каждому пикселу [7]. Учитывая, что для многих современных задач нужны наборы из десятков тысяч примеров, создание баз данных в нужном объеме может потребовать

значительных финансовых вложений. Более того, при таком подходе возможны различные дефекты при составлении семантической разметки. Другой метод требует использования предобученной сегментирующей искусственной нейронной сети (как например [8, 9]) и соответствующего набора данных или разработки собственной системы при наличии собственной базы изображений для каждого из проектов. Помимо этого необходимо, чтобы данная система сегментировала изображения с достаточной для поставленных задач точностью.

Вместе с тем, различные преобразования и манипуляции с изображениями, содержащими семантическую разметку, могут приводить к цветовым трансформациям, снижающим качество разметки. Это может происходить, например, при переходе из формата RGB в форматы с индексированной цветовой палитрой или формат JPEG (с целью сжатия изображений при их передаче или хранении) и обратно в формат RGB. Особенно критично это может быть на изображениях из наборов, содержащих значительное число различных классов объектов.

Отслеживать вручную аномалии и дефекты на семантической разметке также представляется трудоемкой задачей, а автоматизированное устранение дефектов может быть сопряжено с рядом сложностей.

Соответственно различные факторы и модификации входных данных могут влиять на качество синтезируемых генеративными моделями нейронных сетей изображений как в худшую, так и в лучшую сторону [10]. Актуальность и распространенность описанной проблемы подтверждается наличием ряда работ в этом направлении, в частности, разработкой программного комплекса [11], определяющего некорректность и несбалансированность наборов данных.

Таким образом, целью настоящей работы является исследование проблем, возникающих при генерации изображений Условной генеративно-состязательной сетью при работе с некорректными входными данными. В первом разделе описана разработанная и обученная Условная генеративно-состязательная сеть, а также методика оценки синтезируемых изображений. Во втором разделе описаны особенности практически выполненных

экспериментов и полученные в них результаты. В заключении подведены итоги выполненного исследования.

## 1. УСЛОВНАЯ ГЕНЕРАТИВНО-СОСЯЗАТЕЛЬНАЯ СЕТЬ И МЕТОДЫ ОЦЕНКИ ПАРАМЕТРОВ КАЧЕСТВА ГЕНЕРИРУЕМЫХ ИЗОБРАЖЕНИЙ

Рассмотрим сначала архитектуру и свойства нейронной сети Pix2Pix, используемой в настоящей работе. Условная генеративно-состязательная сеть является расширением архитектуры Генеративно-состязательной сети: она так же состоит из двух частей — генератора и дискриминатора, однако при обучении Условных генеративных нейросетей обеспечивается контроль над создаваемыми данными.

Генерирующая часть Условной генеративно-состязательной сети получает изображение в качестве входных данных и переводит его в другое изображение, как это описано, например в работе [12]. Входное изображение является дополнительной информацией для генерации данных — условием, представляющим собой в настоящей работе семантическую разметку сцены. Классифицирующая часть сети (дискриминатор) получает входное изображение и парное ему реальное или сгенерированное изображение и исходя из этих данных определяет, является ли парное изображение реальным или поддельным.

В качестве генератора используется модификация модели автоэнкодера (кодера-декодера) — модель U-Net [13]. Архитектура модели U-Net подразумевает сначала понижение дискретизации входного изображения до уровня «узкого места», а затем сжатое представление интерполируется до окончательного изображения с желаемым размером. Также между слоями одного размера в кодере и декодере устанавливаются пропускающие соединения, позволяющие обходить «узкое место». Пропускающие соединения гарантируют, что общая для входа и выхода низкоуровневая информация передается по сети непосредственно [14].

В качестве дискриминатора используется модель PatchGAN в отличие от традиционной модели Генеративно-состязательной сети, которая использует сверточную нейронную сеть для классификации изображений. Эта

сеть предназначена для классификации не всего изображения, а его фрагментов [14]. В этом случае классификатор предсказывает не класс изображения целиком (является ли оно настоящим или искусственным), а классы, соответствующие его фрагментам. Такой результат представляется в виде матрицы, а затем усредняется. Разработанная сеть Pix2Pix предназначена для работы с цветными изображениями разрешения  $256 \times 256$  пикселей.

Рассмотрим дискриминатор данной Условной генеративно-состязательной сети. Дискриминатор имеет нейроны с размером рецептивного поля  $70 \times 70$  [14]. Сверточные слои дискриминатора создаются таким образом, что число выходных фильтров (представляющее размерность выходного пространства) следующее: 64 (для входного слоя), 128, 256, 512, 1 (последнее — для выходного слоя). Согласно публикации [14], для первого слоя не применяется пакетная нормализация. В качестве функции активации для сверточных слоев дискриминатора выбрана функция LeakyReLU со значением углового коэффициента 0,2.

Дискриминатор представляет собой модель бинарного классификатора, т.е. ее выходные данные — значение вероятности (0 или 1), в данном случае — это вероятность того, является ли соответствующий фрагмент входного изображения реальным или синтетическим. Такая модель дискриминатора реализует процесс обработки двух входных матриц (изображений) с созданием единственной результирующей. Выходная матрица модели представляет собой карту активаций размером  $16 \times 16$  пикселей, или другими словами — 256 значений вероятности достоверности фрагментов разрешением  $70 \times 70$  пикселей для изображений разрешением  $256 \times 256$  пикселей.

В качестве функции ошибки для дискриминатора выбран средний квадрат ошибки. Для оптимизации сети используется метод адаптивной оценки моментов Adam [15].

Далее рассмотрим генератор данной Условной генеративно-состязательной сети. Как отмечалось ранее, генератор строится на основе модели U-Net. Такая архитектура состоит из двух частей — кодера для понижения дискретизации и декодера для повышения дискретизации, а их соответствующие уровни связаны пропускающими соединениями. Пропускающие соединения добавляются

между слоями с картами признаков одинакового размера, т.е. первый слой кодера связан с последним слоем декодера, второй слой кодера связан с предпоследним слоем декодера и т.д.

Кодирующая часть модели (кодер) представляет собой набор сверточных слоев, которые используют шаг свертки  $2 \times 2$  и размер ядра свертки  $4 \times 4$  для субдискретизации входного исходного изображения до уровня «узкого места». Сверточные слои создаются таким образом, что число выходных фильтров следующее: 64 (для первого слоя), 128, 256 и 512 (для последних трех слоев). Как и в случае дискриминатора, для сверточных слоев кодера используется активационная функция LeakyReLU (со значением углового коэффициента 0,2), а пакетная нормализация не применяется на первом слое.

Декодирующая часть модели (декодер) считывает данные из «узкого места» и использует операцию обратной свертки для повышения дискретизации до требуемого размера выходного изображения. Ввиду наличия пропускающих соединений и объединения таких слоев с соответствующими слоям кодера число выходных фильтров становится в два раза больше. Слои декодера также используют шаг  $2 \times 2$  и размер ядра  $4 \times 4$ , однако в качестве функции активации в данном случае используется ReLU [14].

Последний слой генератора представляет собой сверточный слой, размерность выходного пространства которого зависит от количества каналов изображения, которых в данном случае 3. В качестве функции активации используется гиперболический тангенс, что характерно для большинства генераторов Генеративно-состязательной сети [16]. Модель генератора имеет один вход и один выход, размеры которых совпадают, скрытые слои кодера и декодера связаны промежуточными соединениями, а «узкое место» представляет собой слой размером  $2 \times 2$ .

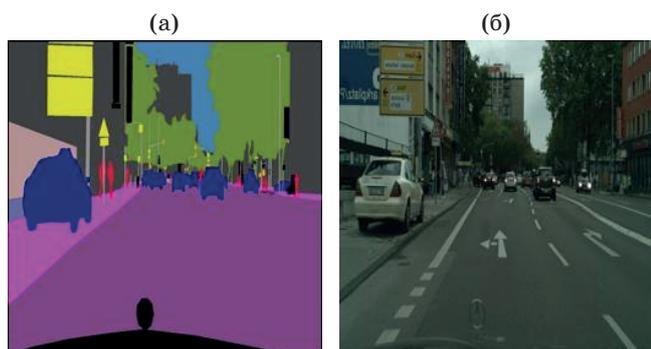
Модели генератора и дискриминатора объединяются в составную модель для совместной работы. При обучении веса генератора обновляются как через состязательную потерю в соответствии с решением дискриминатора о достоверности изображения, так и через среднюю абсолютную потерю, вычисляемую с помощью метрики L1 в соответствии с получае-

мым изображением. Предпочтение отдается вычислению функции потерь именно с помощью метрики L1, а не метрики L2, поскольку согласно [14] последняя приводит к размытию изображений. Состязательная и средняя абсолютная потери суммируются, где потеря по метрике L1 рассматривается как регуляризирующий член и взвешивается с помощью гиперпараметра, равного в данном случае 100. В качестве оптимизатора для составной модели также использовался метод адаптивной оценки моментов.

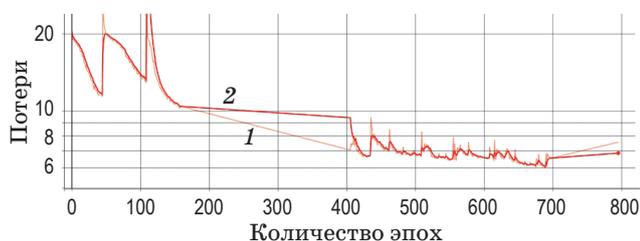
Набор данных для обучения и тестирования разработанной системы состоит из парных изображений — целевого изображения и соответствующей ему семантической разметки. В данном исследовании используется база городских пейзажей «Cityscapes» [17], содержащая 20 различных классов объектов. Каждый пример набора состоит из двух частей: фотография уличной сцены, содержащей элементы городского пейзажа (дороги, здания, транспорт, деревья и др.), и попиксельная аннотация этой фотографии, представляющая семантическую разметку. Именно наглядность представления карты разметки в виде изображения и обусловленная этим ее распространенность стали причиной выбора базы «Cityscapes» и использования рассматриваемой сети Pix2Pix в настоящем исследовании. Набор данных содержит 3475 цветных фотографий, записанных в 50 городах в светлое время суток в разное время года. База разделена на две выборки — обучающую или тренировочную (2975 примеров) и тестовую (500 примеров). Пример пары изображений из базы приведен на рис. 1.

Для обучения разработанной нейронной сети тренировочная выборка была увеличена с помощью аугментации «на лету» [18]. Ввиду характера фотографий набора в качестве преобразований использовалось только зеркальное отражение относительно вертикальной оси, применяемое к каждому изображению. Яркостные (или цветовые) и другие искажения для увеличения количества примеров не применялись. В итоге количество данных для обучения было увеличено не более чем в два раза.

Использованная сеть Pix2Pix обучалась в течение 800 эпох на основе неискаженной обучающей части набора «Cityscapes». На рис. 2 приведен график зависимости показателя



**Рис. 1.** Пример пары изображений из датасета «Cityscapes». Семантическая карта — (а), целевое изображение — (б).



**Рис. 2.** График зависимости функции потерь генератора от количества эпох обучения. Экспериментальные данные — 1, результат сглаживания — 2.

потерь на выходе сети генератора от номера эпохи. На данном графике значения потерь были сглажены методом «экспоненциальное скользящее среднее» с коэффициентом 0,7.

Выбор количества эпох обучения Условной генеративно-сопоставительной сети обусловлен методами оценки результатов работы данной нейронной сети. Однако необходимо добавить, что наименьшее значение функция потерь генератора принимает на 690-й эпохе — 6,008.

Поскольку на сегодняшний день проблема оценки результатов работы Условной генеративно-сопоставительной сети, создающей изображения сцен, остается открытой [19], в настоящем исследовании используются базовые метрики задач обработки изображений. Так, для оценки качества изображений, сгенерированных разработанной в настоящей работе моделью, используются следующие метрики:

- средний квадрат ошибки,
- пиковое отношение сигнала к шуму,
- универсальный индекс качества изображения,
- достоверность визуальной информации.

Средний квадрат ошибки показывает, насколько не совпадают два изображения (реально сгенерированное и целевое). Соответственно более низкий показатель (т.е. меньшая ошибка генерации изображения) означает, что исследуемое и целевое изображения более «схожи».

Пиковое отношение сигнала к шуму (ПОСШ) используется как мера оценки сходства между исходным и преобразованными изображениями: тогда как мера среднего квадрата ошибки вычисляется по квадратам отличий преобразованных и исходных изображений, ПОСШ является отношением максимально возможного значения сигнала к значению шума. Таким образом, чем выше ПОСШ, тем лучше качество сжатого или восстановленного изображения. Данная метрика применяется ввиду того, что преобразования генерирующей части разработанной модели включают в себя сжатие и восстановление изображения.

Универсальный индекс качества (УИК) изображения [20] учитывает такие искажения изображения, как искажение яркости или контраста. Данный индекс используется для определения степени преобразования (искажения) соответствующих данных из целевого изображения (принимаемого за эталонное) в исследуемом изображении. Диапазон значений этого показателя составляет от  $-1$  до  $1$ , а значение  $1$  указывает, что изображения идентичны [21].

Метрика достоверности визуальной информации основана на статистике естественной сцены и понятии информации изображения, извлекаемой зрительной системой человека. Изображения (и видео-данные) трехмерной визуальной среды происходят из общего класса — класса естественных сцен. Большинство процессов искажения данных в реальном мире искажают эту статистику данного класса изображений и делают их неестественными. Индекс достоверности визуальной информации (ИДВИ) использует статистические модели естественной сцены в сочетании с моделью искажений данных для количественной оценки информации, сравнения исследуемого и эталонного (в данном случае — целевое) изображения. Диапазон значений данной метрики лежит в пределах от  $0$  до  $1$ , а в случае сравнения двух идентичных изображений метрика ИДВИ также равна  $1$ . Метрика ИДВИ

отличается от традиционных методов оценки качества в том, что усиление контраста эталонного изображения без добавления шума приведет к тому, что значение ИДВИ будет больше единицы, означая, что улучшенное изображение имеет лучшее визуальное качество, чем эталонное изображение [22].

## 2. РЕЗУЛЬТАТЫ ПРАКТИЧЕСКИ ВЫПОЛНЕННЫХ ЭКСПЕРИМЕНТОВ

Для проведения экспериментов использовалась тестовая выборка набора данных «Cityscapes» (на его обучающей выборке исследуемая сеть Pix2Pix была обучена, как описано выше). В экспериментах использовались карты разметки из тестового набора, подвергнутые рассмотренным ниже преобразованиям, и определялось согласно описанным выше выбранным метрикам, как исказились при этом изображения, сгенерированные сетью Pix2Pix, обученной на основе неискаженной разметки.

В исследованиях был выбран класс «легковой автомобиль», представленный на семантической разметке темно-синим цветом — (0,0,142) в формате RGB или (240°, 100%, 55,7%) в формате HSV.

Из 500 изображений тестовой выборки для дальнейших экспериментов были отобраны те, на которых данный класс занимает не менее 10% от площади всего изображения. В выборке таких изображений 275, среди которых класс «легковой автомобиль» покрывает в среднем 33,3% площади изображения.

Для изучения функциональных возможностей разработанной нейронной сети в условиях неопределенности входных данных было решено исследовать ее реакцию по каждому цветовому параметру схемы HSV меток классов независимо друг от друга, чтобы оценить вклад каждого параметра в возможные изменения синтезируемых изображений. Для моделирования некорректных входных данных проводилась модификация значений цветового тона, насыщенности и яркости цвета, описывающего метку выбранного класса. Так, значения цветового тона и яркости цвета понижались и повышались на 4 и 20%, а значение насыщенности — понижалось на 4 и 20%. Рисунки 3–6 иллюстрируют модификацию семантической разметки: на рис. 3 представлено исходное изо-

бражение, на рис. 4 — изменение цветового тона на цвета метки класса «легковой автомобиль», на рис. 5 и 6 — изменение насыщенности и яркости данного цвета, соответственно.

Из примеров на рис. 4 видно, что значительное изменение цветового тона заметно меняет цветовую метку класса и возможно совпадение с цветовыми метками других определенных в наборе классов.

Полученные модифицированные семантические разметки использовались в качестве входного условия для нескольких моделей нейронных сетей: обученных в течение 150, 240, 510, 690 и 800 эпох. Сгенерированные



Рис. 3. Пример корректной семантической разметки.

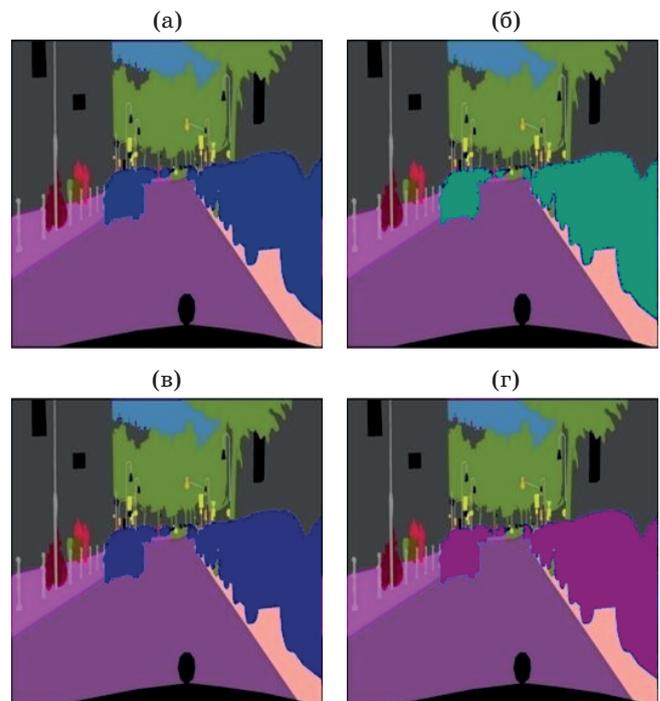
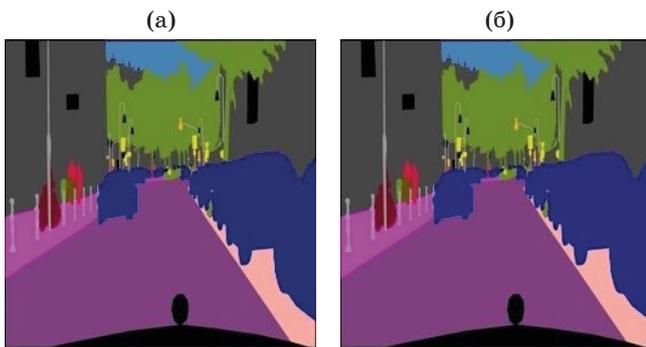
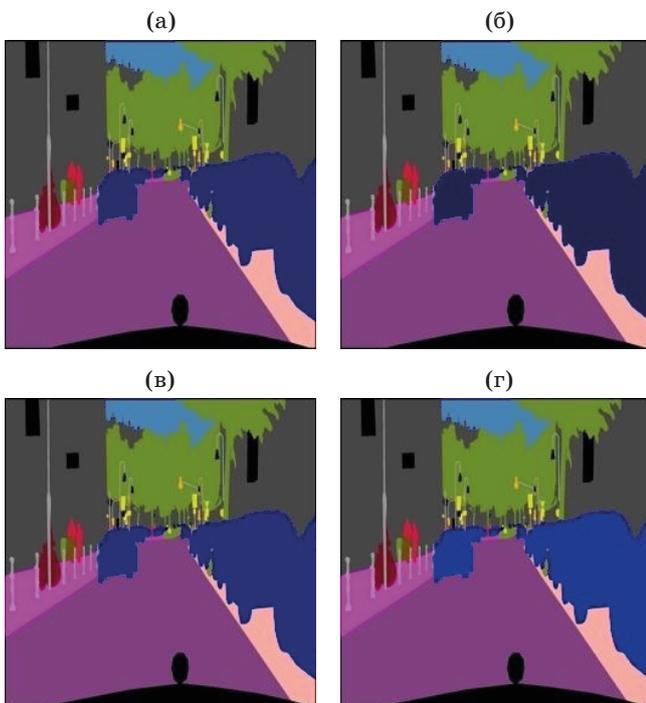


Рис. 4. Изменение цветового тона метки класса. Понижение на 4% — (а), понижение на 20% — (б), повышение на 4% — (в), повышение на 20% — (г).



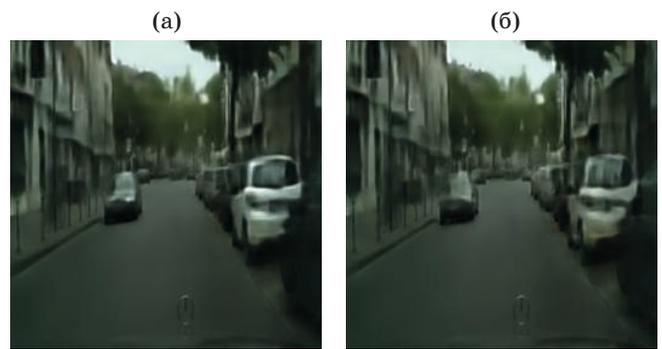
**Рис. 5.** Изменение насыщенности метки класса. Понижение на 4% — (а), понижение на 20% — (б).



**Рис. 6.** Изменение яркости метки класса. Понижение на 4% — (а), понижение на 20% — (б), повышение на 4% — (в), повышение на 20% — (г).

изображения сравниваются с целевыми данными из тестового набора.

Примеры синтезированных нейронной сетью изображений приведены на рис. 7 (использовалась модель, обученная в течение 800 эпох). Так, на нем представлены результаты трансляции немодифицированной семантической разметки (рис. 3) и семантической разметки, яркость выбранного класса на которой была понижена на 20% (рис. 6б). Из примера видно, что объекты класса «легковой автомобиль» на изображении, полученном

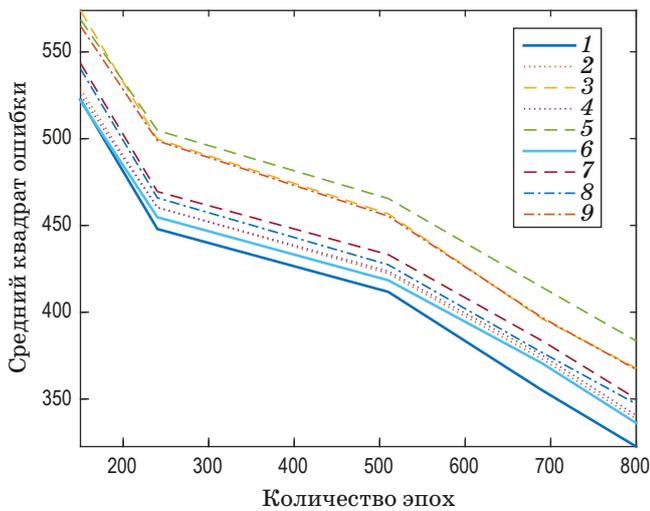


**Рис. 7.** Примеры синтезированных изображений из немодифицированной (а) и из модифицированной семантической разметки (понижение яркости на 20%) (б).

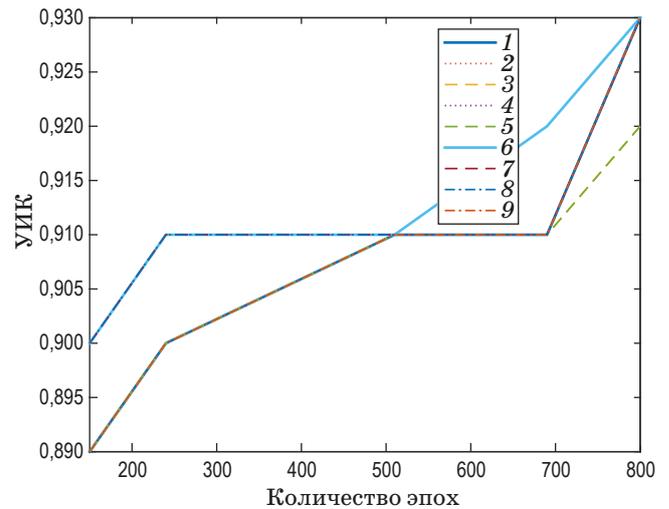
из модифицированной семантической разметки (рис. 7б), визуально менее качественные: такие структурные элементы класса, как «фары» или «окна», размыты (в сравнении с рис. 7а), а их положение не в полной мере соответствует реалистичным изображениям.

На рис. 8–11 приведены результаты оценки качества сгенерированных изображений с описанными модификациями семантической разметки. Для демонстрации динамики качества синтеза изображений приведены оценки для промежуточных моделей разработанной сети, обученных на немодифицированном наборе данных.

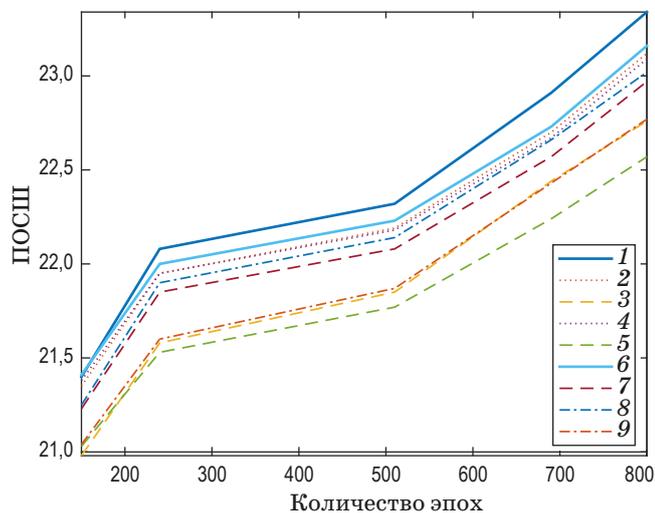
Из рис. 8–11 видно, что показатели, характеризующие качество синтезируемых изображений, улучшаются при увеличении количества эпох обучения, что говорит о достаточности набора обучающих данных и отсутствии эффекта переобучения сети. Так, средний квадрат ошибки сгенерированного изображения по отношению к целевому уменьшается, а значения ПОСШ, УИК изображения и достоверности визуальной информации растут. Такая динамика изменения параметров качества характерна как для использования корректных входных данных, так и для входных данных, содержащих дефекты. На рис. 8–11 не представлены значения метрик, вычисленные для изображений, синтезированных из семантических разметок, цветовой тон которых был изменен на 20%. Это связано с тем, что данная модификация семантической разметки представляет собой существенный дефект, и характеристики генерируемых изображений



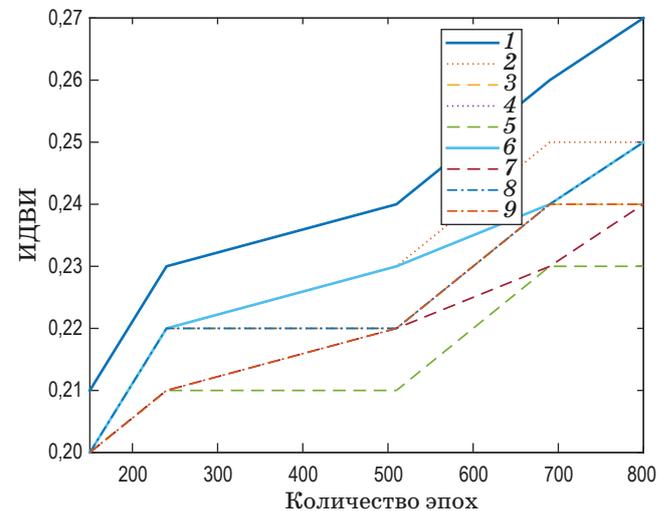
**Рис. 8.** Зависимости среднего квадрата ошибки генерируемого изображения от количества эпох обучения на немодифицированном наборе. Исходная семантическая разметка — 1, понижение насыщенности на 4% — 2, понижение насыщенности на 20% — 3, понижение яркости на 4% — 4, понижение яркости на 20% — 5, повышение яркости на 4% — 6, повышение яркости на 20% — 7, понижение цветового тона на 4% — 8, повышение цветового тона на 4% — 9.



**Рис. 10.** Зависимости универсального индекса качества генерируемого изображения от количества эпох обучения на немодифицированном наборе. Исходная семантическая разметка — 1, понижение насыщенности на 4% — 2, понижение насыщенности на 20% — 3, понижение яркости на 4% — 4, понижение яркости на 20% — 5, повышение яркости на 4% — 6, повышение яркости на 20% — 7, понижение цветового тона на 4% — 8, повышение цветового тона на 4% — 9.



**Рис. 9.** Зависимости ПОСШ генерируемого изображения от количества эпох обучения на немодифицированном наборе. Исходная семантическая разметка — 1, понижение насыщенности на 4% — 2, понижение насыщенности на 20% — 3, понижение яркости на 4% — 4, понижение яркости на 20% — 5, повышение яркости на 4% — 6, повышение яркости на 20% — 7, понижение цветового тона на 4% — 8, повышение цветового тона на 4% — 9.



**Рис. 11.** Зависимости индекса достоверности визуальной информации генерируемого изображения от количества эпох обучения на немодифицированном наборе. Исходная семантическая разметка — 1, понижение насыщенности на 4% — 2, понижение насыщенности на 20% — 3, понижение яркости на 4% — 4, понижение яркости на 20% — 5, повышение яркости на 4% — 6, повышение яркости на 20% — 7, понижение цветового тона на 4% — 8, повышение цветового тона на 4% — 9.

значительно ухудшаются. Например, средний квадрат ошибки при понижении цветового тона на 20% в среднем в 4 раза превышает средний квадрат ошибки для немодифицированного тона на каждой из рассматриваемых эпох, а повышение тона на 20% соответствует двукратному ухудшению результатов генерации.

Из рис. 8–11 также следует, что когда для синтеза изображений применяется модифицированная семантическая разметка, показатели рассматриваемых метрик в большинстве случаев заметно снижаются в сравнении с изображениями, получаемыми из семантической разметки без дефектов. В частности, как и следовало ожидать, согласно значениям среднего квадрата ошибки, ПОСШ и достоверности визуальной информации качество изображений, генерируемых из модифицированной семантической разметки, ниже, чем при синтезе из семантической разметки, содержащей корректные цветовые значения. Наблюдается и динамика ухудшения результатов генерации изображений при возрастающем изменении значений цветового тона, насыщенности и яркости цвета метки класса. В среднем повышение яркости цвета метки выбранного класса на 4% представляет собой наименее влияющую на результат генерации аномалию. Стоит также подчеркнуть, что в случае изменения насыщенности универсальный индекс качества не изменяется, а в случае изменения яркости на 4% даже превосходит значения, получаемые для исходной семантической разметки. В целом, модификация насыщенности и яркости цветовой метки класса вносит меньший вклад в изменение качества создаваемых изображений, и генерируемое изображение остается в рамках целевого класса, в то время как модификация цветового тона приводит к значительным и критичным искажениям.

## ЗАКЛЮЧЕНИЕ

В приведенных выше материалах описывается проблема синтеза изображений генеративной моделью искусственных нейронных сетей — Условной генеративно-состязатель-

ной сетью, а именно, проблема подготовки баз данных для обучения и непосредственного использования таких сетей. В качестве входных данных для работы Условной генеративно-состязательной сети, в частности, могут выступать изображения, содержащие семантическую разметку сцены. Обработка фотографий для получения семантической разметки сцены сопряжена с рядом сложностей, кроме того, преобразование графических форматов файлов изображений для экономного хранения и передачи таких данных может вызывать на них дополнительные отклонения, приводящие к ошибкам определения класса при наличии значительного числа различных классов объектов в анализируемой сцене. В связи с этим исследовано поведение обученной системы генерации фотореалистичных изображений при использовании некорректных входных данных.

Результаты практически выполненных экспериментов показывают, что модификации цветового тона, насыщенности и яркости меток, используемых для описания классов объектов на семантической разметке, снижают качество синтезируемых изображений при использовании предварительно обученной нейронной сети. Изменение параметров яркости и насыщенности в системе HSV может вызывать визуально неразличимые изменения цвета метки. В то же время, особенно сильное отрицательное влияние оказывает изменение цветового тона метки, которое может приводить к смене класса, которому она соответствует. Вследствие этого при выборе цветового набора для аннотаций пикселей следует руководствоваться не только визуальными предпочтениями разработчика, но также и равномерностью распределения цветовых параметров по цветовой шкале (разнесенностью меток классов в цветовом пространстве). Помимо этого, следует вводить дополнительные требования к точности представления цветовых меток (допустимым погрешностям их кодирования) в семантической разметке сцены, используемой в качестве входных данных Условной генеративно-состязательной сети.

## ЛИТЕРАТУРА

1. *Yamauchi H., Haber J., Seidel H.-P.* Image restoration using multiresolution texture synthesis and image inpainting // Proc. Computer Graphics International. 2003. Tokyo, Japan. P. 120–125.

2. *Roberts M., Paczan N.* Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. [Электронный ресурс] / arxiv.org: информационный ресурс. URL: <https://arxiv.org/pdf/2011.02523.pdf> (дата обращения: 27.08.2021).
3. *Nikolenko S.I.* Synthetic data for deep learning. 2019. [Электронный ресурс] / arxiv.org: информационный ресурс. URL: <https://arxiv.org/pdf/1909.11512.pdf> (дата обращения: 26.08.2021).
4. *Zhang H., Xu T., Li H., Zhang S.* StackGAN++: Realistic image synthesis with stacked generative adversarial networks // *IEEE Trans. Pattern Analysis and Machine Intelligence*. 2019. V. 41. № 8. P. 1947–1962.
5. *Shapiro L., Stockman G.* Computer vision. New Jersey: Prentice-Hall, 2001. 580 p.
6. Задача трансляции изображений. [Электронный ресурс] / neerc.ifmo.ru: информационный ресурс. URL: [http://neerc.ifmo.ru/wiki/index.php?title=%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0\\_%D1%82%D1%80%D0%B0%D0%BD%D1%81%D0%BB%D1%8F%D1%86%D0%B8%D0%B8\\_%D0%B8%D0%B7%D0%BE%D0%B1%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D0%B9](http://neerc.ifmo.ru/wiki/index.php?title=%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D1%82%D1%80%D0%B0%D0%BD%D1%81%D0%BB%D1%8F%D1%86%D0%B8%D0%B8_%D0%B8%D0%B7%D0%BE%D0%B1%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D0%B9) (дата обращения: 25.08.2021).
7. *Guo D., Pei Y., Zheng K., Yu H., Lu Y., Wang S.* Degraded image semantic segmentation with dense-gram networks // *IEEE Trans. Image Proc.* 2019. V. 29. P. 782–795.
8. *Long J., Shelhamer E., Darrell T.* Fully convolutional networks for semantic segmentation // *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 7–12, 2015. Boston, MA, USA. P. 3431–3440.
9. *Badrinarayanan V., Kendall A., Cipolla R.* Segnet: A deep convolutional encoder-decoder architecture for image segmentation // *IEEE Trans. Pattern Analysis and Machine Intelligence*. 2017. V. 39. № 12. P. 2481–2495.
10. *Yachnaya V.O., Mikhalkova M.A., Yablokov E.N., Lutsiv V.R.* Noise model effect upon the GAN-synthesized images // *Proc. IEEE Wave Electronics and Its Application in Information and Telecommunication Systems (WECONF-2020)*. 2020. Saint-Petersburg, Russia. P. 1–6.
11. Know your data. [Электронный ресурс] / google.com: информационный ресурс. URL: <https://knowyourdata.withgoogle.com/> (дата обращения: 28.08.2021).
12. *Mirza M., Osindero S.* Conditional generative adversarial nets. 2014. [Электронный ресурс] / arxiv.org: информационный ресурс. URL: <https://arxiv.org/pdf/1411.1784.pdf> (дата обращения: 25.08.2021).
13. *Ronneberger O., Fischer P., Brox T.* U-net: Convolutional networks for biomedical image segmentation // *Proc. 18th Internat. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI-2015)*. October 5–9, 2015. Munich, Germany. P. 234–241.
14. *Isola P., Zhu J.-Y., Zhou T., Efros A.* Image-to-image translation with conditional adversarial networks // *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 21–26 July, 2017. Honolulu, HI, USA. P. 5967–5976.
15. *Kingma D.P., Ba J.* Adam: A method for stochastic optimization // *Proc. 3rd Internat. Conf. Learning Representations*. San Diego. May 7–9, 2015. P. 1–15.
16. *Brownlee J.* How to implement Pix2Pix GAN models from scratch with Keras. [Электронный ресурс] / machinelearningmastery.com: информационный ресурс. URL: <https://machinelearningmastery.com/how-to-implement-pix2pix-gan-models-from-scratch-with-keras/> (дата обращения: 26.08.2021).
17. *Cordts M., Omran M., Ramos S., Rehfeld T.* The cityscapes dataset for semantic urban scene understanding // *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR-2016)*. Las Vegas, NV, USA. June 27–30, 2016. P. 3213–3223.
18. *Гайер А.В., Шешкус А.В., Чернышова Ю.С.* Аугментация обучающей выборки «на лету» для обучения нейронных сетей // *Тр. ин-та системного анализа РАН*. 2018. Т. 68. № S1. С. 150–157.
19. *Borji A.* Pros and cons of GAN evaluation measures // *Computer Vision and Image Understanding*. 2019. V. 179 (February 2019). P. 41–65.
20. *Zhou Wang, Bovik A.C.* A universal image quality index // *IEEE Signal Proc. Lett.* 2002. V. 9. № 3. P. 81–84.
21. *Jagalingam P., Hegde A.* A review of quality metrics for fused image // *Proc. Internat. Conf. Water Resources, Coastal and Ocean Engineering (ICWRCOE 2015)*. Aquatic Procedia. 2015. V.4. P. 133–142.
22. *Sheikh H.R., Bovik A.C.* Image information and visual quality // *IEEE Trans. Image Proc.* V. 15. № 2. P. 430–444. Feb. 2006, doi: 10.1109/TIP.2005.859378.