

Графические нейронные сети и проблемы верификации изображений

Н.А. Бондарева ¹

¹ ИИМ им. М.В. Келдыша РАН, Миусская пл., д.4, Москва, 125047, Россия

Аннотация

Работа предлагает обзор современных графических нейросетей типа text-to-image и получения изображения по текстовому запросу. В работе рассматривается ряд проблем, возникающих при использовании сетей text-to-image и возможные методы их решения. Одной из актуальных задач, связанных с изучением графических нейросетевых технологий, становится изучение нейросетевых изображений и выявления изображений, полученных с помощью нейросетей, среди прочего графического контента. Предлагаются подходы к решению задач, направленных на решение проблемы верификации медиа материалов и разработку алгоритмов для выявления искусственного (нейросетевого) происхождения фото и видео материалов. Стремительное развитие нейросетевых технологий в этой области может оказать значительное влияние на общество, рынок профессий и СМИ, что делает особенно актуальной задачу выявления нейросетевых изображений среди прочего графического контента.

Ключевые слова

Машинное обучение, компьютерное зрение и распознавание образов, нейронная сеть, компьютерная графика, Text-to-image

Graphical Neural Networks and Image Verification Problems

N.A. Bondareva ¹

¹ Keldysh Institute of Applied Mathematics, Miusskaya sq. 4, Moscow, 125047, Russia

Abstract

The paper offers an overview of modern text-to-image graphical neural networks and text-to-image image retrieval. The paper discusses a number of problems arising in the use of text-to-image networks and possible methods for their solution. One of the actual tasks related to the study of graphical neural network technologies becomes the study of neural network images and identification of images obtained with the help of neural networks among other graphical content. We propose approaches to solving problems aimed at solving the problem of verification of media materials and developing algorithms to detect artificial (neural network) origin of photo and video materials. The rapid development of neural network technologies in this area can have a significant impact on society, the market of professions and the media, which makes the task of identifying neural network images among other graphic content particularly relevant.

Keywords

Machine learning, computer vision and pattern recognition, neural network, computer graphics, Text-to-image

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: nicibond9991@gmail.com (Н.А. Бондарева)

ORCID: 0000-0002-7586-903X (Н.А. Бондарева)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

В настоящее время происходит стремительное развитие области нейросетевых технологий, которые с каждым днем становятся все сложнее и обретают все больше умений и возможностей. В частности особенную популярность в наши дни приобретают графические нейросети, способные обрабатывать изображения самыми различными способами от анимирования фотографий до автоматического создания полноценных изображений по текстовому запросу пользователя.

Результаты работы этих нейросетей уже начинают широко применяться во множестве отраслей, от медиа до медицины. Однако такой быстрый прогресс также вызывает немало социальных изменений.

Графические нейросети могут интерпретировать лингвистическую структуру пользовательского запроса, обрабатывать его и на основе этого формировать реалистичные визуальные образы. Их задача включает одновременное управление несколькими объектами, их атрибутами и пространственными отношениями. Для правильной интерпретации текстового запроса алгоритм должен не только правильно составить каждую характеристику объекта, но и установить связи между ними. Нейросети для преобразования текста в изображения основываются на диффузионных нейросетевых моделях [1,2]. Эти модели были представлены еще в 2015 году, но стали популярными только после работы [1], и на сегодняшний день они достигли впечатляющих результатов во многих задачах генерации и изменения изображений, от создания изображений по текстовому запросу (text-to-image) до восстановления деталей изображений (inpainting) и увеличения их разрешения (super-resolution). Самой перспективной альтернативой диффузионным моделям до недавнего времени считались генеративно-состязательные нейросети (GAN), но они слабо применимы в задаче генерации изображений по тексту и нестабильны в обучении [3].

В таких нейросетевых моделях, как text-to-image, входные данные представляют собой лингвистическую конструкцию (текстовый запрос), которая направляет дальнейший процесс обработки. Для этой цели используются языковые модели, которые обучены на парах изображений и соответствующих им текстовых описаний. Эти модели способны одновременно понимать содержание изображений и текстов. Примером такой модели является CLIP (Contrastive Language — Image Pre-training), разработанная OpenAI. CLIP может преобразовывать изображения и тексты в общее латентное векторное пространство, где вектор представляет собой набор значений. В этом пространстве возможно, например, находить ближайшие изображения к заданному текстовому запросу, поскольку это просто операция с векторами. Модель Latent Diffusion [4], представленная в 2021 году, обучает модель векторному пространству текстов для генерации изображений из направленного шума. Эта модель использует свойства общего латентного пространства для текстов и изображений. Аналогичные принципы работы используются в моделях Stable Diffusion, Imagen и других больших нейросетях для преобразования текста в изображения. Основные принципы работы таких моделей описаны в работах [5-8].

2. Графические нейросети

Данный раздел представляет собой краткий обзор наиболее популярных и крупных коммерческих нейросетей типа text-to-image, приобретших широкую известность в последний год. В число их входят:

- коммерческая нейросеть Midjourney,
- DALL-E 2 обновлённая версия нейросети, впервые продемонстрированная в январе 2021-го;
- нейросеть Stable Diffusion с открытым исходным кодом, ставшая основой для десятков новых проектов;
- ruDALL-E – российская нейросеть, основанная на генеративных моделях от фирм SberDevices и Sber AI.

Midjourney открылась для пользователей в марте 2022 нейросеть генерации изображений и быстро завоевала популярность не в последнюю очередь благодаря тому факту, что стала общедоступной раньше, чем подобные ей нейросети DALL-E и Stable Diffusion. Midjourney отличается выразительным и легко узнаваемым стилем, она стремительно развивается и совершенствуется буквально на глазах, позволяя пользователям воссоздавать в графическом виде все более и более сложные запросы.

DALL-E 2 — одна из самых популярных систем нейросетевой графики, разработанная компанией OpenAI с 12 миллиардами параметров на базе GPT-3, обученная генерировать изображения из текстовых описаний с использованием набора данных пар текст-изображение. Она способна генерировать оригинальные изображения из текстовых описаний, а также позволяет пользователям загружать свои изображения и редактировать их, к примеру добавляя какие-то элементы. Каждому пользователю доступны бесплатные 50 запросов в первый месяц и 15 — в каждый следующий месяц. Проект работает через API. Для пользователей из России он доступен только через VPN.

Stable Diffusion выделяется на фоне аналогичных нейросетей в первую очередь открытым исходным кодом и сможет составить конкуренцию DALL-E 2 по качеству. В Stable Diffusion есть ряд параметров для гибкой настройки генератора изображений.

В отличие от Midjourney, которая в большинстве случаев выдает эстетичный и проработанный результат, DALL-E 2 и Stable Diffusion необходимо дополнительно настраивать.

ruDALL-E представляет собой семейство генеративных моделей от SberDevices и Sber AI. Нейросеть разработали и обучили исследователи Sber AI при партнёрской поддержке ученых из Института искусственного интеллекта AIRI на объединённом датасете Sber AI и SberDevices из 1 млрд пар «текст — изображение».

Специалистами было создано и обучено две версии модели, наименованные в честь двух великих российских абстракционистов — Василия Кандинского и Казимира Малевича:

- ruDALL-E Kandinsky (XXL) с 12 миллиардами параметров;
- ruDALL-E Malevich (XL) с 1.3 миллиардами параметров.

Обе модели способны по короткому текстовому описанию генерировать красочные изображения на разнообразные темы. Модель Kandinsky уже обновилась до версии 2.2 и в настоящее время активно развивается.

3. Проблемы и методы решения

Применение нейросетей для построения изображений, как и всякий прогресс, дает обществу большое количество потенциальных возможностей, и в то же время открывается не менее широкий спектр проблем, связанных с недобросовестным применением этой технологии в целях мошеннической, дезинформационной и антигосударственной деятельности.

3.1. Виды проблем

В число самых очевидных проблем, которые можно перечислить, входят:

- неконтролируемые выбросы фейковых новостей;
- мошенничество;
- дискредитация юридической доказательной базы фото и видео источников;
- манипуляция историческими источниками в целях искажения реальной действительности;
- беспрепятственное продуцирование материалов, порочащих репутацию и честь медийной личности, чье лицо достаточно известно в информационном сетевом пространстве;
- манипуляция человеческим сознанием и внедрение заведомо ложных сведений.

Все эти проблемы стремительно перестают иметь чисто теоретический потенциал и с каждым днем все больше становятся частью нашей реальности. Постоянное развитие возможностей нейросетей и следующие одно за другим обновления делают результаты их работы все лучше и

реалистичнее. Это, несомненно, большое открытие для цифрового мира, но также оно может стать серьезной проблемой, если будет использоваться в недобросовестных целях.

В большинстве самых популярных нейросетей, особенно платных, разработчики предпринимают ряд специальных мер, призванных уменьшить потенциальное вредоносное влияние и обезопасить общество. В числе таких мер были специальные слова-маркеры, ограничивающие запросы пользователей в определенных рамках (к недопустимым запросам относятся те, что содержат темы насилия, контента 18+, шокирующего контента). В настоящий момент уже начинают вноситься ограничения, не допускающие использования известных государственных и медийных лиц в изображениях, сгенерированных нейросетями.

Ещё одна из мер – это фильтрация обучающего материала для нейросетей [9]. Команда разработчиков DALLÉ-2 прямо заявляет, что при обучении последних моделей отсекала неподходящий контент из обучающей выборки изображений и пар изображение-текст. Этот подход имеет свои плюсы и минусы: с одной стороны, в случаях, когда пользователи находят способ обмануть «маркеры» и добиваются от нейросетей желаемого изображения (в конкретном случае 18+), оно выглядит максимально нереалистично и не может быть использовано никаким образом. С другой стороны, как отмечают сами разработчики, ограничения на этапе обучения могут негативно сказываться на результатах работы нейросети в дальнейшем.

Ещё одна проблема касается авторского права, так как нейросеть способна обучаться на визуальном материале с ярко выраженным авторским стилем и весьма точно воспроизводить его. Уже известны случаи, когда настоящих художников обвиняли в использовании нейросетевых технологий, потому что их работы были слишком стилистически схожи с результатами генерации нейросети, и визуального различия зрителю уловить не удалось [10].

В обществе активно ведутся эксперименты на тему, может ли человек отличить настоящую фотографию/картину/видео от сгенерированной в нейросети [11].

Как только нейросеть обучается до того уровня, когда в результатах её работы устраняются ярко выраженные артефакты (к примеру, анатомические искажения, нелогичная композиция), то отличить её от оригинальной фотографии становится значительно сложнее.

До появления генеративных нейросетей в области судебной экспертизы и фотокриминалистики был разработан ряд методов для того, чтобы различить, была ли фотография обработана с помощью графических редакторов. В частности можно определить, были ли привнесены на фотографию дополнительные элементы, клонированы или замазаны какие-то области, или удалены какие-либо объекты. Чаще всего применяется совокупность методов ELA (анализ уровня ошибок) и PCA (анализ главных компонент), позволяющих по ряду данных, в том числе повторному сжатию и пересохранению обработанных фотографий, определить постороннее вмешательство и проведенные с изображением манипуляции [12-14].

Поскольку нейросетевое изображение (если не подвергалось дополнительной обработке самим пользователем) не проходит череду преобразований и не подвергается сохранению с изменением форматов изображений, то методы, выявляющие артефакты при преобразовании и сжатии фотографий, с высокой степенью вероятности окажутся малоэффективными и полагаться на них нельзя.

Между тем необходимость в подобных инструментах может возникнуть скорее, чем кажется. Рисунок 1 представляет собой наглядную визуализацию новостей о возможном аресте Дональда Трампа в 2022 году [15]. Это изображение нельзя назвать именно фейком, так как информация была размещена с соответствующей подписью и не подавалась как факт, однако само изображение вызвало определенный резонанс, так как выглядит уже намного реалистичнее:



Рисунок 1 – Нейросеть показала, как мог бы выглядеть арест Дональда Трампа (2022 год) [15]

3.2. Методы решения

Если метод визуального человеческого анализа оказывается неэффективен, то целесообразным кажется задействовать методы компьютерного анализа и искусственного интеллекта уже в целях верификации и возможности отличать сгенерированные материалы от подлинников.

Возникает обратная задача: выявить набор признаков, необходимый и достаточный для того чтобы однозначно определять, является ли анализируемое изображение искусственно сгенерированным или настоящим носителем визуальной информации.

Есть несколько путей решений этой задачи:

- разработка специализированных алгоритмов по определению сгенерированных изображений;
- обучение нейросети-классификатора, способной по набору признаков определять искусственно сгенерированные визуальные материалы.

Эти методы представляются наиболее логичными и потенциально возможными. Однако на сегодняшний момент они всё ещё нереализованы.

Область графических нейросетей появилась совсем недавно и находится в стадии активного развития, поэтому с теоретической точки зрения практически не изучена. В этой области пока ещё не разработано общеизвестных и доступных классификаторов и сравнительно мало теоретического материала. Материалы разработчиков же в большинстве случаев являются проприетарными и отсутствуют в свободном доступе.

Прежде чем станет возможно выработать эффективные алгоритмы противодействия, необходимо провести серьёзную подготовительную работу по исследованию данной области и выявить набор признаков, присущих именно нейросетевой графике, определить основные и главенствующие, а также разработать их классификацию, которая могла бы помочь в дальнейшем автоматизировать процесс подобной экспертизы с привлечением современных технологий.

Данный цикл работ посвящен исследованиям области нейросетевой графики и выявлению набора общих признаков, свойственных именно для визуализации изображений с помощью искусственного интеллекта и нейросетевых технологий.

Говоря о нейросетевой графике, нельзя не упомянуть о её примечательных особенностях. Нейросети способны выдавать бесконечное количество вариантов сочетаний элементов того материала, на котором они обучались и продолжают обучаться в настоящий момент. По сути,

они вбирают в себя огромный поток визуального материала, до недавнего времени не контролируемый никем кроме разработчиков (сейчас многие художественные сайты уже вводят ограничения на использование размещенных там работ для обучения нейросетей) [16, 17]. Графическая нейросеть способна смешивать все собранные материалы между собой, порождая подчас по-настоящему неожиданные сочетания и художественные образы.

В некотором роде её работу можно назвать ускоренной моделью процесса накопления и усвоения человечеством художественных знаний с переработкой их во что-то новое. Однако в отличие от индивидуального человеческого сознания, чаще склонного к творческой переработке чего-то малоизвестного или непопулярного, нейросеть гораздо сильнее зависит от популярных трендов. Она гораздо эффективнее и точнее воспроизводит изображения, распространенные в социальных сетях, нежели что-то специфическое или не имеющее устоявшегося визуального образа.

К примеру, запрос, содержащий слова «девушка», «кот», «здание» [18], заполненный по предлагающемуся шаблону в нейросети Kandinsky 2.2, будет выполнен быстрее, с повышенной вероятностью получения нужного результата (рисунок 2). А вот непривычную нешаблонную позу нейросеть, скорее всего, не сможет воспроизвести без артефактов или постарается подвести её под более распространенную.



Рисунок 2 – Визуализация шаблонных запросов в нейросети Kandinskiy 2.2

Можно с определенной уверенностью утверждать, что графическая нейросеть во многом основывается на «стереотипах», так как в первую очередь она использует для обучения тот материал, который популярен и пользуется спросом среди пользователей сети, а затем преобразует его и сочетает между собой в новом контексте.

В связи с этим можно отметить ещё одно явление: прослеживается прямая зависимость успешного воплощения запросов пользователя от визуального наполнения той или иной тематики в интернете. Чем больше графического контента по заданной теме, тем с большей вероятностью успеха нейросеть сможет визуализировать этот запрос.

По работе нейросети можно определить, какой визуальный информационный сектор не развит в интернет-пространстве, на который она опирается. На рисунке 3 показан результат визуализации нейросетью России, наглядно отображающий многие стереотипы, распространенные за пределами страны [19].

Другой пример, в русскоязычном сегменте очень слабо развита тематика национального фольклора. Запрос «алконост» в Яндекс выдает всего около 2000 изображений, причем большая часть из них пересекается и повторяется, а часть не содержит конкретного визуального образа. Нейросеть Kandinskiy 2.2 весьма четко отражает эту проблему нехватки визуального материала – она не в состоянии воспроизвести по этому слову хотя бы отдаленно похожий образ, а вот запросы конкретных птиц, куда более распространенных в сети, «угод», «снегирь» дают изображения вполне схожие с прототипами, хоть и не лишённые артефактов. На рисунке 4 показаны три попытки нейросети визуализировать образ алконоста и фрагмент с картины В.М. Васнецова "Сирин и Алконост" в левом верхнем углу.



Рисунок 3 – Нейросеть попросили нарисовать, что происходит в России [19]



Рисунок 4 – попытка нейросети визуализировать мифическое существо из русского фольклора «Алконост» и привычный зрителю образ с картины В.М. Васнецова "Сирин и Алконост" (левый верхний угол)

Следующая особенность, которая обнаруживается при взаимодействии с любым образцом искусственного интеллекта от графической нейросети до ChatGPT – это тот факт, что у каждого информационного объекта есть так называемое информационное поле, обладающее определенной плотностью.

Информационное поле можно определить как весь объем неупорядоченной информации, связанной с искомым объектом.

Плотность информационного поля объекта в данном случае – это то количество открытой и публичной информации, которое окружает искомый объект и позволяет воссоздать его образ искусственно.

Информационное поле также можно охарактеризовать такой величиной, как полнота: чем более упорядоченной и разносторонней является информация, тем качественней получится цифровое отображение.

Самое плотное информационное поле на сегодняшний день имеют медиа-персоны и лица, представляющие государство. Информация о них представлена в огромных количествах и самых разнообразных формах: видео, запись голоса, фотографии с множества ракурсов, книги, статьи в прессе и обсуждения в соцсетях. Основываясь на ней, искусственный интеллект способен воспроизвести их голос, видеoversию или предоставить подробное описание их личности.

К примеру, ни ChatGPT, ни графические нейросети не способны описать деятельность ученого, хорошо известного в профессиональных кругах, но о котором в сети представлен минимум информации на специализированных сайтах. И в этом случае ChatGPT не способен дать адекватный ответ на вопрос об области деятельности или темах научных работ. Графическая нейросеть не найдет материала, чтобы сгенерировать хотя бы отдаленное портретное сходство. Это свидетельствует о низкой степени плотности информационного поля.

Другой пример – государственное лицо, имеющее планетарную известность и регулярно появляющееся на публике на протяжении многих лет. В этом случае можно говорить об очень высокой плотности информационного поля с огромным количеством всевозможных информационных источников. Эти источники в совокупности составляют массу недостоверной, эмоционально окрашенной текстовой, медиа и аудио информации, на которую нейросеть будет опираться и генерировать образы сообразно самым популярным трендам и стереотипам, если пользователь предоставит недостаточно конкретный и детализированный запрос.

В связи с нарастающим количеством иллюстративного материала, сгенерированного нейросетями, существуют серьезные основания полагать, что в скором времени проблема определения, является ли медиа-объект подлинным или сгенерированным посредством той или иной нейросети станет остроактуальной. Возникает необходимость постановки и решения задач верификации и обнаружения сгенерированных или обработанных фотореалистичных изображений с целью эффективного противодействия их потенциальному вредоносному использованию.

В целом нейросетевые изображения можно идентифицировать с помощью комбинации прямых и косвенных признаков, так как некоторые артефакты, порождаемые нейросетью, обычно имеют характерные особенности, которые позволяют распознать изображения, созданные с помощью нейросети. В их числе можно назвать:

1. Анахронизмы и явная историческая недостоверность, когда на одной фотографии или картине, отражающей определенный исторический промежуток, появляются объекты, которые не могли быть запечатлены в то время (рисунок 5). Для окончательного определения необходимо опираться на исторические источники. [20].
2. Нелогичность изображения, когда явно нарушена общая логика композиции или совмещены несочетаемые между собой объекты.
3. Нарушения композиции, когда общее построение сгенерированной картины на первый взгляд не нарушено и кажется естественным, однако при более близком рассмотрении окажется, что какие-то объекты не дорисованы, расположены относительно друг друга с искаженной перспективой или перетекают один в другой.
4. Химеризация, когда нейросеть не может правильно отразить запрашиваемый объект или смешивает заданные объекты между собой (наиболее часто встречающийся пример – это пальцы рук на изображениях с людьми).
5. Чрезмерная детализация и проработанность. Некоторые нейросети отличаются особым ярко выраженным стилем и эстетикой изображений, а также с одинаковым качеством и тщательностью прорабатывают все детали изображения (рисунок 6). Это особенно заметно, когда нейросеть просят изобразить что-то в стилистике цифрового рисунка или художественной картины. Одинаковая проработка всех деталей не всегда свойственна для человеческого исполнения, так как часто упущение каких-то деталей, становится художественным приемом, подчеркивающим замысел картины.



Рисунок 5 – Луи Армстронг репетирует первые шаги на Луне [20]

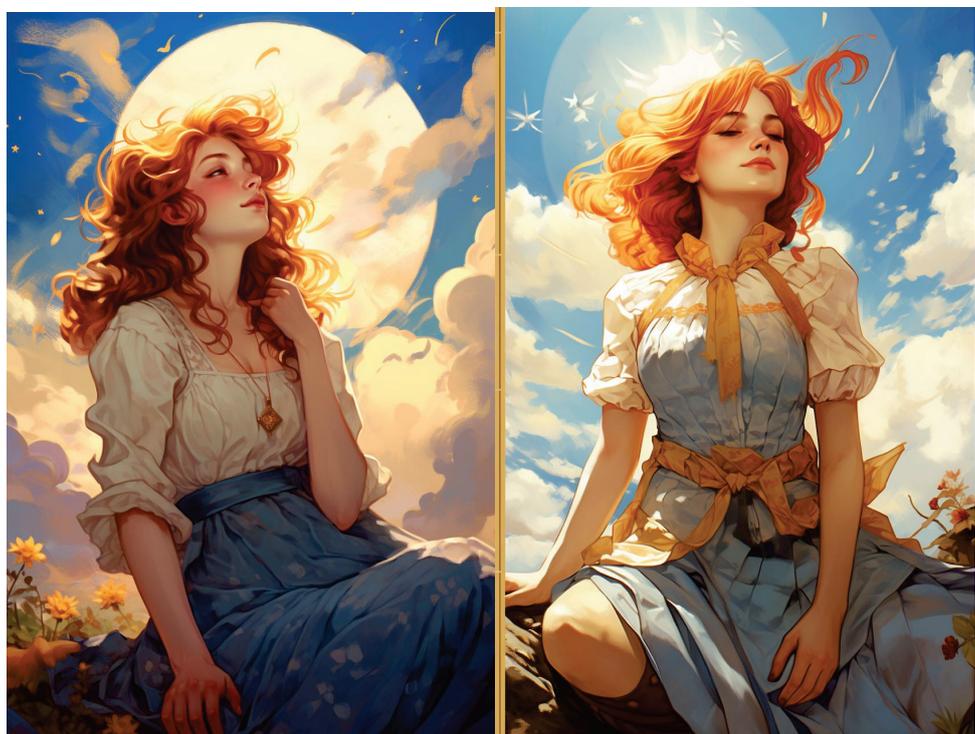


Рисунок 6 – Кейс артов по запросу (prompt: tarot of the sun – the gipfel, in the style of anime aesthetic, charming character illustrations, digital painting, impressionist colorism, jessie arms botke, sky-blue and beige, line art) [21]

6. Артефакты текстуры. Подобные артефакты характерны для всех нейросетей, так или иначе обрабатывающих изображение, от наложения фильтров до генерации изображения с нуля. В этом случае артефакты не влияют на общее изображение и возникают в тех местах, где нейросеть не может в должной мере проработать какой-то участок с высокой детализацией или воссоздать необходимую структуру. Это могут быть волосы, ткань одежды или кожа, и при увеличении можно увидеть видимое отличие поврежденного участка от остальной части изображения.
7. Для портретных фото существует свой набор признаков: идеальные лица, черты которых стремятся к золотому сечению, или расфокусированный взгляд. «Также можно обратить внимание на глаза человека на фотографии. В фейковых фото практически невозможно уловить их выражение, взгляд кажется размытым, несфокусированным» - отмечается в статье [22-24].

Это далеко не все признаки, по которым можно отличить нейросеть, но за счет постоянного развития и улучшения генерационных технологий многие признаки устаревают и могут быть применимы только к нейросетям предыдущих версий.

Необходимо выявлять общие признаки, характерные для нейросетевой графики, выделять основные и разрабатывать их классификацию, чтобы в дальнейшем по набору этих признаков возможно было установить подлинность фото и видео материалов с применением различных математических алгоритмов и верификацией точности их результатов.

4. Заключение

В работе были рассмотрены современные графические нейросети text-to-image и различные методы, используемые для преобразования текстовой информации в изображения. Также был проведен анализ результатов, которые были достигнуты в этой области.

Также рассмотрен ряд проблем, которые порождают такие системы. Были предложены подходы к решению задач, направленных на решение проблемы верификации медиа материалов и разработку алгоритмов для выявления искусственного (нейросетевого) происхождения фото и видео материалов.

5. Благодарности

Вычислительная работа проведена с помощью гибридного суперкомпьютера К-100, установленного в Центре коллективного пользования ИПИМ им. М.В. Келдыша РАН.

6. Список источников

- [1] J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, 2020, <https://doi.org/10.48550/arXiv.2006.11239> (дата обращения 29.03.2023)
- [2] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J. Zhu, S. Ermon, SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, 2022, <https://doi.org/10.48550/arXiv.2108.01073>
- [3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, 2014, <https://doi.org/10.48550/arXiv.1406.2661>
- [4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, 2021, <https://doi.org/10.48550/arXiv.2112.10752>
- [5] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134.
- [6] Koh, J. Y., Baldridge, J., Lee, H., and Yang, Y., 2021. Text-to-image generation grounded by fine-grained user attention. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 237–246.
- [7] Ramesh A., Pavlov M., Goh G., Gray S., Voss C., Radford A., Chen M, Sutskever I., 2021. Zero-Shot Text-to-Image Generation, <https://doi.org/10.48550/arXiv.2102.12092>

- [8] Radford A., Jong W.K., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv preprint arXiv:2103.00020 [cs.CV]. <https://doi.org/10.48550/arXiv.2103>.
- [9] DALL·E 2 Preview - Risks and Limitation, 2022, URL: <https://github.com/openai/dalle-2-preview/blob/main/system-card.md#model> (дата обращения 19.03.2023)
- [10] Stokel-Walker C., 2023, A Professional Artist Spent 100 Hours Working On This Book Cover Image, Only To Be Accused Of Using AI. URL: <https://www.buzzfeednews.com/article/chrisstokelwalker/art-subreddit-illustrator-ai-art-controversy> (дата обращения 26.08.2023).
- [11] O'Donoghue S. The cruellest show on TV? Deep Fake Love goes too far with AI, 2023, URL: <https://www.euronews.com/culture/2023/07/25/the-cruellest-show-on-tv-deep-fake-love-goes-too-far-with-ai> (дата обращения 19.08.2023).
- [12] Wagner J., Forensically, Photo Forensics for the Web, 2015, URL: <https://29a.ch/2015/08/16/forensically-photo-forensics-for-the-web> (дата обращения 27.03.2023).
- [13] Forensically, URL: <https://29a.ch/photo-forensics/#forensic-magnifier> (дата обращения 26.08.2023).
- [14] Tutorial: Error Level Analysis, URL: <https://fotoforensics.com/tutorial-ela.php> (дата обращения 27.03.2023).
- [15] RT на русском, Нейросеть показала, как мог бы выглядеть арест Трампа, 2023, https://t.me/rt_russian/150986 (29.03.2023)
- [16] Художники против нейросетей: справедливый протест или неолуддизм? 2023, URL: <https://habr.com/ru/companies/ruvds/articles/711834/> (дата обращения 26.08.2023).
- [17] We've filed a lawsuit challenging Stable Diffusion, a 21st century collage tool that violates the rights of artists, 2023, URL: <https://stablediffusionlitigation.com/> (дата обращения 26.08.2023)
- [18] Kandinsky 2.2, URL: <https://www.sberbank.com/promo/kandinsky/> (дата обращения 26.08.2023).
- [19] Нейросеть попросили нарисовать, что происходит в России. 2023, URL: <https://t.me/mmmadnet/13049/> (дата обращения 26.08.2023).
- [20] Сафонов С., Нейросеть сгенерировала картинки к историческим событиям, которых не было, 2023, URL: <https://www.maximonline.ru/longreads/neiroset-sgenerirovala-kartinki-k-istoricheskim-sobytiyam-kotorykh-ne-bylo-id875474/> (дата обращения 26.08.2023).
- [21] Арт с телеграм-канала, 2023, URL: https://t.me/natasha_generit/ (дата обращения 26.08.2023)
- [22] Кузьмин Е., Как отличить фото людей, которые сгенерировала нейросеть? 2023, URL: <https://journal-tinkoff-ru.turbopages.org/journal.tinkoff.ru/s/neural-photo/> (дата обращения 26.08.2023).
- [23] Эксперт назвал способ распознать созданное нейросетью изображение, 2023, URL: <https://turbo.ria.ru/20230329/neuroset-1861416305.html> (дата обращения 26.08.2023).
- [24] Румянцева А. Для распознавания созданного нейросетью фото нужно обращать внимание на фон и глаза человека, 2023, URL: <https://russian.rt.com/russia/news/1144375-neiroset-foto-glaza> (дата обращения 26.08.2023).